2025 • 29 • 8

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-25-121

2025 • 29 • 8

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-25-121

СОДЕРЖАНИЕ • 2025 • 29 • 8

Хромосомная и генетическая инженерия

Устойчивость растений к стрессовым факторам

Молекулярные маркеры в генетике и селекции

Структура и эволюция генома растений

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

CONTENTS • 2025 • 29 • 8

Chromosome and gene engineering

Stress resistance in plants

Molecular markers in genetics and breeding

Plant genome structure and evolution

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1161-1168

doi 10.18699/vjgb-25-122

Study of the progeny of sorghum mutants obtained using the CRISPR/Cas9 genetic construct directed at inducing mutations in the α-kafirin *k1C5* gene

L.A. Elkonin [1], G.A. Gerashchenkov [2], N.V. Borisenko [1], S.Kh. Sarsenova [1], V.M. Panin [1]

[1] Federal Centre of Agriculture Research of the South-East Region, Saratov, Russia
[2] Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
lelkonin@gmail.com

Abstract. Site-directed mutagenesis using genetic constructs carrying the CRISPR/Cas system is an effective technology that is actively used to solve a variety of problems in plant genetics and breeding. One of these problems is to improve the nutritional value of grain sorghum, a high-yielding heat- and drought-tolerant cereal crop that is becoming increasingly important in the conditions of climate aridization. The main reason for the relatively low nutritional value of sorghum grain is the resistance of its storage proteins, kafirins, to proteolytic digestion. We have previously obtained mutants with improved kafirin *in vitro* digestibility using the CRISPR/Cas technology in grain sorghum variety Avance. The nucleotide sequence of one of the genes (*k1C5*) of the gene family encoding the signal polypeptide of 22 kDa α-kafirin was used as a target. The aim of this study was to investigate the manifestation of the main agronomically-important traits in the progeny of these mutants and inheritance of high *in vitro* protein digestibility, and also sequencing nucleotide sequences encoding the 22 kDa α-kafirin signal polypeptide in a number of plants from the T_0 generation and their T_1 progeny. It was revealed that four of the six studied T_0 plants, as well as their progeny, had the same mutation: a T→C substitution in the 23rd position of the nucleotide sequence of the *k1C5* gene encoding the signal polypeptide, which led to a substitution of the coding triplet CTC→CCC (Leu→Pro). This mutation is located off-target, 3′ from the PAM sequence. It is suggested that this mutation may have arisen as a result of Cas9 nuclease errors caused by the presence of multiple PAM sequences located close to each other. It was found that the progeny of two of the three studied mutants (T_2 and T_3 families), grown in the experimental field conditions, differed from the original variety by a reduced plant height (by 12.4–15.5 %). The peduncle length, 1,000-grain mass, and grain mass per panicle did not differ from the original variety, with the exception of the progeny of the 2C-1.2.5b mutant, which had a reduced grain yield per panicle. Unlike the original variety, plants from the T_2 and T_3 generations had kernels with a modified type of endosperm (completely floury, or floury with inclusions of vitreous endosperm, or with a thin vitreous layer). The level of grain protein digestibility in the progeny of mutants 2C-2.1.1 #13 and 2C-1.2.5a #14 varied from 77 to 84 %, significantly exceeding the original variety (63.4 ± 2.3 %, $p < 0.05$). The level of protein digestibility from kernels with modified endosperm was higher than that of kernels with normal vitreous endosperm (84–93 %, $p < 0.05$). The reasons for the variation in endosperm texture in the progeny of the mutants and its relationship with the high digestibility of kafirins are discussed.

Key words: *Sorghum bicolor*; CRISPR/Cas; alfa-kafirins; *in vitro* protein digestibility; endosperm

For citation: Elkonin L.A., Gerashchenkov G.A., Borisenko N.V., Sarsenova S.Kh., Panin V.M. Study of the progeny of sorghum mutants obtained using the CRISPR/Cas9 genetic construct directed at inducing mutations in the α-kafirin *k1C5* gene. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1161-1168. doi 10.18699/vjgb-25-122

Funding. The work was carried out with the financial support of the Russian Science Foundation, grant No. 24-16-00063.

Исследование потомства мутантов сорго, полученных с использованием генетической конструкции CRISPR/Cas9, направленной на индукцию мутаций в гене α-кафирина *k1C5*

Л.А. Эльконин [1], Г.А. Геращенков [2], Н.В. Борисенко [1], С.Х. Сарсенова [1], В.М. Панин [1]

[1] Федеральный аграрный научный центр Юго-Востока, Саратов, Россия
[2] Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия
lelkonin@gmail.com

Аннотация. Сайт-направленный мутагенез посредством генетических конструкций, несущих систему CRISPR/Cas, считается эффективной технологией, активно применяемой для решения самых разных задач генетики и селекции у многих возделываемых культур. Зерновое сорго – уникальная по своей засухоустойчивости и жаростойкости зла-

ковая культура, служащая источником кормов и пищевого зерна в аридных регионах земного шара, приобретающая все большее значение в условиях глобального потепления климата. Одной из основных задач селекции сорго является улучшение сравнительно низкой питательной ценности зерна, обусловленной в том числе устойчивостью его запасных белков – кафиринов – к протеолитическому расщеплению. Нами ранее путем использования технологии CRISPR/Cas у зернового сорго сорта Аванс были получены мутанты с улучшенной перевариваемостью кафиринов в системе *in vitro*. При этом в качестве мишени был выбран один из членов генного семейства, кодирующего 22 кДа α-кафирин, а именно ген *k1C5*. Цель данного исследования – изучение потомства полученных нами ранее растений, несущих мутации в гене *k1C5*, отличающихся улучшенной перевариваемостью белков зерна в системе *in vitro*, а именно: анализ наследования высокой перевариваемости, проявления основных селекционно ценных признаков, а также выявление структуры нуклеотидной последовательности, кодирующей сигнальный полипептид 22 кДа α-кафирина у ряда растений из поколения T_0 и их потомства T_1. Обнаружено, что у четырех из шести изученных растений T_0, а также у их потомства присутствует одна и та же мутация: замена Т→С в 23-й позиции нуклеотидной последовательности гена *k1C5*, кодирующей сигнальный полипептид, которая привела к замене кодирующего триплета СТС→ССС (Leu→Pro). Эта мутация располагается за пределами выбранной мишени, в направлении 3′ от последовательности PAM. Высказывается предположение, что данная мутация могла возникнуть в результате ошибок нуклеазы Cas9, обусловленных наличием нескольких последовательностей PAM, расположенных близко друг к другу. Установлено, что потомство двух из трех изученных мутантов (семьи T_2 и T_3), выращенное в условиях опытного поля, отличалось от исходного сорта сниженной высотой растения (на 12.4–15.5 %). Длина цветоножки, масса 1000 зерен и масса зерна с метелки не отличались от исходного сорта, за исключением потомства мутанта 2С-1.2.5b, у которого урожай зерна с метелки был снижен. В отличие от исходного сорта, у растений из поколений T_2 и T_3 присутствовали зерновки с модифицированным типом эндосперма (полностью мучнистым, либо с вкраплениями стекловидного эндосперма, либо с тонким стекловидным слоем). Уровень перевариваемости белков зерна в потомстве мутантов 2С-2.1.1 № 13 и 2С-1.2.5а № 14 варьировал от 77 до 84 %, значимо превышая показатель исходного сорта (63.4 ± 2.3 %, $p < 0.05$). Уровень перевариваемости белков из зерновок с модифицированным эндоспермом был выше, чем у зерновок с обычным стекловидным эндоспермом (84–93 %, $p < 0.05$). Обсуждаются причины варьирования текстуры эндосперма у потомства полученных мутантов и ее связь с высокой перевариваемостью кафиринов.
Ключевые слова: *Sorghum bicolor*; CRISPR/Cas; альфа-кафирин; перевариваемость *in vitro*; эндосперм

Introduction

Modifying the nucleotide sequences of genes using the CRISPR/Cas genome editing technology is one of the most powerful tools in plant genetics and breeding (Zhu et al., 2020; Gao, 2021; Saini et al., 2023). In recent years, the CRISPR/Cas technology has been intensively used in many cultivated plant species, including sorghum, a unique drought- and heat-resistant cereal crop that serves as a source of feed and food grain in arid regions of the globe. Despite the fact that sorghum is one of the most difficult cereal species to transform, many studies have appeared on sorghum genome editing using the CRISPR/Cas technology, which have been summarized in a number of reviews (Balakrishna et al., 2020; Parikh et al., 2021; Wong A.C.S. et al., 2022).

One of the most actual problems in sorghum breeding is improving the digestibility of grain storage proteins. Sorghum grain contains a significant amount of protein (on average 10–12 %, and in some lines up to 16–18 %), represented by different classes of kafirins (α, β, γ, δ), related to alcohol-soluble proteins – prolamins, which make up to 70–80 % of the total protein content, and non-kafirin proteins, the composition of which is poorly studied (Bean et al., 2018). Different classes of kafirins differ in their molecular weight and amino acid composition, and are encoded by different genes. An important feature of kafirins is their resistance to proteolytic digestion. As a result, the level of *in vitro* grain protein digestibility in the vast majority of varieties and hybrids does not exceed 40–60 % (Wong J.H. et al., 2010; Elkonin et al., 2013; Duressa et al., 2018). Such resistance of kafirins to proteolytic digestion also reduces the digestibility of starch, since undigested kafirins prevent complete amylolytic cleavage of starch granules (Zhang, Hamaker, 1998; Ezeogu et al., 2005; Wong J.H. et al., 2009).

The resistance of kafirins to protease digestion is multifactorial (see reviews: Belton et al., 2006; Duressa et al., 2018). These factors include the chemical structure of kafirins, which are rich in sulfur-containing amino acids (especially γ- and β-kafirins) capable of forming intra- and intermolecular cross-links, which are believed to prevent the proteolytic cleavage of kafirins; the interaction of kafirins with polyphenols, which inhibit protease activity. An important factor is the spatial organization of different kafirins in the protein bodies of endosperm cells. In the early stages of endosperm development, γ- and β-kafirins are synthesized and deposited in protein bodies developing in the endoplasmic reticulum. Alpha-kafirin, synthesized at later stages of endospermogenesis, is deposited inside protein bodies, pushing γ-kafirin to the periphery, which forms a kind of "shell" that is difficult for proteases to digest (De Mesa-Stonestreet et al., 2010; Duressa et al., 2018).

As a result of the study of mutants with impaired synthesis of kafirins obtained using RNA interference (see review: Elkonin et al., 2021), it was found that partial suppression of kafirin synthesis significantly improves the digestibility of grain proteins and promotes the synthesis of other proteins with higher nutritional value. In this regard, targeted induction of mutations in the genes encoding kafirin synthesis can contribute to the production of new sorghum lines with improved digestibility of grain proteins, which, unlike lines carrying the genetic construct for RNA silencing, will be devoid of transgenes.

In recent years, several studies have been published reporting successful editing of α-, β-, and γ-kafirin genes (Li A. et al., 2018; Massel et al., 2022, 2023; Elkonin et al., 2023; Li X. et al., 2024). Most of these studies targeted nucleotide sequences encoding signal polypeptides responsible for the

deposition of α- and γ-kafirins in the protein bodies of endosperm cells (Li A. et al., 2018; Elkonin et al., 2023; Li X. et al., 2024). These mutants had improved digestibility of grain proteins, in contrast to mutants with mutations in the β-kafirin gene structure (Massel et al., 2023).

The aim of this study was to explore the progeny of previously obtained plants carrying mutations in the *k1C5* gene, characterized by improved digestibility of grain proteins; namely, to study the inheritance of high digestibility, the manifestation of the main agronomically important traits, and to identify the structure of the nucleotide sequence of the *k1C5* gene encoding the 22 kDa α-kafirin signal polypeptide.

Material and methods

Material and growing conditions. The progenies of T_1 plants with high *in vitro* protein digestibility, which were obtained from the T_0 mutants 2C-2.1.1 [T_1 #11 (86.6 % digestibility) and T_1 #13 (86.7 %)], T_0 2C-1.2.5a [T_1 #11 (92.4 %) and T_1 #14 (77.3 %)], and T_0 2C-1.2.5b [T_1 #14 (91.8 %)], were studied. These mutants were obtained in genome editing experiments with grain sorghum cv. Avance using the binary vector p2C containing the *Cas9* endonuclease gene and gRNA targeted at the nucleotide sequence of the *k1C5* gene encoding the 22 kDa signal polypeptide of α-kafirin (Elkonin et al., 2023). The selected T_1 plants with high protein digestibility did not contain the CRISPR/Cas genetic construct (Elkonin et al., 2023). The studied progenies (T_2 and T_3 generations) were grown in the experimental field of the Federal Centre of Agriculture Research of the South-East Region (Saratov, Russia). Plants were grown in 4-m rows with 70 cm row spacing, with a plant density of 6 plants per 1 m. The panicles of all plants were carefully bagged in parchment bags before flowering. The following traits were analyzed: plant height, peduncle length, 1,000-grain mass, grain yield per panicle, endosperm type, and *in vitro* digestibility of grain proteins. In each family, 10–20 plants were studied.

Grain protein digestibility. To study the digestibility of grain proteins, the method of treating whole-milled flour with pepsin was used (Aboubacar et al., 2001; Wong J.H. et al., 2009). In this case, a weighed sample of flour (60 mg) was incubated in 1 ml of 0.15 % pepsin solution (Sigma-Aldrich, P7000; 250 units/mg) in 0.1 M potassium phosphate buffer (pH 2.0) at 37 °C on a shaker (70 rpm).

A method based on scanning the electrophoretic spectra of proteins obtained in SDS-PAGE was used for quantitative assessment of digestibility (Aboubacar et al., 2001; Nunes et al., 2004; Wong J.H. et al., 2009; Elkonin et al., 2013). For this purpose, flour samples after pepsin digestion, as well as control samples incubated in potassium phosphate buffer without the addition of pepsin, were centrifuged at 13,000 rpm; the pellet was incubated in extraction buffer (0.0125 M $Na_2B_4O_7$, pH 10.0) under reducing conditions (1 % SDS, 2 % 2-mercaptoethanol) at room temperature for 2 h, after which it was boiled (100 °C) for 5 min. Samples were centrifuged at 13,000 rpm and separated by SDS-PAGE on 12.5 % (w/v) polyacrylamide gel according to a modified Laemmli method (Laemmli, 1970). 15 µl of extract were added to each lane. Separation was monitored using protein molecular weight markers, 10–200 kDa (Servicebio, G2058, Wuhan, Hubei, China). Gels were stained with Coomassie R-250. After electrophoresis, gels were scanned using the ChemiDoc system (Bio-Rad Laboratories, Hercules, CA, USA); the protein amount was estimated using Image Lab 6.1 software (Bio-Rad). Digestibility indices were calculated as the percentage difference between the protein volume in the control sample and the digested sample, relative to the control sample. The previously obtained Avance-1/18 mutant with a genetic construct for RNA silencing of the *gKAF1* gene was used as a standard of high *in vitro* digestibility (Elkonin et al., 2021). Experiments were performed in duplicate.

Endosperm texture. The endosperm texture was determined on cross-sections of mature kernels, which were made using a sharp scalpel. The following types of endosperm were distinguished: normal with a thick vitreous layer and modified, which included floury, floury with blurred vitreous endosperm, and floury with a thin rim of vitreous endosperm. Forty kernels were analyzed from each plant.

Sequencing of the *k1C5* gene nucleotide sequence. To identify mutations, PCR amplicons of the *k1C5* gene (primers F: 5′-TTGCCAGGGCTAGTTGACTG-3′ and R: 5′-AGGCTTTGATCCACATGAGCA-3′) were cloned into the pAL2-T vector (Eurogen, Russia). Sanger sequencing was performed by Syntol (Moscow, Russia) on an ABI 3130 genetic analyzer (sequencing primer: 5′-TTGCCAGGGC TAGTTGACTG-3′). Mutations in the sequenced amplicons were identified using Chromas (https://www.technelysium.com.au) and SnapGene Viewer 5.2.4 (https://www.snapgene.com) computer programs.

Methods of biological statistics. To assess differences in *in vitro* protein digestibility of the studied samples, one-way ANOVA was performed using the AGROS software package, version 2.09 (S.P. Martynov, Institute of Genetics of the Russian Academy of Sciences), and Duncan's Multiple Comparisons Test. Differences in the manifestation of morphometric traits between mutant families and the original variety were assessed using Student's *t*-test.

Results

Sequencing of the nucleotide sequence encoding the 22 kDa signal polypeptide of α-kafirin of the *k1C5* gene

Sequencing of the nucleotide sequence encoding the 22 kDa α-kafirin signal polypeptide of one of the genes of the *k1C* gene family (*k1C5*) in two plants from the progeny of the T_0 2C-2.1.1 mutant, #2 and #11 (T_1 generation), characterized by improved digestibility of endosperm proteins (86 and 87 %, respectively), showed that they have the same mutation: a substitution of the 23rd nucleotide, counted from the 5′-end of the nucleotide sequence of the signal polypeptide (in the F-chain: T→C; in the R-chain: A→G) (Fig. 1*b*, *c*). Sequencing of a similar sequence in the original T_0 mutant showed that the same mutation was also present in the parental T_0 plant (Fig. 1*a*). *In silico* analysis showed that this mutation leads to a substitution of the coding triplet CTC→CCC, which should result in a substitution of the eighth amino acid of the α-kafirin signal polypeptide, namely, in the substitution of leucine, an aliphatic non-polar hydrophobic amino acid, for proline, a heterocyclic less hydrophobic amino acid that causes a bend in the α-helix of the protein. Such a substitution could change

the structural and functional properties of the polypeptide and, as a consequence, the nature of α-kafirin deposition in protein bodies, and thereby affect their digestibility.

Sequencing of a similar amplicon in one of the T_1 plants from the progeny of another T_0 mutant 2C-1.2.5a #11, characterized by improved protein digestibility (92 %), also revealed the presence of a T→C mutation at the same site of the nucleotide sequence encoding the signal polypeptide (Fig. 1*d*). Remarkably, we identified the same mutation by sequencing the DNA of another T_0 plant 2C-1.2.5b (Fig. 1*e*), regenerated from the same callus as 2C-1.2.5a. Previously, we found the same mutation in the T_0 plant 2C-1.2.9, while this mutation was absent in two other T_0 plants (1C-2.1.1 and 2C-1.2.4) (Elkonin et al., 2023).

Thus, four out of the six T_0 plants studied have the same mutation: a T→C substitution at position 23 of the nucleotide sequence of the *k1C5* gene, and this mutation is inherited in the T_1 generation.

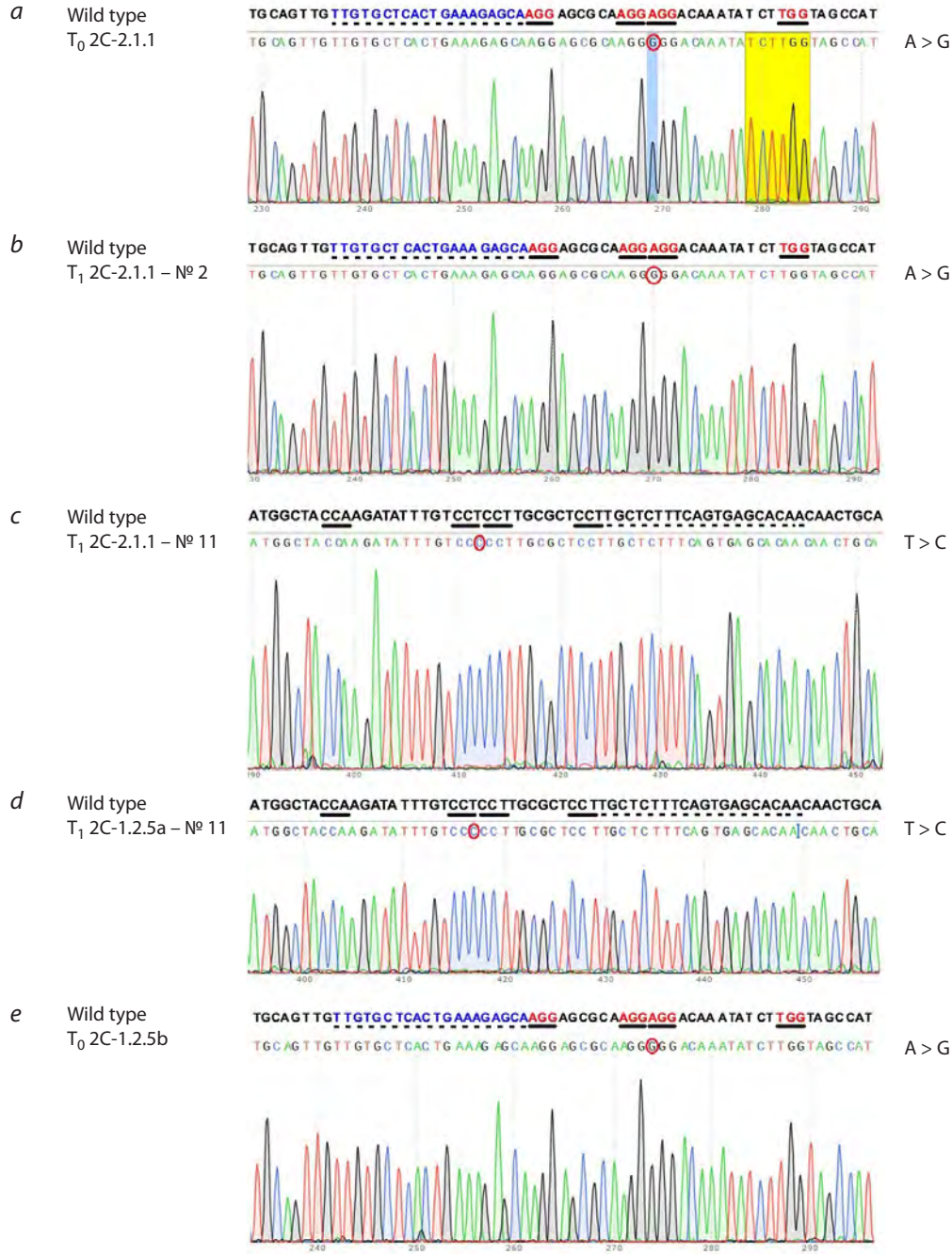


Fig. 1. Results of sequencing of the nucleotide sequences encoding 22 kDa α-kafirin signal polypeptides in T_0 2C-2.1.1 (*a*) and its T_1 progeny 2C-2.1.1, plant #2 (*b*) and plant #11 (*c*); T_1 2C-1.2.5a, plant #11 (*d*); T_0 2C-1.2.5b (*e*).

a, *b*, *e* – R-chain; *c*, *d* – F-chain. PAM sequences are underlined with a solid line; the target sequence is dashed. The nucleotide sequence encoding the 22 kDa α-kafirin signal polypeptide was taken from the Phytozome website, https://phytozome.jgi.doe.gov: Sobic.005G193100, Chr05: 67654898–67655764.

Manifestation of agronomically important traits

An analysis of the manifestation of the main agronomically important traits in the progeny of mutants with improved digestibility of endosperm proteins obtained by us earlier (Elkonin et al., 2023) revealed that in the T_2 generation, two of them, 2C-1.2.5a and 2C-1.2.5b (families 203/23 and 200/23), had reduced plant height compared to the original cv. Avance, by 12.4–15.5 %, respectively (Table 1). The reduced plant height in the 2C-1.2.5a mutant was also inherited in the T_3 generation (by 5.5 %, family 208/23). The length of the peduncle (protrusion of the paniculate internode) did not differ in the progeny of the mutants and the original cv. Avance. The 1,000-grain mass and grain yield per panicle in all families also did not differ from the original cv. Avance, with the exception of the progeny of the 2C-2.1.1 mutant (T_2 195/23 family), which had larger and heavier grains, and the progeny of the 2C-1.2.5b mutant (T_2 200/23 family), which had reduced grain yield per panicle. In plants of all the studied families, most kernels had endosperm of the normal vitreous type, characteristic of the original cv. Avance. However, almost all families contained plants that had kernels with a floury endosperm, or with a blurred or thin vitreous layer (Fig. 2), i. e., endosperm types characteristic of mutants with impaired kafirin synthesis (Elkonin et al., 2021). The proportion of such kernels in some plants from families 197/23 (T_3 generation of the mutant 2C-2.1.1) and 208/23 (T_3 generation of the mutant 2C-1.2.5a) reached 35–40 % (Table 1). Often, such kernels were smaller in size compared to kernels with normal vitreous endosperm.

Analysis of grain protein digestibility in the progeny of mutants 2C-2.1.1 #13 and 2C-1.2.5a #14 (both from T_1) showed that increased values of this trait, compared to the original cv. Avance, were manifested in plants from generations T_2 and T_3 (Fig. 3). For example, in the progeny of mutant 2C-2.1.1 #13 (Table 2, families T_2 195/23 and T_3 197/23), as well as mutant 2C-1.2.5a #14 (family T_3 208/23), the digestibility level reached 77–84 %, exceeding the original cultivar by 10–20 % ($p < 0.05$), while a significantly higher digestibility level was observed in kernels with a normal vitreous type of endosperm, characteristic of the original cultivar. At the same time, the level of protein digestibility from kernels with floury or blurred vitreous endosperm was significantly higher than that of kernels with normal vitreous endosperm, reaching 84–93 % and significantly exceeding the level of digestibility in the original cultivar ($p < 0.05$), which did not have such kernels.

Discussion

The CRISPR/Cas technology is considered one of the most effective tools for inducing mutations at strictly defined loci of plant genome. However, in some cases, the precision of editing gene nucleotide sequences using the CRISPR/Cas9

Table 1. Manifestation of agronomically valuable traits in the progeny of sorghum mutants obtained using the CRISPR/Cas genetic construct targeting the 22 kDa α-kafirin gene (*k1C5*)

T_0 mutant	Progeny					
	Generation, family	Plant height, cm	Peduncle length, cm	Endosperm type[1]	$M_{1,000}$, g	Grain yield per panicle, g
cv. Avance	191/23	133.1 ± 2.0	13.2 ± 1.4	Vitreous (100 %)	32.8 ± 1.4	32.6 ± 3.1
2C-2.1.1	T_2: 195/23	132.7 ± 3.0	14.1 ± 1.2	Vitreous; modified (up to 4 %)	37.4 ± 0.6*	31.3 ± 2.7
	T_3: 197/23	126.5 ± 2.9	13.9 ± 1.0	Vitreous; modified (up to 24–43 %)	34.6 ± 1.5	27.3 ± 3.2
2C-1.2.5a	T_2: 203/23	116.6 ± 1.2**	15.4 ± 0.6	Vitreous; modified (up to 6–10 %)	31.3 ± 1.3	27.0 ± 2.1
	T_3: 208/23	125.8 ± 2.3*	14.3 ± 0.7	Vitreous; modified (up to 19–71 %)	34.9 ± 1.7	27.7 ± 2.9
2C-1.2.5b	T_2: 200/23	112.5 ± 4.6**	12.8 ± 1.0	Vitreous; modified (up to 5–9 %)	30.9 ± 1.4	26.6 ± 1.3*

*, ** Differs from the original cv. Avance at $p < 0.05$ and $p < 0.01$, respectively. [1] The proportion of kernels with a different endosperm type in different plants from the family.

Fig. 2. Cross-sections of the kernels of the mutant 2C-1.2.5a (plants from family 208/13).

a – normal vitreous endosperm, *b* – floury endosperm, *c* – endosperm with a blurred vitreous layer. Scale bar 1 mm.

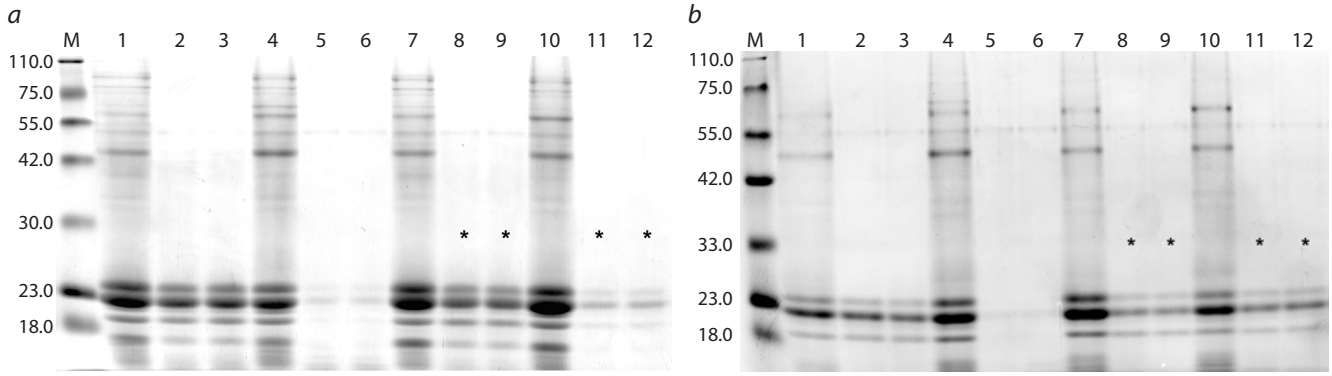


Fig. 3. Electrophoretic spectra of proteins from flour of sorghum kernels from different generations of the mutant 2C-2.1.1 obtained by site-directed mutagenesis of the *k1C5* gene.

a: plant #197-9/23 from the T_3 generation (lanes 7–12) (experiment 03.09-2, see Table S1)[1]. *b*: plant #195-3/23 from the T_2 generation (lanes 7–12) (experiment 13.08-1, see Table S1). On both plates: 1–3 – original cv. Avance; 4–6 – mutant with RNA silencing of the *gKAF1* gene (Elkonin et al., 2021) (standard of high level of *in vitro* protein digestibility); 7–9 – kernels with normal vitreous endosperm; 10–12 – kernels with floury endosperm; 1, 4, 7, 10 – control samples (without pepsin treatment); 2, 3, 5, 6, 8, 9, 11, 12 – samples after pepsin treatment (two replicates for each sample); M – molecular weight markers (Servicebio, G2058). The spectra of samples characterized by significantly higher digestibility compared to the Avance variety (Table 2, Table S1) are marked with asterisks.

Table 2. *In vitro* digestibility of flour proteins from kernels with different endosperm types in some plants from the progeny of sorghum mutants obtained by site-directed mutagenesis of the *k1C5* gene

Mutant	Plant, generation	Protein digestibility (%) from kernels with different endosperm types	
		Normal vitreous	Floury or blurred vitreous
cv. Avance		63.4 ± 2.3	–
2C-1.2.5a, T_1 #14, selection in T_2 from a kernel with blurred vitreous endosperm	208-2/23, T_3	72.8*	79.5*
	208-11/23, T_3	84.3*	84.1*
	208-6/23, T_3	71.7 ns	84.8*
2C-1.2.5a, T_1 #1	203-4/23, T_2	65.2 ns	–
2C-2.1.1, T_1 #13	195-3/23, T_2	73.9*	76.1*
	195-4/23, T_2	71.3*	79.7*
	195-8/23, T_2	62.3 ns	73.9*
2C-2.1.1, T_1 #13, selection in T_2 from a kernel with blurred vitreous endosperm	197-9/23, T_3	68.6*	92.9*
	197-6/23, T_3	64.7 ns	81.0*
	197-11/23, T_3	77.1*	78.3*

Note. * Differs from the original cv. Avance at $p < 0.05$, in accordance with the F-criterion (Table S1); ns – no significant differences from cv. Avance when analyzing the corresponding SDS-PAGE.

system may be flawed (Sturme et al., 2022; Guo et al., 2023; Movahedi et al., 2023).

In our experiments, four out of six studied T_0 plants had the same mutation: a T→C substitution at position 23 of the *k1C5* gene nucleotide sequence. This mutation is located outside the selected target, 3′-end to the PAM (protospacer-adjacent motif) sequence, and is therefore off-target. A detailed analysis of the nucleotide sequence of this region of the *k1C5* gene revealed the presence of several PAM sites located close to each other: two 5′-AGG and one 5′-TGG (Fig. 1). It is possible that due to such proximity, the Cas9 nuclease could make errors and introduce breaks between the two PAM sites: AGG ↓ AGG. Therefore, one of the reasons for the occurrence of off-target mutations, as our data show, may be a significant number of closely located PAM sites. Similar examples of off-target Cas9 activity, where a mutation occurs in the target gene but outside the chosen target, were previously found in a study editing the *Phytoene desaturase* (*PDS*) gene in two cassava varieties (*Manihot esculenta* Crantz) (Odipio et al., 2017). Notably, in another work in sorghum on editing the nucleotide sequence encoding the signal polypeptide of the γ-kafirin gene using a CRISPR/Cas9 genetic construct, mutations occurred not at the

[1] Supplementary Table S1 is available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Elkonin_Engl_29_8.pdf

canonical site, between the 3rd and 4th nucleotides 5′-end to the PAM, but after the 15th nucleotide of the target and beyond, 5′-end to the PAM site, but within the target gene (Li X. et al., 2024). These facts raise questions about the accuracy of genome editing using Cas9 nuclease in sorghum.

Another important point worthy of discussion is the decrease in the level of kafirin digestibility in the progeny of the mutants we obtained. Previously, we found a significant increase in the level of grain protein digestibility in a number of mutants obtained in the T_1 generation: up to 80–87 % in the 2C-2.1.1 mutant, up to 86 and 92 % in the 2C-1.2.5b and 2C-1.2.5a mutants, respectively (Elkonin et al., 2023). In the T_3 generation, the digestibility level decreased to 68–74 % in the 2C-2.1.1 mutant and 72–84 % in the 2C-1.2.5a mutant, significantly exceeding, however, the same indicator in the original cv. Avance (Table 2, Table S1); in the plants from the progeny of the 2C-1.2.5b mutant, there were no significant differences from cv. Avance.

A possible reason for such a decrease in digestibility may be different growing conditions of the plants: the T_1 generation was grown in a climate chamber under conditions of regular watering and high air humidity, while the T_3 plants were grown in an experimental field plot. It is known that under drought stress conditions, the digestibility of sorghum grain proteins is significantly reduced in some cultivars (Impa et al., 2019). In addition, a possible compensatory increase in the expression of other genes controlling the synthesis of kafirins, which led to the restoration of their content and a decrease in the level of digestibility of grain proteins, cannot be ruled out. Such a compensatory increase in the content of γ-kafirin was previously found in sorghum mutants with impaired synthesis of β-kafirin, which restored the overall balance of kafirins in the grain and did not lead to an improvement in the digestibility index (Massel et al., 2023).

Of particular interest is the variation in endosperm texture in mutants from different generations. In T_0 plants, the formation of kernels with impaired development of vitreous endosperm was observed (Gerashchenkov et al., 2021). Such kernels are characteristic of sorghum mutants with partially suppressed kafirin synthesis and high protein digestibility (see reviews: Duressa et al., 2018; Elkonin et al., 2021). In T_1 plants, kernels with normally developed vitreous endosperm and high protein digestibility were formed, which was an unusual phenomenon, given the close correlation between high digestibility and floury endosperm in sorghum (Duressa et al., 2018). In the T_2 and T_3 generations, plants from a number of families again had kernels with a modified type of endosperm (floury or with a thin, often blurred vitreous layer along the periphery of the endosperm), which were distinguished by a significantly higher level of digestibility (Table 2). As a result of the selection of such kernels, we obtained two T_3 families, 208/23 and 197/23, in which plants contained kernels with normal vitreous endosperm with a higher level of protein digestibility than in the original cultivar. Such variations in the endosperm texture may be a consequence of the influence of environmental conditions on the expression of the induced mutation, or another mutation that we have not yet identified affects the modification of the endosperm type. A more definitive conclusion can be made in the future as a result of additional studies.

Conclusion

In summary, as a result of studying the progeny of sorghum mutants obtained using the CRISPR/Cas9 genetic construct aimed at inducing mutations in the *k1C5* gene encoding the synthesis of α-kafirin, we identified two T_3 families, 208/23 and 197/23, in which the plants contained kernels with normal vitreous endosperm and a higher level of grain protein digestibility in comparison with the original cultivar (up to 72–84 %, compared to 62–64 % in the original cv. Avance). Plants from these families do not have significant differences in the manifestation of agronomically valuable traits from the original cultivar, with the exception of reduced height (by 5–15 %), and do not contain the CRISPR/Cas genetic construct. Four of the six T_0 plants studied harbor the same mutation: a T→C substitution at position 23 of the *k1C5* gene sequence, and this mutation is inherited by the T_1 generation. This mutation is located 3′-end to the PAM sequence, and may be a consequence of off-target Cas9 activity, in which the mutation occurs in the target gene but off-target due to the presence of multiple PAM sites located close to each other.

References

Aboubacar A., Axtell J.D., Huang C.P., Hamaker B.R. A rapid protein digestibility assay for identifying highly digestible sorghum lines. *Cereal Chem*. 2001;78:160-165. doi 10.1094/CCHEM.2001.78.2.160

Balakrishna D., Singode A., Narasimham D., Venkatesh Bhat B. Current status and future prospects of genetic transformation and gene editing in sorghum. In: Tonapi V.A., Talwar H.S., Are A.K. (Eds) Sorghum in the 21st Century: Food – Fodder – Feed – Fuel for a Rapidly Changing World. Singapore: Springer, 2020;511-535. doi 10.1007/978-981-15-8249-3_21

Bean S.R., Ioerger B.P., Wilson J.D., Tilley M., Rhodes D., Herald T.J. Structure and chemistry of sorghum grain. In: Rooney W. (Ed.) Achieving Sustainable Cultivation of Sorghum. Vol. 2. 2018. doi 10.19103/as.2017.0015.21

Belton P.S., Delgadillo I., Halford N.G., Shewry P.R. Kafirin structure and functionality. *J Cereal Sci*. 2006;44:272-286. doi 10.1016/j.jcs.2006.05.004

De Mesa-Stonestreet N.J., Alavi S., Bean S.R. Sorghum proteins: the concentration, isolation, modification, and food applications of kafirins. *J Food Sci.* 2010;75:90-104. doi 10.1111/j.1750-3841.2010.01623.x

Duressa D., Weerasoriya D., Bean S.R., Tilley M., Tesso T. Genetic basis of protein digestibility in grain sorghum. *Crop Sci*. 2018;58(6):2183-2199. doi 10.2135/cropsci2018.01.0038

Elkonin L.A., Italianskaya J.V., Fadeeva I.Yu., Bychkova V.V., Kozhemyakin V.V. In vitro protein digestibility in grain sorghum: effect of genotype and interaction with starch digestibility. *Euphytica*. 2013;193:327-337. doi 10.1007/s10681-013-0920-4

Elkonin L.A., Panin V.M., Kenzhegulov O.A., Sarsenova S.Kh. RNAi-mutants of *Sorghum bicolor* (L.) Moench with improved digestibility of seed storage proteins. In: Jimenez-Lopez J.C. (Ed.) Grain and Seed Proteins Functionality. London: Intech Open Ltd, 2021. doi 10.5772/intechopen.96204

Elkonin L.A., Gerashchenkov G.A., Borisenko N.V., Sarsenova S.Kh., Panin V.M. Development of sorghum mutants with improved *in vitro* protein digestibility by CRISPR/Cas9 editing of kafirin genes. *The Crop J*. 2023;11:1411-1418. doi 10.1016/j.cj.2023.02.005

Ezeogu L.I., Duodu K.G., Taylor J.R.N. Effects of endosperm texture and cooking conditions on the in vitro starch digestibility of sorghum and maize flours. *J Cereal Sci*. 2005;42:33-44. doi 10.1016/j.jcs.2005.02.002

Gao C. Genome engineering for crop improvement and future agriculture. *Cell*. 2021;184:1621-1635. doi 10.1016/j.cell.2021.01.005

Gerashchenkov G.A., Elkonin L.A., Gerashchenkov K.G., Rozhnova N.A., Hiekel S., Kumlehn J., Chemeris A.V. Binary vector construction for site-directed mutagenesis of kafirin genes in sorghum. *Amer J Plant Sci*. 2021;12:1276-1287. doi 10.4236/ajps.2021.128089

Guo C., Ma X., Gao F., Guo Y. Off-target effects in CRISPR/Cas9 gene editing. *Front Bioeng Biotechnol*. 2023;11:1143157. doi 10.3389/fbioe.2023.1143157

Impa S.M., Perumal R., Bean S.R., Sunoj V.S.J., Jagadish S.V.K. Water deficit and heat stress induced alterations in grain physicochemical characteristics and micronutrient composition in field grown grain sorghum. *J Cereal Sci*. 2019;86:124-131. doi 10.1016/j.jcs.2019.01.013

Laemmli U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;227:680-685.

Li A., Jia S., Yobi A., Ge Z., Sato S., Zhang C., Angelovici R., Clemente T.E., Holding D.R. Editing of an alpha-kafirin gene family increases, digestibility and protein quality in sorghum. *Plant Physiol*. 2018;177:1425-1438. doi 10.1104/pp.18.00200

Li X., Liu W., Wang G., Sai-Ming Sun S., Yuan L., Wang J. Improving digestibility of sorghum proteins by CRISPR/Cas9-based genome editing. *Food Energy Secur*. 2024;13:e506. doi 10.1002/fes3.506

Massel K., Lam Y., Hintzsche J., Lester N., Botella J.R., Godwin I.D. Endogenous U6 promoters improve CRISPR/Cas9 editing efficiencies in *Sorghum bicolor* and show potential for applications in other cereals. *Plant Cell Rep*. 2022;41:489-492. doi 10.1007/s00299-021-02816-z

Massel K., Hintzsche J., Restall J. CRISPR-knockout of β-kafirin in sorghum does not recapitulate the grain quality of natural mutants. *Planta*. 2023;257:8. doi 10.1007/s00425-022-04038-3

Movahedi A., Aghaei-Dargiri S., Li H., Zhuge Q., Sun W. CRISPR variants for gene editing in plants: biosafety risks and future directions. *Int J Mol Sci*. 2023;24:16241. doi 10.3390/ijms242216241

Nunes A., Correia I., Barros A., Delgadillo I. Sequential *in vitro* pepsin digestion of uncooked and cooked sorghum and maize samples. *J Agric Food Chem*. 2004;52:2052-2058. doi 10.1021/jf0348830

Odipio J., Alicai T., Ingelbrecht I., Nusinow D.A., Bart R., Taylor N.J. Efficient CRISPR/Cas9 genome editing of *Phytoene desaturase* in Cassava. *Front Plant Sci*. 2017;8:1780. doi 10.3389/fpls.2017.01780

Parikh A., Brant E.J., Baloglu M.C. CRISPR/Cas-mediated genome editing in sorghum – recent progress, challenges and prospects. *In Vitro Cell Dev Biol – Plant*. 2021;57:720-730. doi 10.1007/s11627-021-10215-y

Saini H., Thakur R., Gill R., Tyagi K., Goswami M. CRISPR/Cas9-gene editing approaches in plant breeding. *GM Crops & Food*. 2023;14:1-17. doi 10.1080/21645698.2023.2256930

Sturme M.H.J., van der Berg J.P., Bouwman L.M.S., De Schrijver A., de Maagd R.A., Kleter G.A., Battaglia-de Wilde E. Occurrence and nature of off-target modifications by CRISPR-Cas genome editing in plants. *ACS Agric Sci Technol*. 2022;2:192-201. doi 10.1021/acsagscitech.1c00270

Wong A.C.S., Lam Y., Hintzsche J., Restall J., Godwin I.D. Genome editing towards sorghum improvement. In: Zhao K., Mishra R., Joshi R.K. (Eds) Genome Editing Technologies for Crop Improvement. Singapore: Springer Nature, 2022;295-321. doi 10.1007/978-981-19-0600-8_14

Wong J.H., Lau T., Cai N., Singh J., Pedersen J.F., Vensel W.H., Hurkman W.J., Wilson J.D., Lemaux P.G., Buchanan B.B. Digestibility of protein and starch from sorghum (*Sorghum bicolor*) is linked to biochemical and structural features of grain endosperm. *J Cereal Sci*. 2009;49:73-82. doi 10.1016/j.jcs.2008.07.013

Wong J.H., Marx D.B., Jeff D., Wilson J.D., Buchanan B.B., Lemaux P.G., Pedersen J.F. Principal component analysis and biochemical characterization of protein and starch reveal primary targets for improving sorghum grain. *Plant Science*. 2010;179:598-611. doi 10.1016/j.plantsci.2010.08.020

Zhang G., Hamaker B.R. Low α-amylase starch digestibility of cooked sorghum flours and the effect of protein. *Cereal Chem*. 1998;75:710-713. doi 10.1094/CCHEM.1998.75.5.710

Zhu H., Li C., Gao C. Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat Rev Mol Cell Biol*. 2020;21:661-677. doi 10.1038/s41580-020-00288-9

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

Received December 4, 2024. Revised July 15, 2025. Accepted July 28, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1169-1175

doi 10.18699/vjgb-25-123

Silencing of the *Nicotiana benthamiana* phytoendesaturase gene by root treatment of exogenous dsRNA

T.S. Golubeva[1, 2], V.A. Cherenko[1], E.A. Filipenko[1], I.V. Zhirnov[1], A.A. Ivanov[1, 3], A.V. Kochetov[1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Immanuel Kant Baltic Federal University, Kaliningrad, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
frolova@bionet.nsc.ru

Abstract. RNA interference (RNAi) is a powerful tool for gene silencing. It has recently been used to design promising plant protection strategies against pests such as viruses, insects, etc. This generally requires modifying the plant genome to achieve *in planta* synthesis of the double-stranded RNA (dsRNA), which guides the cellular RNA interference machinery to silence the genes of interest. However, given Russian legislation, the approach in which dsRNA is synthesized by the plant itself remains unavailable for crop protection. The use of exogenously produced dsRNA appears to be a promising alternative, allowing researchers to avoid genetic modification of plants, making it possible to implement potential results in agriculture. Furthermore, exogenous dsRNAs are superior to chemical pesticides (fungicides, insecticides, etc.), which are widely used to control various plant diseases. The dsRNA acts through sequence-specific nucleic acid interactions, making it extremely selective and unlikely to harm off-target organisms. Thus, it seems promising to utilize RNAi technology for agricultural plant protection. In this case, questions arise regarding how to produce the required amounts of pathogen-specific exogenous dsRNA, and which delivery method will be optimal for providing sufficient protection. This work aims to utilize exogenous dsRNA to silence the *Nicotiana benthamiana* phytoene desaturase gene. Phytoene desaturase is a convenient model gene in gene silencing experiments, as its knockdown results in a distinct phenotypic manifestation, namely, leaf bleaching. The dsRNA synthesis for this work was performed *in vivo* in *Escherichia coli* cells, and the chosen delivery method was root treatment through watering, both techniques being as simple and accessible as possible. It is surmised that the proposed approach could be adapted for broader use of RNAi technologies in agricultural crop protection.

Key words: RNA interference; gene silencing; phytoene desaturase; exogenous dsRNA; *Nicotiana benthamiana*; root treatment

For citation: Golubeva T.S., Cherenko V.A., Filipenko E.A., Zhirnov I.V., Ivanov A.A., Kochetov A.V. Silencing of the *Nicotiana benthamiana* phytoendesaturase gene by root treatment of exogenous dsRNA. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1169-1175. doi 10.18699/vjgb-25-123

Funding. The research was carried out within the state assignment of the Ministry of Science and Higher Education of the Russian Federation FWNR-2022-0017.

Сайленсинг гена фитоендесатуразы табака Бентхама *Nicotiana benthamiana* с помощью корневой обработки экзогенной дцРНК

Т.С. Голубева[1, 2], В.А. Черенко[1], Е.А. Филипенко[1], И.В. Жирнов[1], А.А. Иванов[1, 3], А.В. Кочетов[1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Балтийский федеральный университет им. И. Канта, Калининград, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
frolova@bionet.nsc.ru

Аннотация. РНК-интерференция – мощный инструмент для генного сайленсинга, благодаря чему используется при разработке новых подходов с большим потенциалом для защиты растений от вирусов, насекомых и других патогенов. Как правило, в таких системах геном растений подвергается модификациям с целью синтеза двуцепочечных РНК (дцРНК), необходимых для РНК-интерференции и последующего сайленсинга непосредственно в растительных клетках. Однако с учетом законодательства Российской Федерации такой подход не может использоваться на сельскохозяйственных растениях, что делает невозможным его применение при условии синтеза дцРНК самим растением. Применение экзогенно синтезированной дцРНК может стать перспективным способом защиты растений, так как позволяет избежать создания генетически модифицированных организмов и внедрить полученную разработку в сельском хозяйстве. Также экзогенные дцРНК имеют

преимущество по сравнению с химикатами (фунгицидами, инсектицидами и т. д.), используемыми для защиты растений, так как дцРНК действуют посредством своей специфической нуклеотидной последовательности, что делает описанный подход крайне избирательным к патогену и безопасным для других организмов. В совокупности вышеперечисленные факторы делают методы РНК-интерференции весьма перспективными для применения в сельском хозяйстве с целью защиты растений, поэтому встает вопрос о крупномасштабном синтезе экзогенных молекул дцРНК, специфичных к определенному патогену, и выборе оптимального способа их доставки для достижения защитного эффекта. Целью настоящей работы является сайленсинг гена фитоендесатуразы табака Бентхама (*Nicotiana benthamiana*) с применением экзогенно синтезированной дцРНК. Ген фитоендесатуразы – очень удобная модель в экспериментах по регуляции генной активности, так как его сайленсинг сопровождается ярким фенотипическим проявлением в виде побеления листьев. Синтез дцРНК осуществляли *in vivo* в клетках *Escherichia coli*; в качестве способа доставки выбрана корневая обработка через полив растения – максимально простые и доступные манипуляции. Предполагается, что предложенный подход может быть масштабирован и адаптирован для защиты растений в сельском хозяйстве с помощью методов, в основе которых лежит РНК-интерференция.

Ключевые слова: РНК-интерференция; генный сайленсинг; фитоендесатураза; экзогенная дцРНК; *Nicotiana benthamiana*; корневая обработка

Introduction

In order to protect plants against pathogens and pests, agriculture relies on the widespread use of chemical pesticides that are applied to the environment in large amounts. These intense applications of chemical pesticides pose potential risks for human health, beneficial organisms, and the environment (Niehl et al., 2018). Therefore, it is imperative that new alternative methods of controlling plant diseases be developed. Thus, there is a need for novel tools and alternative methods to control disease epidemics. A promising new approach with strong potential for protecting plants against viruses and other pathogens involves the application of dsRNA.

dsRNA applications can be highly advantageous over chemical compounds. Whereas chemical compounds act by a structure-dependent mechanism, dsRNAs act through their specific nucleotide sequence. Hence, once engineered to affect a specific pathogen target with a homologous sequence, dsRNAs and small interfering RNA (siRNA) derivatives should act only against the targeted pathogen. It is worth noting that, unlike chemical pesticides, dsRNA agents are biocompatible and biodegradable compounds, natural and universally found inside and outside organisms as well as in food. In this way, applying dsRNA proves to be a much more flexible and environmentally friendly approach.

However, the wide application of dsRNA treatment in agriculture is hampered by the lack of efficient and cost-effective methods to synthesize large quantities of dsRNA. The main approach to obtain dsRNA has been the physical annealing of two enzymatically synthesized single-stranded RNAs (ssRNA) *in vitro* (Laurila et al., 2002). Also, the efficiency of exogenously administered dsRNA in plants can be influenced by several factors: dsRNA concentration/dose and length/size, method of application, method of delivery, plant organ-specific activity, and stability under unfavorable environmental conditions. It is these factors that eventually determine the rate of uptake of exogenous dsRNA by plant cells for RNAi triggering. Given all these factors, it is necessary to develop an efficient method for the large-scale synthesis of dsRNA molecules and to choose the best method for their delivery (Carthew, Sontheimer, 2009).

In this work, we suggest a method for regulating the activity of the phytoene desaturase gene in tobacco (*Nicotiana benthamiana*) using dsRNA synthesized in *Escherichia coli* cells: root treatment of plants with crude bacterial lysate containing the target dsRNA led to photobleaching of plant leaves by RNA interference (RNAi).

Phytoene desaturase is a key enzyme in chlorophyll synthesis, with its silencing known to result in the phenotypic manifestation of leaf photobleaching. Thus, this gene is often used as a model for developing new approaches to regulate gene activity. Root treatment with crude lysate allows scaling up dsRNA production and minimizing delivery time and resources. The approach developed is expected to be used to protect agricultural plants by regulating gene activity by RNAi.

Materials and methods

Characterization of the bacterial strain. This study used the *E. coli* HT115 (DE3) strain [F-, mcrA, mcrB, IN(rrnD-rrnE)1, rnc14::Tn10(DE3 lysogen: lacUV5 promoter – T7 polymerase) (IPTG-inducible T7 polymerase) (RNAse III minus)]. This strain is deficient in RNase III and therefore can be used to generate dsRNA. The strain was cultured in standard Luria-Bertani (LB) liquid media or LB agar with tetracycline (12.5 μg/mL).

Characterization of the L4440 plasmid. Plasmid L4440 (Plasmid #1654, Addgene, USA) was used in this study. The specific feature of this plasmid (i. e. the presence of two strong T7 promoters in opposite directions) makes it suitable for the production of dsRNA fragments. Also, the plasmid carries an ampicillin resistance gene.

Selection of primers. Primers were selected using the Primer-BLAST resource (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi) on the *N. benthamiana* mRNA matrix. The selected primer variants were tested for the presence of hairpins, self- and heterodimers using the OligoAnalyzer tool (https://eu.idtdna.com/calc/analyzer). As a result, the following primers were synthesized: forward 5′-GGCACTC AACTTTATAAACC-3′ and reverse 5′-CTTCAGTTTTCT GTCAAACCATATATGGAC-3′ (Syntol, Russia).

Production of cDNA. mRNA was isolated using the RNeasy Plant kit (Qiagen, USA) according to the manufacturer's protocol, with the RNA sample obtained analyzed on a Bioanalyzer Instrument 2100 RNA analyzer (SB RAS Genom-

ics Core Facility, ICBFM SB RAS). Reverse transcription was performed with the BiolabMix R01-250 kit (BiolabMix, Russia). The HS-qPCR SYBR Blue kit (BiolabMix, Russia) was used to generate fragments, and amplification was performed in a Bio-Rad IQ instrument (Bio-Rad, USA).

Cloning of the *pds* gene fragment. The *pds* gene was cloned in two steps. Initially, the gene was cloned into the pCR2.1 intermediate T-vector (Invitrogen, USA). Bacterial colonies containing the gene fragment were first selected using the blue-white screening. The final selection of clones was carried out based on the presence of the complete gene fragment in the restriction spectra after treatment with *Eco*RI endonuclease.

Cloning into the final vector L4440 was performed using the ligase-restriction method with the restriction endonucleases *Pst*I and *Nco*I. For this purpose, the corresponding restriction sites were introduced into the *pds* gene fragment during its amplification (the same sites are present in the structure of the L4440 vector). Recombinant *E. coli* clones were selected on a selective medium containing ampicillin. To verify the presence of the *pds* gene fragment in the L4440 vector, PCR was performed using primers specific for the *pds* gene. The final presence of the *pds* gene fragment and its nucleotide sequence were confirmed by Sanger sequencing.

Transformation of the *E. coli* strain HT115 (DE3) with plasmid L4440. Competent cells were prepared according to the Nishimura protocol (Nishimura et al., 1990). For transformation, a 0.1 ml aliquot of bacterial suspension was mixed with 5 µl of L4440 plasmid dissolved in TE buffer (100 pg) and incubated on ice for 30 min. Next, the cells were placed in an incubator and maintained at 42 °C for 60 s, then incubated again on ice for 2 min and diluted 10-fold with pre-warmed to 37 °C LB broth. The suspension was incubated at 37 °C for 1 h to develop ampicillin resistance. Then 100 µl of the suspension was rubbed into LB agar with ampicillin (50 µg/ml), and the cups were incubated at 37 °C for 24 h. The grown colonies were formed by transformed bacteria.

The presence of an integrated fragment of the *pds* gene was verified by PCR of selected colonies as well as by Sanger sequencing (SB RAS Genomics Core Facility, ICBFM SB RAS).

VIGS positive control. A model plant, *N. benthamiana*, was used in VIGS experiments. Seeds of *N. benthamiana* were germinated and then the seedlings were planted in plastic pots (10 cm in diameter) containing a mixture of universal potting soil (TerraVita, Russia), perlite and vermiculite (8:1:1, v/v), and cultivated in a growth chamber under continuous lighting at 24 °C.

Vector construction. The VIGS vectors pTRV1 (pYL192) and pTRV2 (pYL279) obtained from the Arabidopsis Biological Resource Center (ABRC, USA) were described previously (Liu et al., 2002; Burch-Smith et al., 2006). Total RNA was extracted from leaf tissues of *Nicotiana tabacum* cv. SR1 (the phytoendesaturase gene sequence in this species is identical to that in *N. benthamiana*) using Trizol reagent (Invitrogen, USA), and first-strand cDNA was synthesized using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, USA) according to the manufacturer's instructions. The cDNA pool was used to amplify the *NtPDS* gene fragment sequence (GenBank accession number: AJ616742.1) by PCR using high-fidelity Phusion polymerase (New England Biolabs, USA) as per the manufacturer's instruction. The selected primers were checked for hairpins, self-dimers, and heterodimers using the OligoAnalyzer tool. As a result, the following primers were synthesized: forward 5′-CACCGGCACTC AACTTTATAAACC-3′ and reverse 5′-CTTCAGTTTTCT GTCAAACCATATATGGAC-3′ (Syntol, Russia).

The resulting 413-bp PCR fragment was cloned into the pENTR/D-TOPO vector (Invitrogen, USA), verified by sequencing, and then recombined into the pTRV2 vector by carrying out an LR recombination reaction using the Gateway system (Invitrogen). The generated vector pTRV2::NtPDS was transformed by heat shock into the *Agrobacterium tumefaciens* GV2260 strain and used in VIGS experiments.

The *A. tumefaciens* GV2260 strain, carrying the pTRV1, pTRV2, and pTRV2::NtPDS vectors, was separately inoculated into Luria-Bertani liquid media containing kanamycin (100 µg/mL) and rifampicin (25 µg/mL). The cultures were incubated overnight with shaking at 28 °C. The cells were harvested from the overnight cultures, re-suspended in the induction buffer (10 mM MES; 10 mM $MgCl_2$; 250 µM acetosyringone; adjusted to pH 5.5 with 1 M KOH), and incubated for 6 h at room temperature in a shaker. After incubation, the cells were harvested from the induced cultures and resuspended in the infiltration buffer (10 mM MES, adjusted to pH 5.5 with 1 M KOH) with dilution to a final absorbance OD_{600} = 1.0.

For leaf infiltration, *A. tumefaciens* strain GV2260 cultures containing pTRV1 and pTRV2 or pTRV2::NtPDS were mixed in a 1:1 (v/v) ratio and infiltrated into lower leaves of 21-days-old *N. benthamiana* plants using a 1-ml needleless syringe (Ratcliff et al., 2001; Liu et al., 2002). Infiltrated plants were maintained under constant light for 12 h at a 20 °C temperature for effective *Agrobacterium* T-DNA insertion (Brigneti et al., 2004).

Root treatment of *N. benthamiana* with a crude lysate of *E. coli* HT115 (DE3) containing dsRNA of phytoendesaturase. After inducing dsRNA synthesis, crude lysates of bacterial suspensions can be used as a source of exogenous dsRNA. Root treatments are also used to deliver target molecules for PTGS in plants (Jiang et al., 2014; Li et al., 2015; Dubrovina, Kiselev, 2019).

Crude lysates of the *E. coli* HT115 strain were prepared for root treatment, transformed with the L4440 plasmid with an integrated fragment of the *pds* gene. Bacterial lysates without plasmid, with incorporation of the native plasmid L4440, buffer, and water were used as controls.

The crude lysate was prepared in the following manner (Gan et al., 2010). An overnight culture of the bacteria was grown for 16 h with shaking at 37 °C in standard LB broth, with ampicillin at a concentration of 50 µg/mL added for strains with the L4440 plasmid insertion. The bacterial suspension was then diluted to OD_{595} = 0.5. Next, dsRNA synthesis was induced by adding IPTG at a final concentration of 0.6 mM, and the bacteria were incubated at 37 °C for 4 h. After the required time, the suspension was centrifuged to obtain a precipitate (1,500g, 15 min, 4 °C), which was then resuspended in ice buffer (50 mM Tris·HCl, 10 mM EDTA, pH 7.5) taken in 1/50 of the original volume. The tubes were then placed in ice and treated with ultrasound (20 kHz, 15 min). The resul-

ting lysate was centrifuged (9,000 rpm, 20 min, 4 °C), and the supernatant was used for plant treatments (2 ml per plant).

Four plants were used for each type of root treatment: water, Tris·HCl/EDTA buffer, *E. coli* HT115 without plasmid, *E. coli* HT115 with the L4440 plasmid, *E. coli* HT115 with the L4440 plasmid with an integrated *pds* fragment. All plants were grown in a sterile mixture of universal soil (TerraVita, Russia), perlite and vermiculite (8:1:1, v/v). The plants were treated three times a week (Monday, Wednesday, Friday) for a period of four weeks. Root treatment was carried out by watering the plants with supernatant resuspended in 2 ml of Tris-EDTA buffer.

Extraction of dsRNA. 1 ml of the bacterial suspension was centrifuged for 5 min at 10,000 rcf, and the supernatant was removed. The precipitate was dissolved in 200 µl of a mixture of 1 M CH_3COONH_4 and 10 mM isoamyl alcohol and extracted with an equivalent volume of isopropanol:phenol:isoamyl alcohol in the ratio (25:24:1). Then it was incubated for 15 min at 65 °C and centrifuged for 15 min at 10,000 rcf. The supernatant was withdrawn and an equivalent volume of isopropanol was added, followed by incubation for 12 hours at –20 °C. After incubation, it was centrifuged for 30 min at 12,000 rcf. The liquid was removed carefully. The precipitate was washed twice with 70 % ethanol, with the precipitate resuspended in 10 µl of RNase-free water, and DNAase buffer was added. Next, the incubation was done for 30 min, followed by the addition of 20 µl of RNase A and 20 µl of DNase. Then, the incubation was performed for 1 hour at 37 °C. 100 µl of a isopropanol:phenol:isoamyl alcohol mixture in the ratio (25:24:1) was extracted. The mixture was stirred vigorously and centrifuged at 12,000 rcf for 15 min. The supernatant was removed, and the precipitate was washed with 200 µl of 70 % ethanol, dried at room temperature, and dissolved with 1× TE buffer. The presence and integrity of dsRNA were checked by gel electrophoresis in 1 % agarose gel.

Results

Production of exogenous dsRNA using an RNase-deficient *E. coli* HT115 strain transformed with the L4440 plasmid

The *E. coli* HT115 (DE3) strain was used to produce dsRNA because it is deficient in RNase III, an enzyme that normally hydrolyzes dsRNA in the bacterial cell. The L4440 vector was used for the transformation. Due to having two strong T7 promoters in opposite directions, this vector can be used to produce target dsRNA fragments.

A 404 bp fragment of the tobacco phytoene desaturase *pds* gene was cloned into the L4440 vector by sticky ends using restriction endonucleases *Pst*I and *Nco*I, with the presence of the required fragment in the vector confirmed by sequencing. A map of the resulting plasmid is shown in Figure 1.

The presence of the target fragments of the correct length was confirmed by electrophoretic analysis (Fig. 2).

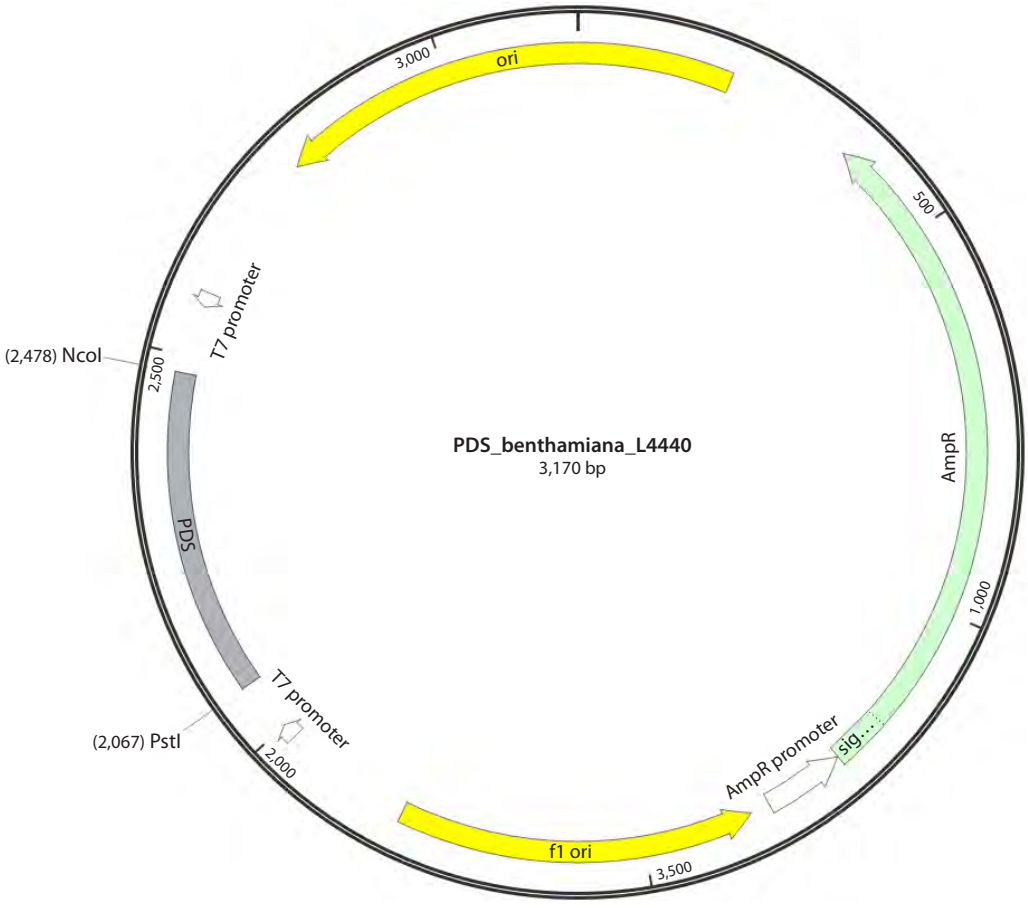


Fig. 1. Map of vector L4440 containing a fragment of the *pds* gene of *N. benthamiana* (PDS_benthamiana_L4440).

Preparation of a crude bacterial lysate containing dsRNA fragments for silencing phytoene desaturase *N. benthamiana*

A strain in *E. coli* HT115 (DE3) transfected with the PDS_benthamiana_L4440 vector was used to produce target dsRNA fragments for silencing *pds* genes. Induction of dsRNA synthesis was performed using IPTG. Next, a crude lysate was obtained from the bacterial suspension (Gan et al., 2010) containing dsRNA. 2-ml aliquots of the crude lysate were prepared for treatment of *N. benthamiana* plants.

Root treatment of *N. benthamiana* with a crude bacterial lysate containing dsRNA for silencing the pds gene

In this study, root treatment of tobacco with a crude bacterial lysate containing exogenous dsRNA was performed for the first time. Bentham tobacco (*N. benthamiana*) was used as a model plant for the experiment. Experimental plants were treated with the *E. coli* HT115 (DE3) crude lysate with the PDS_benthamiana_L4440 plasmid. After four weeks of root treatment, *N. benthamiana* leaves treated with the bacterial lysate with insertion of a phytoene desaturase gene fragment showed photobleaching phenotypes of young leaves typical for *pds* gene silencing (Fig. 3). The leaves of *N. benthamiana* plants from the negative control did not change their phenotype throughout the entire treatment period.

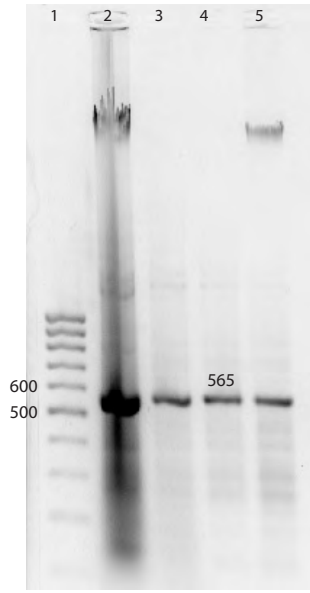


Fig. 2. Electrophoretic analysis of dsRNA fragments synthesized in *E. coli* HT115 (DE3), transformed with the PDS_benthamiana_L4440 vector.

1 – 100 bp DNA marker;
2 – total nucleic acid fraction;
3 – total nucleic acid fraction treated with DNase;
4 – total nucleic acid fraction treated with RNase A and DNase;
5 – total nucleic acid fraction treated with RNase A.

Discussion

Currently, there are numerous studies indicating the possibility of turning off or reducing the expression of certain genes to regulate resistance, growth, and other properties of plants by inducing RNA interference (Kamthan et al., 2015; Tiwari et al., 2017). Yet this approach involves the stage of obtaining a transgenic plant or using constructs based on attenuated plant viruses.

The problem of RNA delivery without using vector systems and modifying the genome became particularly acute after GMOs had been banned in Russia and European countries. Recently, reports have appeared in the scientific literature indicating that exogenously applied double-stranded RNA molecules (e. g. by spraying, spraying under high pressure, using RNA molecule adhesion materials, or using transfer proteins) are capable of penetrating into the plant vascular system and directly into plant cells and then inducing RNA interference, thereby increasing plant resistance to fungal and viral infections (Numata et al., 2014; Koch et al., 2016; Mitter et al., 2017; Wang et al., 2018). There are also few studies on dsRNA delivery aimed at regulating gene function by irrigation. For example, maize resistance to sugarcane mosaic virus (SCMV) infection was enhanced by this delivery method (Gan et al., 2010).

Similarly, we obtained a crude lysate containing a dsRNA fragment of the *N. benthamiana* phytoene desaturase gene, and carried out root treatment of tobacco for four weeks. As a result, all experimental plants exhibited the phenotype of photobleaching of young leaves, which is not uncommon for phytoene desaturase silencing. The degree of whitening was comparable to the positive VIGS control. Thus, we suggest the approach that allows regulating the activity of plant genes without creating GMOs and using only environmentally friendly methods. We believe this approach to be highly promising for implementation in agriculture in order to improve the stress tolerance of cultivated plants.

Our work is based on research by F. Tenllado and co-authors (2003), where tobacco (*N. benthamiana*) was successfully protected from pepper mild mottle virus (PMMoV) by spraying the above-ground parts of the plants with a crude lysate of *E. coli* HT115 bacteria containing dsRNA. The authors note that this method of obtaining target dsRNA molecules is quite simple and economically advantageous compared to *in vitro* dsRNA synthesis. The authors also show that the use of crude lysate may be more cost-effective than the use of a purified preparation, with no significant loss of efficacy observed.

Since plant protection using RNAi methods can be an alternative to the creation of GMOs, which are prohibited by law in the Russian Federation for agricultural purposes, we decided to develop the idea of F. Tenllado and co-authors (2003) and try an even simpler method of delivering dsRNA through root treatment by watering plants with crude lysate. We have shown that the introduction of exogenously synthesized dsRNA in this way can effectively influence the phenotype of plants, confirming that the introduced dsRNA enters the plants through the roots and is transported to the above-ground part of the plant, where the activity of the target *pds* gene is regulated.

Conclusion

Recently, methods based on RNAi have been actively developed for plant protection, especially with the use of exogenously synthesised dsRNA, as they allow avoiding the creation

Fig. 3. Photobleaching of *N. benthamiana* leaves after induced silencing of the *pds* gene using exogenous dsRNA.

1 – water; *2* – Tris-EDTA; *3* – *E. coli* HT115 crude lysate without plasmid; *4* – *E. coli* HT115 crude lysate with the native L4440 plasmid; *5* – VIGS; *6* – crude lysate of *E. coli* HT115 with the L4440 plasmid containing a fragment of the *pds* gene; *7* – comparison of a negative control with an experimental plant with *pds* silencing. Controls in the experiment: treatments *1–4* serve as negative controls, treatment *5* serves as a positive control. Characteristics of leaf photobleaching: L – lightness of color in relative units (the higher the value, the lighter the shade, from 0 to 100); a – green-red channel in relative units (the lower the value, the greener the shade, from –127 to 0).

of GMOs, which are prohibited by law for use in agriculture in a number of countries, including Russia. However, the use of exogenous dsRNA molecules has a number of complications. For example, the method of obtaining dsRNA must be easily scalable and cost-effective, and the delivery of the resulting molecules must be as simple and efficient as possible. In our work on dsRNA synthesis, we developed a system from the *E. coli* HT115 strain, transformed with the L4440 plasmid with a fragment of the Bentham tobacco phytoendesaturase gene (*N. benthamiana*), and for the delivery of the resulting molecules, we used root treatment – watering the plants with a coarse lysate. As a result, we were able to achieve silencing of the phytoendesaturase gene,which confirms the possibility of regulating the work of plant genes without creating GMOs. The approach we propose can also be scaled up and has potential for application in agriculture to protect plants from pathogens.

References

Brigneti G., Martín-Hernández A.M., Jin H., Chen J., Baulcombe D.C., Baker B., Jones J.D.G. Virus-induced gene silencing in *Solanum* species. *Plant J.* 2004;39(2):264-272. doi 10.1111/J.1365-313X.2004.02122.x

Burch-Smith T.M., Schiff M., Liu Y., Dinesh-Kumar S.P. Efficient virus-induced gene silencing in Arabidopsis. *Plant Physiol.* 2006;142(1):21-27. doi 10.1104/pp.106.084624

Carthew R.W., Sontheimer E.J. Origins and mechanisms of miRNAs and siRNAs. *Cell.* 2009;136(4):642-655. doi 10.1016/j.cell.2009.01.035

Dubrovina A.S., Kiselev K.V. Exogenous RNAs for gene regulation and plant resistance. *Int J Mol Sci.* 2019;20(9):2282. doi 10.3390/ijms20092282

Gan D., Zhang J., Jiang H., Jiang T., Zhu S., Cheng B. Bacterially expressed dsRNA protects maize against SCMV infection. *Plant Cell Rep.* 2010;29(11):1261-1268. doi 10.1007/s00299-010-0911-z

Jiang L., Ding L., He B., Shen J., Xu Z., Yin M., Zhang X. Systemic gene silencing in plants triggered by fluorescent nanoparticle-

delivered double-stranded RNA. *Nanoscale*. 2014;6(17):9965-9969. doi 10.1039/c4nr03481c

Kamthan A., Chaudhuri A., Kamthan M., Datta A. Small RNAs in plants: recent development and application for crop improvement. *Front Plant Sci.* 2015;6(APR). doi 10.3389/FPLS.2015.00208/PDF

Koch A., Biedenkopf D., Furch A., Weber L., Rossbach O., Abdellatef E., Linicus L., Johannsmeier J., Jelonek L., Goesmann A., Cardoza V., McMillan J., Mentzel T., Kogel K.H. An RNAi-based control of *Fusarium graminearum* infections through spraying of long dsRNAs involves a plant passage and is controlled by the fungal silencing machinery. *PLoS Pathog.* 2016;12(10):e1005901. doi 10.1371/journal.ppat.1005901

Laurila M.R.L., Makeyev E.V., Bamford D.H. Bacteriophage φ6 RNA-dependent RNA polymerase. Molecular details of initiating nucleic acid synthesis without primer. *J Biol Chem.* 2002;277(19):17117-17124. doi 10.1074/jbc.M111220200

Li H., Guan R., Guo H., Miao X. New insights into an RNAi approach for plant defence against piercing-sucking and stem-borer insect pests. *Plant Cell Environ.* 2015;38(11):2277-2285. doi 10.1111/pce.12546

Liu Y., Schiff M., Marathe R., Dinesh-Kumar S.P. Tobacco *Rar1*, *EDS1* and *NPR1/NIM1* like genes are required for *N*-mediated resistance to tobacco mosaic virus. *Plant J.* 2002;30(4):415-429. doi 10.1046/J.1365-313x.2002.01297.x

Mitter N., Worrall E.A., Robinson K.E., Li P., Jain R.G., Taochy C., Fletcher S.J., Carroll B.J., Lu G.Q., Xu Z.P. Clay nanosheets for topical delivery of RNAi for sustained protection against plant viruses. *Nat Plants*. 2017;3:16207. doi 10.1038/nplants.2016.207

Niehl A., Soininen M., Poranen M.M., Heinlein M. Synthetic biology approach for plant protection using dsRNA. *Plant Biotechnol J.* 2018;16(9):1679-1687. doi 10.1111/pbi.12904

Nishimura A., Morita M., Nishimura Y., Sugino Y. A rapid and highly efficient method for preparation of competent *Escherichia coli* cells. *Nucleic Acids Res.* 1990;18(20):6169. doi 10.1093/nar/18.20.6169

Numata K., Ohtani M., Yoshizumi T., Demura T., Kodama Y. Local gene silencing in plants via synthetic dsRNA and carrier peptide. *Plant Biotechnol J.* 2014;12(8):1027-1034. doi 10.1111/pbi.12208

Ratcliff F., Martin-Hernandez A.M., Baulcombe D.C. Tobacco rattle virus as a vector for analysis of gene function by silencing. *Plant J.* 2001;25(2):237-245. doi 10.1046/J.0960-7412.2000.00942.x

Tenllado F., Martínez-García B., Vargas M., Díaz-Ruíz J.R. Crude extracts of bacterially expressed dsRNA can be used to protect plants against virus infections. *BMC Biotechnol*. 2003;3:3. doi 10.1186/1472-6750-3-3

Tiwari I.M., Jesuraj A., Kamboj R., Devanna B.N., Botella J.R., Sharma T.R. Host Delivered RNAi, an efficient approach to increase rice resistance to sheath blight pathogen (*Rhizoctonia solani*). *Sci Rep.* 2017;7(1):1-14. doi 10.1038/S41598-017-07749

Wang J., Gu L., Knipple D.C. Evaluation of some potential target genes and methods for RNAi-mediated pest control of the corn earworm *Helicoverpa zea*. *Pestic Biochem Physiol.* 2018;149:67-72. doi 10.1016/j.pestbp.2018.05.012

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

Received June 15, 2025. Revised September 19, 2025. Accepted September 23, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1176-1183

doi 10.18699/vjgb-25-124

Effect of RNAi-mediated silencing of the *TaAOS2* gene on phytohormone accumulation, growth and productivity in bread wheat

D.N. Miroshnichenko[1, 2] ✉, A.V. Pigolev[1], E.A. Degtyaryov[1], V.I. Degtyaryova[1, 2, 3], V.V. Alekseeva[2], A.S. Pushin[2], S.V. Dolgov[2], T.V. Savchenko[1]

[1] Institute of Basic Biological Problems of the Russian Academy of Sciences, Pushchino, Moscow region, Russia
[2] Branch of Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Pushchino, Moscow region, Russia
[3] Pushchino Branch of Russian Biotechnological University, Pushchino, Russia
✉ miroshnichenko@bibch.ru

Abstract. Reverse genetics methods are actively used in plant biology to study the functions of specific genes responsible for the adaptation of plants to various environmental stresses. The present study describes the production and primary characterization of transgenic bread wheat with silenced expression of allen oxide synthase (AOS). AOS is a key enzyme involved in the initial step of biosynthesis of stress-related phytohormones known as jasmonates. To induce silencing of *AOS* in wheat, we designed the RNA interference (RNAi) vector containing an inverted repeat region of the *TaAOS2* gene cloned from genome DNA of cv. Chinese Spring. With the help of biolistic-mediated transformation, a number of transgenic Chinese Spring plants have been produced. Real-Time PCR analysis confirmed the suppression of target gene expression, since transgenic dsRNAi lines accumulated only 21–44 % mRNA of *TaAOS2* after leaf wounding compared to the wound-induced level in non-transgenic control. Gas chromatography–mass spectrometry revealed that the silencing of *TaAOS2* substantially reduced the accumulation of jasmonic acid (JA) and jasmonoyl-isoleucine conjugate (JA-Ile), while the production of other phytohormones, such as abscisic acid and salicylic acid, was not affected. *TaAOS2*-silenced lines were characterized by shorter leaves at the juvenile stage, demonstrated a tendency towards reduced plant height and decreased grain weight, while the average flowering time and plant fertility (number of seeds per spike) were not affected. The obtained transgenic lines in combination with *AOS*-overexpressing lines can be used for further detailed analysis of the adaptive responses controlled by the jasmonate hormonal system.

Key words: allene oxide synthase; jasmonates; *Triticum aestivum* L.; RNA interference; wounding

For citation: Miroshnichenko D.N., Pigolev A.V., Degtyaryov E.A., Degtyaryova V.I., Alekseeva V.V., Pushin A.S., Dolgov S.V., Savchenko T.V. Effect of RNAi-mediated silencing of the *TaAOS2* gene on phytohormone accumulation, growth and productivity in bread wheat. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8): 1176-1183. doi 10.18699/vjgb-25-124

Funding. This research was funded by the Russian Science Foundation, grant No. 22-16-00047-P.

Замалчивание гена *TaAOS2* с помощью РНК-интерференции влияет на накопление фитогормонов, рост и продуктивность мягкой пшеницы

Д.Н. Мирошниченко[1, 2] ✉, А.В. Пиголев[1], Е.А. Дегтярёв[1], В.И. Дегтярёва[1, 2, 3], В.В. Алексеева[2], А.С. Пушин[2], С.В. Долгов[2], Т.В. Савченко[1]

[1] Институт фундаментальных проблем биологии Российской академии наук – обособленное подразделение Федерального исследовательского центра «Пущинский научный центр биологических исследований Российской академии наук», Пущино, Московская область, Россия
[2] Филиал Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук, Пущино, Московская область, Россия
[3] Пущинский филиал Российского биотехнологического университета (РОСБИОТЕХ), Пущино, Московская область, Россия
✉ miroshnichenko@bibch.ru

Аннотация. Методы обратной генетики активно используются в биологии растений для изучения функций генов, ответственных за адаптацию растений к различным стрессовым факторам окружающей среды. Данная работа посвящена получению и первичному анализу трансгенных растений мягкой пшеницы с подавленной экспрессией одного из генов алленоксидсинтазы (AOS). AOS является ключевым ферментом, участвующим в начальном этапе биосинтеза стрессовых фитогормонов, известных как жасмонаты. Для подавления экспрессии гена *AOS* в пшенице создан вектор для РНК-интерференционного замалчивания, содержащий инвертированную область *TaAOS2* из генома яровой мягкой пшеницы сорта Chinese Spring. Путем биобаллистической генетической трансформации получен ряд трансгенных растений Chinese Spring,

в которых подавление экспрессии целевого гена подтверждено с помощью количественной ПЦР в режиме реального времени, а изменение уровней накопления отдельных жасмонатов выявлено методом газохроматомасс-спектрометрии. В стрессовых условиях (механическое поранение листовой пластинки молодых растений) трансгенные dsRNAi линии накапливали лишь 21–44 % мРНК *TaAOS2* по сравнению с нетрансгенным пораненным контролем пшеницы. Подавление экспрессии гена *TaAOS2* существенно снизило накопление жасмоновой кислоты и конъюгата жасмоноил-изолейцина, в то время как уровень содержания в листьях других фитогормонов, таких как абсцизовая кислота и салициловая кислота, не изменился. Трансгенные линии пшеницы с подавленной экспрессией гена *TaAOS2* демонстрировали уменьшение длины листьев на ранних стадиях развития растений; также для них была характерна тенденция к сокращению высоты растений и снижению массы зерна. При этом замалчивание не привело к изменению сроков цветения и не повлияло на фертильность растений, так как среднее количество семян в колосе не изменилось. Полученные трансгенные линии пшеницы в сочетании с линиями, сверхэкспрессирующими *AOS*, служат хорошим инструментом для дальнейшего детального анализа адаптивных реакций растений, контролируемых жасмонатным сигналингом.
Ключевые слова: алленоксидсинтаза; жасмонаты; *Triticum aestivum* L.; РНК интерференция; поранение

Introduction

Allen oxide synthase (AOS) is one of the key enzymes in the biosynthesis pathway of jasmonic acid (JA) and its derivatives, lipid-derived plant phytohormones collectively called jasmonates. Jasmonates have various biological functions, including the regulation of developmental and aging processes, activation of the defence system under abiotic stresses, and triggering the immune response upon attacks by fungal pathogens and herbivores. In brief, the role of AOS in the JA biosynthesis pathway is the production of a short-lived intermediate that is converted to various non-volatile oxylipins in plant cells. AOS is a cytochrome P450 enzyme that is localized in chloroplasts and acts as a hydroperoxide dehydratase, converting fatty acid hydroperoxides into allene oxides. Allene oxide synthase is generally divided into three categories depending on the specific catalytic substrate: 13-AOS with activity to 13-hydroperoxy-9,11,15-octadecatrienoic acid (13-HPOT), 9-AOS with activity to 9-hydroperoxy-9,11,15-octadecatrienoic acid (9-HPOT), and 9/13-AOS with a mixed activity to both 9-HPOT and 13-HPOT (Jiang et al., 2009). Allene oxide formed by 13-AOS is rapidly cyclized with the help of allene oxide cyclase to a more stable product, 12-oxo-phytodienoic acid (OPDA), the biologically active precursor of JA.

Due to the involvement of AOS-derived oxylipins in signalling and coordination of diverse adaptive responses, AOS encoding genes have been the focus of researchers' attention for more than 30 years. To date, *AOS* genes have been identified in different dicot and monocot species; the number of *AOS* genes varies from a single copy in Arabidopsis (Laudert et al., 1996) to a family of 36 genes in the genome of sugarcane (*Saccharum officinarum*) (Sun et al., 2023). The *AOS* gene family is relatively conserved in structure, but expression pattern analysis demonstrated that different homologues were involved in various, sometimes distinct, defensive responses across different plant species. In Arabidopsis, the model species for the study of the functioning of the jasmonate system for the last decades, the expression of *AOS* is upregulated by a virulent fungal pathogen, mechanical stress, and the application of jasmonates and salicylic acid (SA) (Laudert, Weiler, 1998). The expression of two *AOS* genes of rice (*OsAOS1* and *OsAOS2*) is activated by mechanical wounding, blast fungus infection, herbivore infestation, and JA treatment (Zeng et al., 2021). In sugarcane, accumulation of *ShAOS1* transcripts was reported to increase due to treatments with methyl jasmonate, SA, and abscisic acid (ABA) (Sun et al., 2023).

At present, many *AOS* genes identified in plant species have been found to play an important role in mediating resistance to fungal pathogens and insects. The functional deficit of mRNA of AOS leads to a decrease in resistance to necrotrophic fungi in Arabidopsis (Chehab et al., 2011), potato (Pajerowska-Mukhtar et al., 2008), onion (Kim et al., 2025), and wheat (Fan et al., 2019). In line with this, the complementation of an Arabidopsis *aos* knockout mutant with overexpressed gingko *HvnAOS1* and *HvnAOS2* genes and potato *StAOS2* alleles allowed restoring the resistance to *Erwinia carotovora* and *Botrytis cinerea* (Pajerowska-Mukhtar et al., 2008; An et al., 2024). Constitutive expression of the *ShAOS1* gene, isolated from sugarcane, enhanced the resistance of transgenic *Nicotiana benthamiana* plants to *Fusarium solani* (Sun et al., 2023). Similarly, transgenic lines of tobacco overexpressing the onion *AOS* gene (*AcAOS*) demonstrated higher resistance to *Botrytis squamosa* (Kim et al., 2025). In rice, silencing of the *OsAOS1* or *OsAOS2* genes negatively affected resistance to the chewing herbivore striped stem borer (*Chilo suppressalis*), resulting in enhanced resistance to the piercing-sucking herbivore brown planthopper (*Nilaparvata lugens*) (Zeng et al., 2021). An Arabidopsis mutant lacking functional AOS demonstrated increased resistance to the nematode *Meloidogyne javanica* (Naor et al., 2018).

Changes in the functional activity of *AOS* genes can have pleiotropic effects on various plant characteristics. Transgenic guayule (*Parthenium argentatum*) with RNAi-mediated silencing of *AOS* showed higher rubber content and biomass (Placido et al., 2019), while the disruption of AOS activity in Arabidopsis caused male sterility (Park et al., 2002). Transgenic tobacco overexpressing the wheat *AOS* gene displayed improved tolerance to excessive zinc stress (Liu et al., 2014). Transgenic emmer wheat plants constitutively expressing the Arabidopsis *AtAOS* gene showed increased tolerance to osmotic stress (Pigolev et al., 2023), whereas overexpression of the native *TaAOS2* gene had a generally negative effect on plant growth and productivity in both emmer and bread wheat (Miroshnichenko et al., 2024a).

In bread wheat (*Triticum aestivum* L.), the most important cereal crop in temperate climate countries, the *AOS* family is represented by 12 genes. According to a recent genome-wide analysis, the AOS family consists of four groups of three homoeologous genes located in each sub-genome (A, B, and D) on chromosomes 2, 5, and 6 (Heckmann et al., 2024). The functioning of the twelve *AOS*-encoded proteins significantly

distinguishes the jasmonate system of wheat from that of the model species, Arabidopsis, operated by a single *AOS* gene. Our data indicate that under non-stress conditions, the total content of jasmonates (12-OPDA, JA, and JA-Ile) in wheat leaves is much lower compared to Arabidopsis (Laudert et al., 1996). Upon mechanical injury of leaves, the level of jasmonates increases in emmer and bread wheat up to 1.5–13 times (Miroshnichenko et al., 2024a, b), whereas in *A. thaliana*, the content of jasmonates increases tens of times (Kimberlin et al., 2022). In addition, it was reported that the expression of *AOS* in the leaves of bread wheat could also be induced in response to salts and various pathogenic fungi (Liu et al., 2014; Heckmann et al., 2024). Constitutive overexpression of *TaAOS2* resulted in pleiotropic effects in bread wheat, including reduced length of the first four leaves, shortened plant height, and reduced number of seeds collected per spike (Miroshnichenko et al., 2024a). A transgenic seedling of tetraploid emmer wheat overexpressing the *AtAOS* gene from Arabidopsis showed an increased length of roots and coleoptiles under osmotic stress (Pigolev et al., 2023).

The published data indicate that the knowledge of the functioning of the AOS branch of oxylipin biosynthesis discovered in various species cannot be directly applied to other species. This suggests that the modification of mRNA levels of *AOS*-encoding genes using reverse genetic techniques (overexpression, RNA interference/silencing, or genome editing) may help to uncover their functional role in specific plant species. To gain new insights into the functions of the *AOS*-encoding gene family in wheat, the present study aimed to study jasmonate signalling in the case of reduced expression of the *TaAOS2* gene, including the modification of hormonal status of transgenic lines and influence of *AOS* silencing on growth characteristics.

Materials and methods

Construction of the RNAi expression cassette. The expression cassette pBAR-GFP-AOSi was constructed using two identical fragments of the *TaAOS2* gene (TraesCS4A02G061800.1) of 333 bp, which were cloned in opposite orientation to each other and separated by a spacer representing the *GUS* gene. To avoid non-selective silencing of other genes belonging to the cytochrome P450 family proteins, especially CYP74 enzymes (AOS, hydroperoxide lyase (HPL), and divinyl ether synthase (DES)), we used the fragment outside of the C-domain encoding heme-binding site. The selected harpin arm sequence also did not contain DNA regions, including sequences identical to the *HPL* and *DES* genes longer than 12–20 nucleotides. The fragment corresponded to the central part of the protein sequence (from L93 to W203) and demonstrated high nucleotide homology between all *AOS* genes.

The RNAi construct containing the sequence 3′-TaAOS-5′-GUS-5′-TaAOS-3′ was cloned in two steps. The first RNAi fragment was obtained by PCR from a plasmid containing the full-length *AOS2* (TraesCS4A02G061800.1) using primers: 5′-rnaI-short-NotI-GATCGCGGCCGCCAGCTGCTCTTCTCCCTCCTCG and 3′-rnaI-EcoRV-GATCGATATCCACTTGGCGGCCTTGGTG. This fragment (335 bp) was then cloned into the Gateway pENTR1A dual plasmids at the NotI and EcoRV sites. The second element of the construct, containing *GUS* and the second RNAi sequence, was obtained by PCR on overlapping DNA templates. The *GUS* gene sequence was amplified using a pair of primers 5′RNAi-GATGACGGTATCGATAAGCTTGATATCTACCCGCTT and 3′GUS-(KpnI)-GATCGGTACCATTCGATCGAGTGAAGATCCCTTTCTTG. The second RNAi fragment was amplified with the 5′RNAI-(DraI)-GATCTTTAAATATCCACTTGGCGGCCTTGGTG and 3′RNAi-GATATCAAGCTTATCGATACCGTCATCCTCTTCACAGGCACCTACATGC primers. After amplification, the two resulting fragments of 933 bp (containing the *GUS* gene sequence) and 335 bp (containing the RNAi arm sequence) were combined by PCR with common primers (containing restriction sites for DraI and KpnI). After that, the combined GUS-RNAi sequence was cloned into the pENTR1A plasmid, containing the first RNAi fragment, at the restriction sites DraI and KpnI. The resulting construct of inverted *TaAOS2* repeats was placed under the control of a strong ubiquitin promoter from maize (*ZmUbi1*) and an octopine synthase transcription terminator by transfer into the vector pANIC5D (ampR, kanR, bar, pporRFP) using the LR clonase enzyme for Gateway cloning. The completed RNAi cassette (ocsT)-pANIC5D::PZmUbi1-RNAi(3′-TaAOS-5′-GUS-5′-TaAOS-3′)-Tocs was cut out from pANIC5D at the SacI–SnaBI sites and finally inserted at the SmaI restriction site into the pBAR-GFP vector for cereal genetic transformation.

Generation of transgenic plants. Transgenic plants of spring bread wheat cv. Chinese Spring were generated using the biolistic-mediated genetic transformation approach as described previously (Miroshnichenko et al., 2024a). After rounds of *in vitro* selection, the herbicide-resistant rooted plantlets were transferred to the greenhouse, and the resulting mature plants were then analysed by PCR for the introduction of the hairpin construct. Transgenic status of T0 plants was confirmed by the amplification of GFP-specific fragment of 606 bp using primers sGFPFor (5′-GCGACGTAAACGGCCACAAG) and sGFPRev (5′-CCAGCAGGACCATcTGTGATCG) as described previously (Pigolev et al., 2018). To verify the presence of left and right arms of the RNAi construct, primers TaAOSRi-1 (5′-ATGAACTCGAAGGAGGTGAAGTCGTTG-3′) and panicGUS-1 (5′-CTCTTCAGCGTAAGGGTAATGCGAGGTA-3′) generating a 500 bp product were designed. The integration of the right hairpin arm was confirmed by amplification of a 466 bp fragment using the TaAOSRi-1 (5′-ATGAACTCGAAGGAGGTGAAGTCGTTG-3′) and panicGUS-2 (5′-CTGCACTCAATGTACACCGACATGTG-3′) primers. T1 seed progeny resulting from self-pollinated primary T0 plants were screened for GFP fluorescence of pollen to identify homozygous transgenic T1 sub-lines stably inheriting the foreign insertion, as previously described (Pigolev et al., 2018). Sets of T3–T4 seeds of four discovered homozygous sub-lines were used for further analysis.

***TaAOS2* expression analysis.** Extraction of total RNA from wheat leaf tissue, subsequent cDNA synthesis, and quantitative real-time RT-PCR analysis were performed as described previously (Pigolev et al., 2023). To detect changes in the expression levels of the *TaAOS2* gene, the forward primer TaAOSshbF (5′-GGCCGGAGAGAAGTTCCAC-3′) and the reverse primer TaAOSshbR (5′-CTTCTCCAGCGCCTC

TATCG-3′) were used. Transcript levels were quantified with QuantStudio™ 5 Real-Time PCR Cycler (Thermo Fisher Scientific) using *TaWIN1* as a reference gene.

JA, JA-Ile, SA and ABA analysis. Intact and mechanically damaged leaves of non-transgenic plants and mechanically damaged leaves of four transgenic lines were used for phytohormone analysis. To induce abiotic stress, blades of fully opened 3rd leaves were wounded with forceps as described earlier (Pigolev et al., 2023). 30 min after mechanical injury, the leaves were collected and frozen in liquid nitrogen. The intact leaves were sampled in parallel with wounded tissues. For phytohormone analysis, the leaf tissues were ground in liquid nitrogen with a mortar and pestle. Extraction of hormones was performed according to the previously described procedure (Degtyaryov et al., 2023). Dihydrojasmonic acid (Merck KGaA, Darmstadt, Germany) and deuterated salicylic acid (Cayman Chemical, USA) were used as internal standards. Extracted samples were treated with trimethylsilyldiazomethane to produce methyl ester derivatives, which were analyzed on a gas chromatograph coupled with a mass spectrometer detector Chromatec-Crystal 5000 (Chromatec, Yoshkar-Ola, Russia) operating in electron ionization mode. One-microliter samples were injected by autosampler in splitless mode at 250 °C, and separated on a CR-5MS column (length 30 m, inner diameter 0.25 mm, film thickness 0.25 μm) with helium being used as a carrier gas (constant flow 0.7 ml/min). Oven temperature programming was as follows: hold at 40 °C for 1 min after injection, ramp at 15 °C/min to 150 °C, then increase at 10 °C/min to 250 °C, hold for 10 min. Mass spectral analysis was done in selective ion monitoring mode (SIM). The fragment ions of esterified forms of phytohormones monitored were as follows: jasmonic acid (JA) 224, jasmonoil-isoleucine conjugate (JA-Ile) 146, dihydro-JA 83, salicylic acid (SA) 152, abscisic acid (ABA) 190, and deuterated SA derivative SA-d4 156. The Chromatec Analytic 3 program was used for the data analysis.

Plant growth and productivity analysis. For the analysis, two homozygous transgenic lines were cultivated together with non-transgenic control in a greenhouse. Two plants were grown in one-liter pot using a 16 h day/8 h night regime. Light intensity was up to 200 $\mu mol \cdot m^{-2}s^{-1}$. The day temperature was 25 ± 2 °C; at night, the temperature was 20 ± 2 °C. A minimum of 15 pots per line were cultivated. The growth parameters analyzed were as follows: the length of the 1st, 2nd, 3rd, and 4th leaves (twenty leaves were measured), the height of plants (15 pots were analyzed), and the date of the anthesis (30 plants were analyzed). The average number of seeds per spike and the weight of 1,000 seeds were recorded at the end of cultivation. Statistically significant differences between control and transgenic lines were confirmed by Student's *t*-test.

Results

Generation of transgenic wheat lines with silenced *TaAOS2* expression

To produce transgenic plants, 630 morphogenic calli of wheat cv. Chinese Spring were transformed using a particle inflow gun with the pBAR-GFP-AOSi plasmid, carrying the hairpin loop sequence for silencing the *TaAOS2* gene (Fig. 1*a*). As a result of the transformation, 43 putative transgenic plantlets were regenerated using a dual selection approach (*GFP*+*BAR*) and adapted to *ex vitro* greenhouse conditions. The efficiency of transgenic plant production was 5.6 %, as PCR analysis detected amplification of a fragment of the *GFP* reporter gene in the DNA extracts of 34 primary independent wheat plants (Fig. 1*b*). The insertion of the complete RNAi sequence (3→5′-TaAOS-GUS-5′→3′-TaAOS) was confirmed for 33 independent T0 plants by amplification of both the left and the right arm fragments using specific primers (Fig. 1*c*, *d*). In order to get a homozygous transgenic population, T1 progenies from T0 plants were analysed for segregation of the introduced construct by detection of *GFP* expression in pollen plants and in T2 embryos. T1 progenies of several transgenic lines showed the segregation for *GFP* expression that fit the Mendelian 3:1 ratio for a single dominant locus. T3 seeds from homozygous T2 individuals of primary transgenic plants CSi3, CSi7, CSi9, and CSi19 were used for further analysis.

To identify RNAi-mediated silencing, we used wounded leaves of transgenic lines CSi3, CSi7, CSi9, and CSi19 and compared the *TaAOS2* expression in these samples with intact and damaged leaves of non-transgenic plants. Quantitative real-time PCR showed that mechanical damage to non-transgenic leaves increased the level of *TaAOS2* mRNA accumulation by about 40 times from 1.6 to 67.0 (Fig. 1*e*). In wounded transgenic lines, *TaAOS2* expression was 2.3–4.7 times lower compared to the damaged leaves of non-transgenic control, whereas the difference with intact non-transgenic control was not statistically confirmed.

Effect of *TaAOS2* silencing on the accumulation of JA, JA-Ile, ABA, and SA in leaves of transgenic wheat plants

The basal levels of jasmonic acid (JA) and jasmonoyl-isoleucine conjugate (JA-Ile) in the intact leaves of Chinese Spring were barely detectable by (GC)–MS chromatography and fluctuated around 10 pmol/g f.w. (Fig. 2). When the leaves of non-transgenic plants were subjected to mechanical wounding, the level of JA and JA-Ile accumulation increased to 64.4 and 40.3 pmol/g f.w., correspondingly. Endogenous concentrations of jasmonates in injured leaves of transgenic plants differed among the four analyzed transgenic lines and did not correlate with the level of *TaAOS2* expression. Two RNAi lines, CSi3 and SCi7, contained only half of the amount of JA-Ile compared to non-transgenic Chinese Spring plants subjected to wounding. Two other transgenic lines, CSi9 and SCi19, showed no difference in JA-Ile content. Measurement of the JA level revealed a similar pattern: JA production in wounded leaves of transgenic lines CSi9 and SCi19 (67.8–68.8 pmol/g f.w.) did not differ from the non-transgenic wounded plants; JA content in transgenic lines CSi3 and SCi7 decreased to 43–44 pmol/g f.w. ($p < 0.05$).

After wound treatment of non-transgenic wheat plants, a significant increase in production of salicylic acid (SA) and abscisic acid (ABA) was observed (Fig. 2): the level of SA increased around 2.5 times from 362.7 to 914.2 pmol/g f.w., wounding stimulated a 4-fold increase of ABA accumulation in leaves (2.9 vs. 12.6 pmol/g f.w.). Levels of both phytohormones (SA and ABA) were not significantly different in wounded transgenic lines and non-transgenic control.

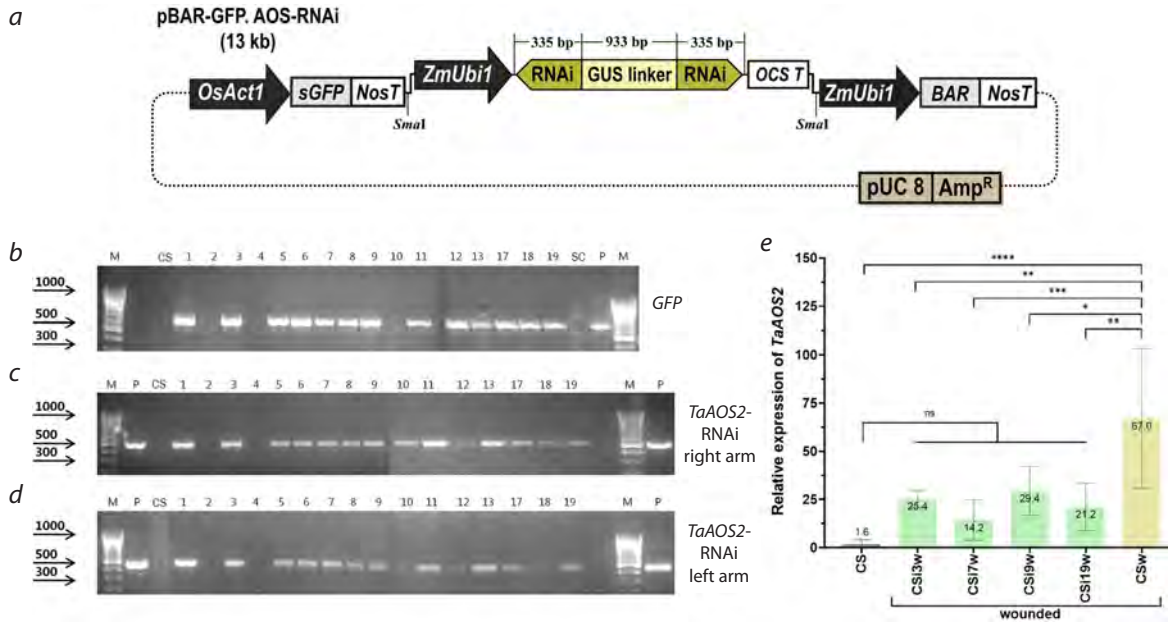


Fig. 1. Generation, selection, and analysis of transgenic wheat plants expressing a construct for silencing of the *TaAOS2* gene.

a, Schematic diagram of the pBAR-GFP-AOSi vector used for the transformation of bread wheat cv. Chinese Spring designed for the induction of RNAi-mediated silencing of the *TaAOS2* gene. *b–d*, Example of PCR analysis of primary putative transgenic plants for the insertion of the heterologous sequences. *b*, Amplification of the *GFP* gene fragment; *c*, *d*, amplification of the right (*c*) and left (*d*) arms of the hairpin sequence for *TaAOS2* silencing; lane M, DNA ladder as a molecular weight marker; lane P, DNA of pBAR-GFP-AOSi; lane CS, non-transgenic wheat plant Chinese Spring; lanes labelled 1–19 represent putative transgenic wheat plants established in the greenhouse. *e*, Relative expression of the *TaAOS2* gene in intact and wounded leaves of non-transgenic plants and wounded transgenic plants. Stars above the graphs indicate statistically significant differences according to Student's *t*-test: * $p \le 0.05$, ** $p \le 0.01$, *** $p \le 0.005$, **** $p \le 0.001$, ns, non-significant.

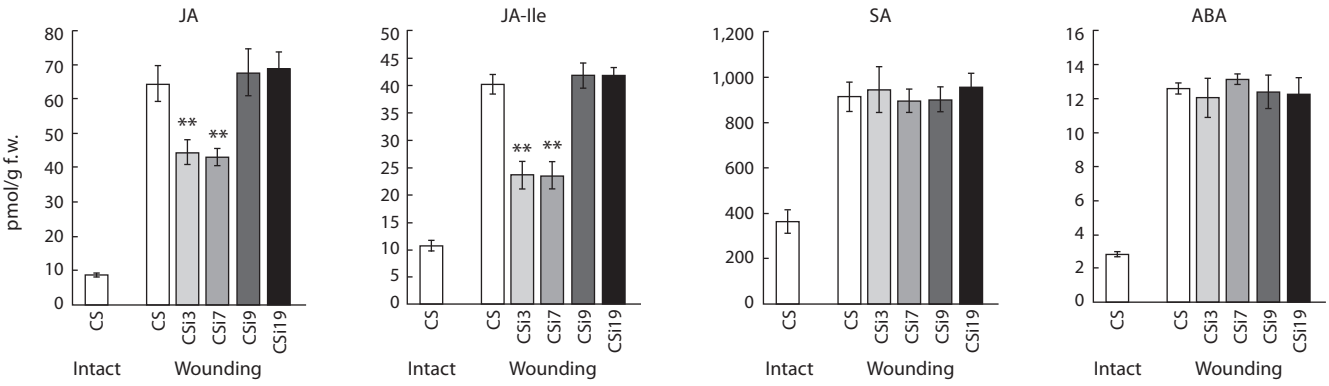


Fig. 2. The content of jasmonic acid (JA), jasmonoyl-isoleucine conjugate (JA-Ile), salicylic acid (SA), and abscisic acid (ABA) in intact and wounded leaf tissues of non-transgenic wheat (Chinese Spring, CS) and wounded transgenic lines with silencing of *TaAOS2* (CSi3, CSi7, CSi9, and CSi19).

Stars indicate a statistically significant difference from the wounded non-transgenic Chinese Spring leaves at $p \le 0.01$, as assessed using Student's *t*-test.

Analysis of plant growth and productivity of RNAi transgenic lines with modified content of jasmonates

In this experiment, two transgenic RNAi lines, CSi3 and SCi7, which showed suppression of *TaAOS2* gene activity resulting in altered JA and JA-Ile content, were compared with the non-transgenic parent variety Chinese Spring. The measurement of the length of four first fully developed leaves of greenhouse-grown plants revealed a decrease in leaf length in both RNAi transgenic lines (Fig. 3). In *AOS* silencing line CSi7, the changes were more pronounced, as the average lengths of the 1st, 2nd, 3rd, and 4th leaves were 2.2 cm ($p < 0.01$), 3.1 cm ($p < 0.005$), 4.8 cm ($p < 0.001$), and 5.8 cm ($p < 0.001$) shorter, correspondingly, compared to the non-transgenic parent plants. The leaves of the transgenic CSi3 line were shorter by 2.5–10 %, and significant leaf length changes were confirmed by Student's *t*-test for 1st, 3rd, and 4th leaves (at $p < 0.05$ to $p < 0.001$).

In addition to a reduction in length of leaves, the transgenic CSi7 line also showed significant changes in plant development. On average, CSi7 plants were 9.9 cm shorter than non-transgenic control, while the transgenic CSi3 line showed almost the same plant height as non-transgenic Chinese Spring plants (Fig. 4*b*).

The average time of the first spike appearance and anthesis was similar in transgenic and non-transgenic plants. A minor change in the average date of flowering observed in RNAi

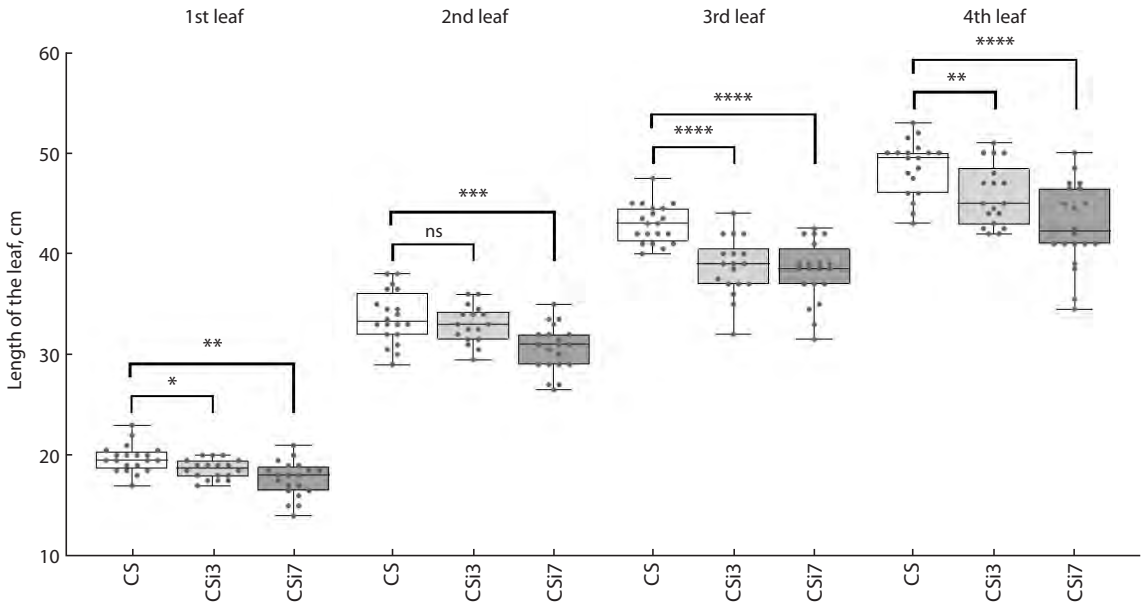


Fig. 3. Analysis of leaf length of non-transgenic bread wheat Chinese Spring and transgenic lines with silencing of *TaAOS2* (CSi3 and CSi7).

Values represent the lengths of 1st, 2nd, 3rd, and 4th leaves measured in 18–20 plants; the line represents the median value, the box margins are SD of the mean, and the bar margins reflect the distribution of values from min to max. Stars indicate statistically significant differences calculated according to Student's *t*-test: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$, ns, non-significant.

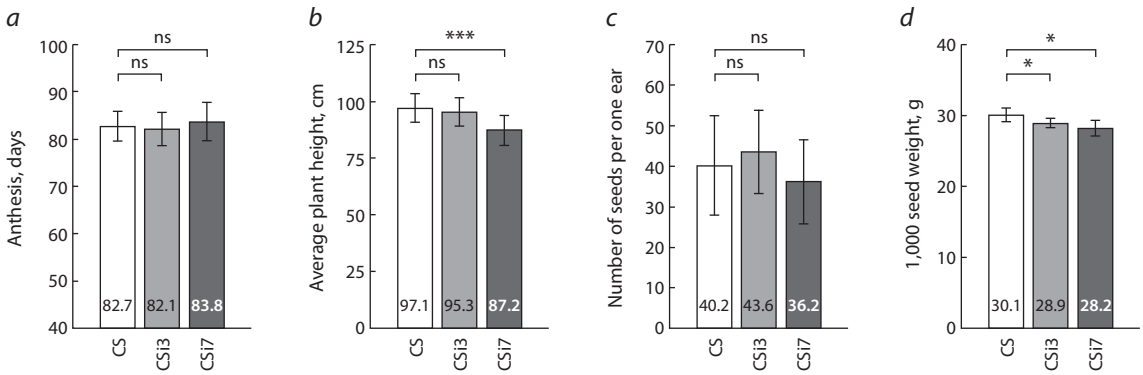


Fig. 4. Analysis of plant growth and productivity of RNAi transgenic wheat lines with modified content of jasmonates.

a, Anthesis date of the first ear; *b*, average plant height; *c*, mean number of seeds per ear; *d*, weight of 1,000 seeds. Stars indicate statistically significant differences of transgenic RNAi lines with silencing of *TaAOS2* (CSi3 and CSi7) from non-transgenic Chinese Spring (SC) plants calculated according to Student's *t*-test: * $p < 0.05$, *** $p < 0.005$, ns, non-significant.

line CSi7 (84 days vs. 83 days in CS) was not proved to be statistically significant (Fig. 4*a*).

Analysis of productivity revealed a variation in the number of seeds collected from one ear of RNAi lines. Compared to control plants (40.2 seeds/ear), the average number of seeds collected from CSi3 plants was slightly higher (43.6 seeds/ear), while in CSi7 plants, it was slightly reduced (36.2 seeds/ear). Statistical analysis, however, did not confirm the differences between the non-transgenic control and RNAi transgenic lines to be significant (Fig. 4*c*).

The measurement of grain weight revealed a decrease in the average weight of 1,000 seeds by 4–7 % ($p < 0.01$) in both transgenic RNAi lines (Fig. 4*d*). The mean values for 1,000 grain weight were reduced from 30.1 g in non-transgenic control plants to 28.9 and 28.2 g in the SCi3 and CSi7 lines, correspondingly.

Discussion

Jasmonates are known as stress-related plant phytohormones, the activity of which is stimulated by various external stressors. Last decade, to study the functioning of jasmonates in plants, reverse genetics methods have been actively used to regulate the expression of specific genes of the jasmonates biosynthesis pathway. In our previous experiments, the functions of *AOS* genes in two wheat species were studied using the overexpression strategy (Pigolev et al., 2023; Miroshnichenko et al., 2024a). The present study focused on the generation and initial characterization of transgenic bread wheat plants in which the endogenous *AOS* gene activity was suppressed by RNA interference.

Previously, producing transgenic lines overexpressing the *AOS* gene was challenging due to the low output of transgenic plants in both emmer and bread wheat (Miroshnichenko et al.,

2024a). It was observed that independently of the *AOS* gene variant used for the design of the overexpression cassette (*TaAOS2* from bread wheat or *AtAOS* from Arabidopsis), morphogenic tissues were characterized by rapid aging and massive necrosis after the biolistic delivery. This resulted in an extremely low rate of transgenic plants production, amounting to 0.3–1.0 % for bread wheat (cv. Saratovskaya 60) and 0.5–1.7 % for emmer (cv. Runo), which was 5–10 times less in comparison with genetic transformation by an empty vector (Miroshnichenko et al., 2024a). In the present study, the down-regulation of *AOS* had no negative effect on the morphogenetic development of wheat tissue culture, and the production of transgenic wheat plants by introducing an RNA interference cassette was not associated with any difficulties. Transgenic plantlets were readily produced from GFP-positive wheat tissues of cv. Chinese Spring achieving the final genetic transformation efficiency of 5.2 % for the hairpin insertions. The resulting transformation rate is significantly higher compared to our previous experiments involving the other pGFP-BAR-based vector (Miroshnichenko et al., 2022), and it is equal to or even higher than the transformation efficiency in other publications, describing biolistic-mediated transformation of Chinese Spring cultivar (Takumi, Shimada, 1997; Harvey et al., 1999).

Overexpression and silencing of *AOS* by the genetic transformation of wheat are supposed to have opposite effects on the capacity of cells to accumulate jasmonates. As it is shown in the present study, the level of *TaAOS2* activity in transgenic wheat tissues expressing the RNAi construct remains low under the induced stress (wounding) and results in weaker accumulation of JA and JA-Ile content (Fig. 2). We suppose that wheat explants transiently expressing the RNAi construct also accumulate less jasmonates when subjected to chemical stress during the selection of transgenic plants on herbicide-containing medium. Considering the previously reported increase of the jasmonate content in transgenic wheat tissues under stress conditions due to the strong constitutive activity of *AOS*, especially in case of *TaAOS2* overproduction (Miroshnichenko et al., 2024a), these findings, taken together, allow suggesting that high *TaAOS2* expression and increased JA content are negative regulators of morphogenesis and somatic embryogenesis in wheat tissue culture, as it was demonstrated for other species (Kamińska, 2021).

The data obtained in the present study suggest that the abiotic stress (wounding) applied to wheat plants of bread wheat cv. Chinese Spring promotes the expression of the *TaAOS2* gene and causes higher production and accumulation of various phytohormones, including JA, JA-Ile, SA, and ABA. The result supports previous studies, where the same stimulus induced a similar response in plants of another bread wheat cultivar, Saratovskaya 60, as well as in tetraploid emmer wheat cv. Runo (Pigolev et al., 2023; Miroshnichenko et al., 2024a, b). The obtained data are consistent with data obtained for other plant species: the wound response in wheat is similar to that in rice (Zeng et al., 2021), but significantly weaker than in Arabidopsis (Kimberlin et al., 2022) and potato (Pajerowska-Mukhtar et al., 2008). Previously, transgenic guayule lines showed increased SA content due to silencing of the *AOS* gene (Placido et al., 2019); in contrast to guayule RNAi lines, there was no alteration in SA content in *AOS*-silenced wheat lines. Such results support the idea that the effect of modified AOS activity is highly dependent on the plant species and functional specialization of *AOS* genes. In line with this, in guayule, which is a dicotyledonous plant species, the silencing of *AOS* was easily detectable in intact plants. In another dicot species, potato, downregulation of *StAOS1* was also not detected in undamaged leaves due to extremely low mRNA levels, while the silencing of the *StAOS2* gene was easily detected as intact dsRNAi transgenic lines accumulated no more than 10 % of the *StAOS2* transcript compared to control plants (Pajerowska-Mukhtar et al., 2008).

Due to the limited number of publications on silencing of *AOS* genes in cereal species (Naor et al., 2018; Fan et al., 2019), information on plant growth and productivity is lacking. In the present study, the *AOS*-silenced wheat lines showed a tendency to grow shorter leaves at the juvenile stage and demonstrated a decrease in final plant height. Surprisingly, a similar trend was observed in both transgenic emmer and bread wheat overexpressing *TaAOS2*. However, despite the growth modification, wheat lines with the knockdown of *TaAOS2* expression showed no change in the mean flowering date. This result is contrasting to a previous study demonstrating that *AOS*-silenced guayule plants showed higher biomass accumulation, enhanced stem branching, increased photosynthetic rate, and increased rubber content (Placido et al., 2019). This increase was associated with elevated rubber transferase enzyme activity, increased SA content, and decreased ABA levels (Placido et al., 2019). In contrast to RNAi lines of guayule, *AOS*-silenced wheat lines showed no significant changes in endogenous SA and ABA accumulation, while similarly to guayule, JA and JA-Ile contents were also reduced, explaining the markedly different effects between the two species. The observed differences can be associated with the presence of a larger number of *AOS* genes in hexaploid wheat, which can perform compensatory functions during *TaAOS2* silencing or with other differences in the complex regulatory system responsible for maintaining the intracellular level of jasmonates and/or jasmonate signaling. This question certainly requires further study.

The downregulation of *AOS* activity did not affect the fertility of transgenic wheat plants. For comparison, the complete disruption of *AOS* function in the aos mutant of Arabidopsis was reported to induce male sterility and failure of seed formation (Park et al., 2002). In the present study, wheat RNAi lines showed normal seed set, probably because there was no complete silencing of *TaAOS2*. This observation is consistent with the previous publications on RNAi-mediated silencing of *AOS* genes in wheat, rice, and guayule, where transgenic homozygous progenies were successfully generated by self-pollination of primary plants (Naor et al., 2018; Fan et al., 2019; Placido et al., 2019). The productivity analysis performed in the present study revealed that the grain weight of RNAi lines tended to decrease, while the mean number of seeds per one ear was not changed. Conversely, when we previously overexpressed the *TaAOS2* gene in bread wheat cv. Saratovskaya 60, this caused a decrease in the seed number per ear (Miroshnichenko et al., 2024a).

Conclusion

Based on these experimental results, we can assume that the downregulation of *TaAOS2* had no clear beneficial effect on the growth and productivity of bread wheat. However, these

results were obtained on plants grown without any temporal or prolonged biotic or abiotic stresses. Given that the functional activity of AOS in wheat is highly dependent on external stimuli, our further research will focus on the analysis of RNAi lines subjected to pathogen attacks and temperature stress. This will help to expand the understanding of the functional effects of *AOS* suppression in wheat and may provide more information on the possibility of precise genetic regulation of the jasmonate biosynthetic pathway to achieve better growth and productivity under various stress conditions.

References

An L., Wang Z., Cui Y., Yao Y., Bai Y., Liu Y., Li X., Yao X., Wu K. Bioinformatics, expression analysis, and functional verification of allene oxide synthase gene *HvnAOS1* and *HvnAOS2* in qingke. *Open Life Sci.* 2024;19(1):20220855. doi 10.1515/biol-2022-0855

Chehab E.W., Kim S., Savchenko T., Kliebenstein D., Dehesh K., Braam J. Intronic T-DNA insertion renders Arabidopsis *opr3* a conditional jasmonic acid-producing mutant. *Plant Physiol.* 2011; 156(2):770-778. doi 10.1104/pp.111.174169

Degtyaryov E., Pigolev A., Miroshnichenko D., Frolov A., Basnet A.T., Gorbach D., Leonova T., Pushin A.S., Alekseeva V., Dolgov S., Savchenko T. *12-Oxophytodienoate* reductase overexpression compromises tolerance to *Botrytis cinerea* in hexaploid and tetraploid wheat. *Plants* (*Basel*). 2023;12(10):2050. doi 10.3390/plants12102050

Fan Y.-H., Hou B.-Q., Su P.-S., Wu H.-Y., Wang G.-P., Kong L.-R., Ma X., Wang H.-W. Application of virus-induced gene silencing for identification of FHB resistant genes. *J Integr Agric.* 2019;18(10): 2183-2192. doi 10.1016/S2095-3119(18)62118-5

Harvey A., Moisan L., Lindup S., Lonsdale D. Wheat regenerated from scutellum callus as a source of material for transformation. *Plant Cell Tissue Organ Cult.* 1999;57:153-156. doi 10.1023/A:1006344615666

Heckmann A., Perochon A., Doohan F.M. Genome-wide analysis of salicylic acid and jasmonic acid signalling marker gene families in wheat. *Plant Biol.* 2024;26(5):691-704. doi 10.1111/plb.13659

Jiang K., Pi Y., Huang Z., Hou R., Zhang Z., Lin J., Sun X., Tang K. Molecular cloning and mRNA expression profiling of the first specific jasmonate biosynthetic pathway gene allene oxide synthase from *Hyoscyamus niger*. *Russ J Genet*. 2009;45:430-439. doi 10.1134/S1022795409040073

Kamińska M. Role and activity of jasmonates in plants under in vitro conditions. *Plant Cell Tissue Organ Cult*. 2021;146:425-447. doi 10.1007/s11240-021-02091-6

Kim S.-J., Park Y.-D., Lee J.-W. Validation of the role of *Allium cepa allene oxide synthase* (*AcAOS*) in resistance to *Botrytis squamosa*. *Hortic Environ Biotechnol.* 2025;66:951-965. doi 10.1007/s13580-025-00706-x

Kimberlin A.N., Holtsclaw R.E., Zhang T., Mulaudzi T., Koo A.J. On the initiation of jasmonate biosynthesis in wounded leaves. *Plant Physiol.* 2022;189(4):1925-1942. doi 10.1093/plphys/kiac163

Laudert D., Pfannschmidt U., Lottspeich F., Holländer-Czytko H., Weiler E.W. Cloning, molecular and functional characterization of *Arabidopsis thaliana* allene oxide synthase (CYP 74), the first enzyme of the octadecanoid pathway to jasmonates. *Plant Mol Biol.* 1996;31:323-335. doi 10.1007/BF00021793

Laudert D., Weiler E.W. Allene oxide synthase: a major control point in *Arabidopsis thaliana* octadecanoid signalling. *Plant J.* 1998;15(5): 675-684. doi 10.1046/j.1365-313x.1998.00245.x

Liu H.-H., Wang Y.-G., Wang S.-P., Li H.-J., Xin Q.-G. Improved zinc tolerance of tobacco by transgenic expression of an allene oxide synthase gene from hexaploid wheat. *Acta Physiol Plant.* 2014;36(9): 2433-2440. doi 10.1007/s11738-014-1616-7

Miroshnichenko D., Timerbaev V., Klementyeva A., Pushin A., Sidorova T., Litvinov D., Nazarova L., Shulga O., Divashuk M., Karlov G., Salina E., Dolgov S. CRISPR/Cas9-induced modification of the conservative promoter region of *VRN-A1* alters the heading time of hexaploid bread wheat. *Front Plant Sci.* 2022;13:1048695. doi 10.3389/fpls.2022.1048695

Miroshnichenko D.N., Pigolev A.V., Pushin A.S., Alekseeva V.V., Degtyaryova V.I., Degtyaryov E.A., Pronina I.V., Frolov A., Dolgov S.V., Savchenko T.V. Genetic transformation of *Triticum dicoccum* and *Triticum aestivum* with genes of jasmonate biosynthesis pathway affects growth and productivity characteristics. *Plants* (*Basel*). 2024a;13(19):192781. doi 10.3390/plants13192781

Miroshnichenko D.N., Pigolev A.V., Tikhonov K.G., Degtyaryov E.A., Leshchenko E.F., Alekseeva V.V., Pushin A.S., Dolgov S.V., Basnet A., Gorbach D.P., Leonova T.S., Frolov A.A., Savchenko T.V. Characteristics of the stress-tolerant transgenic wheat line overexpressing the *AtOPR3* gene encoding the jasmonate biosynthesis enzyme 12-oxophytodienoate reductase. *Russ J Plant Physiol.* 2024b;71(54):1446. doi 10.1134/S1021443724604658

Naor N., Gurung F.B., Ozalvo R., Bucki P., Sanadhya P., Miyara S.B. Tight regulation of allene oxide synthase (AOS) and allene oxide cyclase-3 (AOC3) promote *Arabidopsis* susceptibility to the root-knot nematode *Meloidogyne javanica*. *Eur J Plant Pathol.* 2018; 150:149-165. doi 10.1007/s10658-017-1261-2

Pajerowska-Mukhtar K.M., Mukhtar M.S., Guex N., Halim V.A., Rosahl S., Somssich I.E., Gebhardt C. Natural variation of potato *allene oxide synthase* 2 causes differential levels of jasmonates and pathogen resistance in *Arabidopsis*. *Planta.* 2008;228(2):293-306. doi 10.1007/s00425-008-0737-x

Park J.-H., Halitschke R., Kim H.B., Baldwin I.T., Feldmann K.A., Feyereisen R. A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in *Arabidopsis* due to a block in jasmonic acid biosynthesis. *Plant J.* 2002;31(1):1-12. doi 10.1046/j.1365-313X.2002.01328.x

Pigolev A.V., Miroshnichenko D.N., Pushin A.S., Terentyev V.V., Boutanayev A.M., Dolgov S.V., Savchenko T.V. Overexpression of Arabidopsis *OPR3* in hexaploid wheat (*Triticum aestivum* L.) alters plant development and freezing tolerance. *Int J Mol Sci.* 2018;19: 123989. doi 10.3390/ijms19123989

Pigolev A.V., Miroshnichenko D.N., Dolgov S.V., Alekseeva V.V., Pushin A.S., Degtyaryova V.I., Klementyeva A., Gorbach D., Leonova T., Basnet A., Frolov A.A., Savchenko T.V. Endogenously produced jasmonates affect leaf growth and improve osmotic stress tolerance in emmer wheat. *Biomolecules.* 2023;13(12):121775. doi 10.3390/biom13121775

Placido D.F., Dong N., Dong C., Cruz V.M.V., Dierig D.A., Cahoon R.E., Kang B.-G., Huynh T., Whalen M., Ponciano G., McMahan C. Downregulation of a CYP74 rubber particle protein increases natural rubber production in *Parthenium argentatum*. *Front Plant Sci*. 2019;10:760. doi 10.3389/fpls.2019.00760

Sun T., Chen Y., Feng A., Zou W., Wang D., Lin P., Chen Y., You C., Que Y., Su Y. The allene oxide synthase gene family in sugarcane and its involvement in disease resistance. *Ind Crops Prod.* 2023; 192:116136. doi 10.1016/j.indcrop.2022.116136

Takumi S., Shimada T. Variation in transformation frequencies among six common wheat cultivars through particle bombardment of scutellar tissues. *Genes Genet Syst.* 1997;72(2):63-69. doi 10.1266/ggs.72.63

Zeng J., Zhang T., Huangfu J., Li R., Lou Y. Both allene oxide synthases genes are involved in the biosynthesis of herbivore-induced jasmonic acid and herbivore resistance in rice. *Plants* (*Basel*). 2021; 10(3):30442. doi 10.3390/plants10030442

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

Received September 11, 2025. Revised October 10, 2025. Accepted November 5, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1184-1194

doi 10.18699/vjgb-25-125

Study of the influence of introgression from chromosome 2 of the A_t subgenome of cotton *Gossypium barbadense* L. during backcrossing with the original lines of *G. hirsutum* L.

M.F. Sanamyan [1], Sh.U. Bobokhujayev [1, 3], Sh.S. Abdukarimov [1, 2], J.S. Uralov [1], A.B. Rustamov [1]

[1] National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan
[2] Center of Genomics and Bioinformatics of the Academy of Sciences of the Republic of Uzbekistan, Kibrai district, Tashkent region, Uzbekistan
[3] Chirchik State Pedagogical University, Chirchik, Tashkent region, Uzbekistan
sanam_marina@rambler.ru

Abstract. The creation of chromosome substitution lines containing one pair of chromosomes from a related species is one method for introgression of alien genetic material. The frequency of substitutions in different chromosomes of the genome varies due to the selective transmission of alien chromosomes through the gametes of hybrids. The use of monosomic lines with identified univalent chromosomes and molecular genetic SSR markers at the seedling stage allowed rapid screening of the identity of the alien chromosome in backcross hybrids, significantly accelerating and facilitating the backcrossing process for the creation of new chromosome substitution cotton lines. As a result of studying the process of transmission of chromosome 2 of the A_t subgenome of the cotton plant *G. barbadense* L. during backcrossing of four original monosomic lines of *G. hirsutum* L. with monosomic backcross hybrids with substitution of chromosome 2 of the A_t subgenome, the following specific consequences of the introgression of this chromosome were revealed: decreased crossability, setting and germination of hybrid seeds; differences in the frequency and nature of transmission of chromosome 2 of the A_t subgenome of the cotton plant *G. barbadense*; regularity of chromosome behavior in meiosis; a high meiotic index; a significant decrease in pollen fertility in backcross monosomic hybrids BC_1F_1; specific morphobiological characteristics of monosomic backcrossed plants, such as delayed development of vegetative and generative organs; dwarfism; reduced foliage; and poor budding and flowering during the first year of vegetation. All of these factors negatively impact the study and backcrossing of monosomic hybrids and significantly complicate and delay the creation of chromosome-substituted forms concerning chromosome 2 of the A_t subgenome of cotton, *G. barbadense*. These specific changes likely occurred as a result of hybrid genome reorganization and introgression of alien chromatin. Furthermore, the effectiveness of using molecular genetic microsatellite (SSR) markers to monitor backcrossing processes and eliminate genetic material from the Pima 3-79 donor line of *G. barbadense* for the selection of genotypes with alien chromosome substitutions has been demonstrated.

Key words: cotton; *G. hirsutum*; *G. barbadense*; monosomic lines; chromosome-substituted hybrids; backcrossing; SSR markers

For citation: Sanamyan M.F., Bobokhujayev Sh.U., Abdukarimov Sh.S., Uralov J.S., Rustamov A.B. Study of the influence of introgression from chromosome 2 of the A_t subgenome of cotton *Gossypium barbadense* L. during backcrossing with the original lines of *G. hirsutum* L. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1184-1194. doi 10.18699/vjgb-25-125

Funding. This study was financially supported by the Ministry of Higher Education, Science, and Innovation of the Republic of Uzbekistan as part of the F-OT-2021-155 project.

Изучение влияния интрогрессии хромосомы 2 A_t-субгенома хлопчатника вида *Gossypium barbadense* L. при беккроссировании исходными линиями вида *G. hirsutum* L.

М.Ф. Санамьян [1], Ш.У. Бобохужаев [1, 3], Ш.С. Абдукаримов [1, 2], Ж.С. Уралов [1], А.Б. Рустамов [1]

[1] Национальный университет Узбекистана им. М. Улугбека, Ташкент, Узбекистан
[2] Центр геномики и биоинформатики Академии наук Республики Узбекистан, Кибрайский район, Ташкентская область, Узбекистан
[3] Чирчикский государственный педагогический университет, Чирчик, Ташкентская область, Узбекистан
sanam_marina@rambler.ru

Аннотация. Создание хромосомно-замещенных линий, содержащих одну пару хромосом родственного вида, является одним из способов интрогрессии чужеродного генетического материала. Известно, что частота встречаемости замещений по разным хромосомам генома различается по причине избирательности процесса трансмиссии чужеродных хромосом через гаметы гибридов. Использование моносомных линий с

идентифицированными унивалентными хромосомами и молекулярно-генетических SSR-маркеров на стадии проростков позволило осуществить быстрый скрининг идентичности чужеродной хромосомы у беккроссных гибридов, что значительно ускорило и облегчило процесс беккроссирования при создании новых хромосомно-замещенных линий хлопчатника. В ходе изучения процесса передачи хромосомы 2 A_t-субгенома хлопчатника *Gossypium barbadense* L. при беккроссировании четырех исходных моносомных линий вида *G. hirsutum* L. моносомными беккроссными гибридами с замещением хромосомы 2 A_t-субгенома были выявлены: снижение скрещиваемости, завязываемости и всхожести гибридных семян; отличия в частоте и характере передачи хромосомы 2 A_t-субгенома хлопчатника *G. barbadense*; регулярность поведения хромосом в мейозе и высокий мейотический индекс; значительное снижение фертильности пыльцы у беккроссных моносомных гибридов BC_1F_1; специфические морфобиологические особенности моносомных беккроссных растений в виде задержки развития вегетативных и генеративных органов, низкорослости и сниженной облиственности, а также слабой бутонизации и цветения в течение первого года вегетации. По-видимому, такие специфические изменения происходили вследствие реорганизации гибридного генома и интрогрессии чужеродного хроматина. Кроме того, была продемонстрирована эффективность использования молекулярно-генетических микросателлитных (SSR) маркеров для контролирования процессов беккроссирования и элиминации генетического материала донорной линии Pima 3-79 вида *G. barbadense* с целью отбора генотипов с чужеродным замещением хромосом.
Ключевые слова: хлопчатник; *G. hirsutum*; *G. barbadense*; моносомные линии; хромосомно-замещенные гибриды; беккроссирование; SSR-маркеры

Introduction

Currently, the lack of genetic diversity in the cultivated cotton plant *Gossypium hirsutum* L. ($2n = 52$, AD_1) hinders the development of breeding programs (Wendel, 1989). One method of introducing alien genetic material is the creation of chromosome substitution lines containing one pair of chromosomes from a related species, which substitutes the homeologous pair of chromosomes because the substitution occurs in strict accordance with chromosome homeology (Shchapova, Kravtsova, 1990).

Lines with alien chromosome substitutions were created in the USA via three tetraploid cotton species (*G. barbadense* L., *G. tomentosum* Nutt. ex Seem and *G. mustelinum* Miers ex Watt) (Saha et al., 2004, 2006, 2013), where the largest number of lines were obtained from the *G. barbadense* species, and a study was conducted on the effects of substitution on valuable fibre quality traits (Saha et al., 2004, 2010, 2020; Jenkins et al., 2006, 2007). However, the active use of these lines by other researchers indicated the absence of introgression of the entire chromosome or a chromosomal region in some of these lines (Gutiérrez et al., 2009; Saha et al., 2015; Ulloa et al., 2016; Fang et al., 2023), so only 13 CS-B (chromosome substitution) lines out of 20 had "significant introgression" from the *G. barbadense* species. However, the reasons for the lack of substitution of alien chromosomes in the lines of the American collection have never been clarified.

Earlier, at the National University of Uzbekistan named after Mirzo Ulugbek a collection of new monosomic lines of *G. hirsutum* cotton was created via various irradiation methods. Univalent chromosomes of 35 of these lines, which are deficient in individual chromosomes, were identified via translocation and molecular genetic SSR markers (Sanamyan et al., 2014, 2022). Four monosomic lines were deficient in chromosome 2, 18 lines were deficient in chromosome 4, five lines were deficient in chromosome 6, one line was deficient in chromosome 7, and another line was deficient in chromosome 12 of the A_t subgenome of cotton. There is also one monosomic line on chromosomes 17, 18, 21 and 22 of the D_t subgenome of cotton and two telocentric lines on chromosomes 6 and 11 of the A_t subgenome (Sanamyan et al., 2016a, b; Sanamyan, Bobokhujaev, 2019).

To create chromosome substitution cotton lines, we developed a new scheme based on cytogenetic and molecular genetic methods (Sanamyan et al., 2022). The use of monosomic lines with identified univalent chromosomes and molecular genetic SSR markers at the seedling stage allowed rapid screening of the identity of alien chromosomes in backcross hybrids, significantly accelerating and facilitating the creation of chromosome substitution lines.

Experiments involving many hybrids across multiple generations have revealed new effects of alien chromosome transmission in backcrossed progeny. The aim of our study was to analyse the transmission patterns of chromosome 2 of the A_t subgenome of cotton during backcrossing of four original monosomic lines of *G. hirsutum* with the donor line Pima 3-79 of *G. barbadense* to create *G. hirsutum/G. barbadense* lines with alien chromosome substitutions. We assessed the crossability, seed set, and germination of hybrid seeds, estimated the frequency and pattern of chromosome 2 transmission, studied chromosome behavior during meiosis, and identified specific morphobiological features of backcrossed monosomic plants in the first year of vegetative growth.

Materials and methods

Plant material. The cotton cytogenetic collection of the National University of Uzbekistan named after M. Ulugbek is characterized by the presence of four monosomic lines (Mo11, Mo16, Mo19, and Mo93) deficient in chromosome 2 of the A_t subgenome of the tetraploid cotton species *G. hirsutum* ($2n = 52$, AD_1) (Sanamyan, Bobokhujaev, 2019).

The Pima 3-79 line, which was obtained from a doubled haploid and is the genetic standard for this cotton species in the United States (Endrizzi et al., 1985), was used as the donor parent of the substitution chromosome (CS) from the *G. barbadense* species.

Backcross hybrids obtained from crossing four monosomic lines (Mo11, Mo16, Mo19, and Mo93) deficient in chromosome 2 of the A_t subgenome of the *G. hirsutum* species with previously obtained monosomic F_1 hybrids, which had monosomy for the same chromosomes as the original monosomic plants (Sanamyan et al., 2016b), were studied. All plants of the original monosomic lines and backcross hybrids were

maintained year-round in a film greenhouse at the National University of Uzbekistan.

Cytological analyses, as well as DNA extraction and genotyping, were performed according to methods described previously (Sanamyan et al., 2023). Elimination of *G. barbadense* chromosomes in the monosomic cotton hybrid BC_2F_1 was determined by the absence of marker amplification on the *G. barbadense* chromosomes (paternal) and the presence of only allele-specific PCR products of *G. hirsutum* (maternal).

Results

Crossability, seed set, and germination of BC_1F_1 hybrid seeds involving monosomic lines deficient in chromosome 2 of the A_t subgenome of *G. hirsutum*. Four monosomic lines of *G. hirsutum* deficient in chromosome 2 of the A_t subgenome were crossed with F_1 hybrids (Mo×Pima 3-79), which were monosomic for the same chromosomes as the original monosomic line. All four BC_1F_1 variants were characterized by a significant decrease in the percentage of crossability (from 33.33 to 9.09 %) (Supplementary Material 1)[1] compared with F_1 hybrids (from 68.75 to 50.00 %) (Sanamyan et al., 2022).

The hybrid seed set of the BC_1F_1 plants also decreased (from 52.94 ± 12.11 to 37.31 ± 5.91 %) compared with that of the F_1 hybrids (from 57.69 ± 9.69 to 31.33 ± 5.09 %). Compared with that of the F_1 hybrids, the germination of the backcrossed BC_1F_1 seeds decreased from 88.89 to 51.61 % (Supplementary Material 1) (from 100 to 71.43 %).

Identification of chromosome 2 substitutions in the A_t subgenome of *G. barbadense* in BC_1F_1 hybrids via chromosome-specific molecular genetic markers. For molecular analysis of backcrossed BC_1F_1 plants, the principles of deletion molecular analysis were used (Liu et al., 2000; Gutiérrez et al., 2009). The study was conducted according to our previously proposed scheme for producing chromosome substitution cotton lines (Sanamyan et al., 2022). Molecular genetic analysis of the backcrossed forms was performed on backcrossed cotton seedlings at the 3–5 true leaf stage before they were transplanted into greenhouse soil to accelerate the isolation of monosomic chromosomes from the donor species. Among the plants harboring all the BC_1F_1 variants, those with genomes containing an alien chromosome 2 of the A_t subgenome of the *G. barbadense* species were identified.

Previously, among the BC_1F_1 hybrids in the BC_1F_1(Mo11×$\times F_1(766_3)$) cross variant in four backcross families (9_n, 10_n, 78_n, and 79_n), no hybrid seedlings with polymorphic alleles from the *G. barbadense* species were found, which indicated the absence of substitutions in these monosomes. Later, in a similar variant of the experiment in the backcross family (494_n), one (494_2) monosomic backcross seedling was identified, which was characterized by the presence of polymorphic alleles only from *G. barbadense,* whereas the alleles of the L-458 line of the *G. hirsutum* species were not detected on the basis of the localization of chromosome-specific SSR markers, BNL834, BNL1434, BNL1897, BNL3590, BNL3971, BNL3972, CIR381, and JESPR179. Since all the above-mentioned markers were previously localized on chromosome 2 of the A_t subgenome of cotton (Liu et al., 2000; Gutiérrez et al., 2009; Yu et al., 2011; Saha et al., 2015; Wang et al., 2016), the obtained data indicated the presence of substitutions on this chromosome (Supplementary Materials 2, 3, 19).

Previously, in the BC_1F_1(Mo16×$F_1(98_6)$) variant, four seedlings (922_2, 922_8, 923_7, 923_8) with a substitution on chromosome 2 of the *G. barbadense* species were already found in two hybrid backcross families (922_n and 923_n) (Sanamyan et al., 2022); however, the elimination of the alien chromosome in BC_2F_1 required re-examination of this variant, where three backcross seedlings (495_5, 497_1 and 497_4) were found in two new families, which had only alleles from *G. barbadense,* whereas the alleles of *G. hirsutum* were absent, which indicated the localization of seven chromosome-specific SSR markers: BNL834, TMB0471, JESPR101, JESPR179, CIR376, DPL0674, and NAU2277, previously localized on chromosome 2 of the A_t subgenome (Gutiérrez et al., 2009; Yu et al., 2011; Saha et al., 2015; Wang et al., 2016) and confirmed the presence of a substitution on this chromosome (Supplementary Materials 4–6, 19).

SSR-based deletion analysis in the BC_1F_1(Mo19×$F_1(769_4)$) and BC_1F_1(Mo93×$F_1(516_4)$) combinations with a putative substitution of chromosome 2 of the A_t subgenome of cotton allowed us to detect alleles of the *G. barbadense* species in one backcross seedling (8_8) in the first variant and nine seedlings (7_1, 7_2, 7_3, 88_1, 88_4, 88_7, 89_3, 89_4 and 89_6) in the second, whereas alleles of the *G. hirsutum* species were absent. Since chromosome-specific SSR markers BNL3545, BNL3971, JESPR101, JESPR179, CIR376, and DPL0674 were previously localized on chromosome 2 of the A_t subgenome of cotton (Liu et al., 2000; Gutiérrez et al., 2009; Yu et al., 2011; Saha et al., 2015; Wang et al., 2016), the obtained data indicated the presence of substitution of this chromosome in the studied seedlings (Fig. 1, Supplementary Materials 7–9, 19).

A study of meiosis in BC_1F_1 hybrids with identified univalents. Chromosome pairing at the metaphase I (MI) stage of meiosis was studied in 17 monosomic plants in four backcross variants obtained from crosses of monosomic lines of the *G. hirsutum* species with interspecific monosomic F_1 hybrids (Mo×Pima 3-79). Seven monosomics were found among the backcrossed plants in the BC_1F_1(Mo16×$F_1(98_6)$) variant, one each in the BC_1F_1(Mo11×$F_1(766_3)$) and BC_1F_1(Mo19×$F_1(769_4)$) variants, and eight in the BC_1F_1(Mo93×$F_1(516_4)$) variant.

Analysis of meiotic MI in BC_1F_1 monosomic plants, where all backcrossed monosomic plants had univalent *G. barbadense* chromosomes, revealed a modal chromosome pairing with 25 bivalents and one univalent chromosome, characteristic of tetraploid monosomic cotton plants (Supplementary Material 10). Analysis of the size of univalents in monosomic BC_1F_1 confirmed the large size of chromosome 2 of *G. barbadense* in all four crossing variants (Fig. 2), which indicated that this chromosome belongs to the A_t subgenome of cotton.

Most of the studied BC_1F_1 monosomics were characterized by a high meiotic index (up to 95.74 ± 0.47) and a small number of tetrads with micronuclei (up to 3.86 ± 0.65 %), with the exception of two monosomics (922_8 and 7_3) with a reduced meiotic index (89.04 ± 0.94 and 88.57 ± 1.13, respectively) and an increased number of tetrads with micronuclei (up to 4.75 ± 0.64 and 5.15 ± 0.78 %, respectively) (Fig. 3, Supplementary Material 11).

[1] Supplementary Materials 1–20 are available at: https://vavilovj-icg.ru/download/pict-2025-29/appx45.pdf

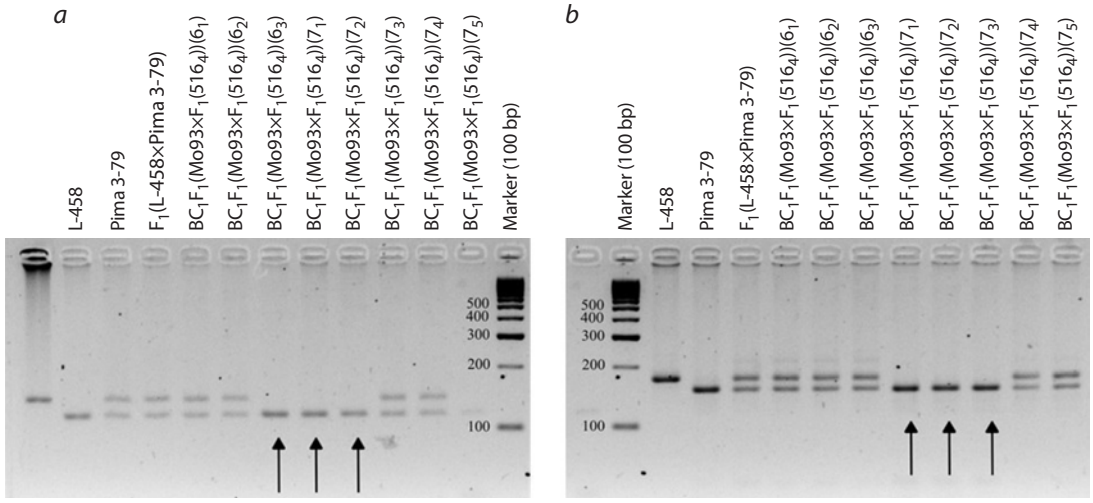


Fig. 1. Electropherogram of DNA amplicons of SSR markers in hybrid seedlings $BC_1F_1(Mo93\times F_1(516_4))$ on chromosome 2 of the A_t subgenome of cotton.

a – BNL3971; *b* – JESPR179.

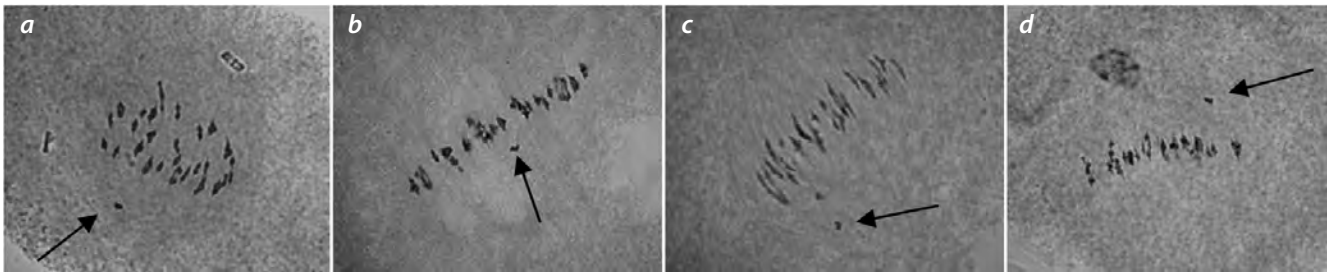


Fig. 2. Chromosome configurations in metaphase I of meiosis in BC_1F_1 hybrid plants obtained from crossing monosomic lines with interspecific monosomic F_1 hybrids: F_1: *a* – $BC_1F_1(Mo11\times F_1(766_3))$ (494_2); *b* – $BC_1F_1(Mo16\times F_1(98_2))$ (497_1); *c* – $BC_1F_1(Mo19\times F_1(769_4))$ (8_8); *d* – $BC_1F_1(Mo93\times F_1(516_4))$ (88_1) with chromosome 2 of the A_t subgenome of *G. barbadense*.

Univalents are indicated by arrows.

Pollen viability was assessed in BC_1F_1 monosomics via acetocarmine staining. Only two monosomic strains (922_8 and 923_7) presented high pollen viability (up to 94.21 ± 1.06 %) (Fig. 4, Supplementary Material 12). Six monosomics methods resulted in a slight reduction in pollen viability (from 88.38 ± 1.46 to 81.15 ± 1.52 %), but five monosomics methods (7_2, 88_1, 88_4, 88_7, and 89_4) involving one variant of BC_1F_1 ($Mo93\times F_1(516_4)$) were characterized by a strong reduction in pollen viability (from 62.65 ± 2.15 to 77.47 ± 1.64 %).

Crossability, seed set, and germination characteristics of BC_2F_1 hybrid seeds with monosomic lines deficient in chromosome 2 of the A_t subgenome of *G. hirsutum*. Three monosomic lines deficient in chromosome 2 of the A_t subgenome of *G. hirsutum* from the cytogenetic collection were crossed with monosomic $BC_1F_1(Mo\times F_1(Mo\times Pima\ 3\text{-}79))$ hybrids that contained substitutions of chromosome 2 of the A_t subgenome of *G. barbadense*. Compared with those of the F_1 and BC_1F_1 hybrids, a strong decrease in crossability was observed (from 18.18 to 3.47 %) (Supplementary Material 13). A study of the seed set rate of hybrid BC_2F_1 plants revealed a significant decrease (from 47.47 ± 5.02 to 27.69 ± 5.55 %) compared with that of the F_1 and BC_1F_1 hybrids, while the germination rate of backcrossed BC_2F_1 seeds also decreased

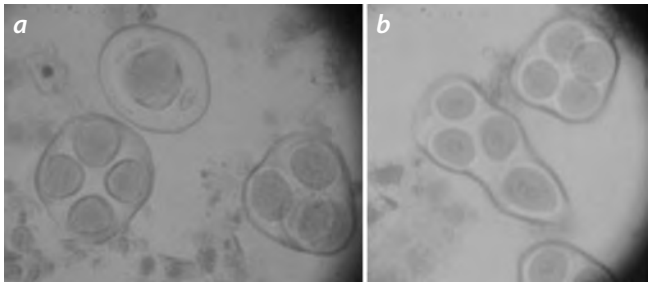


Fig. 3. Sporades in the hybrid plant $BC_1F_1(Mo93\times F_1(516_4))$ (88_7): *a* – monad with three micronuclei and tetrads; *b* – abnormal tetrad and normal tetrad.

(from 69.23 to 44.44 %) compared with that of the same hybrids, with the monosomic line Mo93 exhibiting the strongest decrease in seed set and germination compared with the other two monosomic lines.

Identification of substitutions in the chromosome 2 of the A_t subgenome of *G. barbadense* in BC_2F_1 and BC_3F_1 hybrids via chromosome-specific molecular genetic markers. We previously demonstrated that five monosomic seedlings

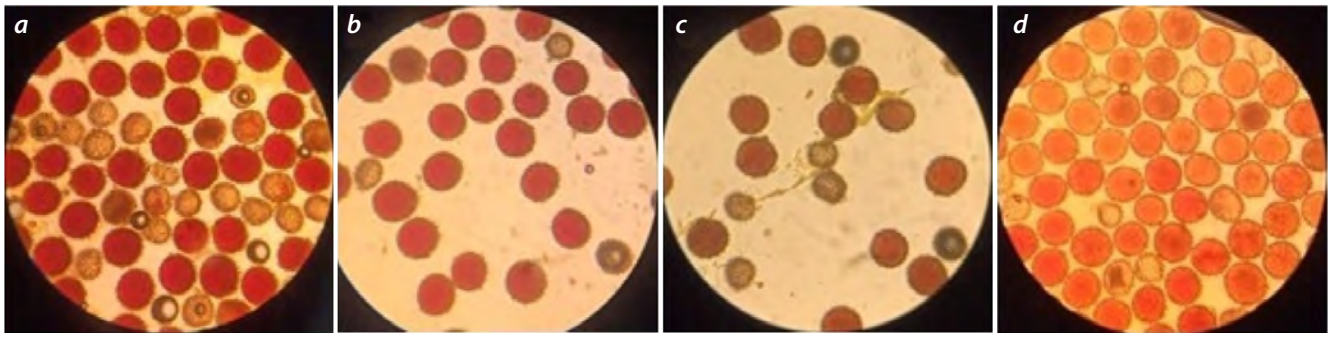


Fig. 4. Sterile (uncolored) and fertile (colored) pollen in monosomic BC_1F_1 hybrids obtained from crossing a monosomic line with a monosomic hybrid F_1(Mo×Pima 3-79): *a*, *b* (88_4) and *c*, *d* (89_4) in the BC_1F_1(Mo93×$F_1(516_4)$) variant.

(21_1, 21_2, 21_4, 21_7 and 22_1) in the BC_2F_1(Mo16×$BC_1F_1(923_7)$) variant with a putative substitution of chromosome 2 in the A_t subgenome of cotton were characterized by the presence of chromosome-specific alleles only from the L-458 line of *G. hirsutum*, whereas alleles from *G. barbadense* were absent, indicating the elimination of the alien chromosome (Sanamyan et al., 2023).

The results of a new study of two variants involving the Mo16 line but two different BC_1F_1 hybrids (BC_2F_1(Mo16×$\times BC_1F_1(922_8)$) and BC_2F_1(Mo16×$BC_1F_1(923_7)$) revealed that in the first variant, three seedlings (817_1, 817_4 and 817_6) had chromosome-specific alleles only from the *G. barbadense* species, whereas alleles from the *G. hirsutum* species were absent; however, in the second variant, four seedlings (818_2, 818_3, 818_4 and 818_7) had alleles only from the *G. hirsutum* species, which indicated the elimination of the alien chromosome. Since previously reported chromosome-specific SSR markers, BNL3971, TMB1194, CIR376, and DPL0674 are located on chromosome 2 of the A_t subgenome of cotton (Liu et al., 2000; Gutiérrez et al., 2009; Saha et al., 2013, 2015; Wang et al., 2016), the obtained data indicate the presence of chromosome 2 substitutions in the first three seedlings (Fig. 5, Supplementary Materials 14, 19).

SSR-based deletion analysis in combinations BC_2F_1(Mo19×$\times BC_1F_1(8_8)$) and BC_2F_1(Mo93×$BC_1F_1(88_4)$) with putative substitution of chromosome 2 of the A_t subgenome of cotton allowed us to detect alleles of the *G. barbadense* species only in 10 backcross seedlings (820_2, 820_3, 820_7, 820_{10}, 820_{14}, 820_{17}, 821_2, 821_3, 821_6, and 821_{10}) of the first variant and in one (515_1) of the second, whereas alleles of the *G. hirsutum* species were absent. Chromosome-specific SSR markers, such as BNL834, BNL1434, BNL1897, BNL3292, BNL3413, BNL3424, BNL3547, BNL3971, BNL3972, JESPR101, JESPR179, JESPR304, TMB0471, TMB1194, CIR376, CIR381, CIR401, DPL0674, and NAU2277 are located on chromosome 2 of the A_t subgenome of cotton (Liu et al., 2000; Gutiérrez et al., 2009; Saha et al., 2013, 2015; Wang et al., 2016; https://www.CottonGenorg/data/download/marker_origin), and the obtained data indicated the presence of substitutions of this chromosome in the studied seedlings (Fig. 6, 7, Supplementary Materials 15–17, 19).

Similarly, in the BC_3F_1(Mo93×$BC_2F_1(515_1)$) combination with a putative substitution of chromosome 2 of the A_t subgenome of cotton, alleles of *G. barbadense* were detected in

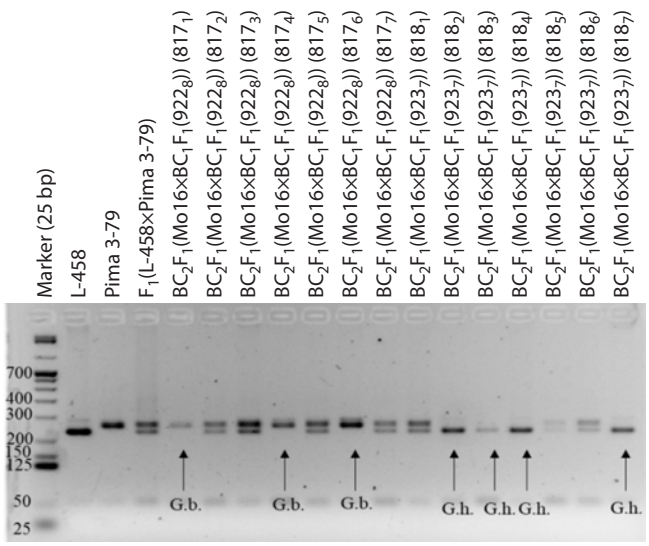


Fig. 5. Electropherogram of DNA amplicons of SSR markers in hybrid seedlings, BC_2F_1(Mo16×$BC_1F_1(922_8)$) on chromosome 2 of the A_t subgenome of cotton.

Marker DPL0674.

seven backcrossed seedlings (839_1, 839_2, 839_4, 839_9, 839_{10}, 839_{11}, and 839_{12}), whereas alleles of the *G. hirsutum* species were absent. Because previously reported chromosome-specific SSR markers, DPL0674 and JESPR179, were located on chromosome 2 of the A_t subgenome of cotton (Liu et al., 2000; Gutiérrez et al, 2009; Yu et al., 2011; Saha et al., 2013, 2015; Wang et al., 2016), the obtained data indicated the presence of a substitution of this chromosome in the studied seedlings (Fig. 8, Supplementary Materials 18, 19).

Study of meiosis in BC_2F_1 hybrids with identified univalent. A study of chromosome pairing at the MI stage of meiosis revealed one backcrossing monosomic in each of the three backcrossing variants of BC_2F_1 (involving the Mo16, Mo19, and Mo93 lines) at the time of writing.

Analysis of meiotic MI in three BC_2F_1 monosomic strains, including one monosomic strain, 21_1, from one cross with univalent chromosome 2 of the A_t subgenome of cotton *G. hirsutum* (Sanamyan et al., 2023) and two other monosomic strains (821_{10} and 515_1) from two crosses with univalent

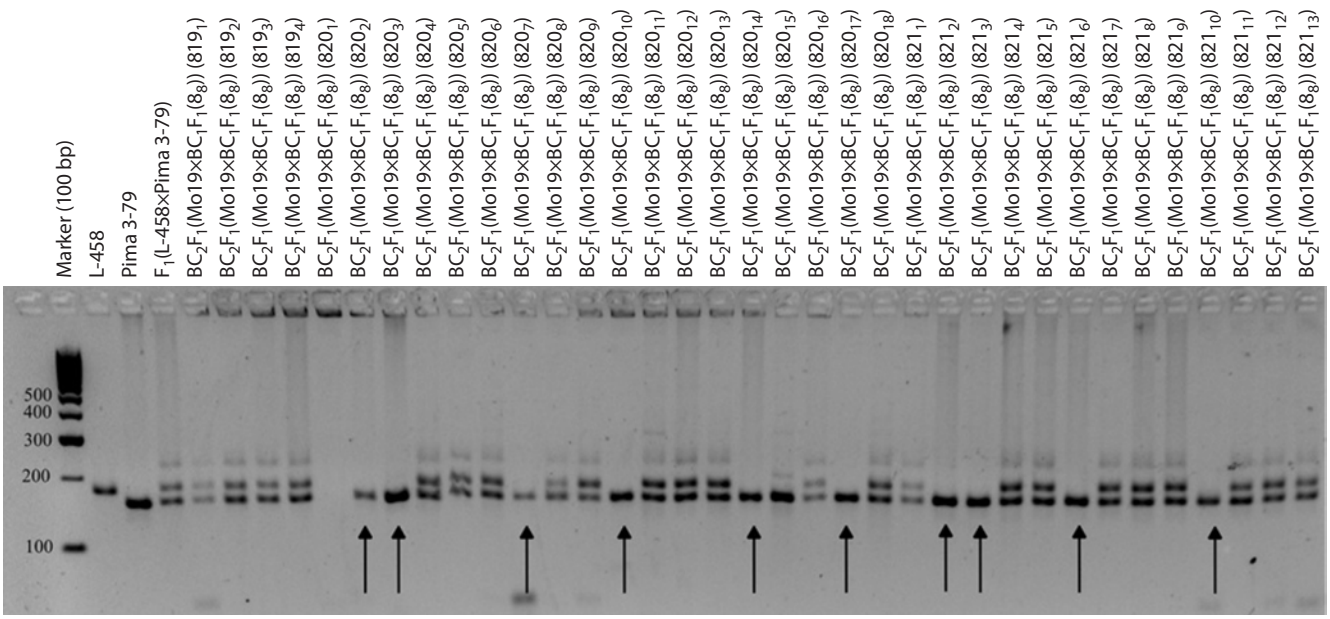


Fig. 6. Electropherogram of DNA amplicons of SSR markers in hybrid seedlings, BC_2F_1(Mo19×$BC_1F_1(8_8)$) on chromosome 2 of the A_t subgenome of cotton.

Marker JESPR179.

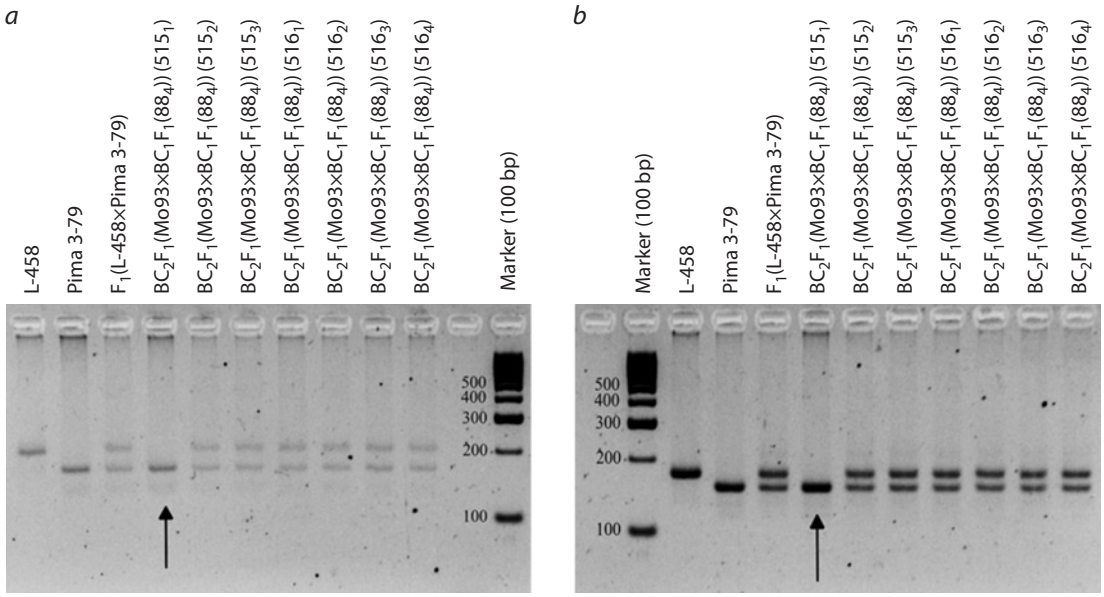


Fig. 7. Electropherogram of DNA amplicons of SSR markers in hybrid seedlings, BC_2F_1(Mo93×$BC_1F_1(88_4)$) on chromosome 2 of the A_t subgenome of cotton.

a – marker TMB0471; *b* – marker JESPR179.

chromosomes 2 of *G. barbadense*, revealed modal pairing of chromosomes with 25 bivalents and one univalent, confirming their monosomic status. Univalent size analysis in BC_2F_1 monosomics revealed a large chromosome 2 of *G. hirsutum* and a similar size in BC_2F_1 monosomic data with chromosome 2 of *G. barbadense*.

Characteristics of the transmission of chromosome 2 of the A_t subgenome of cotton during backcrossing of monosomic lines deficient in this chromosome by monosomic hybrids of different backcross generations. Molecular genetic analysis of backcrossed seedlings (at the stage of 3–5 true leaves), carried out via chromosome-specific molecular markers (SSR), allowed us to estimate the frequency and nature of the transmission of chromosome 2 from the A_t subgenome of cotton in BC_1F_1 hybrid plants. A study of five backcrossed BC_1F_1 families revealed that, in the BC_1F_1(Mo11×$F_1(766_3)$) variant, there was only one seedling (494_2) with the presence of the substituted chromosome, which indicated a very rare

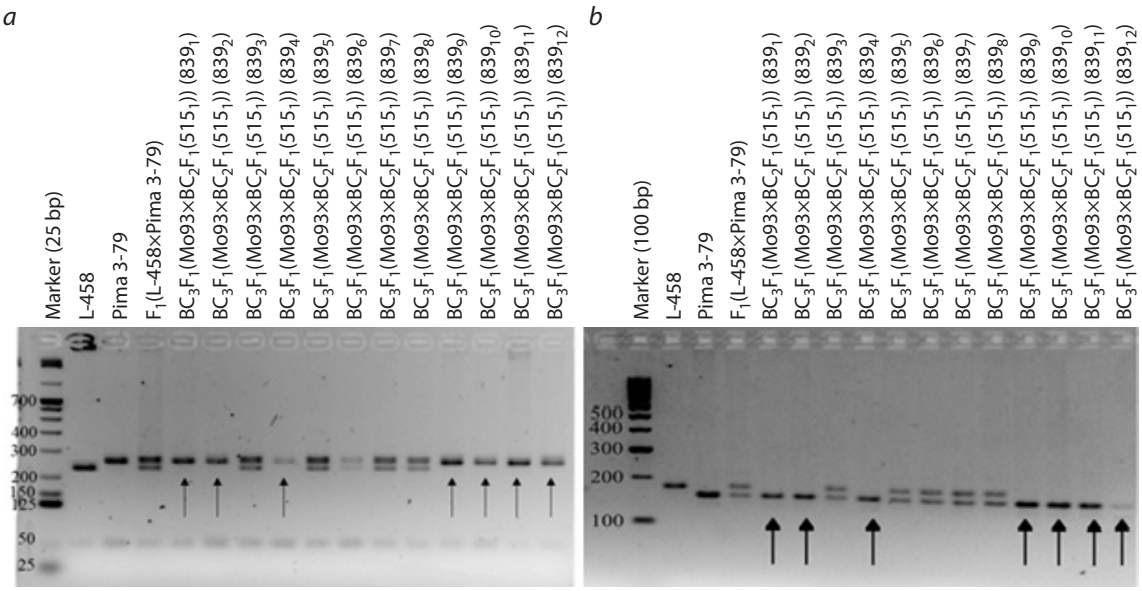


Fig. 8. Electropherogram of DNA amplicons of SSR markers in hybrid seedlings, $BC_3F_1(Mo93\times BC_2F_1(515_1))$ on chromosome 2 of the A_t subgenome of cotton.

a – marker DPL0674; *b* – marker JESPR179.

transmission frequency of chromosome 2 of the A_t subgenome of the *G. barbadense* species (3.13 %) (Fig. 9, Supplementary Material 20).

Among the BC_1F_1 hybrids in the $BC_1F_1(Mo16\times F_1(98_6))$ cross combination, seven seedlings with chromosome 2 substitutions were found, which were characterized by the presence of polymorphic alleles only from the *G. barbadense* species, indicating a relatively high frequency of transmission of chromosome 2 from the A_t subgenome of the *G. barbadense* species (21.21 %) in this variant. However, a study of the transmission of this chromosome in BC_2F_1 hybrids in the $BC_2F_1(Mo16\times BC_1F_1(923_7))$ variant revealed the presence of chromosome-specific alleles only from the *G. hirsutum* species, whereas alleles from the *G. barbadense* species were absent, which revealed the absence of chromosome 2 substitution in five hybrids studied. Repeated analysis of BC_2F_1 hybrids in two cross variants allowed us to identify four hybrids in one $BC_2F_1(Mo16\times BC_1F_1(923_7))$ variant with chromosome-specific alleles only from the *G. hirsutum* species, which drew attention to the repeated elimination of this chromosome, whereas in the other variant (922_8), the presence of chromosome 2 substitution from the *G. barbadense* species was observed. In general, in BC_2F_1 with the participation of the monosomic line Mo16, the frequency of transmission of the substituted chromosome was characterized by a decrease (up to 12.00 %) compared with that in BC_1F_1 hybrids, as well as elimination in the majority of hybrids (36.00 %). Moreover, elimination of the substituted chromosome occurred in one variant of crosses involving the same BC_1F_1 hybrid plant, which suggested the influence of genotype on the nature of chromosome transmission (Fig. 9, Supplementary Material 20).

Among the hybrids of one family in the $BC_1F_1(Mo19\times F_1(769_4))$ variant with a putative chromosome 2 substitution, a single hybrid plant (8_8) was identified that had an allele from *G. barbadense*, which indicated the transmission of the substituted chromosome 2 with a low frequency (12.50 %). However, when the transmission of this chromosome in BC_2F_1 hybrids in the $BC_2F_1(Mo19\times BC_1F_1(8_8))$ variant was studied, chromosome-specific alleles from the *G. barbadense* species were detected only in ten hybrids in three analysed families, which indicates greater transmission of the substituted chromosome 2 (28.57 %) than in BC_1F_1 hybrids (Fig. 9, Supplementary Material 20).

Nine hybrids with chromosome 2 substitutions were detected in the $BC_1F_1(Mo93\times F_1(516_4))$ variant, which was characterized by the presence of alleles from *G. barbadense* only, indicating a high frequency of transmission of this chromosome (42.86 %). Among the BC_2F_1 hybrids in the $BC_2F_1(Mo93\times BC_1F_1(88_4))$ variant, only one hybrid from two families had a substitution from *G. barbadense*, confirming the low frequency of chromosome 2 substitution (14.29 %). However, the study of the transmission of this chromosome in BC_3F_1 hybrids in the $BC_3F_1(Mo93\times BC_2F_1(515_1))$ variant made it possible to characterize the presence of chromosome-specific alleles from the *G. barbadense* species in seven hybrids out of 12 studied, i. e., the transmission of chromosome 2 was observed in more than half of the hybrids (58.33 %) (Fig. 9, Supplementary Material 20).

Morphobiological analysis of monosomic hybrids BC_1F_1, BC_2F_1, and BC_3F_1 obtained from crosses of the monosomic cotton lines *G. hirsutum* with chromosome 2 deficiency of the A_t subgenome with monosomic hybrids of different generations of backcrosses with chromosome substitution. Morphobiological analysis of backcross monosomic hybrids BC_1F_1 with chromosome 2 substitution of the A_t subgenome of cotton revealed short stature (up to 50 cm on average); reduced size of three-lobed leaves, buds, and flowers; low foliage; shortened internodes; and weak budding and flowering during the first year of vegetation (only 2–3 buds

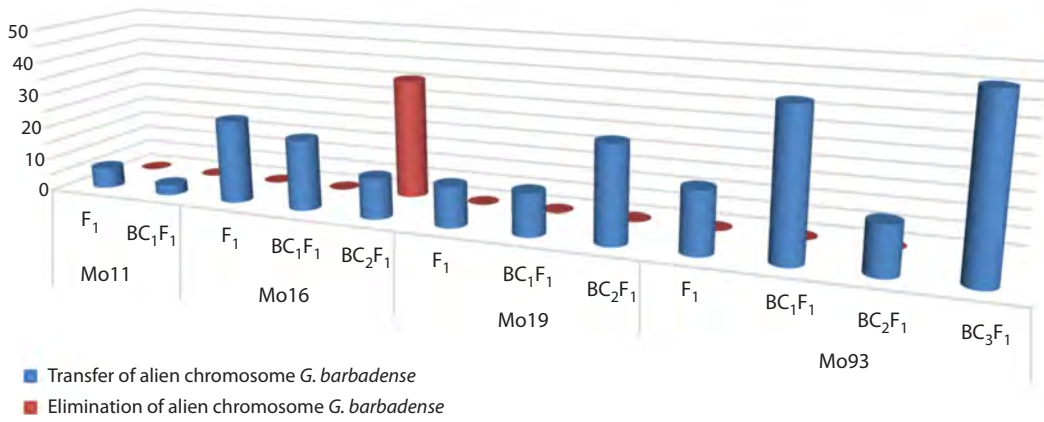


Fig. 9. Frequency of occurrence of monosomic hybrids with alien chromosome 2 of the A_t subgenome of cotton *G. barbadense*.

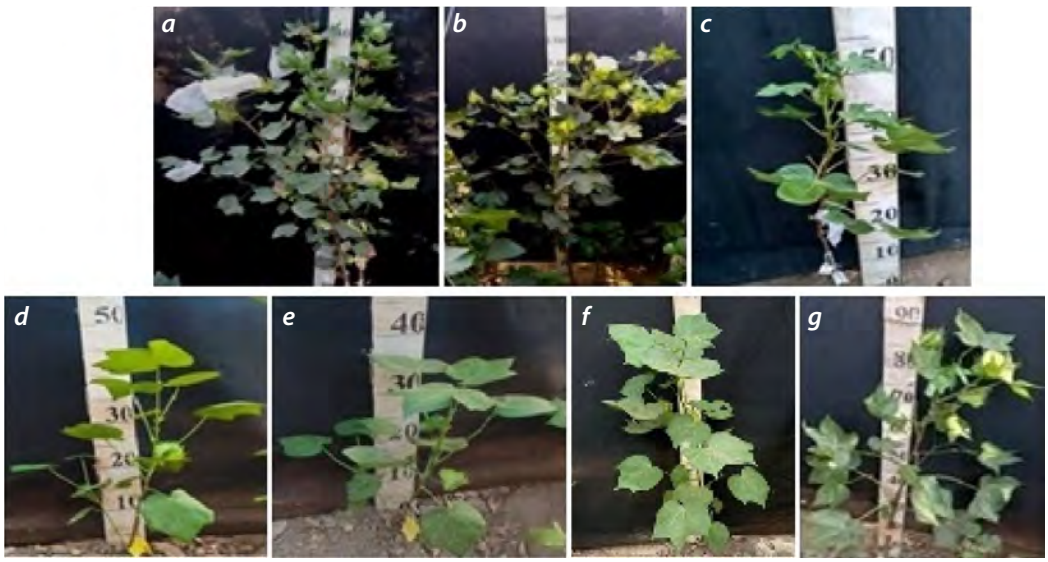


Fig. 10. Plants of the original lines and monosomic hybrids of cotton F_1 and BC_2F_1 obtained from crossings of the recurrent parent with monosomic hybrids.

a – monosomic line Mo16 for chromosome 2 of the A_t subgenome; *b* – F_1(Mo16×Pima 3-79) (98_2); *c–e* – BC_1F_1(Mo16×$F_1 98_6$) with substitution of chromosome 2 of the A_t subgenome of cotton *G. barbadense*: *c* – 923_7, *d* – 495_5, *e* – 497_4; *f* – BC_2F_1(Mo16×$BC_1F_1 922_8$) (817_4) with substitution of chromosome 2; *g* – BC_2F_1(Mo16×$BC_1F_1 923_7$) (21_1) without substitution of chromosome 2 of the A_t subgenome of cotton *G. barbadense*.

and flowers per plant) compared with the recurrent parent. The backcross hybrids with the participation of the monosomic line Mo16 (Fig. 10) were characterized by particularly short stature and a decrease in growth rates, whereas in the second and third years of vegetation, there was an increase in the size of the leaves, the number of buds and the number of opened bolls.

Compared with the monosomic hybrid BC_1F_1, with the substitution of chromosome 2, the complemented hybrids BC_2F_1 and BC_3F_1, with the substitution of chromosome 2 of the A_t subgenome of cotton, were distinguished by taller growth (up to 80 cm), medium foliage, larger three-lobed leaves, an increased number of flowers and medium-sized spherical bolls.

Discussion

Interspecific crosses of various tetraploid cotton species are characterized by high crossability and fertility of first-generation hybrids. However, in later generations, decreases in all these parameters and mass sterility are observed (Mauer, 1954; Abdullaev, 1974).

A study of the crossability and seed set of backcrossed seeds obtained from crosses of monosomic lines deficient in chromosome 2 of the A_t subgenome of cotton *G. hirsutum* with interspecific monosomic F_1 hybrids and backcrossed monosomic BC_1F_1 hybrids revealed a significant decrease in crossability in all the studied backcrossed hybrids, whereas hybrid seed set and germination decreased in most hybrids. This decrease was explained by the decrease in these para-

meters in the original monosomic lines due to the presence of a significant number of nullisomic gametes, as well as the hemizygosity of the maternal and paternal plants, which resulted in the presence of numerous unfertilized eggs in the form of motes and low germination of nullisomic seeds.

Molecular genetic methods using microsatellite sequence markers (SSRs), which are widespread in eukaryotic genomes and exhibit a high level of polymorphism, are widely used to identify alien introgression. To date, several collections of such markers have been created for cotton (BNL, JESPR, CIR, DPL, and NAU), and molecular genetic linkage maps have been constructed due to chromosomal specificity of many SSR markers (Liu et al., 2000; Gutiérrez et al., 2009; Yu et al., 2011; Saha et al., 2013, 2015; Wang et al., 2016; https://www.cottongen.org/data/download/ marker_origin). The presence or absence of microsatellite marker amplification products after polymerase chain reaction allows us to judge the presence or absence of an alien chromosome in the studied genotype. Therefore, our molecular genetic screening of backcross hybrids at the seedling stage allowed us to quickly identify the substitution of chromosome 2 of the A_t subgenome of cotton of the *G. barbadense* species.

Compared with direct cytological observation, PCR-based screening for chromosome-specific markers has been shown to be more productive and significantly more effective (Polgári et al., 2019). This is especially true when it is impossible to detect and monitor alien genetic material directly on cytological preparations via FISH and GISH hybridization.

The originality of our approach lies in the development of a new scheme for the targeted, planned substitution of specific chromosomes via monosomic lines with previously established univalent chromosome identities and chromosome-specific microsatellite markers (SSRs) at the seedling stage. During the creation of new chromosome substitution lines via this scheme, significant differences were identified between specific chromosomes in the cotton genome in terms of the frequency of transmission of alien chromosomes, their elimination at different stages of backcrossing, and the morphobiological characteristics of the backcrossed hybrids. This study presents data on the specific consequences of the transfer of genetic material from chromosome 2 of the A_t subgenome of cotton *G. barbadense* into the genome of cotton *G. hirsutum* via four monosomic lines.

The results of the study of backcross progenies BC_1F_1, BC_2F_1 and BC_3F_1 (in one variant of crossing) revealed that the frequency of transmission of chromosome 2 of the A_t subgenome of cotton, *G. barbadense* depended on both the genotype of the monosomic line and the genotype of the backcross hybrid used in the crossing. Thus, the level of transmission of chromosome 2 of the A_t subgenome of *G. barbadense* in the first backcross generation, with the participation of four monosomic lines, varied within a wide range (from 3.13 to 42.86 %), whereas in the BC_2F_1 generation, where only three monosomic lines were involved, the transmission frequency was noticeably reduced and varied within a narrower range (from 12.00 to 28.57 %). However, in the BC_3F_1 generation, which included only one monosomic line (Mo93), the highest frequency of chromosome 2 transmission (58.33 %) was observed in more than half of the studied hybrids. Moreover, this line was characterized by the highest average number of substitutions per hybrid genome among all four studied lines. In addition, the elimination of chromosome 2 of the A_t subgenome was detected in the BC_2F_1 variant (Mo16×BC_1F_1 (923_7)), which involved the same paternal backcross monosomic hybrid plant (923_7) during two experiments, indicating the preferential elimination of chromosome 2 in the studied backcross progenies and the influence of a specific paternal genotype on this process. Therefore, the analysis of the transmission characteristics of alien chromosome 2 revealed differences in the competitiveness of this chromosome when it is transmitted to offspring in the genotypic environment of four monosomic lines.

The success of substitution depends on how well the alien chromosome compensates for the missing chromosome, since it is difficult to assume that each alien chromosome could compensate equally for the absence of homeologous chromosomes (Morris, Sears, 1970).

Previously, a dependence of rye chromosome introgression on the genotypic environment of the recipient was discovered, since differences in the frequency of rye chromosomes were observed in the F_2 hybrid population between a wheat-rye substitution line for chromosomes 1R+2R and the winter wheat varieties Holme and Kraka (Merker, Forsstrom, 2000). Notably, the genotype of the variety also influenced the frequency of telocentric formation on the rye chromosome and T2R.2DL translocation (Krasilova et al., 2011).

Cases of complete uniparental elimination of chromosomes from the entire genome are widely known (Ishii et al., 2016; Dedukh, Krasikova, 2021). They result from hybridization between distantly related species as an element of protecting the integrity of the genome from "genomic shock" (McClintock, 1984). An example is the preferential elimination of chromosomes from the entire D genome in first-generation hybrids during wheat-rye hybridization (Li et al., 2015).

Examples of partial elimination of individual chromosomes are less numerous, but it is known that in each specific case, different mechanisms of chromosome elimination operate in interspecific hybrids, depending on the specific species involved. Thus, an assessment of a large population of wheat-barley hybrids via genomic *in situ* hybridization (GISH) and simple sequence repeat (SSR) markers revealed the absence of preference for the elimination of individual barley chromosomes compared with wheat chromosomes (Polgári et al., 2019). A study of the transmission of chromosome 5R through the gametes of wheat-rye dimonosomics 5R5D revealed a significantly lower competitive ability of this chromosome in transmission to offspring and its preferential elimination (Silkova et al., 2011). Selective elimination of chromosomes in *Hordeum bulbosum* L. is associated with the loss of one of the types of histone protein H3 (CENH3) in the centromere, leading to its inactivity and absence of chromosome attachment to the mitotic spindle, as well as the formation of micronuclei and their degeneration (Sanei et al., 2011). A comparative analysis of the nucleotide sequences of the centromeric histone CENH3 genes in wheat-rye allopolyploids of different ploidy levels revealed increased expression of rye CENH3 variants, which is associated with the maintenance of a viable hybrid genome (Evtushenko et al., 2019). Using bread wheat as an example,

a wide range of features of alien chromatin introgression was demonstrated, which represents significant potential for gene pool enrichment (Adonina et al., 2021).

Although screening for the presence of alien chromosomes in backcrossed cotton progenies via molecular genetic markers made it possible to detect specific consequences of the introgression of chromosome 2 of the A_t subgenome of the *G. barbadense* species at different stages of backcrossing, the study of the behavior of univalent chromosomes at the MI stage of meiosis revealed the similarity of pairing in backcrossed monosomes with the univalent chromosome 2 of the A_t subgenome of the cotton species *G. hirsutum* and the univalent chromosome 2 of the A_t subgenome of the *G. barbadense* species.

The backcrossed BC_1F_1 monosomic strains we examined were characterized by a general delay in the development of vegetative and generative organs, manifested by stunted growth, reduced foliage, and poor budding and flowering during the first year of vegetative development. This hampered their study and backcrossing and significantly complicated and delayed the creation of chromosome substitution forms. However, during the second year of vegetative development, the backcrossed plants showed normalization of vegetative and generative organ development.

Similar dwarfism was observed in the monosomic alien complemented cotton line *G. hirsutum/G. bickii* Proch. with a 10 Gb chromosome substitution (MAAL), created through crosses with an amphidiploid ($2n$ = 78, $AADDG_1G_1$) and chromosome-specific SSR markers (Tang et al., 2018). It can be assumed that such changes in morphobiology occurred due to hybrid genome reorganization and introgression of alien chromatin.

Conclusion

This study demonstrated the negative consequences of the introgression of chromosome 2 of the A_t subgenome of cotton, *G. barbadense*, into the genome of cotton, *G. hirsutum*, involving four monosomic lines. These negative consequences include decreased crossability, seed set, and germination of hybrid seeds; a wide variation in the transmission level of alien chromosome 2 (from 3.13 to 42.86 %) in BC_1F_1, but a less narrow one (from 12.00 to 28.57 %) in BC_2F_1; elimination of chromosome 2 of the A_t subgenome of the *G. barbadense* species in the $BC_2F_1(Mo16\times BC_1F_1(923_7))$ variant with the same paternal genotype, which indicated the influence of a specific paternal genotype on this process; dwarfism; and reduced foliage, weak budding and flowering in the first year of vegetative development but an increase in the number of buds, flowers and bolls in monosomic hybrids of the following backcross generations.

References

Abdullaev A.A. Evolution and Taxonomy of Polyploid Species of Cotton. Tashkent, 1974 (in Russian)

Adonina I.G., Timonova E.M., Salina E.A. Introgressive hybridization of common wheat: results and prospects. *Russ J Genet*. 2021;57(4):390-407. doi 10.1134/S1022795421030029

Dedukh D., Krasikova A. Delete and survive: strategies of programmed genetic material elimination in eukaryotes. *Biol Rev Camb Philos Soc*. 2021;97(1):195-216. doi 10.1111/brv.12796

Endrizzi J.E., Turcotte E.L., Kohel R.J. Genetics, cytology and evolution of *Gossypium*. *Adv Genet.* 1985;23:271-375. doi 10.1016/S0065-2660(08)60515-5

Evtushenko E.V., Lipikhina Y.A., Stepochkin P.I., Verchinin A.V. Cytogenetic and molecular characteristics of rye genome in octoploid triticale (× *Triticosecale* Wittmack). *Comp Cytogenet*. 2019;13(4):423-434. doi 10.3897/compcytogen.v13i4.39576

Fang D.D., Thyssen G.N., Wang M., Jenkins J.N., McCarty J.C., Jones D.C. Genomic confirmation of *Gossypium barbadense* introgression into *G. hirsutum* and a subsequent MAGIC population. *Mol Genet Genomics*. 2023;298(1):143-152. doi 10.1007/s00438-022-01974-3

Gutiérrez O.A., Stelly D.M., Saha S., Jenkins J.N., McCarty J.C., Raska D.A., Scheffler B.E. Integrative placement and orientation of non-redundant SSR loci in cotton linkage groups by deficiency analysis. *Mol Breeding*. 2009;23:693-707. doi 10.1007/s11032-009-9266-y

Ishii T., Karimi-Ashtiyani R., Houben A. Haploidization via chromosome elimination: means and mechanisms. *Annu Rev Plant Biol*. 2016;67:421-438. doi 10.1146/annurev-arplant-043014-114714

Jenkins J.N., Wu J., McCarty J.C., Saha S., Gutiérrez O., Hayes R., Stelly D.M. Genetic effects of thirteen *Gossypium barbadense* L. chromosome substitution lines in top crosses with upland cotton cultivars: I. Yield and yield components. *Crop Sci*. 2006;46(3):1169-1178. doi 10.2135/cropsci2005.08-0269

Jenkins J.N., McCarty J.C., Wu J., Saha S., Gutiérrez O., Hayes R., Stelly D.M. Genetic effects of thirteen *Gossypium barbadense* L. chromosome substitution lines in top crosses with upland cotton cultivars: II. Fiber quality traits. *Crop Sci*. 2007;47(2):561-572. doi 10.2135/cropsci2006.06.0396

Krasilova N.M., Adonina I.G., Silkova O.G., Shumny V.K. Transmission of rye chromosome 2R in backcrosses of wheat-rye 2R (2D substitution lines to various common wheat varieties. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2011;15(3):554-562 (in Russian)

Li H., Guo X., Wang Ch., Ji W. Spontaneous and divergent hexaploid triticales derived from common wheat × rye by complete elimination of D-genome chromosomes. *PLoS One*. 2015;10(3):e0120421. doi 10.1371/journal.pone.0120421

Liu S., Saha S., Stelly D.M., Burr B., Cantrell R.G. Chromosomal assignment of microsatellite loci in cotton. *Hered.* 2000;91(4):326-332. doi 10.1093/jhered/91.4.326

Mauer F.M. Origin and Taxonomy of Cotton. Vol. 1. Cotton. Tashkent, 1954 (in Russian)

McClintock B. The significance of responses of the genome to challenge. *Science*. 1984;226(4676):792-801. doi 10.1126/science.15739260

Merker A., Forsstrom P.O. Isolation of mildew resistant wheat-rye translocations from a double substitution line. *Euphytica*. 2000;115:167-172. doi 10.1023/A:1004018500970

Morris E.R., Sears E.R. Cytogenetics of Wheat and Related Forms. Wheat and its Improvement. Moscow: Kolos Publ., 1970 (in Russian)

Polgári D., Mihók E., Sági L. Composition and random elimination of paternal chromosomes in a large population of wheat-barley (*Triticum aestivum* L. × *Hordeum vulgare* L.) hybrids. *Plant Cell Rep*. 2019;38:767-775. doi 10.1007/s00299-019-02405-1

Saha S., Wu J., Jenkins J.N., McCarty J.C., Jr., Gutiérrez O.A., Stelly D.M., Percy R.G., Raska D.A. Effect of chromosome substitutions from *Gossypium barbadense* L. 3-79 into *G. hirsutum* L. TM-1 on agronomic and fiber traits. *J Cotton Sci.* 2004;8(3):162-169

Saha S., Raska D.A., Stelly D.M. Upland cotton (*Gossypium hirsutum* L.) × Hawaiian cotton (*G. tomentosum* Nutt. ex Seem.) F_1 hybrid hypoaneuploid chromosome substitution series. *J Cotton Sci*. 2006;10(4):263-272

Saha S., Wu J., Jenkins J.N., McCarty J.C., Hayes R., Stelly D.M. Genetic dissection of chromosome substitution lines of cotton to discover novel *Gossypium barbadense* L. alleles for improvement of agronomic traits. *Theor Appl Genet*. 2010;120(6):1193-1205. doi 10.1007/s00122-009-1247-3

Saha S., Raska D.A., Stelly D.M., Manchali Sh., Gutiérrez O.A. Hypoaneuploid chromosome substitution F_1 hybrids of *Gossypium hirsutum* L. × *G. mustelinum* Miers ex Watt. *J Cotton Sci*. 2013; 17(2):102-114

Saha S., Stelly D.M., Makamov A.K., Ayubov M.S., Raska D., Gutiérrez O.A., Shivapriya M., Jenkins J.N., Dewayne D., Abdurakhmonov I.Y. Molecular confirmation of *Gossypium hirsutum* chromosome substitution lines. *Euphytica*. 2015;205:459-473. doi 10.1007/s10681-015-1407-2

Saha S., Bellaloui N., Jenkins J.N., McCarty J.C., Stelly D.M. Effect of chromosome substitutions from *Gossypium barbadense* L., *G. tomentosum* Nutt. ex Seem and *G. mustelinum* Watt into *G. hirsutum* L. on cotton seed protein and oil content. *Euphytica*. 2020;216:118. doi 10.1007/s10681-020-02644-4

Sanamyan M.F., Bobokhujaev Sh.U. Identification of univalent chromosomes in monosomic lines of cotton (*Gossypium hirsutum* L.) by means of cytogenetic markers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2019;23(7): 836-845. doi 10.18699/VJ19.557 (in Russian)

Sanamyan M.F., Petlyakova J., Rakhmatullina E.M., Sharipova E. Cytogenetic collection of Uzbekistan. In: Abdurakhmonov I.Y. (Ed.) World Cotton Germplasm Resources. InTech, 2014;247-287. doi 10.5772/58589

Sanamyan M.F., Bobokhujaev Sh.U., Makamov A.X., Achilov S.G., Abdurakhmonov I.Y. The creation of new aneuploid lines of the cotton (*Gossypium hirsutum* L.) with identification of chromosomes by translocation and SSR markers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2016a;20(5): 643-652. doi 10.18699/VJ16.186 (in Russian)

Sanamyan M.F., Makamov A.K., Bobokhujaev Sh.U., Usmonov D.E., Buriev Z.T., Saha S., Stelly D.M. The utilization of translocation lines and microsatellite markers for the identification of unknown cotton monosomic lines. In: Abdurakhmonov I.Y. (Ed.) Cotton Research. InTech, 2016b;167-183. doi 10.5772/64558

Sanamyan M.F., Bobokhujaev Sh.U., Abdukarimov Sh.S., Makamov Kh.A., Silkova O.G. Features of chromosome introgression from *Gossypium barbadense* L. into *G. hirsutum* L. during the development of alien substitution lines. *Plants*. 2022;11(4):542. doi 10.3390/plants11040542

Sanamyan M.F., Bobokhujaev Sh.U., Abdukarimov Sh.S., Silkova O.G. Molecular-genetic and cytogenetic analyses of cotton chromosome introgression from *Gossypium barbadense* L. into the genome of *G. hirsutum* L. in BC_2F_1 hybrids. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(8):958-970. doi 10.18699/VJGB-23-110

Sanei M., Pickering R., Kumke K., Nasuda S., Houben A. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc Natl Acad Sci USA*. 2011;108(33):E498-E505. doi 10.1073/pnas.1103190108

Shchapova A.I., Kravtsova L.A. Cytogenetics of Wheat-Rye Hybrids. Novosibirsk, 1990 (in Russian)

Silkova O.G., Leonova I.N., Krasilova N.M., Dubovets N.I. Preferential elimination of chromosome 5R of rye in the progeny of 5R5D dimonosomics. *Russ J Genet*. 2011;47(8):942-950. doi 10.1134/S1022795411080151

Tang D., Feng S., Li S., Chen Yu., Zhou B. Ten alien chromosome additions of *Gossypium hirsutum–Gossypium bickii* developed by integrative uses of GISH and species-specific SSR markers. *Mol Genet Genomics*. 2018;293(4):1-11. doi 10.1007/s00438-018-1434-5

Ulloa M., Wang C., Saha S., Hutmacher R.B., Stelly D.M., Jenkins J.N., Burke J., Roberts P.A. Analysis of root-knot nematode and fusarium wilt disease resistance in cotton (*Gossypium* spp.) using chromosome substitution lines from two alien species. *Genetica*. 2016;144:167-179. doi 10.1007/s10709-016-9887-0

Wang P.Z., Bian L., Cao C. Mapping of nuclear male-sterile gene ms_{14} using SSR markers in cotton. *Mol Plant Breed*. 2016;7:35. doi 10.5376/mpb.2016.07.0035

Wendel J.F. New World tetraploid cottons contain Old World cytoplasm. *Proc Natl Acad Sci USA*. 1989;86(11):4132-4136. doi 10.1073/pnas.86.11.4132

Yu Yu., Yuan D., Liang Sh., Li X., Wang X., Lin Zh., Zhang X. Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC_1 population between *Gossypium hirsutum* and *G. barbadense*. *BMC Genomics*. 2011;12:15. doi 10.1186/1471-2164-12-15

Conflict of interest. The authors declare no conflict of interest.
Received August 19, 2025. Revised November 7, 2025. Accepted November 7, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1195-1202

doi 10.18699/vjgb-25-126

Intrapopulation changes in *Puccinia hordei* induced by two-component fungicides from different chemical classes

M.S. Gvozdeva, O.A. Kudinova, V.D. Rudenko, G.V. Volkova

Federal Research Center of Biological Plant Protection, Krasnodar, Russia
alosa@list.ru

Abstract. Fungicide resistance is a global problem that reduces the effectiveness and duration of action of these compounds due to changes in the racial composition and virulence of phytopathogen populations. Currently, resistance to 100 active substances has been registered in more than 230 fungal plant pathogens. Leaf rust of barley (*Puccinia hordei* Otth.) is one of the most widespread and harmful pathogens in the barley pathocomplex; it is recorded in southern Russia every year. There are very few studies on the effect of fungicides on the characteristics of rust fungi populations, and none have been carried out on *P. hordei* in Russia. This research aimed to analyze the effect of fungicides belonging to the chemical classes of triazoles and strobilurins on intrapopulation changes in *P. hordei* in terms of pathogenicity (virulence and aggressiveness) under the conditions of the North Caucasus region of Russia. Two-component fungicides approved for use in the Russian Federation were selected for the study: Delaro, SC; Amistar Extra, SC; Amistar Gold, SC. Plants were treated using several application rates: 50, 100, 150 and 200 % (the recommended application rate was determined to be 100 %). Treatment of winter barley plants with fungicides with different application rates revealed intrapopulation changes in the virulence structure of *P. hordei*. In all treatment variants, the frequency of isolates virulent to the *Rph4*, *Rph5*, *Rph6+2*, *Rph12* genes decreased with increasing fungicide application rate and the frequency of isolates virulent to *Rph14* increased. No isolates virulent to *Rph7* were found in either the original population or the experimental variants. The average virulence of the fungal populations treated with the fungicides in all experimental variants was lower compared to the original population (no treatment (48.5 %)) and depending on the application rate varied from 33.8 % (Amistar Gold, 50 %) to 28.5 % (Amistar Gold, 200 %). Under the influence of the increased application rates of the fungicides, an increase in the duration of the latent period was observed: from 168 h (original population) to 216 h (Delaro, Amistar Gold, 200 %). A decrease in sporulation ability (spore mass per pustule ranged from 0.013 mg (original population) to 0.002 mg (Delaro, Amistar Gold, 200 %)) and in the viability of *P. hordei* (from 100 % for the original population to 22.5 % in Amistar Gold, 200 % treatment) was found under the action of the fungicides. Thus, a fungicide-treated *P. hordei* population is characterized by intrapopulation changes in aggressiveness and virulence, which can significantly increase barley yield losses due to a decrease in the effectiveness of chemical protection, as well as an increase in the harmfulness of the pathogen.

Key words: barley; leaf rust of barley; fungicides; *Puccinia hordei*; resistance; pathogenicity; population

For citation: Gvozdeva M.S., Kudinova O.A., Rudenko V.D., Volkova G.V. Intrapopulation changes in *Puccinia hordei* induced by two-component fungicides from different chemical classes. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1195-1202. doi 10.18699/vjgb-25-126

Funding. The research was carried out within the framework of the grant of the Russian Science Foundation No. 23-76-10063, https://rscf.ru/project/23-76-10063/

Внутрипопуляционные изменения *Puccinia hordei* под воздействием двухкомпонентных фунгицидов различных химических классов

М.С. Гвоздева, О.А. Кудинова, В.Д. Руденко, Г.В. Волкова

Федеральный научный центр биологической защиты растений, Краснодар, Россия
alosa@list.ru

Аннотация. Резистентность к фунгицидам является мировой проблемой, обуславливающей снижение их эффективности и срока действия вследствие изменения расового состава и вирулентности популяций фитопатогенов. В настоящее время зарегистрирована резистентность более чем у 230 грибных патогенов растений к 100 действующим веществам. Возбудитель карликовой ржавчины (*Puccinia hordei* Otth.) – один из распространенных и вредоносных в патокомплексе ячменя, в Южном федеральном округе, фиксируется на посевах ежегодно. Исследований по влиянию фунгицидов на характеристики популяции ржавчинных грибов

крайне мало, а в отношении *P. hordei* в России не проводилось. Цель данной работы – проанализировать влияние фунгицидов из химических классов триазолов и стробилуринов на внутрипопуляционные изменения *P. hordei* по показателям патогенности (вирулентности и агрессивности) в условиях Северо-Кавказского региона России. Для исследования были выбраны двухкомпонентные фунгициды, разрешенные к применению на территории Российской Федерации: Деларо, КС; Амистар Экстра, СК; Амистар Голд, СК. Обработку растений проводили с использованием нескольких норм применения препаратов: 50, 100, 150 и 200 % (рекомендуемая норма применения принята за 100 %). При обработке растений ячменя озимого фунгицидами с различной нормой применения выявлены внутрипопуляционные изменения в структуре *P. hordei* по вирулентности. Во всех вариантах с увеличением нормы применения фунгицидов происходили снижение частоты изолятов, вирулентных к генам *Rph4*, *Rph5*, *Rph6+2*, *Rph12*, и увеличение частоты изолятов, вирулентных к *Rph14*. Не обнаружено изолятов, вирулентных к *Rph7*, как в исходной популяции, так и в вариантах опыта. Средняя вирулентность популяций гриба, обработанных фунгицидами, во всех вариантах опыта была ниже по сравнению с вирулентностью исходной популяции (без обработки) (48.5 %) и в зависимости от нормы применения варьировала от 33.8 (Амистар Голд, 50 %) до 28. 5 % (Амистар Голд, 200 %). Под действием повышенных норм применения препаратов отмечено увеличение длительности латентного периода развития инфекции – от 168 ч (исходная популяция) до 216 ч (Деларо, Амистар Голд, 200 %). Под действием фунгицидов установлено снижение спорулирующей способности (масса спор с одной пустулы варьировала от 0.013 (исходная популяция) до 0.002 мг (Деларо, Амистар Голд, 200 %) и жизнеспособности *P. hordei* (от 100 % для исходной популяции до 22.5 % при обработке Амистар Голд, 200 %). Таким образом, популяция *P. hordei*, подвергавшаяся обработке фунгицидами, характеризуется внутрипопуляционными изменениями по показателям агрессивности и вирулентности, что может существенно увеличивать потери урожая ячменя вследствие как снижения эффективности химической защиты, так и повышения вредоносности патогена.

Ключевые слова: ячмень; карликовая ржавчина; фунгициды; *Puccinia hordei*; резистентность; патогенность; популяция

Introduction

Southern Russia is the leader in terms of winter grain crops area. According to the Federal State Statistics Service (https://rosstat.gov.ru/storage/mediabank/posev-4cx_2024.xlsx), in 2024 their share in total crops was 51.3 %. Winter barley is a promising crop, valuable for livestock farming, capable of producing a consistently high yield even in the extremely dry conditions of southern Russia (Ereshko et al., 2012). In 2024, the Krasnodar Region was the leader in barley harvesting: 1,159.5 thousand tons were harvested, which is 7 % of the total amount in the country (https://graininfo.ru/news/yachmen-ploshchadi-sbory-i-urozhaynost-v-rossii-v-2024-godu/).

Barley crops are affected by various pathogens that cause yield failure and loss of grain quality (Abbas, 2022). The leaf rust of barley pathogen, the biotrophic basidiomycete *Puccina hordei* Otth., is one of the widespread and harmful pathogens in the barley pathocomplex (Sapkota et al., 2023). In the North Caucasus region of Russia, leaf rust on barley crops is recorded annually, and epiphytoties occur once or twice every 10 years (Volkova et al., 2018). Ensuring high yields is impossible without chemical protection. In addition to the high pesticide load on agrobiocenosis, a serious problem is the emergence of pathogen forms resistant to the active substances of fungicides (Shcherbakova, 2019).

Fungicide resistance has become a major global problem, reducing the efficacy and shelf-life of some promising fungicides (Brent, Hollomon, 1995; Thind, 2021). Currently, resistance to 100 active substances has been reported in more than 230 fungal plant pathogens in various crops and geographic regions (FRAC, 2020). According to P.E. Russell (2003), rust fungi are pathogens with a low risk of emerging resistance, but their rapid life cycle, airborne dispersal of spores, and mixed mode of reproduction can cause intrapopulation changes (Ji et al., 2023) due to the rapid spread and accumulation of resistant forms, which will lead to decreased sensitivity to fungicides.

Traditionally, fungicides of the triazole and strobilurin classes are used to control rust diseases of wheat and barley (Walters et al., 2012). Triazoles belong to the largest class of fungicides with the demethylation inhibitor (DMI) group, suppressing the biosynthesis of ergosterol, a key component of the plasma membrane of fungal cells (Lass-Flörl, 2011). According to FRAC (2020), there is a moderate risk of emerging resistance to such fungicides. At the same time, a number of studies have noted changes in the population structure of biotrophic pathogens. For example, G. Zhan et al. (2022), studying the susceptibility to triademiphon in 446 isolates of *Puccinia graminis*, found that fungicide-resistant isolates showed strong adaptive traits in terms of urediniospore germination rate, latent period, sporulation intensity and lesion spread rate.

Strobilurins are an equally large class of fungicides that accumulate in the waxy layer of the leaf cuticle after plant treatment (Krupen'ko, 2023). Globally, strobilurins account for 20–25 % of total fungicide sales, a third of which is attributed to azoxystrobin, the best-selling fungicide in the world (Leadbeater, 2012). The first strobilurin-resistant isolates were detected in 1998 in *Blumeria graminis* [DC.] in Germany two years after their use, and strobilurin resistance has now been reported among both biotrophic (Dodhia et al., 2021) and hemibiotrophic (Ölmez et al., 2023) pathogens worldwide.

There is also a risk of emerging fungicide resistance for barley leaf rust (Walters et al., 2012). Like all rust fungi, *P. hordei* is a rapidly evolving phytopathogen (Çelik Oğuz, Karakaya, 2021). There are very few studies on the effect of fungicides on the characteristics of the rust fungi population. For the first time in Russia, such work was carried out at the Federal State Budgetary Scientific Institution "Federal Research Center of Biological Plant Protection" (FSBSI FRCBPP) for the "wheat–yellow rust pathogen" pathosystem (Volkova, 2007). The effect of a triadimefon-based fungicide on the virulence and aggressiveness of *P. striiformis* was studied, the rate of

emergence of resistant forms of the pathogen was calculated, and an anti-resistant strategy for the use of fungicides with this active substance on wheat crops was developed and proposed. Further studies were carried out on the wheat leaf rust pathogen (Gvozdeva, Volkova, 2022). It was found that the population of *P. triticina* treated with a chemical fungicide based on tebuconazole is characterized by a change in the structure of aggressiveness and virulence and a decrease in sensitivity to the toxicant. Similar studies on barley leaf rust have not been carried out either in the world or in Russia. Given the high virulence and variability of the North Caucasian population of the pathogen (Danilova, Volkova, 2022), and the need to develop an anti-resistant strategy for each pathogen and cenosis (Corkley et al., 2021), research on this issue is extremely relevant.

This research is aimed at analyzing the effect of fungicides of the chemical classes of triazoles and strobilurins on intrapopulation changes in *P. hordei* in terms of pathogenicity (virulence and aggressiveness) under the conditions of the North Caucasus region of Russia.

Materials and methods

The studies were carried out in the laboratory and greenhouse of the FSBSI FRCBPP on the winter barley variety Vivat, susceptible to leaf rust. Originator of the variety is FSBSI "Donskoy Agrarian Scientific Center"; it is recommended for growing in the North Caucasus region and resistant to lodging and frost.

To obtain spore material of the North Caucasian population of the barley leaf rust pathogen (further in the text – fungal urediniospores, or population), infected leaves were collected during a route survey of industrial barley crops in the Krasnodar Region, Stavropol Region, Rostov Region and the Republic of Adygea. Then, the susceptible Vivat variety was inoculated with a mixture of herbarium samples (Fungal Pathogens of Grain Ear Crops…, 2024). The spore material collected in the required quantity was stored at a temperature of +4 °C.

Two-component fungicides approved for use in the Russian Federation were selected for the study: Delaro, SC (175 g/l prothioconazole + 150 g/l trifloxystrobin); Amistar Extra, SC (200 g/l azoxystrobin + 80 g/l cyproconazole); Amistar Gold, SC (125 g/l azoxystrobin + 125 g/l difenoconazole).

Inoculation of winter barley plants with a spore suspension of the pathogen was carried out in the seedling phase. The infected plants were kept in a humid chamber for 18 hours, then they were grown under controlled conditions at a temperature of +20–22 °C, air humidity of 70–80 %, and light intensity of 10–15 thousand lux with a day and night cycle (16/8 hours) (Fungal Pathogens of Grain Ear Crops…, 2024). When the first signs of the disease appeared, spraying was carried out using several application rates of the preparations: 50, 100, 150 and 200 % (the recommended application rate was determined to be 100 %).

Urediniospores of *P. hordei*, collected from barley plants treated with fungicides at various application rates, were transferred to intact plants for reproduction and determination of aggressiveness indices (spore viability, latent period duration, sporulating capacity, sporulation duration). Aggressiveness indices were determined for the mixed pathogen populations obtained as described above.

The viability of barley leaf rust spores was tested in a humid chamber under a microscope by counting the total number of fungal spores and germinated spores (Sanin et al., 1975). The duration of the latent period was counted from the moment of inoculation until the first signs of the disease appeared (Pyzhikova, 1972). The sporulating capacity was determined by calculating the ratio of the number of pustules to the mass of the collected biomaterial. The duration of sporulation was determined from the beginning of the pustules opening until the end of sporulation (Sanin et al., 1975).

The virulence of the *P. hordei* populations treated with fungicides at different application rates was determined by the reaction of 15 barley differentiators from the International and Australian sets containing known genes of resistance to the pathogen: Sudan (*Rph1*), Peruvian (*Rph2*), Estate (*Rph3*), Gold (*Rph4*), Magnif 104 (*Rph5*), Bolivia (*Rph6+2*), Cebada Capa (*Rph7*), Egypt 4 (*Rph8*), Abyssinian (*Rph9*), Triumph (*Rph12*), PI 531849 (*Rph13*), PI 584760 (*Rph14*), Prior (*Rph19*), Ricardo (*Rph21+2*), Fong Tien (*Rph25*). The infection types were recorded 10–12 days after inoculation using the Levin and Cherevik scale in points (Volkova et al., 2018). The frequencies of virulence genes were calculated using the probabilistic method (Wolfe, Schwarzbach, 1975) based on the ratio of the number of pustules with infection type of 3–4 points on the lines with known resistance genes to the number of pustules on the susceptible variety Vivat. To evenly distribute *P. hordei* spores on barley leaves, inoculation was carried out in a precipitation tower using lycopodium, which was mixed with fungal spores in a ratio of 400:1 (Parlevliet, 1980).

Based on the results of differentiation, the average virulence was calculated (Mihajlova et al., 2003). Differences between the original population (without treatment) and populations treated with different application rates of the preparation in terms of virulence gene frequencies were determined using the Nei index (Nei, 1972). The statistical reliability of the aggressiveness indicators was assessed using the Fisher criterion ($\alpha = 0.05$) using STATISTICA 10.0 software.

The research used the material and technical base of the Unique Scientific Installation (USI) "Phytotron for Isolation, Identification, Study and Maintenance of Races, Strains, Phenotypes of Pathogens" (https://fncbzr.ru/brk-i-unu/unique-installation-2/) and objects of the bioresource collection of the Federal State Budgetary Scientific Institution FNCBZR "State Collection of Entomoacariphages and Microorganisms" (https://fncbzr.ru/brk-i-unu/unique-installation-1/).

Results

When treating winter barley plants with fungicides at different application rates, intrapopulation changes in the structure of *P. hordei* by virulence were revealed (Table 1).

In the option with the Amistar Gold, SC preparation, with an increase in the rate to 200 % of the recommended one, a decrease in the virulence of the pathogen was observed on varieties containing the resistance genes *Rph2*, *Rph6+2*, *Rph9*, *Rph14*. On the line with the *Rph13* and *Rph25* genes, when

treating the barley plants with the preparations at an application rate of 50 % of the recommended one, no damage was observed. With an increase in the rate to 200 %, the type of reaction to infection increased to 3 points.

Treatment of the barley plants with the fungicide Amistar Extra, SC affected the fungus population towards increasing virulence to the *Rph14* gene, while a heterogeneous infection type was observed, pustules of different sizes, areas of dead and chlorotic tissue were present on the leaf.

In the option with the fungicide Delaro, SC, a decrease in the virulence of the *P. hordei* population to the *Rph14* gene was noted. The reaction type decreased from 3 (application rate 50 % of the recommended one) to 0 points (application rate 50 % of the recommended one), for the original population this indicator was 1 and 2 points. When using the recommended application rate of the preparation (100 %), the infection type of the varieties containing the resistance genes *Rph19* and *Rph9* was 0;, 1 point; for the original population this indicator was 3 points.

The type of damage to the varieties containing the resistance genes *Rph1*, *Rph3*, *Rph4*, *Rph5*, *Rph7*, *Rph8*, *Rph12*, *Rph21+2*, in all options with the fungicides, regardless of the application rate, corresponded to the type of the original population (no treatment) and amounted up to 3 points.

This research studied the effect of the fungicides on the virulence of the population of the barley leaf rust pathogen (Table 2).

In all experimental options with an increase in the rate of fungicide application, there was a decrease in the virulence of the fungal population to *Rph4*, *Rph5*, *Rph6+2*, *Rph12*; in the option with Amistar Extra, SC and Delaro, SC – to *Rph2*; Amistar Gold, SC and Amistar Extra, SC – to *Rph9*. The use of the studied fungicides at a rate of 50–150 % of the recommended one contributed to an increase in the virulence of the population to *Rph14*. Under the influence of fungicides Amistar Gold, SC and Delaro, SC with an application rate of 150 % of the recommended one, an increase in the virulence of the population to *Rph19* was noted. Treatment of plants with Amistar Extra, SC and Delaro, SC with an increase in the application rate to 150 % of the recommended one contributed to a decrease in the virulence of the population to *Rph1*, and an increase in the occurrence of isolates virulent to *Rph8*.

According to the Nei index, the maximum differences in the frequency of isolates virulent to lines with *Rph* genes were obtained between the original population (no treatment) and the population treated with Amistar Extra, SC and Delaro, SC with an application rate of 50 % of the recommended one (N = 0.28; N = 0.24, respectively). At the same time, in the option with the fungicide Amistar Gold, SC, the maximum differences were obtained when using an application rate of 200 % of the recommended one (N = 0.22).

The average virulence of the original *P. hordei* population (no treatment) was 48.4 % (Fig. 1). In all experimental options, with an increase in the rate of fungicide application, a change in the average virulence was noted.

Treatment of the plants with Delaro, SC contributed to a decrease in this indicator from 45.9 % (application rate 50 % of the recommended one) to 34.7 % (application rate 200 %), with Amistar Extra, SC – from 44.3 % (application rate 50 % of the recommended one) to 34.7 % (application rate 200 %), with Amistar Gold, SC – from 33.8 % (application rate 50 % of the recommended one) to 28.5 % (application rate 200 %).

Table 1. Infectious type of differentiating varieties in response to inoculation with *P. hordei* population under the influence of various fungicide application rates (greenhouse of the FRCBPP, 2024)

No.	Variety	Gene	Amistar Gold, SC				Amistar Extra, SC				Delaro, SC				Original population (no treatment)
			Fungicide application rate, % of the recommended rate												
			50	100	150	200	50	100	150	200	50	100	150	200	
			Reaction type of varieties to infection, points												
1	Sudan	*Rph1*	3	3	3	3	3	3	3	3	3	3	3	3	3
2	Peruvian	*Rph2*	3	3	3	0;	3	3	3	3	3	3	3	3	3
3	Estate	*Rph3*	3	3	3	3	3	3	3	3	3	3	3	3	3
4	Gold	*Rph4*	3	3	3	3	3	3	3	3	3	3	3	3	3
5	Magnif 104	*Rph5*	3	3	3	3	3	3	3	3	3	3	3	3	3
6	Bolivia	*Rph6+2*	3	3	3	0	3	3	3	3	3	3	3	3	3
7	Cebada Capa	*Rph7*	0;	0;	0;	0;	0;	0;	0;	0;	0;	0;	0;	0;	0;
8	Egypt 4	*Rph8*	3	3	3	3	3	3	3	3	3	3	3	3	3
9	Abyssinian	*Rph9*	3	3	3	0;	3	3	3	3	3	0; 1	3	3	3
10	Triumph	*Rph12*	3	3	3	3	3	3	3	3	3	3	3	3	3
11	PI 531849	*Rph13*	0	0;	0;	3	3	3	3	3	3	3	3	3	3
12	PI 584760	*Rph14*	3	3	3	0;	1, 2, 3	2, 3	1, 2, 3	2, 3	3	3	3	0	1, 2
13	Prior	*Rph19*	3	3	3	3	3	3	3	3	3	0; 1	3	3	3
14	Ricardo	*Rph21+2*	3	3	3	3	3	3	3	3	3	3	3	3	3
15	Fong Tien	*Rph25*	0;	3	3	3	3	3	3	3	3	3	3	3	3

Table 2. Virulence (%) of the *P. hordei* population treated with the fungicides at different application rates to lines carrying *Rph* genes (greenhouse of the FSBSI FRCBPP, 2024)

No.	Variety	Gene	Original population (no treatment)	Amistar Gold, SC				Amistar Extra, SC				Delaro, SC			
				Fungicide application rate, % of the recommended											
				50	100	150	200	50	100	150	200	50	100	150	200
1	Sudan	*Rph1*	43	33	22	20	20	76	45	27	34	67	45	39	44
2	Peruvian	*Rph2*	59	62	67	51	0	87	76	49	44	76	76	68	53
3	Estate	*Rph3*	80	46	72	50	49	60	54	53	55	58	44	45	34
4	Gold	*Rph4*	57	22	21	20	20	52	45	38	38	59	45	39	32
5	Magnif 104	*Rph5*	35	27	13	10	10	72	43	45	41	63	43	43	23
6	Bolivia	*Rph6+2*	44	17	11	10	0	61	55	44	43	73	85	70	64
7	Cebada Capa	*Rph7*	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Egypt 4	*Rph8*	66	16	15	10	40	42	50	98	93	21	59	67	57
9	Abyssinian	*Rph9*	59	36	34	27	0	28	25	22	20	10	0	10	9
10	Triumph	*Rph12*	42	25	13	10	8	92	24	17	14	30	34	22	10
11	PI 531849	*Rph13*	47	0	0	0	37	10	57	65	59	21	27	20	46
12	PI 584760	*Rph14*	0	27	29	28	0	9	10	4	2	4	40	20	0
13	Prior	*Rph19*	51	46	59	67	64	11	20	20	18	80	0	78	57
14	Ricardo	*Rph21+2*	86	99	79	89	99	44	41	33	34	49	78	63	57
15	Fong Tien	*Rph25*	57	41	29	35	21	21	32	28	25	78	62	45	35
Nei index (Nei distance)			–	0.11	0.10	0.13	0.22	0.28	0.10	0.14	0.14	0.24	0.10	0.14	0.10

The results of the influence of different fungicide application rates on the aggressiveness indices of barley leaf rust were obtained. An increase in the duration of the latent period of the disease under the influence of high application rates of preparations was noted. For the original population (no treatment), this indicator was 168 hours, which corresponded to the values obtained when treating barley plants with the studied fungicides at a reduced application rate (50 % of the recommended one). In the option with the fungicides Delaro, SC and Amistar Gold, SC, at an application rate of 200 % of the recommended one, the duration of the latent period increased to 216 hours, with Amistar Extra, SC, up to 192 hours.

A decrease in the sporulating capacity of the *P. hordei* population treated with fungicides was found. In the options with the preparations Delaro, SC and Amistar Gold, SC, with an increase in the application rate, the spore mass from one pustule decreased from 0.007 (application rate 50 % of the recommended one) to 0.002 mg (application rate 200 %). In the option with the fungicide Amistar Extra, SC, this indicator changed from 0.006 (application rate 50 % of the recommended one) to 0.005 mg (application rate 200 %). For the original population (no treatment), the sporulating capacity was 0.013 mg from one pustule.

A change in the viability of *P. hordei* spores was noted to be influenced by the fungicides. Thus, for the original population (no treatment), the value of this indicator was determined as 100 % (Fig. 2). In the option with the fungicide Delaro, SC, the viability of spores decreased from 86.7 (application rate 50 % of the recommended one) to 51.7 % (application rate 200 %); in the option with the preparation Amistar Extra, SC, from 80.0 (application rate 50 % of the recommended one) to 51.7 % (application rate 200 %); with Amistar Gold, SC, from 71.7 (application rate 50 % of the recommended one) to 22.5 % (application rate 200 %).

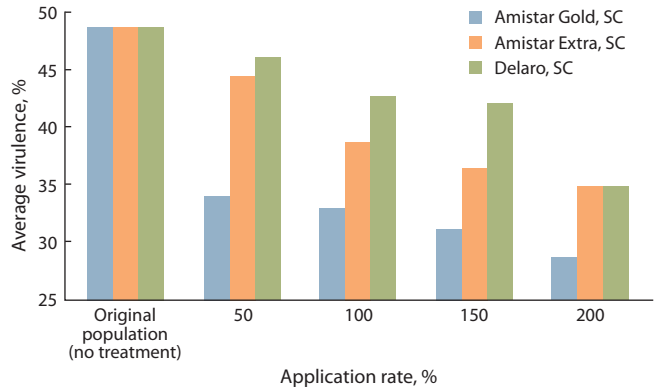


Fig. 1. Average virulence of the population of the barley leaf rust pathogen under the influence of different rates of fungicide application (%) (laboratory and greenhouse of the FSBSI FRCBPP, 2024).

Discussion

The results obtained during the research are consistent with the studies of other scientists. Thus, C. Zhao et al. managed to obtain two flutolanil-resistant isolates of *Rhizoctonia* spp., which were characterized by a lower mycelial growth rate and reduced virulence towards sugar beet sprouts compared to the original isolate (Zhao et al., 2019).

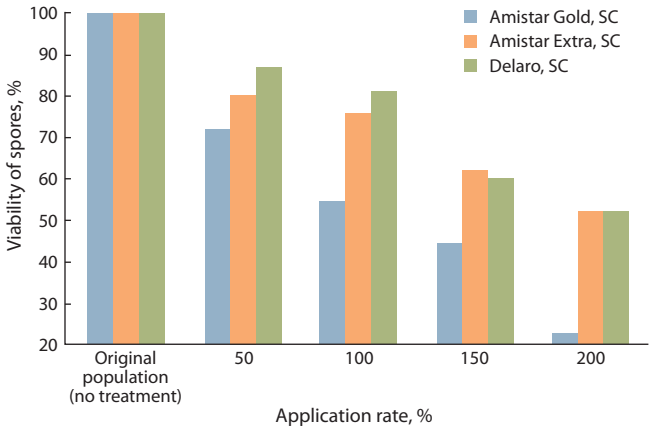


Fig. 2. Influence of different application rates of the fungicides on the viability of barley leaf rust spores (laboratory and greenhouse of the FSBSI FRCBPP, 2024).

An inhibitory effect on the sporulating ability of the *Phytophthora infestans* fungus was noted with an increase in the application rate of the fungicides with difenoconazole and fludioxonil as active ingredients (Myca, 2015).

Under the influence of an aqueous solution of benzimidazole (40 mg/l), a decrease in the sporulating capacity and the number of pustules of the rye leaf rust pathogen *P. dispers* was determined in comparison with the control, and the avirulence of the fungus was also noted (Tyryshkin, 2017).

G.V. Volkova's studies (2007) have proven that the acquisition of resistance to a fungicide from the triazole class was accompanied by a decrease in sporulating capacity by 4.8 times, the rate of diffuse spread of the pathogen mycelium in the tissues of the host plant, by 4 times. The incubation period was extended by 3 days, the period of pustule formation, by 8–12 days.

Previously, data were obtained on the effect of a tebuconazole-based fungicide on the aggressiveness of the population of the wheat leaf rust pathogen. With an increase in the application rate of the preparation, the viability of spores decreased to 21.5 %, the sporulating capacity, to 0.02 mg of spores, and the duration of sporulation, to 8 days (application rate 0.7 l/ha). At the same time, the duration of the latent period increased to 233 hours (Gvozdeva, Volkova, 2022).

Changes in the population structure under the influence of triazole fungicides have been recorded. Thus, X. Wu et al. (2020) described the sensitivity of 89 *P. graminis* isolates to triademifon, a triazole fungicide. It was found that isolates with resistance to triademifon may have cross-resistance to carbendazim. Resistant isolates to azole fungicides have also been recorded in *P. striiformis* (Tian et al., 2019), *B. graminis* (Cao et al., 2008). For the *P. triticina* population from Brazil, a decrease in sensitivity to triazoles was noted for the five most common races of the pathogen (Ardium et al., 2012). We studied the change in the *P. triticina* population structure under the influence of a tebuconazole-based fungicide (Gvozdeva, Volkova, 2022). According to the data obtained, the maximum of the studied changes in the genetic structure of the population (according to Nei, N) were noted at reduced rates of fungicide application. The average virulence of the pathogen population decreased with an increase in the rate of fungicide application.

Under the influence of the studied fungicides, a decrease in the virulence of the fungal population to *Rph4*, *Rph5*, *Rph6+2*, *Rph12* was found, which have been ineffective for the North Caucasian population for more than 10 years (Volkova et al., 2019; Danilova, Volkova, 2023). At the same time, virulence to *Rph14* increased compared to the original population. In 2021, the frequency of fungal isolates to this line was low (Danilova, Volkova, 2023). This may indicate a decrease in the efficacy of the *Rph14* gene under the influence of fungicides.

The development of resistance for pathogens often resembles the effect of vertical resistance: in the first years of fungicide use, complete suppression of infection is observed; over time, the emergence of individual tolerant strains and their accumulation in the population is observed; and, finally, there is a complete loss of fungicide efficacy (Dyakov, 1998). One of the factors reducing the sensitivity of phytopathogens to the active substances of preparations is a change in their intrapopulation structure (Tyuterev, 2001). Mutations that cause resistance of phytopathogen isolates to fungicides can lead to a decrease in their adaptability and virulence (Hawkins, Fraaije, 2018), but later, an increase in the aggressiveness of the pathogen may be observed (Dyakov, 1998). In our studies for the North Caucasian population of the barley leaf rust pathogen under the influence of two-component preparations of the class of triazoles and strobilurins, changes in the intrapopulation structure in terms of aggressiveness and virulence were noted, which justifies the need for constant study of this issue to control the accumulation of resistant forms of *P. hordei* in the fungal population.

Conclusion

A comparative assessment of pathogenicity indicators under the influence of fungicides of the chemical classes of triazoles and strobilurins revealed a decrease in the aggressiveness and virulence of the North Caucasian population of the barley leaf rust pathogen. So, in all options with an increase in the rate of fungicide application, a change in the intrapopulation structure and average virulence of populations was noted. In all options with an increase in the rate of fungicide application, there was a decrease in the virulence of the fungal population to *Rph4*, *Rph5*, *Rph6+2*, *Rph12*; in the option with Amistar Extra, SC and Delaro, SC, to *Rph2*; with Amistar Gold, SC and Amistar Extra, SC, to *Rph9*.

The use of the studied fungicides at a rate of 50–150 % of the recommended one contributed to an increase in the virulence of the population to *Rph14*. The line with the *Rph7* gene showed no signs of infection both in the original population and in the experimental options. The average virulence of the fungal populations treated with the fungicides in all experimental options was lower compared to the original population (no treatment) (48.5 %). Significant changes in this indicator were noted under the influence of the fungicide Amistar Gold, SC, the average virulence decreased from 33.8 to 28.5 %.

In comparison with the original population (no treatment), an increase in the duration of the latent period of the disease was noted under the influence of high application rates of the preparations. In the Delaro, SC and Amistar Gold, SC options at an application rate of 200 % of the recommended one, the

value of this indicator varied from 168 to 216 hours, with Amistar Extra, SC, from 168 to 192 hours.

A decrease in the sporulating capacity of the *P. hordei* population treated with the fungicides was determined. In the options with the Delaro, SC and Amistar Gold, SC preparations with an increase in the application rate, the spore mass from one pustule decreased from 0.013 (original population (no treatment)) to 0.002 mg (application rate 200 %), in the Amistar Extra, SC option, to 0.005 mg (application rate 200 %). A significant decrease in the viability of *P. hordei* spores in comparison with the original population (no treatment) was noted under the influence of the fungicide Amistar Gold, SC. With an application rate of the preparation of 200 % of the recommended one, this indicator decreased from 100 to 22.5 %.

Thus, our studies allow us to identify changes in the population structure by virulence and aggressiveness of the *P. hordei* population under the influence of the studied fungicides, which will make it possible to promptly adjust the winter barley protection system and, in the future, can contribute to the development of an anti-resistant strategy to control *P. hordei*.

References

Abbas H.S. Barley diseases: introduction, etiology, epidemiology, and their management. In: Abd-Elsalam K.A., Mohamed H.I. (Eds) Cereal Diseases: Nanobiotechnological Approaches for Diagnosis and Management. Singapore: Springer, 2022;97-117. doi 10.1007/978-981-19-3120-8_6

Arduim G.D.S., Reis E.M., Barcellos A.L., Turra C. *In vivo* sensitivity reduction of *Puccinia triticina* races, causal agent of wheat leaf rust, to DMI and QoI fungicides. *Summa Phytopathol.* 2012;38(4):306-311. doi 10.1590/S0100-54052012000400006

Bai Q., Liu T., Wan A., Wang M., See D.R., Chen X. Changes of barley stripe rust populations in the United States from 1993 to 2017. *Phytopathology*. 2022;112(11):2391-2402. doi 10.1094/PHYTO-04-22-0135-R

Brent K.J., Hollomon D.W. Fungicide Resistance in Crop Pathogens: How Can it Be Managed? Bristol, UK: Fungicide Resistance Action Committee, 2007

Cao X.R., Zhao W.J., Zhou Y.L. Monitoring of resistance of *Blumeria graminis* f. sp. *tritici* isolates to triadimefon in 2007. *Plant Protection*. 2008;34:74-77

Corkley I., Fraaije B., Hawkins N. Fungicide resistance management: maximizing the effective life of plant protection products. *Plant Pathology*. 2021;71(1):150-169. doi 10.1111/ppa.13467

Çelik Oğuz A., Karakaya A. Genetic diversity of barley foliar fungal pathogens. *Agronomy*. 2021;11(3):434. doi 10.3390/agronomy11030434

Danilova A.V., Volkova G.V. Virulence of barley leaf rust in the South of Russia in 2017–2019. *Span J Agric Res*. 2022;20(1):e10SC01. doi 10.5424/sjar/2022201-18337

Danilova A., Volkova G. Efficiency of *Rph genes* against *Puccinia hordei* in Southern Russia in 2019–2021. *Agronomy*. 2023;13(4):1046. doi 10.3390/agronomy13041046

Dodhia K.N., Cox B.A., Oliver R.P., Lopez-Ruiz F.J. Rapid in situ quantification of the strobilurin resistance mutation G143A in the wheat pathogen *Blumeria graminis* f. sp. *tritici*. *Sci Rep*. 2021;11(1):4526. doi 10.1038/s41598-021-83981-9

Dyakov Y.T. Population Biology of Phytopathogenic Fungi. Moscow, 1998 (in Russian)

Ereshko A.S., Hronyuk V.B., Repko N.V. State and prospects of barley production in the Russian Federation. *Vestnik Agrarnoj Nauki Dona = Don Agrarian Bulletin*. 2012;3(19):57 (in Russian)

FRAC. List of first confirmed cases of plant pathogenic organisms resistant to disease control agents. CropLife International. Brussels, 2020. Available at: https://www.frac.info/media/3oils1zl/list-of-first-confirmed-cases-of-plant-pathogenic-organisms-resistant-to-disease-control-agents_05_2020.pdf

Gvozdeva M.S., Volkova G.V. Influence of the Kolosal fungicide on the population structure of the wheat leaf rust pathogen by signs of pathogenicity and sensitivity. *Mikologiya i Fitopatologiya = Mycology and Phytopathology*. 2022;56(1):52-63. doi 10.31857/S0026364822010044 (in Russian)

Fungal Pathogens of Grain Ear Crops: Biology, Distribution, Harmfulness, Methods of Detection, Collection and Storage of Biomaterial. Creation of Artificial Infectious Backgrounds: Scientific and Practical Recommendations. Krasnodar, 2024 (in Russian)

Hawkins N.J., Fraaije B.A. Fitness penalties in the evolution of fungicide resistance. *Annu Rev Phytopathol*. 2018;56:339-360. doi 10.1146/annurev-phyto-080417-050012

Ji F., Zhou A., Liu B., Liu Y., Feng Y., Wang X., Huang L., Kang Z., Zhan G. Sensitivity of *Puccinia triticina* f. sp. *tritici* from China to triadimefon and resistance risk assessment. *Plant Dis*. 2023;107(12):3877-3885. doi 10.1094/PDIS-02-23-0277-RE

Krupen'ko N.A. Classification and mode of action of chemical fungicides applied for cereal crops protection in Belarus. *Vestnik Zashchity Rastenij = Plant Protection News*. 2023;106(2):84-92. doi 10.31993/2308-6459-2023-106-2-15690 (in Russian)

Lass-Flörl C. Triazole antifungal agents in invasive fungal infections: a comparative review. *Drugs*. 2011;71(18):2405-2419. doi 10.2165/11596540-000000000-00000

Leadbeater A. Resistance risk to QoI fungicides and anti-resistance strategies. In: Thind T.S. (Ed.) Fungicide Resistance in Crop Protection: Risk and Management. CABI, 2012;144-154. doi 10.1079/9781845939052.0141

Mihajlova L.A., Gul'tyaeva E.I., Mironenko N.V. Methods of studying the Population Structure of the Wheat Leaf Rust Pathogen *Puccinia recondita* Rob. ex Desm. f. sp. *tritici*. Immunogenetic Methods of Creation of Varieties Resistant to Pests: Methodical Recommendations. Saint Petersburg, 2003 (in Russian)

Myca E.D. Effect of some pesticides on fungal pathogens of potato (*Solanum tuberosum* L.) and tomato (*Lycopersicon esculentum* Mill.): Cand. Sci. Biol. Dissertation. Moscow, 2015 (in Russian)

Nei M. Genetic distance between populations. *Am Nat*. 1972;106:283-292. doi 10.1086/282771

Ölmez F., Turgay E.B., Mustafa Z., Büyük O., Kaymak S. Screening the *Zymoseptoria tritici* population in Turkey for resistance to azole and strobilurin fungicides. *J Plant Dis Prot*. 2023;130(5):991-998. doi 10.1007/s41348-023-00761-5

Parlevliet J.E. Variation for latent period one of the component of partial resistance in barley to yellow rust by *Puccinia striiformis*. *Cereal Rust Bull*. 1980;8:2-17

Pyzhikova G.V. Effect of temperature on infection and severity of leaf rust of wheat. *Mikologiya i Fitopatologiya = Mycology and Phytopathology*. 1972;6(3):51-53 (in Russian)

Russell P.E. Sensitivity baselines in fungicide resistance research and management. FRAC Monograph No. 3. CropLife International, 2003. Available at: https://www.frac.info/media/cj3hvben/monograph-3.pdf

Sanin S.S., Shinkarev V.P., Kaydash A.S. Methods for determining the number of spores produced by rust and other phytopathogenic fungi. *Mikologiya i Fitopatologiya = Mycology and Phytopathology*. 1975;9(3):443-445 (in Russian)

Sapkota R., Jørgensen L.N., Boeglin L., Nicolaisen M. Fungal communities of spring barley from seedling emergence to harvest during a severe *Puccinia hordei* epidemic. *Microb Ecol*. 2023;85(2):617-627. doi 10.1007/s00248-022-01985-y

Shcherbakova L.A. Fungicide resistance of plant pathogenic fungi and their chemosensitization as a tool to increase anti-disease effects of triazoles and strobilurines. *Agric Biol*. 2019;54(5):875-891. doi 10.15389/agrobiology.2019.5.875eng

Thind T.S. Changing trends in discovery of new fungicides: a perspective. *Indian Phytopathol*. 2021;74:875-883. doi 10.1007/s42360-021-00411-6

Tian Y., Meng Y., Zhao X., Chen X., Ma H., Xu S., Zhan G. Trade-off between triadimefon sensitivity and pathogenicity in a selfed sexual population of *Puccinia striiformis* f. sp. *tritici*. *Front Microbiol*. 2019;10:2729. doi 10.3389/fmicb.2019.02729

Tyryshkin L.G., Sidorov A.V. The Changing virulence and aggressiveness of the pathogen of leaf rust of rye under the influence of abiotic factors and the possibility of its practical application. *Izvestiya Saint-Petersburg State Agrarian University*. 2017;46:58-63 (in Russian)

Tyuterev S.L. Problems of phytopathogen resistance to new fungicides. *Vestnik Zashchity Rastenij = Plant Protection News*. 2001;1:38-53 (in Russian)

Volkova G.V. An antiresistant strategy for the triadimefon-based fungicides application against the leaf-rust pathogen on wheat in North Caucasus. *Nauka Kubani = Science of Kuban*. 2007;2:39-43 (in Russian)

Volkova G.V., Kudinova O.A., Gladkova E.V., Vaganova O.F., Danilova A.V., Matveeva I.P. Virulence of Populations of Cereal Rust Pathogens of Grain Ear Crops. Krasnodar, 2018 (in Russian)

Volkova G.V., Danilova A.V., Kudinova O.A. The virulence of the barley leaf rust pathogen in the North Caucasus in 2014-2017. *Agric Biol*. 2019;54(3):589-596. doi 10.15389/agrobiology.2019.3.589eng

Walters D.R., Avrova A., Bingham I.J., Burnett F.J., Fountaine J., Havis N.D., Hoad S.P., Hughes G., Looseley M., Oxley S.J.P., Renwick A., Topp C.F.E., Newton A.C. Control of foliar diseases in barley: towards an integrated approach. *Eur J Plant Pathol*. 2012; 133:33-73. doi 10.1007/s10658-012-9948-x

Wolfe M.S., Schwarzbach E. The use of virulence analysis in cereal mildews. *J Phytopathol*. 1975;82(4):297-307. doi 10.1111/j.1439-0434.1975.tb03495.x

Wu X., Bian Q., Lin Q., Sun Q., Ni X., Xu X., Qiu Y., Xuan Y., Cao Y., Li T. Sensitivity of *Puccinia graminis* f. sp. *tritici* isolates from China to triadimefon and cross-resistance against diverse fungicides. *Plant Dis*. 2020;104(8):2082-2085. doi 10.1094/PDIS-01-20-0009-RE

Zhan G., Ji F., Zhao J., Liu Y., Zhou A., Xia M., Zhang J., Huang L., Guo J., Kang Z. Sensitivity and resistance risk assessment of *Puccinia striiformis* f. sp. *tritici* to triadimefon in China. *Plant Dis*. 2022;106(6):1690-1699. doi 10.1094/PDIS-10-21-2168-RE

Zhao C., Zhang X., Hua H., Han C., Wu X. Sensitivity of *Rhizoctonia* spp. to flutolanil and characterization of the point mutation in succinate dehydrogenase conferring fungicide resistance. *Eur J Plant Pathol*. 2019;155:13-23. doi 10.1007/s10658-019-01739-6

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

Received March 31, 2025. Revised May 14, 2025. Accepted May 15, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1203-1212

doi 10.18699/vjgb-25-127

Effect of the biopesticide Novochizol on the expression of defense genes during wheat infection with stem rust *Puccinia graminis* f. sp. *tritici*

A.B. Shcherban [1], A.V. Razuvaeva [1], E.S. Skolotneva [1], V.V. Fomenko [2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
atos@bionet.nsc.ru

Abstract. Stem rust, caused by the fungus *Puccinia graminis* f. sp. *tritici* (*Pgt*), is a harmful disease affecting grain crops. The traditional way to combat this and other infectious plant diseases is to use chemical pesticides. Biopesticides, as well as plant disease resistance inducers – in particular those based on chitosan, a derivative of chitin – are increasingly being considered as an effective and safe alternative. Recently, a globular form of chitosan, Novochizol, has been developed, which has a number of advantages and has shown its effectiveness in preliminary field and laboratory experiments. However, there are no works devoted to the effect of this preparation on the expression of defense genes. Therefore, the aim of this work was to search for genes involved in the response of common wheat (*Triticum aestivum* L.) to stem rust infection and to evaluate the effect of Novochizol treatment on their transcription during the infection process. The wheat line ISr6-Ra with the stem rust resistance gene *Sr6* and two *Pgt* isolates – an avirulent one, Avr6, and a virulent one, vr6 – were used as a model, allowing us to compare the effects of Novochizol depending on the genetic compatibility in the plant–pathogen pathosystem. To analyze the transcription level of defense genes, leaf material was collected at different time points from 3 to 144 h after inoculation of plants with the pathogen. Quantitative PCR analysis showed an increase in the transcription levels of the *CERK1*, *PR3*, *PR4*, *PR5*, *PR6* and *PR9* genes in plants treated with Novochizol and infected with various *Pgt* isolates compared to untreated infected plants. *Pgt* isolate Avr6 induced the highest expression of some defense genes (primarily *CERK1*), which is consistent with the phytopathology data showing the maximum degree of resistance (IT1) to stem rust in Novochizol-treated plants with a combination of *Sr6–Avr6* genes. The data obtained confirm that one of the optimal strategies for increasing the resistance of grain crops to fungal pathogens is a combination of selection for specific resistance genes with the use of biological control agents.

Key words: biopesticide; chitosan; Novochizol; defense genes; stem rust, transcription; common wheat

For citation: Shcherban A.B., Razuvaeva A.V., Skolotneva E.S., Fomenko V.V. Effect of the biopesticide Novochizol on the expression of defense genes during wheat infection with stem rust *Puccinia graminis* f. sp. *tritici*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1203-1212. doi 10.18699/vjgb-25-127

Funding. The study was supported by the grant of the Russian Science Foundation No. 23-16-00119.

Acknowledgements. The wheat line for laboratory testing was propagated at the Center for Collective Use of Plant Reproduction of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences with the support of the budget project FWNR-2022-0017.

Влияние биопестицида Новохизоль на экспрессию генов защиты при заражении пшеницы стеблевой ржавчиной *Puccinia graminis* f. sp. *tritici*

А.Б. Щербань [1], А.В. Разуваева [1], Е.С. Сколотнева [1], В.В. Фоменко [2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский институт органической химии им. Н.Н. Ворожцова Сибирского отделения Российской академии наук, Новосибирск, Россия
atos@bionet.nsc.ru

Аннотация. Стеблевая ржавчина, вызываемая грибом *Puccinia graminis* f. sp. *tritici* (*Pgt*), является вредоносным заболеванием, поражающим посевы зерновых культур. Традиционный способ борьбы с этим и другими инфекционными болезнями растений – использование химических средств защиты. В качестве их эффективной и безопасной альтернативы все чаще рассматриваются биопестициды, а также индукторы болезнеустойчивости растений, в частности на основе хитозана, производного хитина. Недавно разработана глобулярная форма хитозана – Новохизоль, имеющая ряд преимуществ и показавшая свою эффективность в предварительных полевых и лабораторных экспериментах. Однако в настоящее время отсутствуют работы, посвященные влиянию

данного препарата на экспрессию генов защиты. Поэтому целью данной работы стали поиск генов, принимающих участие в реакции растений мягкой пшеницы *Triticum aestivum* L. на заражение стеблевой ржавчиной, и оценка влияния обработки препаратом Новохизоль на их транскрипцию в ходе инфекционного процесса. В качестве модели были задействованы линия пшеницы с геном устойчивости к стеблевой ржавчине *Sr6* и два отобранных изолята *Pgt* для этой линии: авирулентный (Avr6) и вирулентный (vr6), позволяющие сопоставить эффекты препарата Новохизоль в зависимости от генетической совместимости в патосистеме растение–патоген. Для анализа уровня транскрипции генов защиты использовали листовой материал в различных временных точках, от 3 до 144 ч после инокуляции растений патогеном. Количественный ПЦР-анализ показал повышение уровня транскрипции генов *CERK1*, *PR3*, *PR4*, *PR5*, *PR6* и *PR9* у растений, обработанных изучаемым биопестицидом и инфицированных различными изолятами *Pgt*, по сравнению с необработанными инфицированными растениями. Полученные данные подтверждают, что одна из оптимальных стратегий повышения устойчивости зерновых культур к грибным патогенам с точки зрения экологической безопасности – сочетание методов селекции по генам специфической устойчивости с применением биологических средств защиты.

Ключевые слова: биопестицид; хитозан; Новохизоль; гены защиты; стеблевая ржавчина; транскрипция; мягкая пшеница

Introduction

Common wheat *Triticum aestivum* L. is the most popular and largest grain crop in terms of sown area. One of the main problems in its cultivation is the decrease in its yield due to diseases caused by various pathogens, especially fungi (Hafeez at al., 2021). Almost 90 % of the world's wheat sown area is at risk of being affected by at least one fungal disease (Chai et al., 2022). Among the most dangerous are leaf-stem diseases that reduce the assimilation surface of leaves, and destroy chlorophyll, which leads to a decrease in photosynthesis, premature aging and death of plant tissues. Crop losses can range from 20 to 70 %. Often, such fungal infections take the form of epiphytoties. Since the beginning of 2000, the incidence of stem rust in wheat crops in Europe has increased sharply and continues to grow (Patpour et al., 2022). In the Russian Federation (RF), epiphytoties of stem and brown rust, septoria, and powdery mildew are periodically observed (Sanin, 2012).

In response to pathogen attacks, plants have developed a complex immune system that includes different levels of defense. The perception of pathogen-associated molecular patterns (PAMPs) by membrane pattern recognition receptors (PRRs) leads to PAMP-triggered immunity (PTI) (Bigeard et al., 2015). In an attempt to overcome this level of defense, the pathogen secretes specific avirulence effectors (*Avr* gene products) that suppress PTI, thus promoting further disease development. In response, plants have developed the next, specific level of defense based on intracellular recognition of effectors by NOD-like receptors (NLRs), products of plant *R* genes. This effector-triggered immunity (ETI) defense mechanism often results in programmed cell death in infected tissue or a hypersensitive response (HSR) that creates a barrier to pathogen dissemination (Jones, Dangl, 2006; Bent, Mackey, 2007).

PTI and ETI have been shown to be closely interrelated and jointly mediate systemic resistance to disease (Ngou et al., 2021). This resistance is based on the activation of a number of signaling cascades, including hormonal ones, which are controlled by hormones such as salicylic (SA), jasmonic (JA) and abscisic acids, ethylene, etc., depending on the type of nutrition of the pathogen: biotrophic, hemibiotrophic or necrotrophic (Jones, Dangl, 2006).

The salicylate-controlled cascade characteristic of the response against biotrophic pathogens, including the stem rust pathogen *Puccinia graminis* f. sp. *tritici* (Pgt), leads to the accumulation of reactive oxygen species (ROS) – an oxidative burst that provokes HSR, synthesis of phenolic substances and specific protective PR (pathogenesis-related) proteins (Tada et al., 2008; Ding et al., 2018).

PR proteins are the most important factors in plant resistance mechanisms. They participate in signaling systems, catalyzing the formation of hormones, strengthen the cell walls of plants, and are capable of causing damage to the cell walls and cytoplasmic membranes of pathogens (Tarchevskij, 2002). PR proteins are divided into 17 families (PR1–PR17) (Van Loon, Van Strien, 1999; Sels et al., 2008). Many of them are hydrolases. Thus, proteins of the PR2 family are β-1,3-endoglucanases; they accumulate around the site of infection, exhibiting fungicidal activity. Proteins of the PR3, PR4, PR8 and PR11 families have endochitinase activity and are capable of destroying the cell wall of fungi; their pool increases during infection. Representatives of the PR7, PR8, PR9 and PR10 families exhibit proteinase, lysozyme, peroxidase and ribonuclease activities, respectively. Proteins of the PR6 family are proteinase inhibitors that suppress proteolytic enzymes secreted by pathogens to penetrate plant tissues. The activity of other PR proteins is associated with an increase in membrane permeability (families PR5, PR12, PR13, PR14), as well as with the accumulation of hydrogen peroxide (families PR15 and PR16); the synthesis of these proteins is significantly increased by treatment with elicitors (Van Loon, Van Strien, 1999; Scherer, 2005; Van Loon et al., 2006).

One of the well-established long-term strategies for plant protection against pathogens is based on genetic protection using *R* genes, the inducers of ETI (see above). The gene catalog of the International Symposium on Wheat Genetics contains about 58 stem rust resistance genes (*Sr*) identified in wheat (McIntosh et al., 2011). Most of them, such as *Sr25*, *Sr26*, *Sr31*, *Sr35*, *Sr38*, *Sr39*, *Sr44*, *Sr45*, provide specific resistance mechanisms, including HSR and PR protein expression. However, the emergence of new pathogen races overcoming resistance in crop varieties increases losses from diseases. Therefore, the use of chemical pesticides became

an additional line of defense, although it can have a negative impact on human health, food safety and the environment, and can also be toxic to non-target beneficial organisms. This stimulates the search for safe alternative plant protection products – biopesticides and resistance inducers (Ali et al., 2023).

Chitin is one of the most studied PAMPs that induce plant immunity. It is the main component of fungal cell walls, exoskeletons of arthropods, including insects. Chitosan, a biopesticide, was obtained by hydrolysis and deacetylation of chitin, the effect of which as a growth stimulator and inducer of non-specific resistance to fungal, bacterial and viral diseases has been confirmed in many studies (Maluin, Hussein, 2020; Shcherban, 2023). However, there are a number of disadvantages that complicate the use of chitosan: insolubility in water at physiological pH, low stability, variability of effects depending on the weight of the molecules, the degree of their deacetylation and the composition of the monomers (Katiyar et al., 2014; Varlamov et al., 2020).

Since there are currently no standard protocols for determining these parameters, assessments of the effects of chitosan preparations in different laboratories are often poorly comparable. Recently, scientists from the RF (patent US 20230096466A1; https://www.freepatentsonline.com/y2023/0096466.html) obtained a globular chitosan form, Novochizol, which has a number of advantages: increased solubility in aqueous solutions, high adhesion to plant tissues, chemical stability, the ability to form complexes with other biologically active substances (Fomenko, Loroch, 2021). As a result of treating common wheat seeds with this preparation, their germination was accelerated, and an increase in root mass and total plant mass was observed (Teplyakova et al., 2022).

In field conditions, pre-sowing seed treatment and plant treatment with Novochizol in combination with natural fungicides had a positive effect on the growth and productive properties of wheat, and also increased its resistance to a number of fungal diseases (Orlova et al., 2025). At the same object, the inhibitory effect of Novochizol on the development of stem rust was demonstrated in laboratory conditions (Shcherban et al., 2024). It was shown that under the influence of treatment with this drug, ROS, in particular hydrogen peroxide (H_2O_2), accumulate in the leaf tissue of plants after infection with *Pgt* due to the modulation of the activity of antioxidant enzymes. It should be noted that the latter work studied the reaction of a variety with a complex genotype infected with a mixed population of the fungus (Shcherban et al., 2024).

It is of interest to study Novochizol-induced mechanisms of resistance to isolated *Pgt* races with contrasting virulence using a common wheat line carrying the corresponding *Sr* gene. For this purpose, we selected an isogenic line for the *Sr6* gene, on which avirulent and virulent *Pgt* isolates were effectively differentiated (Skolotneva et al., 2020).

The aim of this work was to assess the susceptibility level of plants of this line treated with Novochizol and separately infected with one or another *Pgt* isolate, as well as to conduct a comparative analysis of the dynamics of expression of protection genes (PR) in the leaf tissue of plants within 144 hours after inoculation. To date, data have been accumulated on the effect of chitosan preparations on the expression of PR genes after plant infection with various pathogenic fungi (Manjunatha et al., 2008; Maluin, Hussein, 2020; Shcherban, 2023); however, in our work, the study of such an effect of the Novochizol preparation was carried out for the first time.

Materials and methods

Plant material and stem rust pathogen. To obtain seed material, the common wheat line ISr6-Ra, kindly provided by Prof. J. Colmer (St. Paul, USA), was grown hydroponically in the artificial climate laboratory of the Institute of Cytology and Genetics SB RAS. To confirm the presence of the stem rust resistance gene *Sr6* in this line, the plant material was genotyped using the molecular marker Xcfd43 to this gene (Tsilo et al., 2009). In the laboratory experiment, plants were grown in vessels with soil, under conditions recommended for work with rust fungi by international protocols (Roelfs et al., 1992). For treatment with Novochizol, 10-day-old seedlings were used, and as a control, uninfected plants without treatment, grown in parallel with the experimental samples, were used.

Previously, the West Siberian population of *Pgt* was genotyped for races of the fungus containing different combinations (pathotypes) of *Avr* and *vr* genes (Skolotneva et al., 2020). In our experiment, we used the LKCSF (vr6)/LCCSF (Avr6) pathotype system with the highest yield of urediniospores. The selected isolates were propagated on the susceptible variety of common wheat Khakasskaya to obtain spore material in the amount required to infect an experimental batch of plants at the rate of 10 mg per 1,000 plants. The spore material was stored at a temperature of −70 °C until the experiment.

Novochizol treatment and *Pgt* infection. Assessment of plant susceptibility. Novochizol with a deacetylation degree of ≥90 % and a molecular weight of ~500 kDa was provided by Novochizol SA (Monte, Switzerland). Aqueous suspensions of this preparation were prepared as described previously (Teplyakova et al., 2022). A concentration of 0.125 %, which had previously been shown to be effective against stem rust (Shcherban et al., 2024), was chosen for the treatment of 10-day-old plant seedlings. Novochizol suspension was applied to plants using a household sprayer (15 ml/100 plants) four days before *Pgt* inoculation.

For plant inoculation, *Pgt* urediniospores were premixed with Novec 7100 mineral oil (3M Novec™, St. Paul, MN, USA) at a concentration of 25 mg spores per 5 ml oil and applied to seedlings using a sprayer. Control plants were treated with bidistilled water. Inoculated plants were incubated for 24 h in a humidified chamber, in the dark, at 15–20 °C to ensure maximum spore germination. The plants were then transferred to growth chambers and incubated for 16 h under 10,000 lux phytolamp illumination and at a temperature of 26–28 °C. These conditions are necessary for the formation of appressoria, pathogen penetration into the stomata, and

Primer sequences and expected sizes of PCR products

Primers	Gene product	Sequence (5′–3′)	Size of PCR product (bp)	References
GAPDHF GAPDHR	Glyceraldehyde-3-phosphate dehydrogenase	TTGCTCTGAACGACCATTTC GACACCATCCACATTTATTCTTC	176	Hao et al., 2022
CERK1F CERK1R	Chitin elicitor receptor kinase 1	TTACCCCATCGACGCCATTC TCTGCCGGACATGAGGTTCA	190	Lee et al., 2014
PR3F PR3R	Endochitinase	AGAGATAAGCAAGGCCACGTC GGTTGCTCACCAGGTCCTTC	116	Desmond et al., 2005
PR4F PR4R	Hevein-like protein	CGAGGATCGTGGACCAGTG GTCGACGAACTGGTAGTTGACG	128	Bertini et al., 2003
PR5F2 PR5R2	Thaumatin-like protein	GCCGCAAGCCTACCAACA CGCGGTGCGACGTATAGAG	107	Ray et al., 2003
PR6F PR6R	Proteinase inhibitor	TTCAGGGCTTCAGGCTCAG ATTGACCAAGACACGACGATAG	141	Gao et al., 2013
PR9F PR9R	Peroxidase	GAGATTCCACAGATGCAAACGAG GGAGGCCCTTGTTTCTGAATG	102	Pritsch et al., 2000

the development of infectious hyphae in the intercellular spaces of the plant (Roelfs et al., 1992).

The degree of plant susceptibility to stem rust was assessed taking into account the qualitative (type of reaction) and quantitative components of this indicator. Infection types of plant reaction (IT) were determined 12–14 days after inoculation using the modified Steckman scale (Roelfs et al., 1992), where IT "0", ";", "1" and "2" were interpreted as resistant (Low, L), and "3", "3+" and "4", as susceptible (High, H).

RNA extraction, reverse transcription, RT-PCR and qPCR. For sample collection at different time points after infection (3, 6, 24, 48, 72, 144 h), three plants were planted as biological replicates for each experimental sample. Individual leaves were cut from each plant and immediately placed in liquid nitrogen. The collected material was stored at –80 °C before RNA extraction. Total RNA was extracted using the R-PLANTS-kit (Biolabmix, Novosibirsk, Russia) according to the manufacturer's instructions. Genomic DNA was removed using heat-labile DNase (Biolabmix). Reverse transcription was performed using the OT-M-MuLV-RH first-strand cDNA synthesis kit (Biolabmix) using 1 μg of total RNA and oligo$(dT)_{16}$ as a primer.

For reverse transcription-PCR (RT-PCR), the conditions were the same as for qPCR (see below). RT-PCR products were analyzed by electrophoresis in 1.5 % agarose gel followed by staining with ethidium bromide.

The BioMaster HS-qPCR Lo-ROX SYBR (2×) reagent kit (Biolabmix) and the QuantStudio 5 Real-Time PCR System (Thermo Scientific, USA) were used for qPCR. The sequences of gene-specific primers are listed in the Table. Amplification conditions were 5 min at 95 °C, followed by 40 cycles of 10 s at 95 °C, 20 s at 60 °C (55 °C for the *GAPDH* gene), and 10 s at 72 °C. qPCR data were collected during each round of amplification. Melting curve analysis between 60 and 95 °C was applied to all products to determine the fidelity of the PCR reaction.

Gene expression measurements were performed in three biological replicates, each with three technical replicates. Relative mRNA amounts were determined using Pfaffl's formula (Pfaffl, 2001). PCR efficiency was assessed using LinReg software (Ruijter et al., 2009). mRNA expression levels were normalized to the level of the housekeeping gene *GAPDH*. This gene was chosen as a reference because no differences in its mRNA content were detected between *Pgt*-infected and control samples (data not shown). Analysis of variance with Tukey's post hoc test was used to compare expression levels.

To assess the transcription level of common wheat genes induced during *Pgt* infection using qPCR, we selected six genes for which preliminary RT-PCR revealed the presence of a monomorphic amplification product of the expected length with varying intensity at individual time points after inoculation with the pathogen (see the Table).

Results

Effect of Novochizol on stem rust development

To study the effect of Novochizol treatment on the susceptibility of common wheat plants to stem rust and the expression of defense genes, we used a model including the *T. aestivum* ISr6-Ra line and the corresponding Avr6 and vr6 *Pgt* isolates. This model allowed us to study various resistance mechanisms depending on the recognition of pathogen ligands by the host's immune receptors.

Fourteen days after inoculation, the susceptibility level of plants infected with the Avr6 isolate corresponded to the resistant type (IT = 1–2) (Fig. 1). This scenario corresponds to cases of interaction of plant/pathogen genotypes carrying complementary genes of resistance and avirulence,

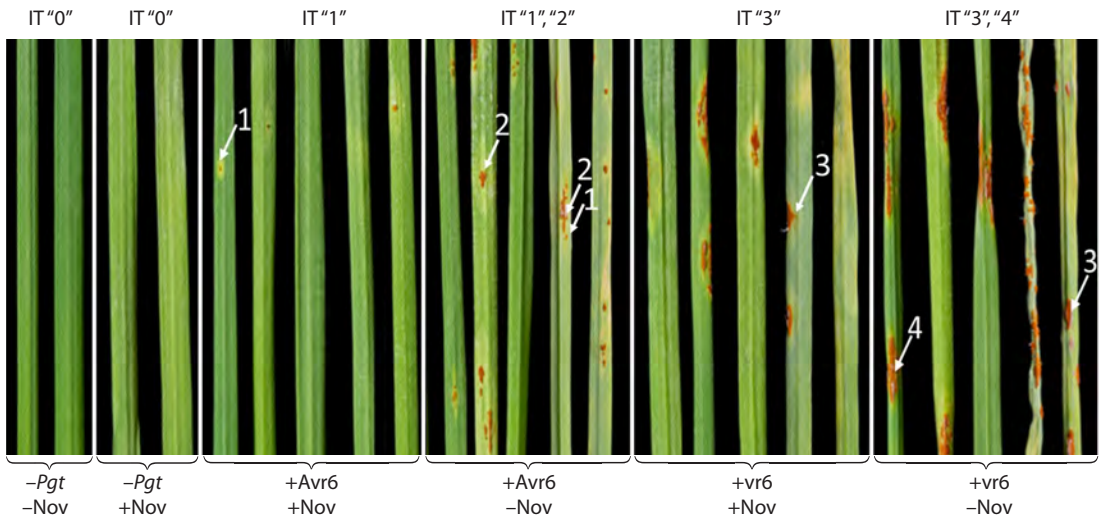


Fig. 1. Effect of Novochizol on the development of stem rust symptoms on leaves of wheat of the *T. aestivum* ISr6-Ra line.

Avr6 and vr6 are avirulent and virulent isolates of the stem rust pathogen (*Pgt*). Infectious types of plant response (IT) are indicated above the figure. Pustules characteristic of each IT are indicated by arrows. Nov – Novochizol. Photofixation was carried out on the 14th day after inoculation with the pathogen. The age of plants at the time of fixation was 28 days.

i. e. encoding the immune receptor and its specific ligand. Infection with the vr6 isolate resulted in a susceptibility reaction (IT = 3–4). An earlier assessment of the quantitative component of resistance, determined by the average number of pustules per leaf blade, showed a multifold decrease in the number of pustules in plants pretreated with Novochizol compared to untreated plants, in the case of both isolates (data not published). As a result of treatment, plants infected with Avr6 and vr6 were similar in terms of the quantitative indicator (~1.5 pustules per leaf on average); however, in general, they retained the specific IT of resistance and susceptibility for each isolate, determined by the size of the pustules (Fig. 1).

Transcription level assessment of defense genes

Induction of the *CERK1* gene during *Pgt* infection of wheat. Chitin elicitor receptor kinase 1 (CERK1) has been identified as a PAMP receptor for chitin, a major structural component of fungal cell walls (Gong et al., 2020; Wang L. et al., 2024). Upon recognition of chitin, this receptor protein activates various defense-related signaling pathways, playing a key role in plant immunity.

In untreated plants infected with the vr6 isolate, a slight increase in transcription of this gene was detected at 24 hours post inoculation (h/i), while at other points the transcription level did not exceed the control. In the group of untreated plants infected with the Avr6 isolate, no increase in the amount of *CERK1* transcripts was observed throughout the experiment, compared to the control.

Novochizol-pretreated plants showed different responses to *Pgt* infection depending on the isolate type (Fig. 2*a*). In vr6-infected plants, we observed an increase in *CERK1* transcript levels at 3 and 48–144 h/i relative to untreated plants infected with the same isolate, although the significance of this increase at individual points was low. At the same time, Avr6-infected plants pretreated with the drug showed a significant increase in *CERK1* transcription by approximately 2–3 times relative to untreated plants in the range of 24–72 h/i. In uninfected plants treated with the preparation, the level of *CERK1* transcription did not significantly exceed that of the control plants, and at some points even decreased (Fig. 2*a*). Thus, although *CERK1* induction under the influence of the biopesticide Novochizol was observed for both isolates, the degree of this induction was higher in the case of the Avr6 isolate.

Induction of the *PR3* and *PR4* genes encoding various chitinases. The *PR3* and *PR4* genes encode proteins with chitinase activity that cleave the chitin of pathogenic fungi, herbivorous insects, and nematodes (Van Loon et al., 2006). These enzymes do not have an endogenous substrate and play an important role in the lysis of the cell walls of many pathogens.

Transcription of both genes in control plants and plants treated with Novochizol without infection was approximately at the same level, without any significant changes (Fig. 2*b*, *c*). In untreated plants infected with vr6, a slight increase in the *PR3* transcription level was observed at 24 h/i, followed by a decrease to the initial level on the third day after inoculation. In untreated Avr6-infected plants, an elevated mean *PR3* transcription level was observed at 3 h/i; no significant changes were observed relative to controls at other time points. Regarding *PR4*, the expression of this gene in untreated vr6-infected plants showed an increase in the mean level at 144 h/i, whereas induction was observed as early as 3 h/i after Avr6 infection.

Treatment with the preparation induced higher levels of *PR3* and *PR4* gene transcription in response to *Pgt* infection compared to untreated plants. Thus, in the case of vr6 infection of treated plants, we observed increased transcription of the *PR3* gene after 3–24 h/i with a maximum level at

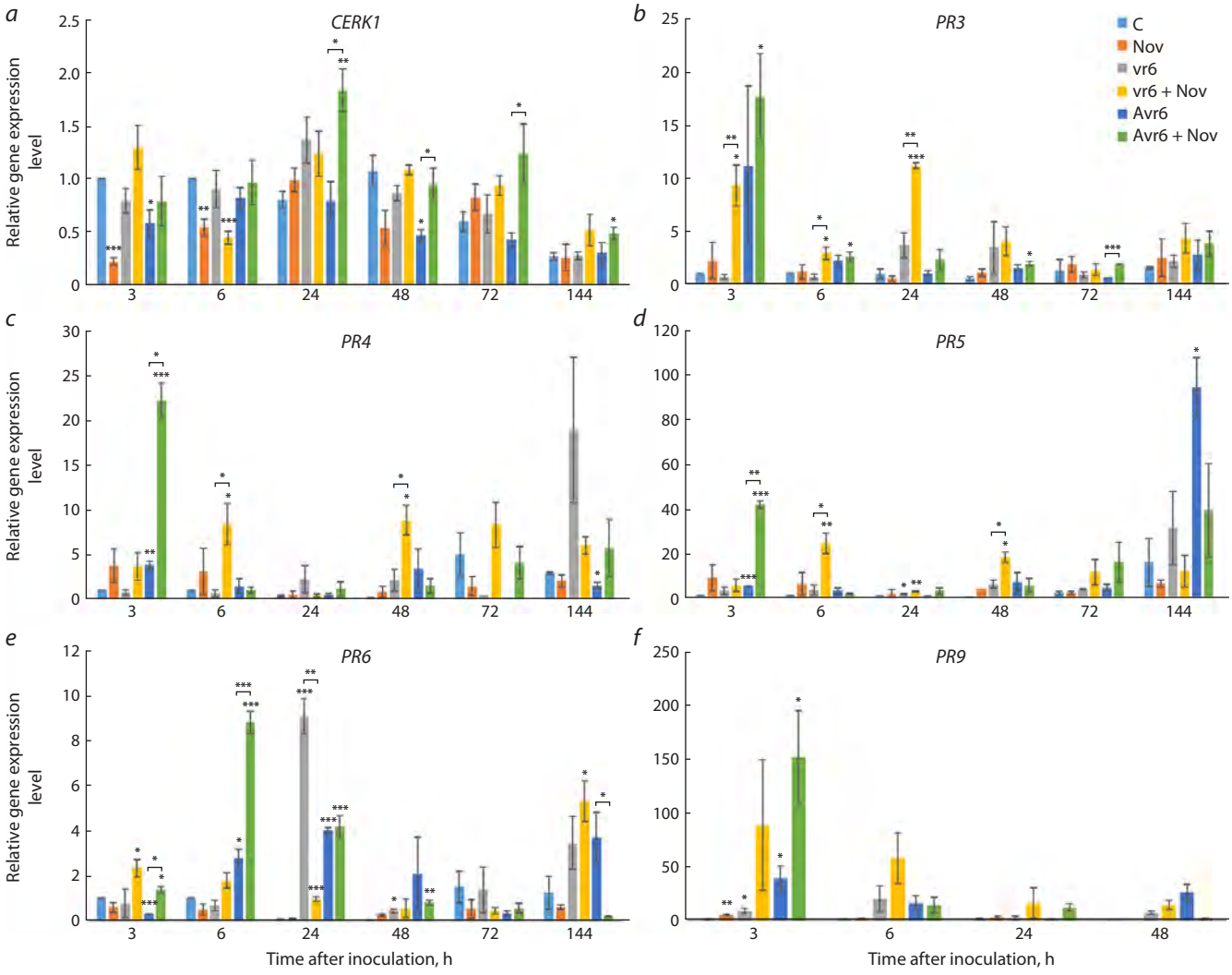


Fig. 2. Dynamics of gene transcription of some protective proteins in leaves of common wheat line ISr6-Ra after their treatment with 0.125 % Novochizol and inoculation with the vr6 or Avr6 – *Pgt* isolate.

C – control; Nov – treatment with the preparation; vr6 – inoculation with isolate vr6; vr6 + Nov – treatment with the preparation and inoculation with isolate vr6; Avr6 – inoculation with isolate Avr6; Avr6 + Nov – treatment with the preparation and inoculation with isolate Avr6. Columns represent the average level of gene transcripts relative to the reference (± standard error of the mean). Comparisons were made with control plants at individual time points and pairwise comparisons of vr6/vr6 + Nov and Avr6/Avr6 + Nov (in parentheses); *p*-values: * $p \le 0.05$, ** $p \le 0.01$, *** $p \le 0.001$.

24 h/i, and for PR4, after 6 and 48–72 h/i. Transcription of both genes increased after 3 h/i when plants were infected with the Avr6 isolate, and the statistically most significant increase compared to untreated plants was found for the *PR4* gene (Fig. 2*b*, *c*).

Expression of *PR5*, *PR6* and *PR9* genes. In this work, we also studied the dynamics of transcription of the *PR5*, *PR6* and *PR9* genes, which play an important role in plant defense against pathogens. The *PR5* gene encodes a thaumatin-like protein with antifungal activity and is also an osmoregulator that protects plants from abiotic stress (Van Loon et al., 2006). The *PR6* gene product is a proteinase inhibitor that suppresses the pathogen's proteolytic enzymes. The peroxidase encoded by the *PR9* gene plays an important role in strengthening the plant cell wall and in the formation of ROS to create a toxic environment for the pathogen (Almagro et al., 2009).

The transcription levels of the *PR5*, *PR6* and *PR9* genes in the control and uninfected plants treated with Novochizol did not undergo significant changes during the experiment (Fig. 2*d–f*). In untreated plants infected with one or another *Pgt* isolate, accumulation of *PR5* transcripts was observed only on the 6th day of infection with the pathogen (Fig. 2*d*). Under the influence of the preparation, an earlier increase in the *PR5* transcription level was observed in plants infected with vr6, especially noticeable after 6 and 48 h/i. In plants infected with Avr6, pretreatment with the preparation caused a significant induction of *PR5* transcription after 3 h/i with a subsequent decrease in the expression level. It should be noted that there are signs of similarity in the dynamics of *PR5* transcription with that of the *PR3* and *PR4* genes (see above).

In untreated vr6-infected plants, *PR6* transcription peaked at 24 h/i, whereas in untreated Avr6-infected plants, tran-

scription of this gene increased less significantly at 6 and 24 h/i (Fig. 2*e*). Interestingly, biopesticide treatment caused a significant decrease in *PR6* transcript levels in vr6-infected plants at 24 h/i. In the case of infection with Avr6, maximum *PR6* expression in treated plants, significantly higher than in untreated plants, was observed at 6 h/i, and then the levels of both groups of plants became similar except at 144 h/i.

In untreated vr6 infected plants, *PR9* transcript levels increased between 3 and 6 h/i and then decreased (Fig. 2*f*). In treated vr6 infected plants, *PR9* transcript levels increased between 3 and 24 h/i with high interplant variability. A lesser increase in expression of this gene was observed in treated Avr6 infected plants, except for 3 h/i, which was the time point at which the highest *PR9* transcript level was recorded. *PR9* transcripts were not detected at 72 and 144 h/i in control and infected samples.

Discussion

The effect of chitosan preparations on defense reactions and signaling pathways that control immunity has been studied in detail in various plant objects, although not enough to fully understand the mechanisms of this effect (review: Shcherban, 2023). However, so far there have been only a few studies devoted to the analysis of the effects of Novochizol, a globular derivative of chitosan (Teplyakova et al., 2022; Shcherban et al., 2024; Orlova et al., 2025). Our work provides new information on the changes in the expression of defense genes induced by this preparation during infection of wheat with the causative agent of a dangerous fungal disease, stem rust.

In response to pathogen attacks, plants have developed a complex immune system, including PRRs, PTI-mediated receptors that represent the first line of plant defense against infections. PRRs are central elements of plant immunity, triggering a cascade of defense reactions, including ROS accumulation, *PR* gene expression, and other responses (Jones and Dangl, 2006).

In our study, we analyzed the effect of Novochizol on the expression of the gene encoding PRR CERK1 in leaf tissue of wheat plants infected with different *Pgt* isolates. In response to infection with avirulent and virulent isolates of *Pgt*, plants not treated with the preparation did not show significant changes in the expression level of this gene throughout the experiment (Fig. 2*a*). Also, no significant changes in the expression level were detected in uninfected plants treated with the preparation.

Previously, activation of *CERK1* gene transcription in ear tissue was shown under the influence of chitin and during infection with the causative agent of fusarium *Fusarium graminearum* Schwabe (Wang L. et al., 2024). Apparently, the causative agents of *Fusarium* and stem rust provoke different reactions of the defense system in various plant tissues, and chitin, a natural elicitor, can provoke a more pronounced reaction compared to the Novochizol preparation.

Plants treated with the preparation demonstrated a more pronounced response to infection with *Pgt*, which depended on the isolate type. So, when plants were infected with vr6, we observed a slight increase in the *CERK1* transcription level after 3 h/i and in the range of 48–144 h/i, whereas when plants were infected with the avirulent isolate Avr6, the degree of *CERK1* induction was higher than in the previous case and exceeded the expression level of this gene in untreated plants by 2–3 times in the range of 24–72 h/i. This increase in Avr6-infected plants can be explained by the close relationship between PTI and ETI.

Previous studies have shown that these levels of defense, activated in plants by surface and intracellular receptors, respectively, are mutually reinforcing (Ngou et al., 2021). From this perspective, it can be speculated that the presence of an intracellular plant receptor (*Sr6* gene product) complementary to the Avr6 product of the *Pgt* gene is an additional trigger capable of enhancing PTI by increasing *CERK1* gene expression. Recently, *CERK1* overexpression has been shown to confer broad-spectrum resistance to fungal diseases in wheat (Wang L. et al., 2024).

Genes encoding the PR3 and PR4 proteins have endochitinase activity, are capable of destroying the cell wall of fungi, and are activated when plants are exposed to various pathogens (Van der Bulcke et al., 1990; Ward et al., 1991; Hammond-Kosack, Gones, 1996; Gao et al., 2013). According to the data obtained in the work, upon infection with vr6, the level of *PR3* gene transcription increased after 24 h/i, while the *PR4* gene was induced much later, on the 6th day after inoculation (Fig. 2*b*, *c*). During Avr6 infection, a synchronous increase in the expression of both genes occurred at 3 h/i.

As in the case of *CERK1*, treatment with the drug caused an increase in the transcription of both chitinase genes, compared to untreated plants, which was most pronounced in the case of the Avr6 isolate at the same point as in untreated Avr6-infected plants – 3 h/i. Thus, the *PR3* and *PR4* genes are characterized by early accumulation of transcripts during infection with the Avr6 isolate, while Novochizol has an additional stimulating effect on the expression of these genes. It has been previously shown that treatment with chitosan preparations causes early accumulation of chitinases, which non-specifically destroy the chitin of fungal cell walls and thereby increase plant resistance to a wide range of fungal diseases (Manjunatha et al., 2008; Ma et al., 2013; Elsharkawy et al., 2022).

Numerous studies have shown that overexpression of thaumatin-like proteins (TLPs) belonging to the PR5 family provides significant plant resistance to various pathogenic fungi (Wang X. et al., 2010; Cui et al., 2021). We showed that, similar to the *PR4* gene, the accumulation of *PR5* transcripts occurred late – at 144 h/i in untreated plants infected with vr6 (Fig. 2*d*). Treatment with the drug resulted in earlier accumulation of *PR5* transcripts at 6 and 48 h/i in plants infected with the same isolate. Finally, similar to *PR4*, *PR5* showed significant induction of its expression at 3 h/i in treated plants infected with Avr6. Such similarity in the dynamics of *PR4* and *PR5* expression indicates a common mechanism of regulation of these genes, depending on the pathogen genotype.

Regarding the possible target of PR5 protein action, it was shown that overexpression of TLP in grape *Vitis vinifera* L.

leads to a decrease in hyphal growth of the pathogenic fungus *Plasmopara viticola* (He et al., 2017). The exact mechanism of TLP action has not been established, however, PR5 was reported to degrade β-1,3-glucans, which, like chitin, are an important structural component of fungal cell walls (Grenier et al., 1999).

An important role of the *PR6* gene product is to inhibit pathogen proteinases that degrade plant proteins (Ryan, 2000). According to our data, the expression of this gene in untreated plants showed a pattern similar to that of other *PR* genes studied in this work: a slight increase in the transcription level at 24 h/i in the case of vr6 and at 6, 24 and 144 h/i in the case of Avr6 (Fig. 2*e*). A special feature of this gene is its different response to Novochizol, depending on the type of isolate, at 24 h/i. The vr6 isolate at this point showed a sharp decrease in *PR6* transcription as a result of treatment, whereas for Avr6, the transcription levels in treated and untreated plants were approximately the same. In the previous 6 h/i, we observed a more typical picture for Avr6, namely a significant increase in *PR6* transcription in treated plants compared to untreated ones. It can be assumed that the peak of *PR6* functional activity occurs in the period up to 24 h/i, after which its role decreases, which causes variability in the transcription level.

The *PR9* gene encodes peroxidase, which is involved in hydrogen peroxide detoxification and lignin synthesis. The latter substance, accumulating in the plant cell wall, creates a barrier to pathogen spread (Andreeva, 1988). In vr6-treated infected plants, we observed an increase in *PR9* transcription after 3–24 h/i (Fig. 2*f*). Treated plants infected with Avr6 showed an increased average transcription level after 3 h/i and no increase at other time points relative to untreated plants infected with the same isolate. The increased dispersion of data observed for both isolates can be explained by the uneven distribution of pustules on individual leaves (Fig. 1); the highest level of peroxidase activity and, accordingly, transcription of this gene can be expected in the areas of formation of morphological structures of the fungus (see below).

Nevertheless, these data correlate with a recent biochemical analysis, which showed an increase in the enzymatic activity of peroxidase in the 6–144 h/i period in Novochizol vr6-infected plants (data not published). We assume that the increase in peroxidase activity in this group of plants under the influence of the preparation is associated with the acceleration of the lignification process. Previous histochemical analysis revealed earlier and more intensive accumulation of phenolic substances in the cytoplasm and lignin on the cell walls in the *Pgt* colony zones in plants treated with the preparation compared to untreated plants (Shcherban et al., 2025).

Thus, the induction of defense genes in plants treated with Novochizol is enhanced in almost all cases after pathogen inoculation, which is consistent with the previously proposed hypothesis of priming, a rapid response of the immune system to a subsequent pathogen attack under the influence of elicitors (Conrath, 2011).

Conclusion

The data obtained indicate that the biopreparation Novochizol increases the quantitative component of resistance of the isogenic line ISr6-Ra to stem rust both in the case of the Avr6 isolate *Pgt* and in the case of the vr6 isolate. Analysis of the transcription of the defense genes *CERK1*, *PR3*, *PR4*, *PR5*, *PR6* showed an increase in the transcription level of these genes in plants treated with the preparation and infected with various isolates of *Pgt* compared to untreated infected plants. The data obtained indicate that the use of the biopesticide Novochizol in combination with selection for resistance genes (pyramiding) is an effective way to increase the resistance of common wheat to stem rust.

References

Ali S., Ahmad N., Dar M.A., Manan S., Rani A., Alghanem S.M.S., Khan K.A., Sethupathy S., Elboughdiri N., Mostafa Y.S., Alamri S.A., Hashem M., Shahid M., Zhu D. Nano-agrochemicals as substitutes for pesticides: prospects and risks. *Plants* (*Basel*). 2023; 13(1):109. doi 10.3390/plants13010109

Almagro L., Gómez Ros L.V., Belchi-Navarro S., Bru R., Ros Barceló A., Pedreño M.A. Class III peroxidases in plant defence reactions. *J Exp Bot*. 2009;60(2):377-390. doi 10.1093/jxb/ern277

Andreeva V.A. The Enzyme Peroxidase: Participation in the Protective Mechanism of Plants. Moscow: Nauka Publ., 1988 (in Russian)

Bent A.F., Mackey D. Elicitors, effectors, and *R* genes: the new paradigm and a lifetime supply of questions. *Annu Rev Phytopathol*. 2007;45:399-436. doi 10.1146/annurev.phyto.45.062806.094427

Bertini L., Leonardi L., Caporale C., Tucci M., Cascone N., Di Berardino I., Buonocore V., Caruso C. Pathogen-responsive wheat PR4 genes are induced by activators of systemic acquired resistance and wounding. *Plant Sci*. 2003;164(6):1067-1078. doi 10.1016/S0168-9452(03)00112-2

Bigeard J., Colcombet J., Hirt H. Signaling mechanisms in pattern triggered immunity (PTI). *Mol Plant*. 2015;8(4):521-539. doi 10.1016/j.molp.2014.12.022

Chai Y., Senay S., Horvath D., Pardey P. Multi-peril pathogen risks to global wheat production: a probabilistic loss and investment assessment. *Front Plant Sci*. 2022;13:1034600. doi 10.3389/fpls.2022.1034600

Conrath U. Molecular aspects of defence priming. *Trends Plant Sci*. 2011;16(10):524-531. doi 10.1016/j.tplants.2011.06.004

Cui Z., Liang F., Zhang J., Wang F., Liu D., Wang H. Transgenic expression of *TaTLP1*, a thaumatin-like protein gene, reduces susceptibility to common root rot and leaf rust in wheat. *Crop J*. 2021;9(5): 1214-1218. doi 10.1016/j.cj.2021.03.021

Desmond O.J., Edgar C.I., Manners J.M., Maclean D.J., Schenk P.M., Kazan K. Methyl jasmonate induced gene expression in wheat delays symptom development by the crown rot pathogen *Fusarium pseudograminearum*. *Physiol Mol Plant Pathol*. 2005;67(3-5):171-179. doi 10.1016/j.pmpp.2005.12.007

Ding Y., Sun T., Ao K., Peng Y., Zhang Y., Li X., Zhang Y. Opposite roles of salicylic acid receptors NPR1 and NPR3/NPR4 in transcriptional regulation of plant immunity. *Cell*. 2018;173(6):1454-1467.e15. doi 10.1016/j.cell.2018.03.044

Elsharkawy M.M., Omara R.I., Mostafa Y.S., Alamri S.A., Hashem M., Alrumman S.A., Ahmad A.A. Mechanism of wheat leaf rust control using chitosan nanoparticles and salicylic acid. *J Fungi* (*Basel*). 2022;8(3):304. doi 10.3390/jof8030304

Fomenko V., Loroch V. Novochizol: a new type of cross-linked chitosan particles for formulation and parsimonious delivery of copper compounds. In: 6th European Copper Conference in Plant Protection, Berlin, Germany, 17–18 November 2021. Berlin: IFOAM Organics Europe, 2021. Available online: https://www.boelw.de/fileadmin/user_upload/Dokumente/Veranstaltungen/Kupfertagung_2021/copper_novochizol.pdf

Gao C., Kou X., Li H., Zhang J., Saad A., Liao Y. Inverse effects of *Arabidopsis NPR1* gene on fusarium seedling blight and fusarium head blight in transgenic wheat. *Plant Pathol*. 2013;62(2):383-392. doi 10.1111/J.1365-3059.2012.02656.x

Gong B.Q., Wang F.Z., Li J.F. Hide-and-seek: chitin-triggered plant immunity and fungal counterstrategies. *Trends Plant Sci*. 2020;25(8): 805-816. doi 10.1016/j.tplants.2020.03.006

Grenier J., Potvin C., Trudel J., Asselin A. Some thaumatin-like proteins hydrolyse polymeric β-1,3-glucans. *Plant J*. 1999;19(4):473-480. doi 10.1046/j.1365-313x.1999.00551.x

Hafeez A.N., Arora S., Ghosh S., Gilbert D., Bowden R.L., Wulff B.B.H. Creation and judicious application of a wheat resistance gene atlas. *Mol Plant.* 2021;14(7):1053-1070. doi 10.1016/j.molp.2021.05.014

Hammond-Kosack K.E., Gones J.D.G. Resistance gene dependent plant defence responses. *Plant Cell.* 1996;8(10):1773-1791. doi 10.1105/tpc.8.10.1773

Hao G., Tiley H., McCormick S. Chitin triggers tissue-specific immunity in wheat associated with Fusarium head blight. *Front Plant Sci*. 2022;13:832502. doi 10.3389/fpls.2022.832502

He R., Wu J., Zhang Y., Agüero C.B., Li X., Liu S., Wang C., Walker M.A., Lu J. Overexpression of a thaumatin-like protein gene from *Vitis amurensis* improves downy mildew resistance in *Vitis vinifera* grapevine. *Protoplasma*. 2017;254:1579-1589. doi 10.1007/s00709-016-1047-y

Jones J.D.G., Dangl J.L. The plant immune system. *Nature*. 2006; 444(7117):323-329. doi 10.1038/nature05286

Katiyar D., Hemantaranjan A., Singh B., Bhanu A.N. A future perspective in crop protection: chitosan and its oligosaccharides. *Adv Plants Agric Res*. 2014;1(1):23-30. doi 10.15406/apar.2014.01.00006

Lee W.S., Rudd J.J., Hammond-Kosack K.E., Kanyuka K. *Mycosphaerella graminicola* LysM effector-mediated stealth pathogenesis subverts recognition through both CERK1 and CEBiP homologues in wheat. *Mol Plant Microbe Interact*. 2014;27(3):236-243. doi 10.1094/ MPMI-07-13-0201-R

Ma Z., Yang L., Yan H., Kennedy J.F., Meng X. Chitosan and oligochitosan enhance the resistance of peach fruit to brown rot. *Carbohydr Polym*. 2013;94(1):272-277. doi 10.1016/j.carbpol.2013.01.012

Maluin F.N., Hussein M.Z. Chitosan-based agronanochemicals as a sustainable alternative in crop protection. *Molecules*. 2020;25(7): 1611. doi 10.3390/molecules25071611

Manjunatha G., Roopa K.S., Prashanth G.N., Shetty H.S. Chitosan enhances disease resistance in pearl millet against downy mildew caused by *Sclerospora graminicola* and defence-related enzyme activation. *Pest Manag Sci*. 2008;64(12):1250-1257. doi 10.1002/ps.1626

McIntosh R.A., Dubcovsky J., Rogers W.J., Morris C., Appels R., Xia X.C. Catalogue of gene symbols for wheat: 2011 supplement. IWGS, 2011. Available online: http://www.shigen.nig.ac.jp/wheat/komugi/genes/macgene/supplement2011.pdf

Ngou B.P.M., Ahn H.K., Ding P., Jones J.D. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature*. 2021;592(7852):110-115. doi 10.1038/s41586-021-03315-7

Orlova E.A., Bekhtold N.P., Shcherban A.B., Fomenko V.V. The effect of new biological products based on Novochizol on the condition of spring common wheat crops. *Zernovoe Hozyajstvo Rossii = Grain Economy of Russia*. 2025;17(2):86-93. doi 10.31367/2079-8725-2025-97-2-86-93 (in Russian)

Patpour M., Hovmøller M.S., Rodriguez-Algaba J., Randazzo B., Villegas D., Shamanin V.P., Berlin A., … Meyer K.J.G., Valade R., Thach T., Hansen J.G., Justesen A.F. Wheat stem rust back in Europe: diversity, prevalence and impact on host resistance. *Front Plant Sci*. 2022;13:882440. doi 10.3389/fpls.2022.882440

Pfaffl M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29(9):e45. doi 10.1093/nar/29.9.e45

Pritsch C., Muehlbauer G.J., Bushnell W.R., Somers D.A., Vance C.P. Fungal development and induction of defence response genes during early infection of wheat spikes by *Fusarium graminearum*. *Mol Plant Microbe Interact*. 2000;13(2):159-169. doi 10.1094/MPMI.2000.13.2.159

Ray S., Anderson J.M., Urmeev F.I., Goodwin S.B. Rapid induction of a protein disulfide isomerase and defense-related genes in wheat in response to the hemibiotrophic fungal pathogen *Mycosphaerella graminicola*. *Plant Mol Biol*. 2003;53(5):701-714. doi 10.1023/B:PLAN.0000019120.74610.52

Roelfs A.P., Singh R.P., Saari E.E. Rust diseases of wheat: concepts and methods of disease management. Mexico: CIMMYT, 1992

Ruijter J.M., Ramakers C., Hoogaars W.M.H., Karlen Y., Bakker O., van den Hoff M.J.B., Moorman A.F.M. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res*. 2009;37(6):e45. doi 10.1093/nar/gkp045

Ryan C.A. The systemin signaling pathway: differential activation of plant defensive genes. *Biochim Biophys Acta*. 2000;1477(1-2):112-121. doi 10.1016/s0167-4838(99)00269-1

Sanin S.S. Epiphytoties of Diseases of Grain Crops: Theory and Practice. Moscow, 2012 (in Russian)

Scherer N.M., Thompson C.E., Freitas L.B., Bonatto S.L. Patterns of molecular evolution in pathogenesis-related proteins. *Genet Mol Biol*. 2005;28(4):645-653. doi 10.1590/S1415-47572005000500001

Sels J., Mathys J., De Coninck B.M., Cammue B.P., De Bolle M.F. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol Biochem*. 2008;46(11):941-950. doi 10.1016/j.plaphy.2008.06.011

Shcherban A.B., Skolotneva E.S., Fedyaeva A.V., Boyko N.I., Fomenko V.V. Effect of biopesticide Novochizol on development of stem rust *Puccinia graminis* f. sp. *tritici* in wheat, *T. aestivum* L. *Plants*. 2024;13(23):3455. doi 10.3390/plants13233455

Shcherban A.B. Chitosan and its derivatives as promising plant protection tools. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(8):1010-1021. doi 10.18699/VJGB-23-116

Shcherban A.B., Plotnikova L.Ya., Knaub V.V., Skolotneva E.S., Fomenko V.V. Cyto-physiological manifestations of protective reactions of wheat against stem rust, induced by the biofungicide Novochizol. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(4):539-548. doi 10.18699/vjgb-25-57

Skolotneva E.S., Kelbin V.N., Morgunov A.I., Boyko N.I., Shamanin V.P., Salina E.A. Races composition of the Novosibirsk population of *Puccinia graminis* f. sp. *tritici*. *Mykologiya i Fitopatologiya = Mycology and Phytopathology*. 2020;54(1):49-58. doi 10.31857/S0026364820010092 (in Russian)

Tada Y., Spoel S.H., Pajerowska-Mukhtar K., Mou Z., Song J., Wang C., Zuo J., Dong X. Plant immunity requires conformational changes [corrected] of NPR1 via S-nitrosylation and thioredoxins. *Science*. 2008;321(5891):952-956. doi 10.1126/science.1156970

Tarchevskij I.A. Signal systems of plant's cells. Moscow: Nauka Publ., 2002 (in Russian)

Teplyakova O.I., Fomenko V.V., Salakhutdinov N.F., Vlasenko N.G. Novochizol™ seed treatment: effects on germination, growth and development in soft spring wheat. *Nat Prod Chem Res*. 2022; 10(5):1-4. doi 10.35248/naturalproducts.10.5.1-04

Tsilo T.J., Chao S., Jin Y., Anderson J.A. Identification and validation of SSR markers linked to the stem rust resistance gene *Sr6* on the short arm of chromosome 2D in wheat. *Theor Appl Genet.* 2009; 118(3):515-524. doi 10.1007/s00122-008-0917-x

Van der Bulcke C., Bauw G., De Rucke R., Castresana C. The role of vacuolar and secreted pathogenesis-related B (1-3)-gluconases and chitinases in the defense response of plants. *Bull Soc Bot Fr*. 1990;137(3-4):51-63. doi 10.1080/01811789.1990.10827029

Van Loon L.C., Van Strien E.A. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol Mol Plant Pathol*. 1999;55(2):85-97. doi 10.1006/pmpp.1999.0213

Van Loon L.C., Rep M., Pieterse C.M.J. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phyto-*

pathol. 2006;44:135-162. doi 10.1146/annurev.phyto.44.070505.143425

Varlamov V.P., Ilyina A.V., Shagdarova B.Ts., Lunkov A.P., Mysyakina I.S. Chitin/chitosan and its derivatives: fundamental problems and practical approaches. *Biochemistry*. 2020;85:154-176. doi 10.1134/s0006297920140084

Wang L., He Y., Guo G., Xia X., Dong Y., Zhang Y., Wang Y., Fan X., Wu L., Zhou X., Zhang Z., Li G. Overexpression of plant chitin receptors in wheat confers broad-spectrum resistance to fungal diseases. *Plant J*. 2024;120(3):1047-1063. doi 10.1111/tpj.17035

Wang X., Tang C., Deng L., Cai G., Liu X., Liu B., Han Q., Buchenauer H., Wei G., Han D., Huang L., Kang Z. Characterization of a pathogenesis-related thaumatin-like protein gene TaPR5 from wheat induced by stripe rust fungus. *Physiol Plant*. 2010;139(1):27-38. doi 10.1111/j.1399-3054.2009.01338.x

Ward E.R., Uknes S.J., Williams S.C., Dincher S.S., Wiederhold D.L., Alexander D.C., Ahl-Goy P., Metraux J.P., Ryalset J.A. Coordinate gene activity in response to agents that induce systemic acquired resistance. *Plant Cell*. 1991;3(10):1085-1094. doi 10.1105/tpc.3.10.1085

Conflict of interest. The authors declare no conflict of interest.
Received June 5, 2025. Revised July 17, 2025. Accepted July 30, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1213-1220

doi 10.18699/vjgb-25-128

Differential expression profile of DREB2 subfamily transcription factor genes in the dynamics of salt stress and post-stress recovery in tomato plants

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

Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia
michel7753@mail.ru

Abstract. In response to stress, epigenetic modifications occur in the plant genome, which together form a stress memory that can be inherited and increases the efficiency of the plant's defense response to repeated stress events. Genes whose expression becomes the target of epigenetic modifications serve as biomarkers of stress memory. Their characteristic features are considered to be an expression profile that differs between responses to primary and repeated stress events, as well as long-term retention of changes after the stress is canceled. Tomato (*Solanum lycopersicum* L.) is an important vegetable crop whose yield decreases with soil salinity. Genes induced by salt stress include genes encoding transcription factors of the DREB2 (DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEIN 2) subfamily. In this work, we evaluated the *SlDREB2* genes of tomato as possible marker genes of salt stress memory. The expression of the genes *SlDREB16*, *20*, *22*, *24*, *43*, *44* and *46* was determined in the leaves of two plant varieties (Gnom, Otradnyi) with different degrees of salt tolerance in response to 24 h of NaCl exposure and in the dynamics of a long-term (14 days) post-stress recovery period. Significant genotype-specific fluctuations in the levels of gene transcripts were revealed both in the control and in the stressed plants. It was shown that during the long-term memory phase, gene expression returns to the control values either temporarily (*SlDREB24*, *44* and *46* in the moderately resistant Gnom variety after 7 days; after 14 days, the expression changed again) or slowly (*SlDREB16* and *43* in the highly resistant Otradnyi variety after 14 days of recovery). Only two genes (*SlDREB22* and *46*) showed a similar between varieties pattern of expression fluctuations in the dynamics of stress and recovery, and the *SlDREB20* gene was not expressed in either the control or the experiment. The data obtained suggest that the *SlDREB2* subfamily genes (except *SlDREB20*) are involved in the response of *S. lycopersicum* to salt stress in a genotype-specific manner and can serve as markers of stress memory linked to the epigenetic regulation of tomato adaptation to salt stress. The *SlDREB16*, *28*, *43* and *44* genes may contribute to the determination of differences in the mechanism of regulation of plant response to salt stress between salt-tolerant genotypes of *S. lycopersicum*. The obtained results can form the basis for further studies of the role of *SlDREB2* genes in the epigenetic regulation of tomato plant adaptation to salt stress, which can be used in breeding salt-tolerant varieties.

Key words: tomato; *Solanum lycopersicum* L.; salt stress; stress memory; memory phase; SlDREB2 transcription factors; gene expression; potential stress memory genes

For citation: Filyushin M.A., Shchennikova A.V., Kochieva E.Z. Differential expression profile of DREB2 subfamily transcription factor genes in the dynamics of salt stress and post-stress recovery in tomato plants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1213-1220. doi 10.18699/vjgb-25-128

Funding. The study was supported by the RSF grant No. 24-16-00043 (analysis of gene expression) and the Ministry of Science and Higher Education of the Russian Federation (plant material).

Дифференциальный профиль экспрессии генов транскрипционных факторов подсемейства DREB2 в динамике солевого стресса и послестрессового восстановления растений томата

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

Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия
michel7753@mail.ru

Аннотация. В ответ на стрессовое воздействие в геноме растения происходят эпигенетические модификации, вместе формирующие стрессовую память, которая может наследоваться и повышает эффективность защитной реакции растения на повторные стрессовые события. Гены, чья экспрессия становится мишенью эпигенетических модификаций, служат биомаркерами стрессовой памяти. Их характерными признаками считаются профиль экспрессии, различающийся между ответами на первичное и повторное стрессовые события, а также длительное

удерживание изменений после отмены стресса. Томат (*Solanum lycopersicum* L.) – важная овощная культура, урожайность которой снижается при засолении почв. К генам, индуцируемым солевым стрессом, относятся гены транскрипционных факторов подсемейства DREB2 (DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEIN 2). В настоящей работе проведена оценка генов *SlDREB2* томата как возможных генов-маркеров памяти о солевом стрессе. Для этого в листьях растений двух сортов, Гном и Отрадный, с разной степенью солеустойчивости была определена экспрессия генов *SlDREB16*, *20*, *22*, *24*, *43*, *44* и *46* в ответ на 24-ч воздействие NaCl и в динамике продолжительного (14 дней) послестрессового периода восстановления. Выявлены значительные генотип-специфичные колебания уровней транскриптов генов как в контроле, так и в подвергнутых стрессу растениях. Показано, что в процессе длительной фазы памяти экспрессия генов возвращается к контрольным показателям либо временно (*SlDREB24*, *44* и *46* у среднеустойчивого сорта Гном через 7 дней; через 14 дней экспрессия снова меняется), либо медленно (*SlDREB16* и *43* у высокоустойчивого сорта Отрадный через 14 дней восстановления). Было определено, что только два гена, *SlDREB22* и *46*, имеют сходный между сортами паттерн колебаний экспрессии в динамике стресса и восстановления, а ген *SlDREB20* не экспрессируется ни в контроле, ни в опыте. Полученные данные позволили предположить, что гены подсемейства *SlDREB2*, кроме *SlDREB20*, участвуют в ответе *S. lycopersicum* на солевой стресс генотип-специфичным образом и могут служить маркерами стрессовой памяти, сцепленными с эпигенетической регуляцией адаптации томата к солевому стрессу. Гены *SlDREB16*, *28*, *43* и *44* могут вносить вклад в определение различий в механизме регуляции ответа растений на солевой стресс между солеустойчивыми генотипами *S. lycopersicum*. Полученные результаты могут стать основой для дальнейших исследований роли генов *SlDREB2* в эпигенетической регуляции адаптации растений томата к солевому стрессу, что может быть использовано в селекции солеустойчивых сортов.

Ключевые слова: томат; *Solanum lycopersicum* L.; солевой стресс; стрессовая память; фаза памяти; транскрипционные факторы SlDREB2; экспрессия генов; гены-кандидаты стрессовой памяти

Introduction

The plant phenotype is formed through the combined action of the genotype and the epigenome, where the latter determines the plasticity of the phenotype depending on environmental conditions, including in response to various stress factors, which are often recurrent (Villagómez-Aranda et al., 2022). The initial (during the plant's life cycle) experience of stress (priming) induces changes in the epigenome (DNA methylation, post-translational histone modifications, non-coding RNA activity, etc.), which enable a more effective response to repeated stress (stimulus) (Villagómez-Aranda et al., 2022).

The set of epigenetic marks that emerge during priming is called the plant's stress memory, which can persist throughout the organism's life cycle and be inherited (Villagómez-Aranda et al., 2022; Zuo et al., 2023). That is, the plant's stress memory is the initial experience of effectively regulating the stress response, imprinted in the epigenome, which, upon a repeated stress event, can quickly trigger the transcriptomic and metabolomic changes necessary for protection (Villagómez-Aranda et al., 2022; Zuo et al., 2023).

Biomarkers of stress memory are generally considered to be individual genes (metabolites), the expression (metabolism) of which becomes the target of epigenetic modifications after priming (Aina et al., 2024). There may be many such markers. For example, drought stress memory in the model species *Arabidopsis thaliana* L. is associated with more than 2,000 genes (Ding et al., 2013). A comparison of this list with a similar set in *Zea mays* L. reduced the list to 556 genes as possible interspecific markers of plant memory about drought (Ding et al., 2014; Virlouvet et al., 2018; Jacques et al., 2021). When selecting candidate memory markers, the principle is that the level and/or direction of changes in gene expression (metabolite content) differs between responses to priming and stimulus, while genes (metabolites) not associated with memory respond equally to priming and stimulus (Friedrich et al., 2019; Bäurle, Trindade, 2020; Jacques et al., 2021). Another important criterion is that during the period between stress repeats (recovery, or memory phase), the expression of marker genes (metabolite content) is maintained at an altered level for a long time, while the expression of genes (metabolite content) not associated with memory quickly returns to control values (Friedrich et al., 2019; Bäurle, Trindade, 2020; Jacques et al., 2021).

An example of the criteria use is a metabolomic analysis of the halophyte *Limonium angustebracteatum*'s response to repeated drought and salt stresses, which identified various organic osmolytes and antioxidant compounds (including flavonoids) as potential markers of stress memory (Calone et al., 2023). Transcriptomic studies of potato (*Solanum tuberosum* L.) under recurrent drought conditions identified potential memory genes, including genes involved in photosynthesis, carbohydrate metabolism, flavonoid metabolism, and others (Chen et al., 2019).

Given the observed associations of various important processes with plant stress memory, studying the effects of stress on the expression of genes of specific metabolic or signaling pathways can help identify marker genes. For example, analysis of the expression dynamics of *AsCHS* genes of the chalcone synthase family (flavonoid pathway) in garlic (*Allium sativum* L.) exposed to abiotic stressors identified only one out of eight genes as a potential biomarker (Anisimova et al., 2025). Another example: tracking changes in the expression of various *PR* genes in garlic cloves in response to priming with an elicitor (chitosan) and a biotic stimulus (infection with *Fusarium proliferatum*) identified candidate genes for markers of *A. sativum* memory of *Fusarium* infection (Filyushin et al., 2022).

Selected stress memory markers (both genes and metabolites) can be used to identify donors of a trait of the desired conditional (epigenetic) resistance to target stressors in crop plants (Aina et al., 2024). In plant genetic engineering, altering the expression of marker genes can facilitate the production of stress-resistant genotypes. For example, overexpression of individual genes from the *WRKY* family increases the resistance of tomato plants (*Solanum lycopersicum* L.) to phytopathogens (Bai et al., 2018), while overexpression of

the *DREB1A* and *OsPIL1* genes increases drought tolerance in *A. thaliana* (Kudo et al., 2017).

Tomato (*S. lycopersicum*) is an important vegetable crop, mainly grown in protected ground; soil salinity is considered one of the main factors reducing tomato crop yield (Guo et al., 2022). Epigenetic marks associated with the formation of salt stress memory in plants (Gallusci et al., 2023) and the mechanisms of salt tolerance in tomato are known (Guo et al., 2022). Among the genes, the expression of which is induced by salt stress, there are genes encoding transcription factors (TFs) of the DREB family (APETALA2/Ethylene Responsive Factor (AP2/ERF) superfamily), in particular the DREB2 (DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEIN 2) subfamily (Bai et al., 2018; Guo et al., 2022). The tomato genome contains seven *SlDREB2* genes (Maqsood et al., 2022).

The aim of this study was to evaluate *SlDREB2* genes as possible marker genes for salt stress memory by profiling gene expression in two *S. lycopersicum* cultivars in response to the NaCl stimulus and during long-term post-stress recovery (memory phase).

Materials and methods

The study involved plants of two salt-tolerant tomato (*S. lycopersicum*) varieties: the highly tolerant cv. Otradnyi and the moderately tolerant cv. Gnom, bred at the Federal Scientific Vegetable Center (FSVC, Moscow Region). Seeds were sown in the soil, and plants were grown until 6–8 leaves developed (experimental climate control facility, Federal Research Center for Biotechnology, Russian Academy of Sciences; day/night cycle – 16/8 h, 23/21 °C).

The obtained plants were exposed to salt stress. Namely, the plants (experimental and control) were transferred from soil to water (after shaking off and washing the roots) and after 1 h transferred to a liquid MS nutrient medium containing (experimental) and not containing (control) 100 mM NaCl. After 24 h, the experimental samples were returned to the MS medium without NaCl and kept for two weeks in parallel with the control. Leaf samples (all leaves from one plant; two biological replicates) were collected at the following time points: S24 (experimental, 24 h of stress exposure) and 24K (control); R7 and R7K (week of the post-stress period); R14 and R14K (two weeks of recovery) (Fig. 1*a*).

The collected samples were ground in liquid nitrogen and used for analysis of the expression of SlDREB2 subfamily genes using quantitative real-time PCR (RT-qPCR). Total RNA was isolated from 0.2–0.5 g of tissue (RNeasy Plant Mini Kit and RNase-free DNase set; QIAGEN, Germany) and used for cDNA synthesis (GoScript Reverse Transcription System; Promega, USA). The concentration of the preparation was determined (Qubit® Fluorometer, Thermo Fisher Scientific, USA; Qubit RNA HS Assay Kit, Invitrogen, USA), and 3 ng of cDNA was used for RT-qPCR with gene-specific primers (Table 1). Primers were designed based on available *S. lycopersicum* genome and transcriptome data (https://www.solgenomics.net/; https://www.ncbi.nlm.nih.gov/). The

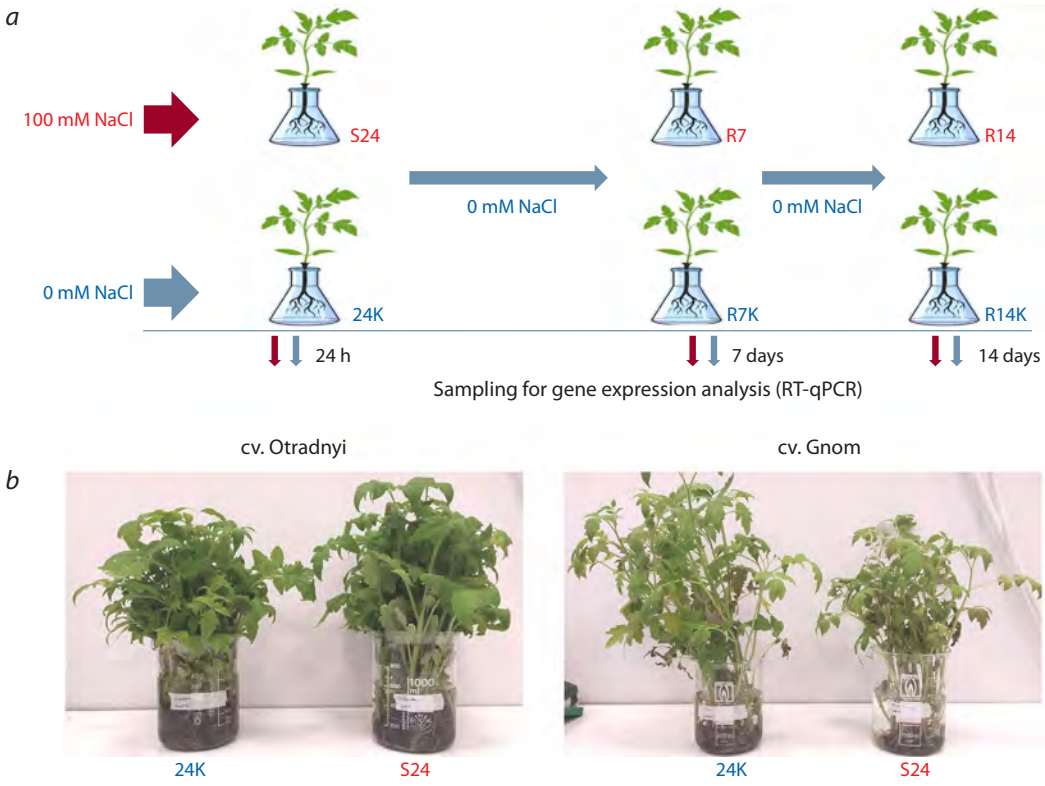


Fig. 1. Experimental design:

a – 24-h salt stress (100 mM NaCl (S24) and 0 mM NaCl (24K)) and post-stress recovery (7 (R7K, R7) and 14 (R14K, R14) days); *b* – photo of experimental (S24) and control (24K) plants of the Otradnyi and Gnom tomato varieties after 24 h of stress.

Table 1. Primer sequences for RT-qPCR

Gene[1]	Gene ID[1]	Sequences (5′→3′) of the forward and reverse primers
SlDREB16	Solyc04g050750	TGAGAGGTAAAGGTGGACCAG CTCTGGCTGCTTCCACTGAA
SlDREB20	Solyc04g080910 (LOC101246344)	GGTAAATGGGTAGCCGAGATC AGTTTGGCTTCGGCACCATAG
SlDREB22	Solyc05g052410 (LOC101261712)	GATACATTGGAAGGTCTGCAGC CATCCAACTCATCCAAGCAGAAG
SlDREB24	Solyc06g050520	GGTATCAGATTCGCAATGTCAGG GAATGCAAGGTCATCCGAACTG
SlDREB43	Solyc10g076370 (LOC101245410)	GGTACTTACTCGACTGCTGGT CAATCGACGATTGACCACTCG
SlDREB44	Solyc10g076380 (LOC101268444)	TCGCCTGCTTGTTCCTGGAA CGCACCAACATCTTCATTCACG
SlDREB46	Solyc10g080310 (LOC101268444)	CAATGTAGCCGTTCGTGGTG TGACTCTGTGAAACTACTGATGC
Expressed	(LOC101263039)	GCTAAGAACGCTGGACCTAATG TGGGTGTGCCTTTCTGAATG
actin-7	(LOC101262163)	CATTGTGCTCAGTGGTGGTTC TCTGCTGGAAGGTGCTAAGTG

[1] The numbering and Solyc_IDs of genes are given according to (Maqsood et al., 2022); the corresponding NCBI_IDs of genes (if Solyc-protein homologs are present in the NCBI database) are given in brackets.

reaction was carried out using the "2.5× Reaction Mixture for Real-Time PCR in the Presence of SYBR Green I and ROX" kit (Synthol LLC, Russia) on a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, USA). The RT-qPCR program was as follows: 5 min at 95 °C, 40 cycles (15 s at 95 °C; 40 s at 60 °C). *SlDREB2* gene expression was normalized to the reference genes *Expressed* and *actin-7* (Efremov et al., 2020). The analysis was performed in two biological and three technical replicates. The obtained data were statistically processed and visualized using GraphPad Prism v. 9.5.1 (Two-Way ANOVA: multiple comparisons corrected with the Bonferroni test; GraphPad Software Inc., USA; https://www.graphpad.com/scientific-software/prism/). Excel was used to construct heatmaps and linear graphs based on expression data.

Results

In this study, we characterized the effects of salt stress and prolonged post-stress recovery on *SlDREB2* gene expression in the leaves of tomato plants with high (cv. Otradnyi) and moderate (cv. Gnom) salt tolerance. After 24 h of NaCl exposure, as well as 7 and 14 days post-stress, plants of both varieties were visually indistinguishable from control, unstressed samples (Fig. 1*b*).

Leaves from plants (control and experimental) at time points S24/24K, R7/R7K, and R14/R14K were collected and used to analyze the expression of *SlDREB2* genes, the homologs of which in other plant species are known to be involved in the response to osmotic stress (Akbudak et al., 2018; Guo et al., 2022; Filyushin et al., 2023; Sun et al., 2025). Genes for analysis were identified based on the published characterization of the *S. lycopersicum DREB* gene family, in which the *SlDREB2* subfamily is represented by seven intronless genes: *SlDREB16*, *20*, *22*, *24*, *43*, *44*, and *46* (Maqsood et al., 2022). RT-qPCR analysis revealed that the *SlDREB20* gene was not expressed in leaves in either the experimental or control plants, while the expression pattern of the remaining six genes was genotype-dependent (Fig. 2).

The varieties differed in gene expression under control conditions, both in terms of the level at the 24K baseline (*SlDREB16*, *24*, *43*, and *44*) and in the tendency to change over the measurement period (*SlDREB16*, *44*, and *46*). Only *SlDREB22* showed a similar expression pattern between varieties under control conditions (Fig. 2). A heatmap was constructed based on the expression data (Fig. 3), clearly showing that in the case of the highly resistant cv. Otradnyi, only three genes (*SlDREB22*, *24*, and *44*) retained control expression levels after 24 h of stress. However, their transcript levels increased after one (~1.7, 8.2, and 2.4-fold) and two (~5.7, 3.4, and 1.4-fold) weeks of the recovery period. The expression of the remaining three genes decreased (*SlDREB16*, and *46*) or increased (*SlDREB43*) after 24 h of stress and increased significantly at point R7. After two weeks of recovery, only two genes (*SlDREB16*, and *43*) were expressed similarly to the control (Fig. 3).

In the leaves of the moderately resistant cv. Gnom, after 24 h of stress, gene expression increased (*SlDREB43*, and *44*), decreased (*SlDREB22*, and *46*), or remained unchanged (*SlDREB16*, and *24*). After a week of recovery (R7), changes in expression were observed for three genes (*SlDREB16*, *22*, and *43*), whereas after two weeks (R14), all six genes were expressed differently from the control (Fig. 3).

Thus, a return to control expression was observed only long after the stress and only for *SlDREB24*, *44*, and *46* (cv. Gnome, point R7), the expression of which at point R14 changed again (*vs.* control), as well as for the *SlDREB16*, and *43* (cv. Otradnyi, point R14) (Fig. 3).

To more clearly compare *SlDREB2* expression patterns between cultivars, linear graphs were drawn using the expression data, expressed as the ratio of gene expression levels between the experimental and control conditions (Fig. 4). The graphs show that two genes (*SlDREB22*, and *46*) have similar patterns of response to salt stress and memory phase in two analyzed cultivars. *SlDREB16*, *28*, *43*, and *44* genes showed different response patterns between varieties (Fig. 4).

To assess the possible dependence of the expression of the *SlDREB16*, *20*, *22*, *24*, *43*, *44*, and *46* genes on the variability of their regulatory regions in tomato varieties, an *in silico* analysis of the promoters (1 kb) of these genes was performed in 10 tomato accessions (sequences were taken from the NCBI database). It was shown that the promoters of *SlDREB16*, *22*, *24*, *43*, and *44* are highly conserved (0–2 polymorphisms (SNPs) per 1 kb), while the promoters of the *SlDREB20* and *46* genes contain indels/SNPs (5/58 and 5/13, respectively).

Discussion

In this study, we evaluated genes of the tomato *SlDREB2* subfamily as potential marker genes for salt stress memory by profiling gene expression in two *S. lycopersicum* cultivars in response to NaCl and during the long-term post-stress

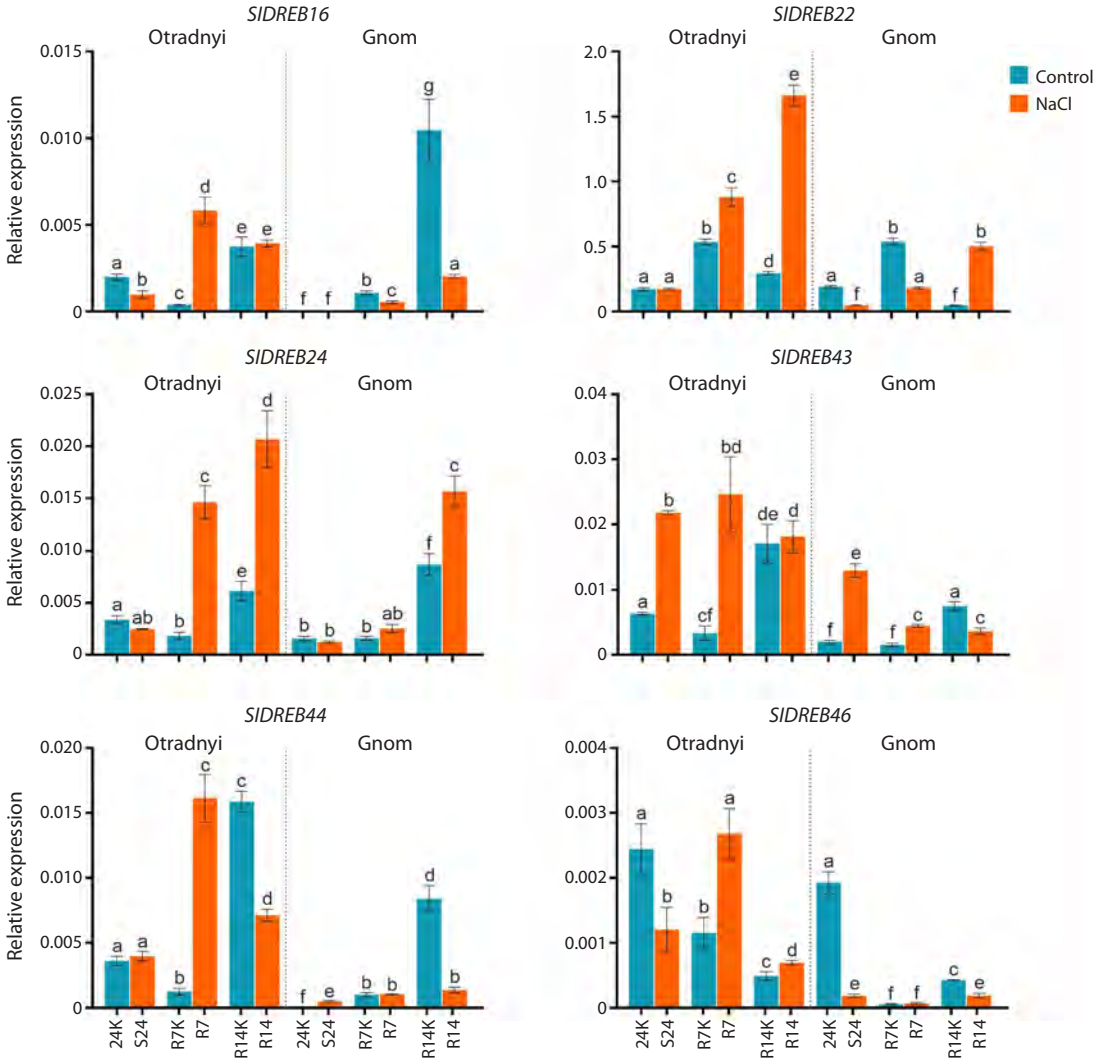


Fig. 2. Expression pattern of *SlDREB2* genes in the leaves of cv. Otradnyi and cv. Gnom tomato plants in response to salt stress for 24 h (24K and S24) and in the dynamics of post-stress recovery after 7 (R7K and R7) and 14 (R14K and R14) days.
[a–g] Significant differences between expression levels within the variety ($p < 0.05$).

recovery period (memory phase). The cultivars differed in their tolerance to salt stress (moderate in cv. Gnom and high in cv. Otradnyi). Cultivar tolerance can be regulated both by genetic variations governing gene expression in response to salt stress and by conditionally inherited epigenetic modifications, previously acquired as a result of salt priming and maintained by stress memory.

In the first case, genetic variations may be represented by genes and loci associated with the salt tolerance trait (Ismail, Horie, 2017). Differences in genes/loci may determine the degree of plant adaptability, as demonstrated by tomato genotypes carrying mutant *TSS1* and *TSS2* loci, which confer contrasting sensitivity to general osmotic stress and different mechanisms of salt tolerance (Borsani et al., 2001). Given the genetic regulation of NaCl tolerance, our experiment can be considered a primary stress for plants. In the second case, given the presumed presence of salt stress memory, the simulated salt stress conducted in this study will activate this memory. The third possible scenario involves genetic regulation of salt tolerance in one variety and epigenetic regulation in another.

0.49*	15.35*	1.07	1.00	0.51*	0.20*	*SlDREB16*
1.01	1.65*	5.65*	0.25*	0.34*	10.66*	*SlDREB22*
0.74	8.22*	3.41*	0.80	1.60	1.82*	*SlDREB24*
3.43*	7.88*	1.08	6.54*	2.90*	0.48*	*SlDREB43*
1.10	13.27*	0.45*	25.42*	1.07	0.16*	*SlDREB44*
0.50*	2.39*	1.43*	0.10*	1.22	0.44*	*SlDREB46*
24 h NaCl	7 days Recovery	14 days Recovery	24 days NaCl	7 days Recovery	14 days Recovery	
cv. Otradnyi			cv. Gnom			

Fig. 3. Heatmap of *SlDREB2* gene expression in the leaves of cv. Otradnyi and cv. Gnom tomato plants in response to salt stress (24 h) and during post-stress recovery (7 and 14 days). Numerical data are presented as the ratio of values for experimental samples to the control (taken as 1).
* Significant differences in expression levels between the experiment and the control ($p < 0.05$).

Various transcriptome studies of NaCl exposure in plants suggest that key genes involved in salt stress memory are represented by TF genes of various families (Zhu et al., 2023), including the DREB family (Hassan et al., 2022). The importance of the latter is highlighted by the differential

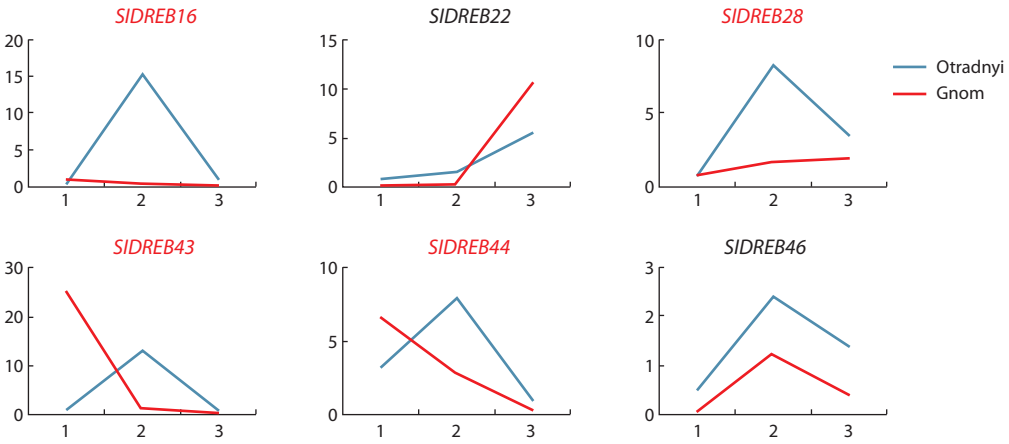


Fig. 4. *SlDREB2* gene expression patterns in the leaves of tomato cultivars Otradnyi and Gnom in response to salt stress (24 h) and during post-stress recovery (7 and 14 days), presented as linear graphs. Genes, the expression patterns of which show significant intervarietal differences in fluctuation trends, are highlighted in red.

expression of *DREB* genes in response to salinity in wheat *Triticum aestivum* L. (Hassan et al., 2022), pepper *Capsicum annuum* L. (Sun et al., 2025), garlic *A. sativum* (Filyushin et al., 2023), and other species.

The choice of DREB2 subfamily genes from the two largest DREB subfamilies for analysis was determined by the fact that DREB1/CBF proteins play the greatest role in regulating cold tolerance (Shi et al., 2018), whereas DREB2 TFs are mainly involved in the response to osmotic stresses (Akbudak et al., 2018; Baillo et al., 2019). In the tomato genome, the *DREB2* subfamily consists of seven genes: *SlDREB16*, *20*, *22*, *24*, *43*, *44*, and *46* (Maqsood et al., 2022) (Table 1).

During the experiment, two tomato varieties were subjected to salt stress (24 h), followed by a long-term (14-day) memory phase (Fig. 1). Subsequent gene expression profiling (S24/24K–R7/R7K–R14/R14K) revealed significant genotype-specific variations in gene transcript levels in both control and stressed plants (Fig. 2), suggesting intervarietal differences in the mechanism of salt tolerance regulation.

It was determined that during the long-term post-stress recovery period, gene expression values returned to control values either temporarily (*SlDREB24*, *44*, and *46* in the Gnom variety at point R7; they changed again at point R14) or extremely slowly (*SlDREB16* and *43* in the Otradnyi variety at point R14) (Fig. 3). This gene response in the case of both varieties corresponds to the feature of stress memory marker genes, the expression of which is maintained at an altered level for a long time during the recovery phase, while the expression of genes not associated with memory quickly returned to control values (Friedrich et al., 2019; Bäurle, Trindade, 2020; Jacques et al., 2021). This suggests that all six genes, *SlDREB16*, *22*, *24*, *43*, *44*, and *46*, may function as salt stress memory marker genes in tomato plants.

Only two genes (*SlDREB22* and *46*) were shown to have a similar pattern of expression fluctuations between cultivars during the measurement period (S24/24K–R7/R7K–R14/R14K) (Fig. 4). This suggests that the remaining four genes (*SlDREB16*, *28*, *43*, and *44*) may play a role in determining differences in the mechanism of regulation of plant responses to salt stress between salt-tolerant genotypes of *S. lycopersicum*.

Overall, the performed assessment of the expression pattern of *SlDREB2* subfamily genes in the leaves of two salt-tolerant tomato cultivars in response to NaCl and during the long-term memory phase suggests that these genes (except for *SlDREB20*) participate in the response of *S. lycopersicum* to excess salt in a genotype-specific manner. These genes may potentially serve as markers of stress memory linked to epigenetic regulation of plant adaptation to salt stress. The response of *SlDREB2* genes to salt stress may also depend on genetic variations in the promoter regions of both the *SlDREB2* subfamily genes themselves and the potential targets of the SlDREB2 TFs in *S. lycopersicum* accessions.

The invariability in the regulatory sequences of the *SlDREB16*, *22*, *24*, *43*, and *44* genes that we found (using *in silico* analysis of the promoters of the analyzed genes in 10 tomato cultivars/accessions) suggests that the conservation of these promoters may also extend to the cultivars used in this study. This suggests that the response of *SlDREB16*, *22*, *24*, *43*, and *44* to salt stress is independent on intervarietal variations in their regulatory sequences. The *SlDREB20* gene, the promoter of which is the most variable between accessions (58 SNPs), was not expressed in leaves; thus, the question of the dependence under consideration for this gene does not arise. At the same time, the expression level of *SlDREB46* can be regulated by polymorphisms (13 SNPs), which requires additional studies of the *SlDREB46* promoter in the tomato varieties used in the work, with a search for correlations between expression and the SNPs found.

The expression level of some *DREB2* subfamily genes is positively associated with plant resistance to various abiotic stresses, as demonstrated by *A. thaliana* plants overexpressing the rice (*Oryza sativa* L.) *OsDREB2B* gene and exhibiting increased tolerance to drought and heat (Matsukura et al., 2010). It is suggested that in response to abiotic stress, the expression of DREB1/2 TFs is altered, which in turn regulate the transcription of target genes involved in plant defense. To date, data are available on 10 possible target genes of the DREB1/2 TF (*A. thaliana*) containing DRE/DRE-related *cis*-regulatory elements in their promoters, and six of these genes may be involved in the plant's response to salt stress (Table 2) (Dubouzet et al., 2003; Matsukura et al., 2010).

Table 2. List of putative target genes of the *A. thaliana* DREB family TFs

Gene	NCBI_ID (*A. thaliana*/ *S. lycopersicum*)	Protein	Involvement in stress responses, according to NCBI
COR15A	At2g42540	Cold-inducible protein	Cold, heat, salinity, drought
FL05-21-F13 (*LEA14*)	At1g16850	Late embryogenesis abundant protein	Salinity
RD29A	At5g52310	Drought-responsive protein, low-temperature-responsive protein 78 (LTI78) / desiccation-responsive protein 29A	Cold, salinity, drought
RD17 (*COR47*)	At1g20440	Dehydrin, COR47 cold-regulated 47	Cold, heat, osmotic stresses, drought
AtGOLS3	At1g09350	Glycosyl transferase family 8, GolS3 galactinol synthase 3	Cold
FL05-20-N18 (*COR15B*)	At2g42530	COR15B cold regulated 15b	Cold
KIN1	At5g15960	Cold-inducible protein	Cold, osmotic stresses, drought
FL06-16-B22 (*COR413-PM1*)	At2g15970	Cold-acclimation protein, COR413-PM1 cold regulated 413 plasma membrane 1	Cold, drought
KIN2 (*COR6.6*)	At5g15970	KIN2 stress-responsive protein (KIN2) / stress-induced protein (KIN2) / cold-responsive protein (COR6.6) / cold-regulated protein (COR6.6)	Cold, osmotic stresses, drought
ERD10	At1g20450	ERD10 Dehydrin family protein	Cold, drought

Conclusion

Thus, further studies of the structure and expression of *SlDREB2* genes and their possible targets using repeated stress events interspersed with memory phases of varying duration, accompanied by expression analysis of genes presumably not linked to stress memory, are needed. The results of such studies can be used in breeding salt-tolerant tomato varieties.

References

Aina O., Bakare O.O., Fadaka A.O., Keyster M., Klein A. Plant biomarkers as early detection tools in stress management in food crops: A review. *Planta*. 2024;259(3):60. doi 10.1007/s00425-024-04333-1

Akbudak M.A., Filiz E., Kontbay K. *DREB2* (dehydration-responsive element-binding protein 2) type transcription factor in sorghum (*Sorghum bicolor*): genome-wide identification, characterization and expression profiles under cadmium and salt stresses. *3 Biotech*. 2018;8(10):426. doi 10.1007/s13205-018-1454-1

Anisimova O.K., Shchennikova A.V., Kochieva E.Z., Filyushin M.A. Identification of chalcone synthase genes from garlic (*Allium sativum* L.) and their expression levels in response to stress factors. *Acta Nat.* 2025;17(2):4-14. doi 10.32607/actanaturae.27639

Bai Y., Sunarti S., Kissoudis C., Visser R.G.F., van der Linden C.G. The role of tomato *WRKY* genes in plant responses to combined abiotic and biotic stresses. *Front Plant Sci.* 2018;9:801. doi 10.3389/fpls.2018.00801

Baillo E.H., Kimotho R.N., Zhang Z., Xu P. Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement. *Genes* (*Basel*). 2019;10(10):771. doi 10.3390/genes10100771

Bäurle I., Trindade I. Chromatin regulation of somatic abiotic stress memory. *J Exp Bot*. 2020;71(17):5269-5279. doi 10.1093/jxb/eraa098

Borsani O., Cuartero J., Fernández J.A., Valpuesta V., Botella M.A. Identification of two loci in tomato reveals distinct mechanisms for salt tolerance. *Plant Cell*. 2001;13(4):873-887. doi 10.1105/tpc.13.4.873

Calone R., Mircea D.M., González-Orenga S., Boscaiu M., Zuzunaga-Rosas J., Barbanti L., Vicente O. Effect of recurrent salt and drought stress treatments on the endangered halophyte *Limonium angustebracteatum* Erben. *Plants* (*Basel*). 2023;12(1):191. doi 10.3390/plants12010191

Chen Y., Li C., Yi J., Yang Y., Lei C., Gong M. Transcriptome response to drought, rehydration and re-dehydration in potato. *Int J Mol Sci*. 2019;21(1):159. doi 10.3390/ijms21010159

Ding Y., Liu N., Virlouvet L., Riethoven J.J., Fromm M., Avramova Z. Four distinct types of dehydration stress memory genes in *Arabidopsis thaliana*. *BMC Plant Biol.* 2013;13:229. doi 10.1186/1471-2229-13-229

Ding Y., Virlouvet L., Liu N., Riethoven J.J., Fromm M., Avramova Z. Dehydration stress memory genes of *Zea mays*; comparison with *Arabidopsis thaliana*. *BMC Plant Biol.* 2014;14:141. doi 10.1186/1471-2229-14-141

Dubouzet J.G., Sakuma Y., Ito Y., Kasuga M., Dubouzet E.G., Miura S., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 2003;33(4):751-763. doi 10.1046/j.1365-313x.2003.01661.x

Efremov G.I., Slugina M.A., Shchennikova A.V., Kochieva E.Z. Differential regulation of phytoene synthase PSY1 during fruit carotenogenesis in cultivated and wild tomato species (*Solanum* section Lycopersicon). *Plants* (*Basel*). 2020;9(9):1169 doi 10.3390/plants9091169

Filyushin M.A., Shagdarova B.T., Shchennikova A.V., Il'ina A.V., Kochieva E.Z., Varlamov V.P. Pretreatment with chitosan prevents *Fusarium* infection and induces the expression of chitinases and β-1,3-glucanases in garlic (*Allium sativum* L.). *Horticulturae*. 2022;8(5):383. doi 10.3390/horticulturae8050383

Filyushin M.A., Anisimova O.K., Shchennikova A.V., Kochieva E.Z. *DREB1* and *DREB2* genes in garlic (*Allium sativum* L.): Genome-wide identification, characterization, and stress response. *Plants* (*Basel*). 2023;12(13):2538. doi 10.3390/plants12132538

Friedrich T., Faivre L., Bäurle I., Schubert D. Chromatin-based mechanisms of temperature memory in plants. *Plant Cell Environ.* 2019;42(3):762-770. doi 10.1111/pce.13373

Gallusci P., Agius D.R., Moschou P.N., Dobránszki J., Kaiserli E., Martinelli F. Deep inside the epigenetic memories of stressed plants. *Trends Plant Sci.* 2023;28(2):142-153. doi 10.1016/j.tplants.2022.09.004

Guo M., Wang X.S., Guo H.D., Bai S.Y., Khan A., Wang X.M., Gao Y.M., Li J.S. Tomato salt tolerance mechanisms and their potential applications for fighting salinity: A review. *Front Plant Sci.* 2022;13:949541. doi 10.3389/fpls.2022.949541

Hassan S., Berk K., Aronsson H. Evolution and identification of DREB transcription factors in the wheat genome: modeling, docking and simulation of DREB proteins associated with salt stress. *J Biomol Struct Dyn.* 2022;40(16):7191-7204. doi 10.1080/07391102.2021.1894980

Ismail A.M., Horie T. Genomics, physiology, and molecular breeding approaches for improving salt tolerance. *Annu Rev Plant Biol.* 2017;68:405-434. doi 10.1146/annurev-arplant-042916-040936

Jacques C., Salon C., Barnard R.L., Vernoud V., Prudent M. Drought stress memory at the plant cycle level: A review. *Plants* (*Basel*). 2021;10(9):1873. doi 10.3390/plants10091873

Kudo M., Kidokoro S., Yoshida T., Mizoi J., Todaka D., Fernie A.R., Shinozaki K., Yamaguchi-Shinozaki K. Double overexpression of DREB and PIF transcription factors improves drought stress tolerance and cell elongation in transgenic plants. *Plant Biotechnol J.* 2017;15(4):458-471. doi 10.1111/pbi.12644

Maqsood H., Munir F., Amir R., Gul A. Genome-wide identification, comprehensive characterization of transcription factors, cis-regulatory elements, protein homology, and protein interaction network of *DREB* gene family in *Solanum lycopersicum*. *Front Plant Sci*. 2022;13:1031679. doi 10.3389/fpls.2022.1031679

Matsukura S., Mizoi J., Yoshida T., Todaka D., Ito Y., Maruyama K., Shinozaki K., Yamaguchi-Shinozaki K. Comprehensive analysis of rice *DREB2*-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Mol Genet Genomics.* 2010;283(2):185-196. doi 10.1007/s00438-009-0506-y

Shi Y., Ding Y., Yang S. Molecular regulation of CBF signaling in cold acclimation. *Trends Plant Sci.* 2018;23(7):623-637. doi 10.1016/j.tplants.2018.04.002

Sun N., Sun X., Zhou J., Zhou X., Gao Z., Zhu X., Xu X., Liu Y., Li D., Zhan R., Wang L., Zhang H. Genome-wide characterization of pepper DREB family members and biological function of *CaDREB32* in response to salt and osmotic stresses. *Plant Physiol Biochem*. 2025;222:109736. doi 10.1016/j.plaphy.2025.109736

Villagómez-Aranda A.L., Feregrino-Pérez A.A., García-Ortega L.F., González-Chavira M.M., Torres-Pacheco I., Guevara-González R.G. Activating stress memory: eustressors as potential tools for plant breeding. *Plant Cell Rep*. 2022;41(7):1481-1498. doi 10.1007/s00299-022-02858-x

Virlouvet L., Avenson T.J., Du Q., Zhang C., Liu N., Fromm M., Avramova Z., Russo S.E. Dehydration stress memory: gene networks linked to physiological responses during repeated stresses of *Zea mays*. *Front Plant Sci.* 2018;9:1058. doi 10.3389/fpls.2018.01058

Zhu Z., Dai Y., Yu G., Zhang X., Chen Q., Kou X., Mehareb E.M., Raza G., Zhang B., Wang B., Wang K., Han J. Dynamic physiological and transcriptomic changes reveal memory effects of salt stress in maize. *BMC Genomics*. 2023;24(1):726. doi 10.1186/s12864-023-09845-w

Zuo D.D., Ahammed G.J., Guo D.L. Plant transcriptional memory and associated mechanism of abiotic stress tolerance. *Plant Physiol Biochem*. 2023;201:107917. doi 10.1016/j.plaphy.2023.107917

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

Received July 8, 2025. Revised September 19, 2025. Accepted September 19, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1221-1234

doi 10.18699/vjgb-25-129

SmartCrop: knowledge base of molecular genetic mechanisms of rice and wheat adaptation to stress factors

P.S. Demenkov [1], T.V. Ivanisenko [1], M.A. Kleshchev [1], E.A. Antropova [1], I.V. Yatsyk [1], A.R. Volyanskaya [1], A.V. Adamovskaya [1], A.V. Maltseva [1], A.S. Venzel [1], H. Chao [2], M. Chen, V.A. Ivanisenko [1] ✉

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China
✉ salix@bionet.nsc.ru

Abstract. The study of molecular genetic mechanisms of plant responses to specific growth conditions and stress factors is a central focus of scientific research aimed at developing new valuable crop varieties, particularly rice and wheat. These factors include abiotic stresses (high or low temperatures, drought, salinity, soil metal contamination), biotic stresses (pathogens, pests), as well as plant responses to regulatory factors (fertilizers, hormones, elicitors, and other compounds). Modern research in plant genetics is based on the understanding that the formation of any phenotypic characteristics (molecular genetic, biochemical, physiological, morphological, etc.) is controlled by gene networks – groups of coordinately functioning genes interacting through their products (RNA, proteins, and metabolites). Previously, we developed the ANDSystem intelligent technology designed to extract knowledge from scientific publication texts for the reconstruction of gene networks in biology and biomedicine. In this work, using an adapted version of ANDSystem for plants, we created the SmartCrop knowledge base designed to address challenges related to studying molecular genetic mechanisms of genotype-phenotype-environment interactions for agriculturally valuable rice and wheat crops. SmartCrop is designed to assist researchers in solving tasks such as interpreting omics technology results (establishing connections between gene sets and biological processes, phenotypic traits, etc.); reconstructing gene networks describing relationships between molecular genetic objects and concepts in breeding, phenomics, seed production, phytopathology, diagnostics, protective agents, etc.; identifying regulatory and signaling pathways of plant responses to specific growth conditions and biotic and abiotic stresses; predicting candidate genes for genotyping; searching for markers for marker-assisted selection; and identifying potential targets for substances (including external factors) affecting plants to ensure timely and uniform germination, better vegetative growth, efficient nutrient uptake, and improved stress resistance.

Key words: SmartCrop knowledge base; ANDSystem; text mining; artificial intelligent; molecular genetic mechanisms; rice; wheat; associative gene networks; abiotic stress; biotic stress; genotype–phenotype–environment interactions; omics technologies; long non-coding RNAs; marker-assisted selection; plant adaptation; stress resistance

For citation: Demenkov P.S., Ivanisenko T.V., Kleshchev M.A., Antropova E.A., Yatsyk I.V., Volyanskaya A.R., Adamovskaya A.V., Maltseva A.V., Venzel A.S., Chao H., Chen M., Ivanisenko V.A. SmartCrop: knowledge base of molecular genetic mechanisms of rice and wheat adaptation to stress factors. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1221-1234. doi 10.18699/vjgb-25-129

Funding. The work of PSD, TVI, MAK, EAA, IVY, ARV, AVA, AVM, ASV, and VAI was supported by the Russian–Chinese grant of the Russian Science Foundation No. 23-44-00030. The work of MCh and HCh was supported by the National Natural Science Foundation of China, grant No. 32261133526.

SmartCrop: база знаний молекулярно-генетических механизмов адаптации риса и пшеницы к стрессовым факторам

П.С. Деменков [1], Т.В. Иванисенко [1], М.А. Клещев [1], Е.А. Антропова [1], И.В. Яцык [1], А.Р. Волянская [1], А.В. Адамовская [1], А.В. Мальцева [1], А.С. Вензель [1], Х. Чао [2], М. Чен [2], В.А. Иванисенко [1] ✉

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Факультет биоинформатики, Колледж естественных наук, Чжэцзянский университет, Ханчжоу, Китай
✉ salix@bionet.nsc.ru

Аннотация. Изучение молекулярно-генетических механизмов реакций растений на специфические условия роста и стрессовые факторы – одно из приоритетных направлений исследований, нацеленных на создание новых сортов сельскохозяйственных культур, в частности риса и пшеницы. К числу таких факторов относятся

абиотические стрессы (высокие или низкие температуры, засуха, засоление, загрязнение почвы металлами), биотические стрессы (патогены, вредители), а также реакции растений на регуляторные факторы (удобрения, гормоны, элиситоры и другие соединения). Современные исследования в области генетики растений основаны на понимании того, что формирование любых фенотипических характеристик (молекулярно-генетических, биохимических, физиологических, морфологических и др.) контролируется генными сетями – группами согласованно функционирующих генов, взаимодействующих через свои продукты (РНК, белки и метаболиты). Ранее с целью реконструкции генных сетей, значимых для биологии и биомедицины, нами была разработана интеллектуальная компьютерная система ANDSystem, предназначенная для автоматизированного извлечения знаний из текстов научных публикаций и баз данных. В настоящей работе, используя адаптированную версию ANDSystem для растений, мы создали базу знаний SmartCrop для решения задач, связанных с изучением молекулярно-генетических механизмов взаимодействий «генотип–фенотип–среда» для сельскохозяйственно ценных культур риса и пшеницы. SmartCrop предназначена для помощи исследователям в решении таких задач, как интерпретация результатов омиксных экспериментов на растениях: установление связей между наборами генов и биологическими процессами, фенотипическими признаками и др.; реконструкция генных сетей, описывающих отношения между молекулярно-генетическими объектами и понятиями в селекции, феномике, семеноводстве, фитопатологии; выявление регуляторных и сигнальных путей, ответных реакций растений на специфические условия роста и биотические и абиотические стрессы; прогнозирование генов-кандидатов для генотипирования; поиск маркеров для маркер-опосредованной селекции; выявление потенциальных мишеней (генов и белков) для субстанций, влияющих на растения (контролирующих процессы прорастания семян, вегетативного роста, эффективного поглощения питательных веществ и улучшения устойчивости к стрессовым факторам).

Ключевые слова: база знаний SmartCrop; ANDSystem; извлечение знаний из текстов; искусственный интеллект; молекулярно-генетические механизмы; рис; пшеница; ассоциативные генные сети; абиотические стрессы; биотические стрессы; взаимодействия генотип–фенотип–среда; омиксные технологии; длинные некодирующие РНК; маркер-опосредованная селекция; адаптация растений; стрессоустойчивость

Introduction

Rice (*Oryza sativa* L.) and wheat (*Triticum aestivum* L.) are among the most important agricultural crops, ensuring food security for a significant portion of the world's population. Both crops are well known for their high nutritional, industrial, and fodder value (Shewry, Hey, 2015). Under current conditions, the production of these crops faces serious challenges. Extreme weather events, adverse climate change, plant diseases, and pests lead to substantial yield losses (Lesk et al., 2016). Overcoming these difficulties is impossible without studying the molecular genetic mechanisms underlying plant resistance to unfavorable biotic and abiotic factors, which requires the analysis of complex systems that include intricate signaling, regulatory, transport, and metabolic pathways (Mittler, 2006; Nykiel et al., 2023).

An effective tool for studying such mechanisms is gene networks, which control molecular genetic processes that determine the formation of phenotypic traits and the functioning of biological processes, including plant stress responses.

The modern concept of gene networks encompasses not only molecular components (RNAs, genes, proteins, and metabolites) but also a wide range of heterogeneous entities, including diseases, biological processes, and environmental factors. This type of gene network is known as an associative gene network (Ivanisenko V.A. et al., 2015). Structurally, such networks represent a knowledge graph integrating information about interactions among diverse objects involved in the functioning of molecular genetic systems or influencing them. In agrobiology and crop science, gene network analysis is successfully used to study economically important traits such as resistance to diseases and pests, tolerance to abiotic stress factors, and yield (Virlouvet et al., 2018; Chen et al., 2020).

The reconstruction of plant gene networks is a complex task that requires processing massive amounts of data and integrating fragmented information from scientific publications, including data on regulatory, transport, and catalytic processes, as well as relationships between genetic features, phenotypic manifestations, and environmental factors. To extract such knowledge, text-mining methods are applied, based both on classical computational approaches (dictionary-based methods, syntactic and linguistic rules and patterns, statistically significant co-occurrence, etc.) and on machine-learning techniques (Ivanisenko T.V. et al., 2014; Shrestha et al., 2024; Zhang et al., 2024).

Machine-learning algorithms used for constructing and analyzing gene networks can be divided into the following categories: supervised learning, unsupervised learning, semi-supervised learning, and hybrid approaches. Supervised learning methods rely on pre-annotated data to build predictive models, for example, to identify key regulators or to predict functional interactions between plant genes (Ni et al., 2016). Unsupervised learning enables the discovery of hidden patterns in large datasets, which is important, for instance, when clustering genes based on expression similarity or identifying gene-network modules. Semi-supervised learning combines the strengths of both approaches, using both labeled and unlabeled data, which is particularly relevant when the amount of well-annotated data is limited (Yan, Wang, 2022). Hybrid approaches integrate various machine-learning methods as well as traditional bioinformat-

ics techniques, allowing them to effectively compensate for the limitations each approach may have when used alone (Guindani et al., 2024; Ivanisenko T.V. et al., 2024). For example, combining dictionary-based named-entity recognition in texts with machine-learning methods improves the accuracy of entity identification (Do et al., 2018; Ivanisenko T.V. et al., 2020).

In recent years, deep machine learning has achieved significant advances through the introduction of the transformer architecture and attention mechanisms, which have enabled substantial progress in natural language processing and the analysis of biological sequences (Vaswani et al., 2017). The analysis of gene networks has also seen considerable development with the application of graph neural networks, among which the GraphSAGE architecture (Hamilton et al., 2017) enables efficient training on large heterogeneous graphs by aggregating features from neighboring nodes. A promising direction is the use of large language models, such as Gemma-2-9b-it (Gemma Team, Google DeepMind, 2024), which provide high-quality semantic analysis of scientific texts and validation of extracted interactions.

A number of specialized resources have been developed for the reconstruction and analysis of plant gene networks. These include PlantRegMap (Tian et al., 2020), designed for analyzing transcription factor regulatory interactions; STRING (Szklarczyk et al., 2021), which enables the exploration of protein–protein interactions; the KEGG PLANT platform (Kanehisa, 2013), which integrates information on metabolic pathways across various plant species; and the Plant Reactome resource (Naithani et al., 2020), containing detailed data on signaling and metabolic pathways in model plant organisms. For visualization and analysis of gene networks, the Cytoscape software environment (Otasek et al., 2019) is widely used, offering an extensive set of plugins for working with biological data. The ncPlantDB database provides comprehensive information for analyzing regulatory networks, including data on cell-type specific expression of noncoding RNAs and their interactions (Cheng et al., 2024; Liu et al., 2025). The integration of such omics resources forms an effective platform for reconstructing gene networks of agricultural crops (Chao et al., 2023).

Earlier, we developed the ANDSystem cognitive software information platform (Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V., 2020, 2022) designed for the full knowledge-engineering cycle in the biomedical domain. The system's knowledge base contains more than 50 million interactions for various organisms.

In the field of plant biology, ANDSystem has been used to create a knowledge base on the genetics of *Solanum tuberosum* (Saik et al., 2017; Ivanisenko T.V. et al., 2018; Demenkov et al., 2019), to reconstruct and analyze the regulatory gene network controlling cell wall functions in *Arabidopsis thaliana* leaves under water deficit (Volyanskaya et al., 2023), and to develop a method for prioritizing biological processes based on the reconstruction and analysis of associative gene networks (Demenkov et al., 2021).

The application of the ANDSystem automated reconstruction of associative gene networks to analyze microRNA-mediated regulation of bread wheat (*Triticum aestivum* L.) adaptation to water deficit made it possible to propose new candidate microRNAs (MIR7757, MIR9653a, MIR9670, MIR9672b) of interest for further experimental studies of plant adaptation mechanisms under insufficient moisture (Kleshchev et al., 2024).

In another study (Antropova et al., 2024), ANDSystem was used to reconstruct the molecular genetic network of rice (*Oryza sativa*) responses to *Rhizoctonia solani* infection under nitrogen excess, which revealed three potential mechanisms explaining reduced plant resistance to the pathogen. Key regulatory pathways were identified: an OsGSK2-mediated cascade, the OsMYB44–OsWRKY6–OsPR1 signaling pathway, and a pathway involving SOG1, Rad51, and the PR1/PR2 genes. In addition, markers promising for breeding were identified: 7 genes regulating a broad range of stress responses and 11 genes that modulate the immune system. Additional analysis of noncoding RNAs (Antropova et al., 2024) identified 30 microRNAs targeting genes within the reconstructed gene network. For two of them (Osa-miR396 and Osa-miR7695), approximately 7,400 unique long noncoding RNAs with differing co-expression indices were found, which may indicate a complex architecture of post-transcriptional regulation under nitrogen stress.

The aim of the present work was to adapt ANDSystem to create the SmartCrop knowledge base, integrating data on molecular genetic mechanisms and associative gene networks of stress responses in rice and wheat based on intelligent analysis of scientific publications and curated factual databases. This work included the development of a domain ontology and the optimization of intelligent knowledge-extraction methods from scientific texts using semantic–linguistic patterns and pretrained large language models. The SmartCrop ontology is represented by a set of interconnected dictionaries describing: molecular genetic entities (genes, proteins, metabolites, microRNAs), biological processes, phenotypic traits and diseases, pathogens, genetic biomarkers, markers of resistance to crop protection products, molecular targets of chemical crop protection agents, biotic and abiotic factors, crop protection products, as well as cultivars with their economically valuable and consumer traits.

As a result of automated analysis of scientific publications, the SmartCrop knowledge base was formed, integrating more than 10 million interactions among the entities defined in the ontology.

Materials and methods

Information resources used in the development of SmartCrop. To create the SmartCrop knowledge base, we used the ANDSystem software information platform (Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V., 2020, 2022) and its information and bioinformatics technologies.

Customization of ANDSystem methods for the subject domain was carried out using an ontology that included specialized dictionaries of entities and a description of the types of their interactions. The main sources of genetic and genomic information for constructing the dictionaries were: the NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene), the rice-specific database Oryzabase (https://shigen.nig.ac.jp/rice/oryzabase), the microRNA database miRBase (https://www.mirbase.org), the long noncoding RNA co-expression database ncPlantDB (https://bis.zju.edu.cn/ncPlantDB/), the single nucleotide polymorphism database dbSNP (https://www.ncbi.nlm.nih.gov/snp), and the database on cereal crops GrainGenes (https://wheat.pw.usda.gov/GG3).

To standardize terminology, we used the following ontologies: Gene Ontology (http://geneontology.org), Crop Ontology for wheat and rice (https://cropontology.org), as well as the genetic resources collection of VIR (https://www.vir.nw.ru).

Chemical compounds and metabolites were annotated using the ChEBI database (https://www.ebi.ac.uk/chebi). Information on herbicide resistance was obtained from the International Herbicide-Resistant Weed Database (http://www.weedscience.org), and data on pesticides were taken from the EU Pesticide Database (https://food.ec.europa.eu/plants/pesticides/eu-pesticides-database_en).

For knowledge extraction from texts, we used ANDSystem's semantic-linguistic templates, as well as newly developed templates tailored to the specifics of the subject domain. In addition, artificial intelligence methods were applied for knowledge extraction, including GraphSAGE graph neural networks (Hamilton et al., 2017) and the large language model Gemma-2-9b-it (Gemma Team, Google DeepMind, 2024).

Evaluation of accuracy. To assess the quality of named-entity annotation in the text, the F1-score was used, which is the harmonic mean of precision (Precision) and recall (Recall):

$$F1 = 2\cdot(\text{Precision}\times\text{Recall})/(\text{Precision}+\text{Recall}),$$

$$\text{Precision} = TP/(TP+FP),$$

$$\text{Recall} = TP/(TP+FN),$$

where TP – are true positives, FP – are false positives, and FN – are false negatives.

Results

A schematic representation of the main components of the SmartCrop software-information system is shown in Figure 1.

SmartCrop domain-specific ontology module

The development of a domain-specific ontology was a key stage in building SmartCrop. The domain-oriented ontology defines a conceptual model of the problem area and includes dictionaries of entities and types of their interactions. Based on these dictionaries, information about interactions between specific entities is extracted from texts and factual databases. The current version of the SmartCrop ontology contains 15 dictionaries of different entity types (Table 1), compiled by extracting entity names from specialized databases and existing ontologies.

Interaction types. In the SmartCrop system, 16 types of relationships between ontology entities are defined. All interactions in the system are directional and can be divided into several main groups. Physical interactions include processes

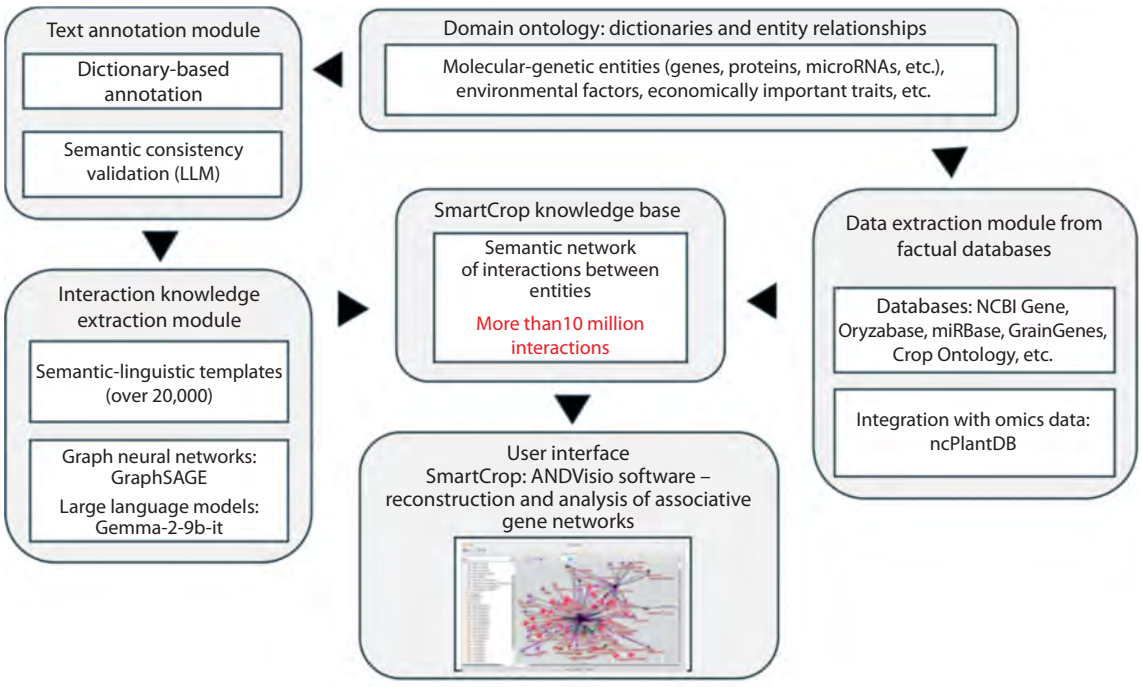


Fig. 1. Schematic representation of the architecture of the SmartCrop software information system.

Table 1. Dictionaries represented in the SmartCrop ontology

Dictionary	Number of objects
Genes/proteins of rice (*O. sativa*)	45,198
Genes/proteins of wheat (*T. aestivum*)	155,761
microRNAs of rice (*O. sativa*)	604
microRNAs of wheat (*T. aestivum*)	122
Genetic biomarkers (*T. aestivum*)	862
QTL polymorphisms (*O. sativa*)	1,987
QTL polymorphisms (*T. aestivum*)	1,266
Rice cultivars (*O. sativa*)	14,377
Wheat cultivars (*T. aestivum*)	25,501
Metabolites	74,838
Biological processes	122,477
Economically important traits	234
Phenotypic traits	2,386
Diseases, pathogens, and pests	1,065
Markers of resistance to crop protection products	861
Biotic factors	710
Abiotic factors	496
Crop protection products and herbicides	1,336
Molecular targets of chemical crop protection agents	14
Long noncoding RNAs	6,546

of forming both short-lived molecular complexes and stable associations between proteins and metabolites.

Chemical interactions comprise catalytic reactions of the substrate–enzyme–product type, protein proteolysis, as well as various post-translational protein modifications such as phosphorylation and glycosylation.

A distinct group is formed by regulatory interactions, which encompass the regulation of gene expression by transcription factors, modulation of protein activity and function, control of protein and metabolite transport, as well as regulation of protein stability and degradation. An important feature is that regulatory interactions also define relationships between molecular genetic entities, biological processes, and phenotypic traits. Each regulatory event may be characterized by an enhancing or attenuating effect on the corresponding process.

Expression and co-expression of genes are distinguished separately. The products of gene expression are proteins and noncoding RNAs. Co-expression is the simultaneous expression of genes driven by shared regulatory mechanisms under changing cellular conditions. Additionally, the system accounts for associative links, which include unclassified interactions between various ontology entities.

Table 2. Accuracy assessment of entity name annotation for the new dictionaries

Dictionary	F1-score
Cultivars (*O. sativa*)	0.103 (0.88)*
Cultivars (*T. aestivum*)	0.274 (0.919)*
Economically important traits	0.789
Abiotic factors	0.748
Biotic factors	0.927
Diseases, pathogens, and pests	0.881

* Values in parentheses indicate the accuracy after filtering the types of recognized names using a neural network.

Text-annotation module based on ontology entities

Recognition of molecular genetic entities in scientific texts is a challenging task due to the specific nature of their nomenclature. Our experience with ANDSystem shows that a substantial portion of errors in automatic reconstruction of associative gene networks is associated with inaccurate identification of named entities (Ivanisenko T.V. et al., 2022). The causes of such errors include the use of abbreviations by authors, semantic ambiguity of terms, and various linguistic features of scientific texts. In publications, standard names of entities are often modified, punctuation and word order are altered, grammatical forms vary, abbreviations are used, or technical typos are introduced (Pearson, 2001; Krallinger et al., 2015; Islamaj et al., 2021).

To improve recognition accuracy, we developed a two-stage process: 1) initial matching of names to the ontology dictionary and 2) subsequent verification of whether each annotated entity name corresponds to its type, based on contextual document analysis using neural networks.

The verification process is implemented as follows: a language model converts the context (about 400 words) containing the analyzed entity, which is replaced with a special mask tag, into a vector representation. Based on this representation, a neural network performs binary classification, determining whether the contextual environment of the term is consistent with its typical usage.

For entities from the ANDSystem ontology (genes, proteins, metabolites, etc.), classification accuracy was reported in a previously published paper (Ivanisenko T.V. et al., 2022). For the new SmartCrop dictionaries, manual expert evaluation of annotation quality was carried out (Table 2) based on the analysis of 1,000 randomly selected documents from the PubMed and PubMed Central databases.

The evaluation results demonstrated high annotation accuracy for most dictionaries, with the exception of rice and wheat cultivar names. The identification of plant cultivar names is a well-known complex task, determined by several factors, including substantial overlap of terms with common

vocabulary and anthroponyms, as well as the lack of a unified standard in the nomenclature of new cultivars (Do et al., 2018; D'Souza, 2024).

To address this problem, a specialized language model was trained, focused on the task of contextual term classification. The training was carried out in accordance with the methodology previously described in our work on improving the accuracy of identifying eight types of molecular genetic entities, including proteins, genes, metabolites, and cellular components (Ivanisenko T.V. et al., 2022). Integration of the developed model made it possible to substantially increase the recognition accuracy (F1-score) of cultivar names to 0.88 for rice and 0.919 for wheat.

Knowledge extraction module

The knowledge extraction module for scientific texts implements three main stages: 1) primary knowledge extraction using semantic-linguistic templates; 2) reconstruction of the initial semantic network; 3) its extension using graph neural networks and large language models. Additionally, to further expand the semantic network, a data extraction module for factual databases containing structured information is used, which makes it possible to obtain additional information about interactions between entities.

Semantic-linguistic templates are structured records containing metadata about the types of entities and the nature of their interactions. They include two main components: 1) syntactic relations that describe the order of entities and keywords in a sentence using regular expressions, and 2) semantic relations that define the type of interaction between entities. Regular expressions are used to search for patterns in the arrangement of entity names in annotated text sentences. When a match is found, specific entity names from the text are mapped to the template identifiers.

For each interaction type, specialized groups of templates with unique syntactic rules were developed. The knowledge base contains more than 18,000 ANDSystem templates for interaction types represented in both the ANDSystem and SmartCrop ontologies, as well as more than 3,000 templates specifically designed for the rice and wheat ontologies. The effectiveness of the template-based interaction extraction method was demonstrated during the development of ANDSystem (Ivanisenko V.A. et al., 2015).

Application of graph neural networks and large language models. At the second stage, based on the knowledge extracted using templates, a primary knowledge graph (semantic network) was constructed and used to train a graph neural network. After training, the network was used to predict missing edges in the knowledge graph. At the third stage, large language models were applied to validate these predictions by analyzing scientific texts in which the annotated entities with the predicted interactions co-occur (Ivanisenko T.V. et al., 2024).

Integration with omics data

Noncoding RNAs (ncRNAs) represent a broad and functionally diverse class of RNA molecules that are not translated into proteins but perform key regulatory functions in the cell. Long noncoding RNAs (lncRNAs) are of particular interest, as they participate in the regulation of gene expression at multiple levels – from modulating mRNA stability and translation to being involved in complex signaling cascades (Statello et al., 2021; Supriya et al., 2024).

A well-known specialized resource on ncRNA co-expression in plants, including rice lncRNAs, is the ncPlantDB database (https://bis.zju.edu.cn/ncPlantDB/). It provides information on tissue-specific ncRNA expression at the single-cell level and their putative interactions, obtained using modern single-cell transcriptomics methods (Cheng et al., 2024; Liu et al., 2025). Integration of SmartCrop with ncPlantDB made it possible to use ncRNA co-expression data, including their relationships with microRNAs, to enrich the reconstructed gene networks.

Module for gene network analysis and visualization

As the graphical user interface of SmartCrop, intended for the reconstruction and analysis of gene networks based on information from the SmartCrop knowledge base, the ANDVisio software is used (Fig. 2).

The ANDVisio program (Demenkov et al., 2012) was originally developed as a component of the ANDSystem platform and was later adapted for integration with SmartCrop. It provides researchers with a wide range of tools for structural and functional analysis of gene networks, including:

- multiple graph layout algorithms;
- a multi-parameter filtering system;
- mechanisms for pathways and cycles finding;
- tools for calculating node centrality measures;
- tools for assessing the enrichment of biological processes with network genes;
- additional methods of network analysis.

SmartCrop knowledge base

The system's knowledge base is implemented as a semantic network (knowledge graph) that integrates data extracted both from scientific publications and from factual databases. In this graph structure, nodes correspond to entities of the domain ontology, and edges represent various types of interactions between them.

The knowledge base was populated through systematic analysis of the scientific literature, including abstracts from PubMed and full-text articles from the open-access resource PubMed Central. The time span of the analyzed publications covered the period from 1970 to 2024, with the main selection criterion being the presence of references to wheat or rice. Detailed statistics on the number of recorded interactions in the SmartCrop knowledge base are presented in Table 3.

Discussion

To demonstrate the capabilities of SmartCrop, we consider two use cases: analysis of experimental omics data and experiment planning.

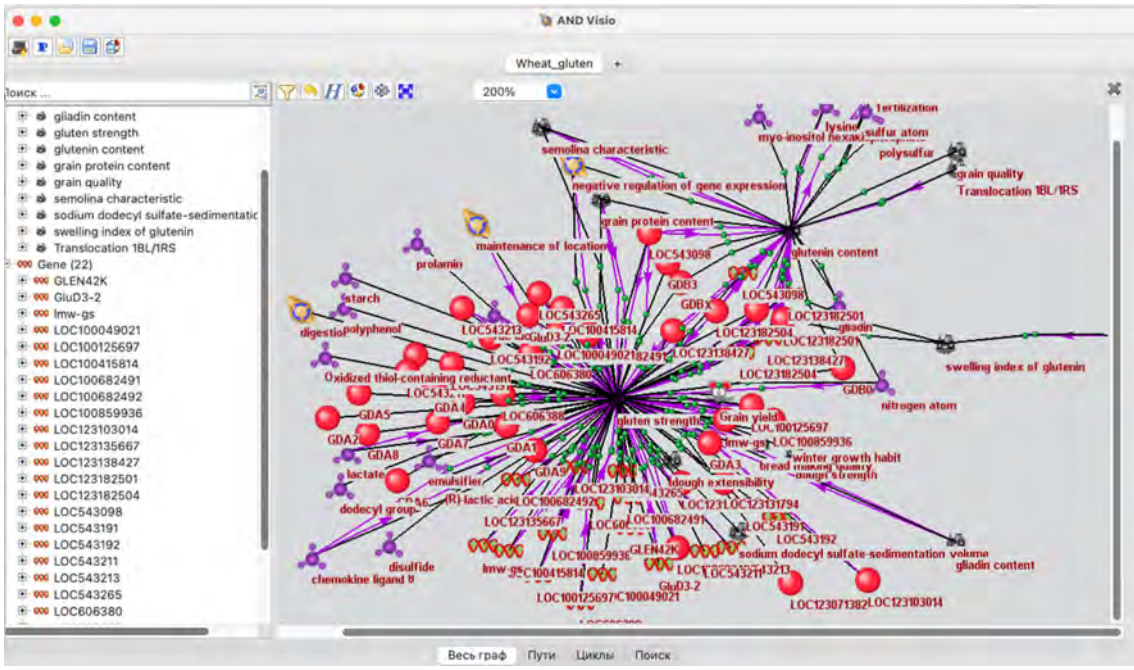


Fig. 2. Screenshot of the ANDVisio program interface.

Table 3. Statistics of the SmartCrop knowledge base on interactions between entities for wheat and rice

Interaction type	Number of interactions	
	Rice (*Oryza sativa*)	Wheat (*Triticum aestivum*)
Association	189,650	207,415
Regulation of expression	22,472	26,474
Regulation of activity	3,991	23,411
Regulation of degradation	1,442	4,415
Regulation of transport	830	1,320
ncRNA/miRNA regulation	2,125,036	5,814
Regulation of processes	23,274	23,766
Catalytic reactions	2,267	5,798
Expression	86,353	311,964
Physical interaction	8,551	11,810
Marker	435	226

Analysis of experimental omics data

As an example of omics data interpretation, we performed functional annotation of differentially expressed genes (DEGs) in bread wheat under salt stress. For the analysis, we used a set of 5,829 DEGs obtained from the NCBI GEO database (GSE225565, Alyahya, Taybi, 2023) for root tissues of bread wheat (*Triticum aestivum* L., cultivar Saudi) in response to salinization.

The results of the overrepresentation analysis of SmartCrop entities (biological processes, phenotypic traits, agronomically important traits, pathogens) for this DEG set and their protein products are presented in Supplementary Table S1[1]. In total, significant overrepresentation (p-value < 0.05, Bonferroni-corrected) was found for 217 terms describing biological processes (entity type Pathway), 50 phenotypic traits (Phenotype), 9 agronomically important traits (Agrophenotype), and 38 pathogenic species. The list of entities belonging to the five groups of the most statistically significant characteristics is given in Table 4.

Analysis of the overrepresented biological processes showed that the DEG set under study is associated not only with the response to salt stress, but also with the response to

[1] Supplementary Table S1 is available at: https://vavilovj-icg.ru/download/pict-2025-29/appx46.xls

Table 4. Bread-wheat characteristics significantly associated with the DEG set and their protein products under salt stress, identified using the SmartCrop system

Object (term)	Number of associated genes and proteins	*p*-value	Object (term)	Number of associated genes and proteins	*p*-value
Biological processes			Agronomically important traits		
Response to water deprivation	37	2.73E–66	Seed longevity	2	3.96E–08
Cell death	31	1.12E–54	Grain thickness	2	1.77E–07
Cold acclimation	23	4.52E–41	Grain length	7	1.62E–20
Hyperosmotic salinity response	23	2.67E–38	Grain protein content	7	1.42E–17
Seed germination	20	7.21E–34	Fiber quality	2	1.23E–06
Phenotypic traits			Pathogens		
Stomatal closure rate	2	8.84E–47	*Fusarium* sp.	22	2.22E–59
Cell membrane stability	2	1.45E–31	*Fusarium culmorum*	10	1.23E–27
Grain moisture content	2	5.46E–31	*Fusarium oxysporum*	10	2.39E–26
Seed length	2	2.32E–30	*Botrytis cinerea*	10	8.63E–26
Na^+ uptake	2	1.98E–29	*Fusarium pseudograminearum*	10	2.45E–25

water deficit. This reflects plant adaptation mechanisms to the state of so-called "physiological drought", which arises when effective water uptake becomes impossible due to high osmotic pressure of the surrounding environment. Among such adaptations is stomatal closure, mediated by a rapid increase in abscisic acid levels (Verma et al., 2016; Zhao et al., 2021). Accordingly, the significantly overrepresented entities included both the phenotypic trait "stomatal closure rate" (Table 4) and signaling pathways associated with abscisic acid (Table S1), which confirms their important role in the response to salinity.

It should be noted that an important advantage of the SmartCrop knowledge base, compared with widely used resources for gene functional annotation (DAVID, Gene Ontology, ShinyGO, etc.), is the ability to analyze relationships between genes/proteins and not only biological processes, molecular functions, cellular components, and KEGG pathways, but also a broad spectrum of abiotic and biotic environmental factors, phenotypic traits, agronomically important properties, and pathogens. This integration makes it possible to assess overrepresentation for different types of entities in the gene set under study, substantially expanding the capabilities of functional annotation and enabling the identification of genes with pleiotropic effects. The latter is particularly important for marker-assisted selection, since selection based on a single target phenotypic trait or genetic marker may simultaneously affect several other, non-target traits.

In particular, the results of functional annotation of DEGs in bread wheat under salt stress showed their association not only with responses to salinity and water deficit, but also with seed germination and with agronomically important traits reflecting grain quality (Table 4). For example, aquaporins (encoded by genes *LOC543267*, *LOC100037645*, *LOC123093445*, and others) provide selective transport of water molecules, participate in maintaining cellular ion balance and in regulating water–salt homeostasis under elevated salinity (Ayadi et al., 2019), and also facilitate the movement of water and solutes within seeds, which plays a key role in the germination process (Hoai et al., 2020).

The functionality of SmartCrop is not limited to overrepresentation analysis. The system also makes it possible to reconstruct associative networks of proteins and genes significantly associated with overrepresented entities and to search for their regulators. This provides a deeper understanding of the molecular mechanisms underlying these relationships and helps to reveal their specificity under experimental conditions.

As an example, a gene network regulating plant tolerance to hyperosmotic stress (GO:0042538 hyperosmotic salinity response) was reconstructed (Fig. 3). According to SmartCrop, the wheat response to hyperosmotic stress involves 95 genes and 119 proteins, including aquaporins and sodium/hydrogen exchangers, which play a key role in regulating intracellular pH, water balance, and sodium-ion homeostasis (Gupta et al., 2021). Excess sodium ions entering from the environment are removed from the cytoplasm into the apoplast and vacuoles in exchange for hydrogen ions via transmembrane Na^+/H^+ exchangers (Zhao et al., 2021).

The network also includes peroxidases and catalases involved in antioxidant defense under abiotic stress; transcription factors of the MYB and WRKY families; dehydrins (LOC123125487, LOC100141381, and others); cold-shock

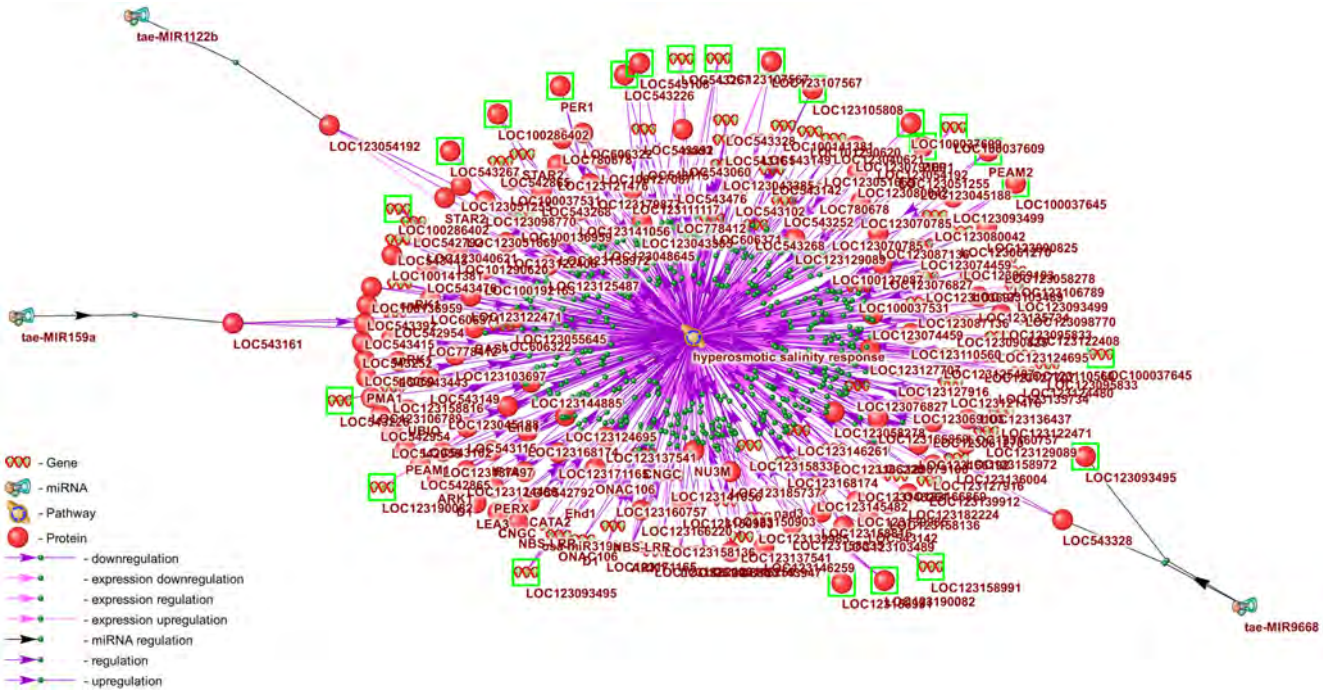


Fig. 3. Genes, proteins, and microRNAs involved in regulating the response of bread wheat to hyperosmotic stress.
Genes differentially expressed in bread-wheat roots in response to soil salinization, as well as their protein products, are highlighted with a green frame.

proteins (LOC123080042, LOC543252, LOC542792); as well as DELLA proteins, which, by suppressing the gibberellin signaling pathway and interacting with jasmonic acid signaling, increase plant tolerance to abiotic stress, including salinity (Colebrook et al., 2014). In addition, the network contains calcium-dependent protein kinases – key components of calcium signaling cascades activated under abiotic stress.

In addition, according to SmartCrop, the regulation of the response to hyperosmotic stress involves the microRNA *tae-MIR159a*, which regulates the expression of the transcription factor TaMyb3 (LOC543161), as well as *tae-MIR1122b* and *tae-MIR9668*, the targets of which are the aquaporins LOC123054192 and LOC123093495, respectively.

Of the full set of genes involved in the regulation of the hyperosmotic stress response, only nine showed differential expression in bread-wheat root tissues under experimental salt stress in the study (Alyahya, Taybi, 2023). This list includes genes encoding aquaporins, peroxidases, catalases, and the serine/threonine protein kinase CTR1 (LOC100286402). Thus, under the experimental conditions described by (Alyahya, Taybi, 2023), signaling pathways associated primarily with antioxidant defense were activated.

The associative network reconstructed in SmartCrop includes these nine DEGs and their protein products, regulatory proteins, as well as two microRNAs: *tae-MIR159a*, which regulates expression of the transcription factor TaMyb3 (LOC543161), and *tae-MIR9668*, targeting the aquaporin LOC123093495. This network is shown in Figure 4.

It is interesting to note that the transcription factor TaMyb3 (LOC543161), which is a target of the microRNA *tae-MIR159a*, in turn acts as a negative regulator of the expression of several genes encoding peroxidases. Suppression of the expression of these enzymes leads to increased accumulation of hydrogen peroxide in tissues and, consequently, to reduced plant tolerance to salinity (Wei et al., 2021). Thus, in this case a "cassette-cascade" regulatory principle involving microRNAs is implemented, in which a microRNA controls the expression of its target transcription factor, and the latter regulates an entire set of genes involved in the response to abiotic stress (Kleshchev et al., 2024).

Transcription factors of the MYB family are well known as regulators of responses to various abiotic stresses, including salinity (Kong et al., 2021; Wang S. et al., 2021). In particular, they participate in the regulation of flavonoid biosynthesis – metabolites required for protecting cells from oxidative stress (Wang X. et al., 2021).

Application of SmartCrop to experimental design

As a second example of SmartCrop use, we performed a search for promising genes and phenotypic markers for subsequent marker-assisted selection and genome editing aimed at increasing rice (*Oryza sativa* L.) tolerance to soil salinity.

According to SmartCrop, the following traits can serve as markers of salinity tolerance: chlorophyll content, seed shape, and the content of the metabolites 3′-methoxyapigenin and 5,7,4′-trihydroxy-3′-methoxyflavone. According to the SmartCrop knowledge base, rice tolerance to salinity is regulated by 30 genes and their corresponding 30 protein products (Fig. 5). In addition to genes, this regulation involves the microRNAs *osa-MIR444f* and *osa-MIR444e*, which target the transcription factor OsMADS23, as well

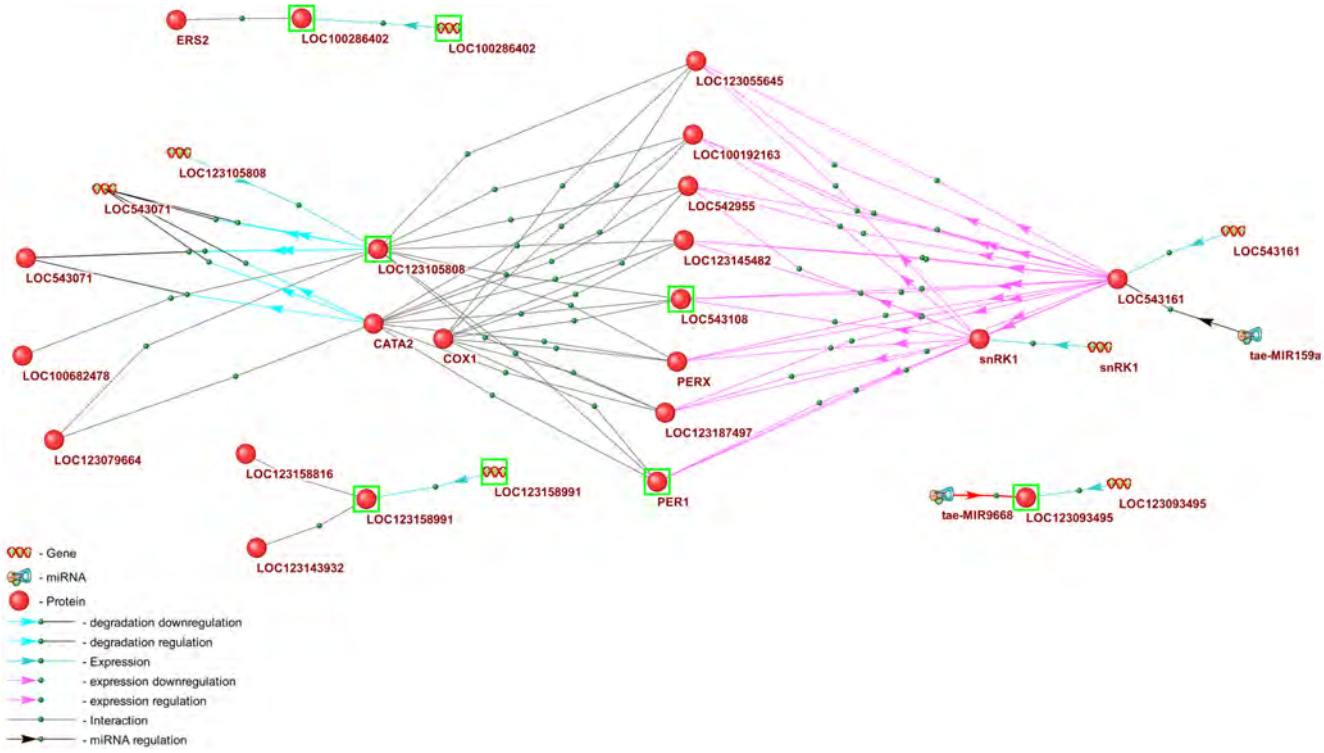


Fig. 4. Gene network regulating the response of bread wheat to salt stress.

Genes differentially expressed in bread-wheat roots in response to soil salinization, as well as their protein products, are highlighted with a green frame.

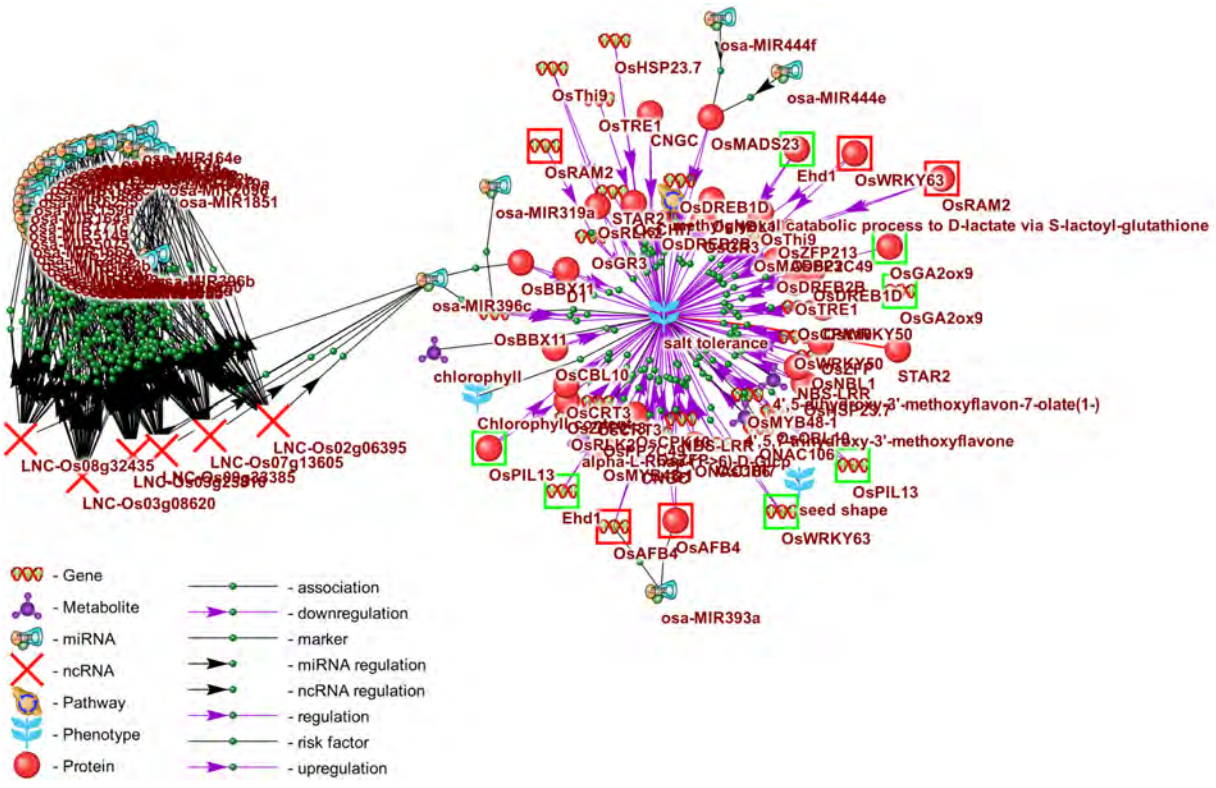


Fig. 5. Associative gene network illustrating the involvement of genes, proteins, microRNAs, and long noncoding RNAs in the regulation of rice (*Oryza sativa* L.) tolerance to salinity and their potential role as phenotypic markers.

Genes in rice and the proteins they encode that positively regulate both salt tolerance and other agronomically important traits are outlined in green. Genes in rice and the proteins they encode that enhance salt tolerance but suppress other agronomically important traits are outlined in red.

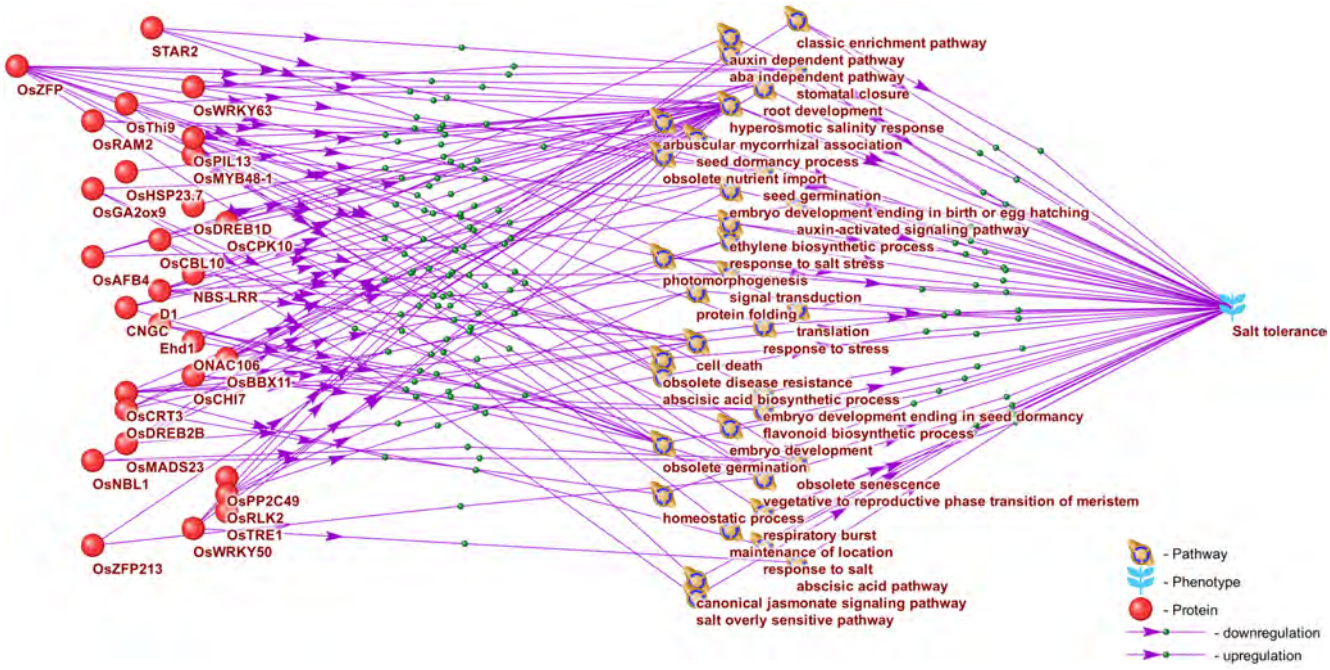


Fig. 6. Regulatory relationships between genes associated with rice salinity tolerance and biological pathways involved in the regulation of this trait.

as *osa-MIR444e*, targeting the auxin receptor OsABF4. The transcription factor OsBBX11, a known regulator of salinity tolerance (Lei et al., 2023), is targeted by the microRNAs *osa-MIR319a* and *osa-MIR396c*.

Long noncoding RNAs (lncRNAs) are molecules longer than 200 nucleotides that regulate gene expression at the transcriptional, post-transcriptional, and epigenetic levels, thereby modulating plant responses to various abiotic and biotic factors, including salinity (Sun X. et al., 2018). lncRNAs can interact with DNA (chromatin, promoters, and enhancers), proteins, mRNAs, and microRNAs. One important mechanism of their action is binding to microRNAs, which prevents the latter from acting on their targets and thus turns lncRNAs into key regulators of microRNA activity (Saha et al., 2025).

According to SmartCrop, the microRNA osa-MIR396c interacts with 508 long noncoding RNAs, six of which (*LNC-Os02g06395*, *LNC-Os03g08620*, *LNC-Os03g25810*, *LNC-Os07g13605*, *LNC-Os08g32435*, *LNC-Os09g33385*) are co-expressed not only with *osa-MIR396c* but also with 42 other rice microRNAs (Fig. 5). This indicates their potential role as key players in the regulation of rice tolerance to abiotic stresses, including salinity.

Of the 30 genes that regulate salinity tolerance, six (*OsPIL13*, *OsNBL1*, *OsABF4*, *OsCPK10*, *OsCRT3*, *OsBBX11*) control chlorophyll content. The remaining 24 genes have not previously been associated with known markers of rice salinity tolerance and therefore represent promising candidates for the discovery of new genetic markers of this trait.

It should be particularly emphasized that prioritizing genes for marker-assisted selection and genome editing requires consideration of the specificity of their regulatory effects, since selection for a single target trait may influence other agronomically important characteristics. The analysis showed that genes and proteins regulating salinity tolerance are associated with 67 other phenotypic traits, including biomass, leaf area, grain morphology, and others, which reflects pleiotropic effects.

Three genes – *OsPIL13*, *Ehd1*, and *OsGA2Ox3* – are positive regulators of both salinity tolerance and such agronomically important traits as grain quality, seed dormancy period, and grain length. This makes them promising candidates for breeding and genome editing, since their modulation may simultaneously increase salt tolerance and improve grain quality. At the same time, the genes *OsWRKY63*, *OsRAM2*, and *OsABF4* enhance rice tolerance to salinity but are associated with negative regulation of seed dormancy period, grain protein content, and plant resistance to *Fusarium graminearum* and *F. pseudograminearum*, which must be considered in breeding programs.

According to SmartCrop, 21 genes are involved exclusively in the regulation of salinity tolerance and are not associated with the regulation of agronomically important traits or resistance to pathogens, which makes them suitable candidates for targeted breeding aimed at increasing salt tolerance.

Another important factor that must be taken into account when selecting genes for marker-assisted selection and/or genome editing is the potential bidirectionality of their regulatory effects, since gene products may either stimulate or suppress biological processes involved in the positive or negative regulation of the target trait. To assess such bidirectionality, the "Pathway Wizard" module of the ANDSystem program was used to identify regulatory relationships between the protein products of the 30 genes associated with rice salinity tolerance and the biological processes that, in turn, participate in regulating this trait (Fig. 6).

Among the genes regulating salinity tolerance, particular interest is drawn to *OsMYB48-1*, *OsCPK10*, *OsCBL10*, *OsDREB2B*, *OsRAM2*, and *NBS-LRR*, which exhibit a unidirectional effect in the form of positive regulation of key processes that ensure tolerance to salt stress (hyperosmotic salinity response, stomatal closure, ABA-independent pathway, etc.). The high degree of connectivity of these genes with the target trait, combined with the unidirectional nature of their regulatory action, suggests that their use in marker-assisted selection or genome editing may have a more direct and pronounced impact on increasing salt tolerance compared with other candidates.

Conclusion

The SmartCrop knowledge base is a specialized version of the ANDSystem software information platform, adapted for the tasks of rice and wheat genetics and breeding. It integrates information on a wide range of entities – genes, proteins, metabolites, noncoding RNAs, biological processes, breeding-relevant and phenotypic traits, pathogens, as well as biotic and abiotic factors – and their relationships. This architecture provides extensive opportunities for studying the molecular genetic mechanisms of plant stress tolerance, as well as for selecting genes, genetic markers, and phenotypic traits within the framework of marker-assisted selection of crop plants.

Examples of SmartCrop applications for the functional annotation of differentially expressed genes in bread wheat under salt stress and for planning experiments to increase rice salinity tolerance using marker-assisted selection have demonstrated the high efficiency of the system and its potential for solving applied problems in breeding and genome editing.

References

Alyahya N., Taybi T. Comparative transcriptomic profiling reveals differentially expressed genes and important related metabolic pathways in shoots and roots of a Saudi wheat cultivar (Najran) under salinity stress. *Front Plant Sci.* 2023;14:1225541. doi 10.3389/fpls.2023.1225541

Antropova E.A., Volyanskaya A.R., Adamovskaya A.V., Demenkov P.S., Yatsyk I.V., Ivanisenko T.V., Orlov Y.L., Haoyu Ch., Chen M., Ivanisenko V.A. Computational identification of promising genetic markers associated with molecular mechanisms of reduced rice resistance to *Rhizoctonia solani* under excess nitrogen fertilization using gene network reconstruction and analysis methods. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2024;28(8):960-973. doi 10.18699/vjgb-24-103

Ayadi M., Brini F., Masmoudi K. Overexpression of a wheat aquaporin gene, *Td*PIP2;1, enhances salt and drought tolerance in transgenic durum wheat cv. Maali. *Int J Mol Sci.* 2019;20(10):2389. doi 10.3390/ijms20102389

Chao H., Zhang S., Hu Y., Ni Q., Xin S., Zhao L., Ivanisenko V.A., Orlov Y.L., Chen M. Integrating omics databases for enhanced crop breeding. *J Integr Bioinform.* 2023;20(4):20230012. doi 10.1515/jib-2023-0012

Chen T., Nomura K., Wang X., Sohrabi R., Xu J., Yao L., Paasch B.C., Ma L., Kremer J., Cheng Y., Zhang L., Wang N., Wang E., Xin X.F., He S.Y. A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature*. 2020;580(7805):653-657. doi 10.1038/s41586-020-2185-0

Cheng M., Zhu Y., Yu H., Shao L., Zhang Y., Li L., Tu H., … Orlov Y.L., Chen D., Wong A., Yang Y.E., Chen M. Non-coding RNA notations, regulations and interactive resources. *Funct Integr Genomics.* 2024;24(6):217. doi 10.1007/s10142-024-01494-w

Colebrook E.H., Thomas S.G., Phillips A.L., Hedden P. The role of gibberellin signalling in plant responses to abiotic stress. *J Exp Biol.* 2014;217(1):67-75. doi 10.1242/jeb.089938

Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol.* 2012;11(3-4):149-161. doi 10.3233/isb-2012-0449

Demenkov P.S., Saik O.V., Ivanisenko T.V., Kolchanov N.A., Kochetov A.V., Ivanisenko V.A. Prioritization of potato genes involved in the formation of agronomically valuable traits using the SOLANUM TUBEROSUM knowledge base. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2019;23(3):312-319. doi 10.18699/VJ19.501

Demenkov P.S., Oshchepkova E.A., Ivanisenko T.V., Ivanisenko V.A. Prioritization of biological processes based on the reconstruction and analysis of associative gene networks describing the response of plants to adverse environmental factors. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2021;25(5):580-592. doi 10.18699/VJ21.065

Do H., Than K., Larmande P. Evaluating named-entity recognition approaches in plant molecular biology. In: Multi-disciplinary Trends in Artificial Intelligence. MIWAI 2018. Lecture Notes in Computer Science. Vol. 11248. Springer, Cham., 2018;219-225. doi 10.1007/978-3-030-03014-8_19

D'Souza J. Agriculture named entity recognition – towards FAIR, reusable scholarly contributions in agriculture. *Knowledge.* 2024;4(1):1-26. doi 10.3390/knowledge4010001

Gemma Team, Google DeepMind. Gemma 2: improving open language models at a practical size. *arXiv*. 2024:2408.00118. doi 10.48550/arXiv.2408.00118

Guindani L.G., Oliveirai G.A., Ribeiro M.H.D.M., Gonzalez G.V., de Lima J.D. Exploring current trends in agricultural commodities forecasting methods through text mining: developments in statistical and artificial intelligence methods. *Heliyon.* 2024;10(23):e40568. doi 10.1016/j.heliyon.2024.e40568

Gupta A., Shaw B.P., Sahu B.B. Post-translational regulation of the membrane transporters contributing to salt tolerance in plants. *Funct Plant Biol*. 2021;48(12):1199-1212. doi 10.1071/FP21153

Hamilton W.L., Ying R., Leskovec J. Inductive representation learning on large graphs. *arXiv.* 2017. doi 10.48550/arxiv.1706.02216

Hoai P.T.T., Tyerman S.D., Schnell N., Tucker M., McGaughey S.A., Qiu J., Groszmann M., Byrt C.S. Deciphering aquaporin regulation and roles in seed biology. *J Exp Bot.* 2020;71(6):1763-1773. doi 10.1093/jxb/erz555

Islamaj R., Wei C.H., Cissel D., Miliaras N., Printseva O., Rodionov O., Sekiya K., Ward J., Lu Z. NLM-Gene, a richly annotated gold standard dataset for gene entities that addresses ambiguity and multi-species gene recognition. *J Biomed Inform.* 2021;118:103779. doi 10.1016/j.jbi.2021.103779

Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Text mining on PubMed. In: Chen M., Hofestädt R. (Eds) Approaches in Integrative Bioinformatics. Springer, 2014;161-170. doi 10.1007/978-3-642-41281-3_6

Ivanisenko T.V., Saik O.V., Demenkov P.S., Khlestkin V.K., Khlestkina E.K., Kolchanov N.A., Ivanisenko V.A. The SOLANUM TUBEROSUM knowledge base: the section on molecular-genetic regulation of metabolic pathways. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2018;22(1):8-17. doi 10.18699/VJ18.325 (in Russian)

Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of

ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics.* 2020;21(S11):228. doi 10.1186/s12859-020-03557-8

Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. The new version of the ANDDigest tool with improved AI-based short names recognition. *Int J Mol Sci.* 2022;23(23):14934. doi 10.3390/ijms232314934

Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. An accurate and efficient approach to knowledge extraction from scientific publications using structured ontology models, graph neural networks, and large language models. *Int J Mol Sci.* 2024;25(21):11811. doi 10.3390/ijms252111811

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol.* 2015;9(Suppl.2):S2. doi 10.1186/1752-0509-9-s2-s2

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics.* 2019;20(S1):34. doi 10.1186/s12859-018-2567-6

Kanehisa M. Molecular network analysis of diseases and drugs in KEGG. In: Mamitsuka H., DeLisi C., Kanehisa M. (Eds) Data Mining for Systems Biology. Methods in Molecular Biology. Vol. 939. Humana Press, 2013;263-275. doi 10.1007/978-1-62703-107-3_17

Kleshchev M.A., Maltseva A.V., Antropova E.A., Demenkov P.S., Ivanisenko T.V., Orlov Y.L., Chao H., Chen M., Kolchanov N.A., Ivanisenko V.A. Reconstruction and computational analysis of the microRNA regulation gene network in wheat drought response mechanisms. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2024;28(8):904-917. doi 10.18699/vjgb-24-98

Kong W., Sun T., Zhang C., Deng X., Li Y. Comparative transcriptome analysis reveals the mechanisms underlying differences in salt tolerance between *indica* and *japonica* rice at seedling stage. *Front Plant Sci.* 2021;12:725436. doi 10.3389/fpls.2021.725436

Krallinger M., Rabal O., Leitner F., Vazquez M., Salgado D., Lu Z., Leaman R., … Alves R., Segura-Bedmar I., Martínez P., Oyarzabal J., Valencia A. The CHEMDNER corpus of chemicals and drugs and its annotation principles. *J Cheminform.* 2015;7(S1):S2. doi 10.1186/1758-2946-7-s1-s2

Lei L., Cao L., Ding G., Zhou J., Luo Y., Bai L., Xia T., … Xie T., Yang G., Wang X., Sun S., Lai Y. *OsBBX11* on *qSTS4* links to salt tolerance at the seeding stage in *Oryza sativa* L. ssp. *Japonica*. *Front Plant Sci.* 2023;14:1139961. doi 10.3389/fpls.2023.1139961

Lesk C., Rowhani P., Ramankutty N. Influence of extreme weather disasters on global crop production. *Nature.* 2016;529(7584):84-87. doi 10.1038/nature16467

Liu L., Liu E., Hu Y., Li S., Zhang S., Chao H., Hu Y., Zhu Y., Chen Y., Xie L., Shen Y., Wu L., Chen M. ncPlantDB: a plant ncRNA database with potential ncPEP information and cell type-specific interaction. *Nucleic Acids Res.* 2025;53(D1):D1587-D1594. doi 10.1093/nar/gkae1017

Mittler R. Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 2006;11(1):15-19. doi 10.1016/j.tplants.2005.11.002

Naithani S., Gupta P., Preece J., D'Eustachio P., Elser J.L., Garg P., Dikeman D.A., … Bolton E., Papatheodorou I., Stein L., Ware D., Jaiswal P. Plant Reactome: a knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Res.* 2020;48(D1):D1093-D1103. doi 10.1093/nar/gkz996

Ni Y., Aghamirzaie D., Elmarakeby H., Collakova E., Li S., Grene R., Heath L.S. A machine learning approach to predict gene regulatory networks in seed development in Arabidopsis. *Front Plant Sci.* 2016;7:1936. doi 10.3389/fpls.2016.01936

Nykiel M., Gietler M., Fidler J., Prabucka B., Labudda M. Abiotic stress signaling and responses in plants. *Plants*. 2023;12(19):3405. doi 10.3390/plants12193405

Otasek D., Morris J.H., Bouças J., Pico A.R., Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 2019;20(1):185. doi 10.1186/s13059-019-1758-4

Pearson H. Biology's name game. *Nature*. 2001;411(6838):631-632. doi 10.1038/35079694

Saha C., Saha S., Bhattacharyya N.P. LncRNAOmics: a comprehensive review of long non-coding RNAs in plants. *Genes*. 2025;16(7):765. doi 10.3390/genes16070765

Saik O.V., Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. Development of methods for automatic extraction of knowledge from texts of scientific publications for the creation of a knowledge base SOLANUM TUBEROSUM. *Agric Biol.* 2017;52(1):63-74. doi 10.15389/agrobiology.2017.1.63eng

Shewry P.R., Hey S.J. The contribution of wheat to human diet and health. *Food Energy Secur.* 2015;4(3):178-202. doi 10.1002/fes3.64

Shrestha A.M.S., Gonzales M.E.M., Ong P.C.L., Larmande P., Lee H.S., Jeung J.U., Kohli A., Chebotarov D., Mauleon R.P., Lee J.S., McNally K.L. RicePilaf: a post-GWAS/QTL dashboard to integrate pangenomic, coexpression, regulatory, epigenomic, ontology, pathway, and text-mining information to provide functional insights into rice QTLs and GWAS loci. *GigaScience*. 2024;13:giae013. doi 10.1093/gigascience/giae013

Statello L., Guo C.J., Chen L.L., Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*. 2021;22:96-118. doi 10.1038/s41580-020-00315-9

Sun X., Zheng H., Sui N. Regulation mechanism of long non-coding RNA in plant response to stress. *Biochem Biophys Res Commun.* 2018;503(2):402-407. doi 10.1016/j.bbrc.2018.07.072

Supriya P., Srividya G.K., Solanki M., Manvitha D., Prakasam V., Balakrishnan M., Neeraja C.N., Rao C.S., Sundaram R.M., Mangrauthia S.K. Identification and expression analysis of long non-coding RNAs of rice induced during interaction with *Rhizoctonia solani*. *Physiol Mol Plant Pathol*. 2024;134:102389. doi 10.1016/j.pmpp.2024.102389

Szklarczyk D., Gable A.L., Nastou K.C., Lyon D., Kirsch R., Pyysalo S., Doncheva N.T., Legeay M., Fang T., Bork P., Jensen L.J., von Mering C. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605-D612. doi 10.1093/nar/gkaa1074

Tian F., Yang D.C., Meng Y.Q., Jin J., Gao G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res*. 2020;48(D1):D1104-D1113. doi 10.1093/nar/gkz1020

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need. *arXiv*. 2017. doi 10.48550/arXiv.1706.03762

Verma V., Ravindran P., Kumar P.P. Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 2016;16(1):86. doi 10.1186/s12870-016-0771-y

Virlouvet L., Avenson T.J., Du Q., Zhang C., Liu N., Fromm M., Avramova Z., Russo S.E. Dehydration stress memory: gene networks linked to physiological responses during repeated stresses of *Zea mays*. *Front Plant Sci*. 2018;9:1058. doi 10.3389/fpls.2018.01058

Volyanskaya A.R., Antropova E.A., Zubairova U.S., Demenkov P.S., Venzel A.S., Orlov Y.L., Makarova A.A., Ivanisenko T.V., Gorshkova T.A., Aglyamova A.R., Kolchanov N.A., Chen M., Ivanisenko V.A. Reconstruction and analysis of the gene regulatory network

for cell wall function in *Arabidopsis thaliana* L. leaves in response to water deficit. *Vavilov J Genet Breed.* 2023;27(8):1031-1041. doi 10.18699/vjgb-23-118

Wang S., Shi M., Zhang Y., Xie X., Sun P., Fang C., Zhao J. FvMYB24, a strawberry R2R3-MYB transcription factor, improved salt stress tolerance in transgenic Arabidopsis. *Biochem Biophysical Res Commun.* 2021;569:93-99. doi 10.1016/j.bbrc.2021.06.085

Wang X., Niu Y., Zheng Y. Multiple functions of MYB transcription factors in abiotic stress responses. *Int J Mol Sci.* 2021;22(11):6125. doi 10.3390/ijms22116125

Wei T., Guo D., Liu J. PtrMYB3, a R2R3-MYB transcription factor from *Poncirus trifoliata*, negatively regulates salt tolerance and hydrogen peroxide scavenging. *Antioxidants*. 2021;10(9):1388. doi 10.3390/antiox10091388

Yan J., Wang X. Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. *Plant J.* 2022;111(6):1527-1538. doi 10.1111/tpj.15905

Zhang D., Zhao R., Xian G., Kou Y., Ma W. A new model construction based on the knowledge graph for mining elite polyphenotype genes in crops. *Front Plant Sci.* 2024;15:1361716. doi 10.3389/fpls.2024.1361716

Zhao S., Zhang Q., Liu M., Zhou H., Ma C., Wang P. Regulation of plant responses to salt stress. *Int J Mol Sci.* 2021;22(9):4609. doi 10.3390/ijms22094609

Conflict of interest. The authors declare no conflict of interest.
Received August 18, 2025. Revised October 1, 2025. Accepted October 6, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1235-1245

doi 10.18699/vjgb-25-133

Validation of markers for resistance to *Pyrenophora teres* f. *teres* loci on barley chromosomes 3H, 4H, and 6H in the polygenic inheritance of the trait

O.S. Afanasenko [1], N.V. Mironenko [1], N.M. Lashina[1], I.V. Rozanova[2], E.I. Kyrova[1], Yu.S. Nikolskaya[1], A.A. Zubkovich[3]

[1] All-Russian Research Institute of Plant Protection (VIZR), St. Petersburg, Russia
[2] Sirius University of Science and Technology, Center of Genetics and Life Sciences, Sochi, Russia
[3] Research and Practical Center of Agriculture of the National Academy of Sciences of Belarus, Zhodino, Belarus
olga.s.afan@gmail.com

Abstract. The causal agent of net blotch *Pyrenophora teres* Drechs. f. *teres* (*Ptt*) is a dangerous pathogen of barley. The development of genetic protection against this disease is a necessary link in resource-saving and environmentally friendly barley cultivation technologies. Effective QTL markers controlling both qualitative and quantitative resistance are required for breeding for resistance to *Ptt*. As a result of GWAS, we identified barley accessions of different origins, the SNP haplotypes of which were associated with resistance loci simultaneously on different barley chromosomes (VIR catalogue numbers: k-5900, k-8829, k-8877, k-14936, k-30341 and k-18552). The aim of the study was to validate SNP markers (MM) of *Ptt* resistance loci on chromosomes 3H, 4H and 6H in F_2 from crossing six resistant accessions with the susceptible variety Tatum. The observed segregation for resistance in all crossing combinations confirmed the presence of several genetic determinants of resistance in the studied accessions. To study the polymorphism of the parents from the crosses and the correspondence between the phenotypes to the presence/absence of the markers in the segregating populations, primers with a specific 3'-end, CAPS markers, and KASP markers were developed. A significant association ($p < 0.05$) between the presence of the CAPS marker JHI-Hv50k-2016-391380 HindIII on chromosome 6H and the phenotype of resistance to *Ptt* in F_2 plants was revealed in crosses between the susceptible cultivar Tatum and accessions k-5900, k-8829, k-8877 and k-18552. On chromosome 4H, a significant association with the resistance phenotype in the F_2 population from the cross with accession k-8877 was revealed for marker JHI-Hv50k-2016-237924, and in that from the cross with accession k-5900, for marker SCRI_RS_181886. The presence of QTL on chromosome 6H, which controls qualitative resistance in four barley accessions, masks the expression of other genes, which explains the discrepancy between the resistance phenotype and the presence of molecular markers in the segregating populations. Resistance donors and molecular markers with proven efficacy can be used in marker-assisted selection (MAS) to develop barley cultivars resistant to net blotch.

Key words: barley; net blotch; landraces; resistance; SNP markers; CAPS markers; KASP markers

For citation: Afanasenko O.S., Mironenko N.V., Lashina N.M., Rozanova I.V., Kyrova E.I., Nikolskaya Yu.S., Zubkovich A.A. Validation of markers for resistance to *Pyrenophora teres* f. *teres* loci on barley chromosomes 3H, 4H, and 6H in the polygenic inheritance of the trait. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1235-1245. doi 10.18699/vjgb-25-133

Funding. This work was supported by Russian Science Foundation grant No. 24-26-00072.

Валидация маркеров, разработанных для выявления локусов устойчивости к *Pyrenophora teres* f. *teres* на хромосомах ячменя 3H, 4H и 6H при полигенном наследовании признака

О.С. Афанасенко [1], Н.В. Мироненко [1], Н.М. Лашина[1], И.В. Розанова[2], Е.И. Кырова[1], Ю.С. Никольская[1], А.А. Зубкович[3]

[1] Всероссийский научно-исследовательский институт защиты растений (ВИЗР), Санкт-Петербург, Россия
[2] Научно-технологический университет «Сириус», Центр генетики и наук о жизни, Сочи, Россия
[3] Научно-практический центр Национальной академии наук Беларуси по земледелию, Жодино, Беларусь
olga.s.afan@gmail.com

Аннотация. Возбудитель сетчатой пятнистости листьев ячменя *Pyrenophora teres* Drechs. f. *teres* (*Ptt*) относится к эпифитотийно опасным патогенам ячменя. Разработка генетической защиты от этой болезни – важное звено ресурсосберегающих и экологически безопасных технологий возделывания ячменя. Для селекции

на устойчивость к *Ptt* необходимы эффективные маркеры QTL, контролирующих как качественную, так и количественную устойчивость. При проведении полногеномного анализа (GWAS) нами выявлены образцы ячменя различного происхождения, SNP-гаплотипы которых ассоциировались с локусами устойчивости одновременно на разных хромосомах. Целью исследований была валидация SNP-маркеров локусов устойчивости к *Ptt* на хромосомах 3H, 4H и 6H с использованием F_2 популяции от скрещивания шести устойчивых образцов к-5900, к-8829, к-8877, к-14936, к-30341 и к-18552 с восприимчивым сортом Tatum. Расщепление по устойчивости во всех комбинациях скрещиваний подтвердило наличие нескольких генетических детерминант устойчивости у изучаемых образцов. Для изучения полиморфизма родительских компонентов скрещиваний и соответствия фенотипа наличию/отсутствию маркера в расщепляющихся популяциях были разработаны праймеры со специфичным 3′-концом, CAPS- и KASP-маркеры. Значимая связь ($p < 0.05$) наличия CAPS-маркера JHI-Hv50k-2016-391380 HindIII на хромосоме 6H и фенотипа устойчивости к *Ptt* у растений F_2 выявлена в комбинациях скрещивания восприимчивого сорта Tatum с образцами к-5900, к-8829, к-8877 и к-18552; на хромосоме 4H при фрагментном анализе значимая связь с фенотипом устойчивости в популяции F_2 с участием образца к-8877 выявлена для маркера JHI-Hv50k-2016-237924, образца к-5900 для маркера SCRI_RS_181886 и образца к-8829 для маркера JHI-Hv50k-2016-166356. Наличие QTL на хромосоме 6H, контролирующего качественную устойчивость у четырех образцов ячменя, маскирует проявление других генов с меньшим фенотипическим проявлением, что и является причиной несоответствия фенотипа устойчивости и наличия молекулярного маркера в расщепляющихся популяциях. Доноры устойчивости и молекулярные маркеры с доказанной эффективностью могут быть использованы в MAS для создания устойчивых к возбудителю сетчатой пятнистости сортов ячменя.

Ключевые слова: образцы ячменя; сетчатая пятнистость; устойчивость; SNP-маркеры; CAPS-маркеры; KASP-маркеры; доноры устойчивости

Introduction

The causal agent of net blotch, *Pyrenophora teres* Drechs. f. *teres* (anamorph: *Drechslera teres* Sacc. (Shoem.) = *Helminthosporium teres*), is a dangerous pathogen of barley. Yield losses from this pathogen on susceptible cultivars under favorable conditions can reach 40 %, with annual losses estimated at 12–17 %. According to our data, the majority of both spring and winter barley cultivars registered in the State Register of Breeding Achievements are susceptible to the net blotch. This is partly due to the difficulties of working with hemibiotrophic pathogens: the strong dependence of resistance expression on environmental factors, incomplete dominance of resistance and, consequently, difficulties in selection in segregating hybrid populations, complex inheritance of resistance traits determined by multiple QTLs, and epistatic interactions between resistance genes.

Genetically protected cultivars are an essential component of resource-saving and environmentally friendly agricultural crop cultivation technologies. The development of effective genetic protection is based on the availability of genetically diverse donors of qualitative and quantitative resistance genes and their rational use, taking into account the ranges of pathogen populations in different climatic regions. Timely rotation of genetically protected cultivars helps stabilize the population composition of plant pathogens and reduce the likelihood of epidemics.

Currently, using biparental mapping populations and genome-wide association study (GWAS) technology, genes and loci for quantitative resistance (QTL) to *P. teres* f. *teres* (*Ptt*) have been identified on all barley chromosomes (Steffenson et al., 1996; Richter et al., 1998; Friesen et al., 2006; Manninen et al., 2006; Yun et al., 2006; Grewal et al., 2008, 2012; Gupta et al., 2010; Cakir et al., 2011; Liu et al., 2011; König et al., 2013, 2014; O'Boyle et al., 2014; Afanasenko et al., 2015, 2022; Richards et al., 2017; Wonneberger et al., 2017; Amezrou et al., 2018; Martin et al., 2018; Dinglasan et al., 2019; Novakazi et al., 2019; Rozanova et al., 2019; Clare et al., 2021; Rehman et al., 2025). In our study, in a collection of 449 barley accessions, genotyped using the 50K Illumina SNP chip for 33,818 markers, 15 loci and 43 SNPs significantly associated with resistance to *Ptt* haplotypes were identified (Novakazi et al., 2019). As a result of this work, a group of resistant barley accessions was identified, the SNP haplotypes of which were associated with resistance loci simultaneously on different barley chromosomes, which apparently indicates the presence of several QTL and a possible additive effect. For example, in six resistant barley accessions included in this study, k-5900, k-8829, k-8877, k-14936, k-30341 (VIR catalogue numbers) and k-18552 (cultivar Zolo), SNP marker haplotypes in each accession were associated with 5–8 resistance loci on chromosomes 3H, 4H, 6H and 7H.

The molecular markers (MMs) of genes and QTLs for resistance to *P. teres* f. *teres* identified in these studies and in the studies of other authors, in most cases, have not been validated in other genetic environments for their effective use in barley breeding.

The aim of this study was to validate the SNP markers for *Ptt* resistance loci on chromosomes 3H, 4H, and 6H, known from the scientific literature, in F_2 populations obtained from crossing six resistant accessions with the susceptible cultivar Tatum.

Materials and methods

Barley genotypes. Six resistant barley accessions were selected for crossing and obtaining segregating F_2 populations (VIR catalogue numbers): k-5900, k-8829, k-8877, k-14936, k-18552 (Zolo cultivar), and k-30341. Their SNP marker haplotypes were associated with resistance loci on different barley chromosomes, including chromosomes 3H, 4H, and 6H. The productive two-row barley cultivar Tatum from Germany was used as the susceptible parent. The characteristics of the barley accessions are presented in Table 1.

Table 1. Origin of barley accessions and chromosomal location of QTLs associated with resistance (Novakazi et al., 2019)

VIR catalogue numbers	Variety	Origin	QTLs
k-5900	*pallidum*	Turkmenistan	6H-1, 6H-2, 6H-3, 7H
k-8829		Italy	4H, 6H-1, 6H-2, 6H-3, 3H-1, 7H
k-8877		Spain	4H, 6H-1, 6H-2, 6H-3, 3H-1, 7H
k-14936		Tajikistan	4H, 6H-1, 6H-2, 6H-3, 7H
k-18552		Australia	4H, 6H-1, 6H-2, 6H-3, 3H-1, 3H-2, 7H
k-30341	*nigrum*	Peru	4H, 6H-1, 6H-2, 3H-1, 7H

Note. Loci are within intervals determined using the Barleymap resource (https://barleymap.eead.csic.es/barleymap): 4H – 58,942,545–67,692,302 bp and 448,603,913–449,611,912 bp, 6H-1 – 64,219,990–67,138,358 bp, 6H-2 – 125,903,650–151,127,756 bp, 6H-3 – 338,755,997–378,210,479 bp, 3H-1 – 119,627,830–130,790,360 bp, 3H-2 – 490,244,247–491,381,651 bp, 7H – 5,165,127 bp. All barley samples had a row count of six.

***P. teres* f. *teres* isolates.** Five *Ptt* isolates were used to assess resistance in a GWAS: No. 13 (Russia), Hoehenstedt (Germany), NFNB 50, NFNB 73, and NFNB 85 (Australia) (Novakazi et al., 2019). In this study, the resistance of these six accessions was assessed in addition to nine *Ptt* isolates of different origins (Table S1)[1]. For all isolates, the virulence formula was determined using a standard set of differentials (Afanasenko et al., 2009) (Table S2).

A study of barley resistance to *P. teres* f. *teres*. Methods for isolating the fungus into pure culture, storing it, grown on modified Chapek medium (KCL – 0.5 g, KH_2PO_4 – 0.5 g, $MgSO_4$ – 0.5 g, urea – 1.2 g, lactose – 20 g, agar-agar – 20 g per 1 l of distilled water), and obtaining a *Ptt* conidial suspension for plant inoculation are described in detail in (Afanasenko et al., 2022; Lashina et al., 2023). The parent accessions and 65 seeds of each F_1 hybrid population were sown in 1-liter containers with Terra Vita® potting soil. The plants were grown under controlled conditions in a VIZR climate room at 20–22 °C and a 16-hour photoperiod for 10–14 days. Barley plants were inoculated at the two- to three-leaf stage by spraying a suspension of single conidia isolates at a rate of 0.2 ml per plant. Conidia were counted with a hemocytometer, and the concentration was adjusted to 6,000 conidia/ml for inoculation. After inoculation, the plants were covered with plastic bags and left for 48 hours at 20–22 °C without light. After two days, the infected plants were transferred to light (TL-FITO VR LED lamps) with a 16-hour photoperiod and maintained at 60–70 % humidity.

Seedling response types were assessed on the second leaf 10–12 days after inoculation using a modified 10-point scale by A. Tekauz (1985), where values 1.0–4.9 indicated resistance; 5.0–5.9, an intermediate response; and 6.0–10, susceptibility.

Primer development. Three approaches were used to validate the identified SNP markers: a) Allele-Specific PCR (AS_PCR): development of primers with a 3′ end located at the position of the SNP of interest. Depending on the correspondence (complementarity) of the 3′ end SNP to the target DNA region, the presence or absence of a PCR amplification product is determined; b) Cleaved Amplified Polymorphic Sequences (CAPS): detection of SNPs using CAPS markers, the SNP of interest is located in the recognition site of a restriction endonuclease. As a result, the polymorphism of the restriction products determines the presence or absence of a restriction site in the amplicon – different genotypes will correspond to restriction fragments of different lengths in an agarose or polyacrylamide gel; c) Kompetitive allele-specific PCR (KASP): use of a PCR-based fluorescent genotyping system.

The candidate SNP position was confirmed using the Barleymap resource (https://barleymap.eead.csic.es/barleymap). The nucleotide sequences flanking the SNP (500 bp on each side) were exported to the Essembl Plants database (http://plants.ensembl.org/index.html). Primer design was developed using the UGENE software package (v 49.1). For CAPS markers, the SnapGene Viewer software package (https://www.snapgene.com) was additionally used for sequence analysis and selection of a restriction endonuclease, differentiating genotypes based on the presence/absence of a restriction site at the SNP position.

To develop KASP markers, nucleotide sequences flanking the resistance-associated SNP (50 bp on each side) were exported from the Essembl Plants database (http://plants.ensembl.org/index.html). Based on these sequences, SNP allele-specific primer sequences were developed, using fluorescent tail sequences according to the protocol described by S. Jatayev et al. (2017).

DNA extraction and PCR conditions. DNA from frozen barley leaves was isolated using CTAB (cetyltrimethylammonium bromide). For this, the first leaf of each plant was ground in a mortar with liquid nitrogen supplemented with 2 % CTAB before inoculation with isolate F18. The homogenate was then lysed at 65 °C for two hours. DNA purification and extraction were performed according to the protocol (Murray, Thompson, 1980). The DNA precipitate was dissolved in deionized bidistilled sterile water to a concentration of 100–150 µg/µl. A C1000 thermal cycler (BIO-RAD) was used for amplification. The reaction was carried out in 25 µl: buffer (×10) – 2.5 µl, $MgCl_2$ (50 mM) – 1.25 µl, dNTP (10 mM) – 0.5 µl, forward and reverse primers (10 pmol) – 0.25 µl each, Taq polymerase – 0.25 µl, water (bidistilled) – 19.0 µl, DNA (10–20 ng) – 1.0 µl. The optimal PCR conditions were selected for each primer. For most primers, the annealing temperature was 60 °C. Primers were purchased from Beagle (St. Peters-

[1] Supplementary Tables S1–S14 and Figures S1–S8 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Afan_Engl_29_8.pdf

burg). Restriction endonuclease digestion was performed according to the manufacturer's protocols (SibEnzyme), and restriction products were visualized on a 2 % agarose gel (for HindIII, NruI, and RsaI).

Statistical data processing. Statistical analysis was performed using the χ^2 test. Calculations were performed using STATISTICA 13.0 (Statsoft, www.statsoft.com) and the methodology described in N. Pandis (2016). For $p < 0.05$, Fisher's exact test was additionally applied to the χ^2 test.

The diagnostic efficacy of the tested markers was determined as the ratio of the sum of true positive and true negative results to the total number of plants tested.

Results

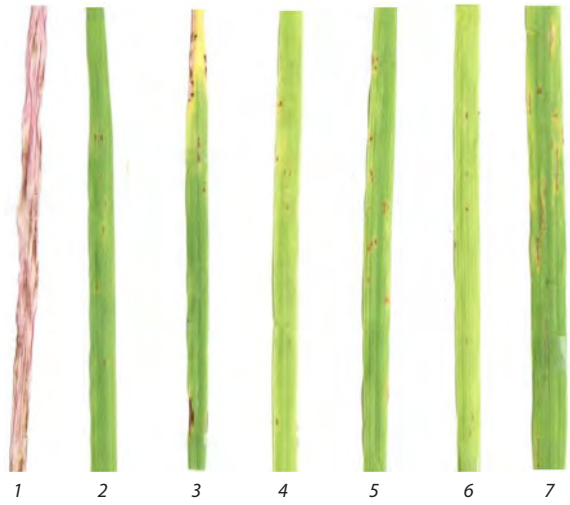
Resistance of parental accessions

The resistance of parental accessions to nine isolates of different origins, belonging to eight *Ptt* pathotypes, was studied (Table S2). All barley accessions exhibited race-specific resistance (Table 2). Of the nine isolates studied, one was virulent against k-8829, k-8877, k-14936, and k-18552, while four isolates were virulent against accession k-30341.

To analyze the segregation of resistance in F_2 hybrid populations from the crossing of resistant barley accessions with the susceptible cultivar Tatum, the F18 isolate was used, since all the studied accessions were resistant to it, and the cv. Tatum demonstrated the maximum type of reaction – 10 (susceptibility) (see the Figure).

Segregation of resistance to *Ptt* in F_2 populations from crosses of resistant barley accessions with the susceptible Tatum cultivar

The results of segregation of resistance in F_2 hybrid populations are presented in Table 3. The actual segregation in all cross combinations does not correspond to simple inheritance of the resistance, whether the class with intermediate reactions is combined with the class of resistant or susceptible plants, confirming the presence of multiple genetic determinants of resistance in the studied accessions (Table 3).



Types of reaction of parental accessions when infected with isolate F18, the damage score is indicated in brackets: *1* – Tatum (10), *2* – k-5900 (3), *3* – k-8829 (3.5), *4* – k-8877 (2), *5* – k-14936 (3), *6* – k-18552 (2), *7* – k-30341 (4).

Parental accession polymorphism for molecular markers on chromosome 4H

To study the polymorphism of parental accessions, primers with a specific 3′ end (Table S3), CAPS markers, and competitive allele-specific PCR (KASP markers) were used. Ten markers on chromosome 4H were studied: five markers, identified from GWAS data, were associated with resistance to isolate No. 13 of *P. teres* f. *teres* in the position of 50.0–50.4 cM (Novakazi et al., 2019), and five markers were associated with *Ptt* resistance in the works of other researchers (Richards et al., 2017; Wonneberger et al., 2017; Amezrou et al., 2018). The positions of all 10 markers are listed in Table 4. CAPS markers were developed for two SNP markers on chromosome 4H.

Using the restriction endonuclease NruI for the JHI-Hv50k-2016-237684 marker, two alleles are distinguished: in the presence of the T allele, which lacks a restriction site,

Table 2. Response of barley accessions to inoculation with *P. teres* f. *teres* isolates

Accession (VIR catalogue number)	Infection responses (IRs) to isolates (score 1–10)									
	F18	S18	B18	V13	Pr2	Ger7	Cz11.1	Can11	SA7	Mean
k-5900	3	5	2	2	3	3	5	2	4	3.2
k-8829	3	4	2	1	4	6	4	3	2	3.2
k-8877	2	7	1	1	2	1	1	3	2	2.2
k-14936	3	2	3	3	2	1	6	2	1	2.9
k-18552	2	3	1	1	4	6	2	3	2	2.7
k-30341	4	4	4	9	7	8	8	5	2	5.7
Resistant and susceptible test varieties										
Canadian Lake Shore (CLS) $QPtt3H_{CLS}$ (R)	1	3	3	2	3	5	1	3	2	2.6
C.I. 5791 $QPtt6H_{5791\ 55CM}$ (R)	1	2	5	5	3	2	3	2	2	2.8
Harrington (S)	10	10	10	10	9	10	10	9	9	9.7
Tatum (S)	10	10	9	10	10	9	10	10	9	9.7

Table 3. Segregation of resistance to isolate F18 in the F_2 population from crossing resistant barley accessions with the susceptible cultivar Tatum

Resistant accession	Reaction of parents		Numbers of resistant/susceptible and intermediate response of plants in F_2 populations		
	P1 resistant	P2 susceptible	Resistant	Intermediate response	Susceptible
k-5900	3.0	10.0	8	7	38
k-8829	3.5	10.0	4	7	54
k-8877	2.0	9.0	28	11	25
k-14936	3.0	10.0	17	14	28
k-18552	2.0	9.0	31	8	24
k-30341	4.0	9.0	13	11	41

Note. Reaction of parents based on the 10-point scale of A. Tekauz (1985).

Table 4. Positions of SNP markers associated with juvenile resistance to *Ptt* on chromosomes 4H, 3H, and 6H

Marker	Chromosome	Position on the genetic and physical maps of the barley genome		Reference
JHI-Hv50k-2016-237471	4H	50.00	58942545	Novakazi et al., 2019
JHI-Hv50k-2016-237347		50.00	57098155	
JHI-Hv50k-2016-237684		50.20	60114530	
JHI-Hv50k-2016-237839		50.30	61872363	
JHI-Hv50k-2016-241935		50.40	67692302	
JHI-Hv50k-2016-237924		50.99	63065507	
SCRI_RS_170494		52.00	448603913	Richards et al., 2017
SCRI_RS_181886		52.20	449611912	
SCRI_RS_153184		97.00	584761404	Amezrou et al., 2018
SCRI_RS_154517		2.00	2772827	Wonneberger et al., 2017
JHI-Hv50k-2016-183463	3H-2	54.53	491373166	Novakazi et al., 2019
JHI-Hv50k-2016-183478		54.53	491381651	
JHI-Hv50k-2016-183207		52.46	490244247	
JHI-Hv50k-2016-165152	3H-1	45.82	73225203	
JHI-Hv50k-2016-166392		47.1	130790360	
JHI-Hv50k-2016-166356		47.2	119627830	
JHI-Hv50k-2016-391380	6H-2	52.2	125903650	
BOPA2_12_31178	6H-3	55.00	378210479	Manninen et al., 2006

Note. The genetic position is determined from the current version of MorexV3 (Mascher et al., 2021).

a 548 bp fragment is formed; in the presence of the C allele, which does have a restriction site, fragments of 197 bp and 351 bp are formed. Using the restriction endonuclease RsaI for the JHI-Hv50k-2016-237924 marker, two alleles can also be distinguished: the G allele is cut into fragments of 177, 105, 38, and 55 bp, and the C allele is cut into fragments of 177, 29, 76, 38, and 55 bp. Four SNP markers on chromosome 4H were converted to KASP marker format (Table S4).

Fragment analysis

The results of testing the developed primers on the parental accessions are presented in Table S5. The criterion for a promising marker was the presence of amplification products for resistant barley genotypes and the absence of them for susceptible ones, or vice versa. Polymorphism for the presence of amplification products in certain barley genotypes was detected for markers JHI-Hv50k-2016-237924 (4H-924), SCRI_RS_153184 (4H-184), and SCRI_RS_181886 (4H-886). Figure S1 shows an example of polymorphism detection in the parental accessions using the SCRI_RS_181886 marker. The presence of the amplification product in both resistant and susceptible barley genotypes was detected using the primers of the remaining eight markers.

Fragment length analysis of marker amplification products after restriction enzyme treatment

Restriction analysis of the amplification products revealed polymorphism for markers JHI-Hv50k-2016-237684 and JHI-Hv50k-2016-237924 (Table 5): marker JHI-Hv50k-2016-237684 (NruI restriction enzyme): two fragments (351 and 197 bp) were detected in four resistant accessions – k-8829, k-8877, k-14936, and k-30341. The amplification

Table 5. Results of detection of polymorphic restriction fragments of marker amplification products on chromosome 4H in parental accession

Markers	Primers	Restriction enzyme	Restriction fragments of amplification products in barley accessions (bp)						
			k-5900	k-8829	k-8877	k-14936	k-18552	k-30341	Tatum
JHI-Hv50k-2016-237684	4H-684	NruI	548	351, 197	351, 197	351, 197	548	351, 197	548
JHI-Hv50k-2016-237924	4H-924	RsaI	177, 105, 55, 38	177, 76, 56, 29, 38	177, 76, 56, 29, 38	177, 76, 56, 29, 38	177, 105, 55, 38	177, 76, 56, 29, 38	177, 105, 55, 38

Table 6. Polymorphism of KASP markers on chromosome 4H in parental accessions

Barley accessions	4H-471	4H-839	4H-935	4H-347
Tatum	CC	GG	GG	CC
k-8829	GG	AA	AA	AA
k-14936	GG	AA	AA	AA
k-8877	AA	AA	AA	AA

product of the susceptible Tatum cultivar and accessions k-5900 and k-18552 was 548 bp (Fig. S2); marker JHI-Hv50k-2016-237924 (restriction enzyme RsaI): five fragments were detected in resistant accessions k-8829, k-8877, k-14936, and k-30341, while four fragments were detected in the susceptible cultivar Tatum and accessions k-5900 and k-18552 (Fig. S2). Thus, to study the co-segregation of MM and the resistance trait in segregating barley populations, markers JHI-Hv50k-2016-237684 and JHI-Hv50k-2016-237924 and the corresponding restriction enzymes NruI and RsaI were used to digest the amplification products of both markers.

The results of the study of polymorphism for KASP markers on chromosome 4H of the parental accessions are presented in Table 6. Allelic polymorphism of resistant samples and the susceptible variety Tatum was detected for four markers: JHI-Hv50k-2016-237471 (4H-471), JHI-Hv50k-2016-237839 (4H-839), JHI-Hv50k-2016-241935 (4H-935), and JHI-Hv50k-2016-237347 (4H-347), which were used to study the co-segregation of the resistance phenotype and the marker genotype.

Molecular markers polymorphism on chromosome 3H in parental accessions used for crossing

According to GWAS data, resistance in accessions k-8829, k-8877, k-18552, and k-30341 was also associated with the 3H-1 and 3H-2 loci (Tables 1 and 4). We previously validated KASP markers for these loci on chromosome 3H in segregating populations, which were highly effective (over 80 %) in the CLS, Morex, and Fox barley genotypes carrying the major resistance gene *qPttCLS* (Afanasenko et al., 2022). These KASP markers in intervals 45.82–47.2 and 52.46–54.53 cM were used to analyze segregating populations obtained from crossing the Tatum cultivar with *Ptt*-resistant accessions (Table S6).

For fragment analysis of marker amplification products on chromosome 3H, the primers proposed in the article by O. Afanasenko et al. (2022) were also used (Table S7). In fragment analysis, polymorphism for the presence of amplification products in three resistant accessions (k-8877, k-5900, and k-8829) and the susceptible Tatum cultivar was detected only for the JHI-Hv50k-2016-166356 marker, which was used to analyze the segregating populations. For the remaining six markers, no polymorphism was observed between the resistant accessions and the susceptible cultivar Tatum.

When using KASP markers, polymorphism for SNP haplotypes was detected only for one resistant accession, k-14936 (GG), and the susceptible cultivar Tatum (CC), and only for marker JHI-Hv50k 2016-165152. KASP markers JHI-Hv50k-2016-166392, JHI-Hv50k-2016-183463, and JHI-Hv50k-2016-183207 exhibited heterozygous SNP haplotypes, making them unsuitable for labeling accessions (Table S8).

Parental accession polymorphism on chromosome 6H

Resistance in the studied accessions was also associated with several loci on chromosome 6H (Tables 1 and 4). Previously, using double haploid mapping populations, the major *RPt5* gene, determining high-quality resistance to *Ptt*, was identified on chromosome 6H in the position 52.00–55.03 cM in barley accessions CI9819, CI5791, and k-23874 (Manninen et al., 2006; Potokina et al., 2010; Koladia et al., 2017). The GWAS results (Novakazi et al., 2019) confirmed the presence of resistance loci in this interval, the markers of which were combined into four groups, depending on their location on the genetic and physical maps of barley (Tables 1 and 4). In previous studies, using barley accessions CI9819, CI5791 and k-23874 as tester genotypes, we demonstrated the effectiveness of two markers of resistance loci on chromosome 6H, which were used in this study (Table 4): JHI-Hv50k-2016-391380 (6H-380) at position 52.20 (6H-2) and BOPA2_12_31178 (6H-178) at position 55.03 cM (6H-3) (unpublished data). A CAPS marker was developed using the 6H-380 marker using the HindIII restriction enzyme. Allele A: restriction site → two fragments of 282 and 254 bp; allele G: undigested fragment of 536 bp.

Primers for markers associated with resistance to *P. teres* f. *teres* on chromosome 6H are given in Table S9. Both markers showed polymorphism for the parental accessions. Marker 6H-380: HindIII restriction enzyme did not digest the amplification product of marker 6H-380 in all six resistant genotypes (one fragment), but digested it in the susceptible cultivar Tatum (two fragments). Marker 6H-178 revealed polymorphism between resistant barley accessions k-5900, k-8829 and the susceptible cv. Tatum (Fig. S3).

Study of co-segregation of resistance to *Ptt* and molecular markers in segregating populations

To study co-segregation for resistance to *Ptt* and identified polymorphic markers on chromosomes 4H, 3H, and 6H, 10 resistant and 10 susceptible lines were selected in each

hybrid population. In some cross combinations, the analyzed sample of hybrid plants was expanded to 40 (20 resistant and 20 susceptible) to confirm segregation results.

Fragment analysis using polymorphic molecular markers

The results of the correlation between the F_2 plant resistance phenotype and the presence/absence of MM amplification products are presented in Table 7. A significant association between the marker and plant resistance using the χ^2 criterion was found for marker 3H-56 in the Tatum × k-8829 combination, but Fisher's exact test did not confirm the significance of the association (Table 7).

A significant association between the F_2 plant resistance trait and marker 4H-924 was found in the Tatum × k-8877 cross, as well as with marker 4H-886 (F × Rin1) in the Tatum × k-5900. For the remaining MMs studied, despite polymorphism in the parental crosses, no significant association with resistance was found in the segregating populations. The obtained data indicate the presence of a QTL for *Ptt* resistance on chromosome 4H in accessions k-8829, k-8877, and k-5900.

Analysis of the correlation between the resistance phenotype of F_2 plants and the restriction products of CAPS markers

Two markers on chromosome 4H were found to be polymorphic in the sizes of restriction products in the parental components of the crosses: 4H-684 NruI and 4H-924 RsaI. Figures S4 and S5 demonstrate the polymorphism of the restriction fragments of marker 4H-684 by endonuclease NruI in the progeny of the crosses Tatum × k-8829 and Tatum × k-8877. No statistically significant association was found between the genotype and phenotype of disease resistance ($p > 0.05$) (Table S10). The significant association of features identified for the 4H-924 marker in fragment analysis was absent when using the CAPS marker 4H-924 RsaI.

On chromosome 6H, polymorphism in the sizes of restriction products was detected for the JHI-Hv50k-2016-391380 HindIII (6H-380 HindIII) marker in the susceptible cv. Tatum and the resistant accessions k-18552, k-8877, k-14936, k-8829, k-5900, and k-30341. A significant correspondence ($p < 0.05$) between the genotype and phenotype of resistance to *Ptt* in F_2 plants was found in the combinations Tatum × k-18552, Tatum × k-8877, Tatum × k-8829, and Tatum × k-5900 (Table 8, Figs. S6–S8). Thus, accessions k-18552, k-8877, k-8829, and k-5900 have a resistance QTL on chromosome 6H at position 52.2 cM.

KASP genotyping results

In three cross combinations, the resistant parents k-8829, k-14936, and k-8877 and the susceptible cv. Tatum were polymorphic for the SNP haplotypes of MM on chromosome 4H. In hybrid combinations involving the accession k-8829, the diagnostic efficiency was greater than 0.5 (0.6) for marker 4H-471 alone; for the remaining markers, this indicator was <0.5. In the Tatum × k-8877 combination, the diagnostic efficiency of markers 4H-471, 4H-935, and 4H-347 was 0.71–0.73 (Table S11).

For the KASP marker JHI-Hv50k-2016-165152 on chromosome 3H, allelic polymorphism was detected in cv. Tatum (CC) and accession k-14936 (GG). In the segregating population from their cross, no correlation was found between plant resistance and haplotypes (Table S11).

A QTL on chromosome 6H, detected by the HindIII marker 6H380, determines high resistance to *Ptt* in four barley accessions and masks the presence of other QTLs (Tables S12–S14). Therefore, the resistance trait does not correlate with the other

Table 7. Reliability of the association between the *Ptt* resistance and molecular markers polymorphic on the parental accessions (fragment analysis)

Crosses of cv. Tatum with resistant accessions	Marker (primers)	χ^2	*p*-value
k-8829	4H-184 (F × R)	2.50	> 0.05
	4H-347 (F × R1)	**4.62**	**< 0.05**
	6H-178 (F × R)	2.50	> 0.05
	3H-56 (F × R)	**4.29**	**< 0.05**
k-8877	**4H-924 (Fin1 × Rout)**	**7.07**	**< 0.05**
	4H 886 (F × R)	0.00	> 0.05
	3H-56 (F × R)	3.14	> 0.05
k-14936	4H-924 (Fin1 × Rout)	0.19	> 0.05
	4H 886	2.33	> 0.05
k-30341	4H-924 (Fin2 × Rout)	0.40	> 0.05
	4H 886 (F × Rin1)	0.00	> 0.05
k-5900	**4H-886 (F × Rin1)**	**4.41**	**< 0.05**
	6H-178 (F × R)	2.61	> 0.05
	3H-56 (F × R)	0.01	> 0.05
k-18552	4H-924	0.02	> 0.05

Note. The relationship between the resistance phenotype and the marker is significant at $p < 0.05$, highlighted in bold.

Table 8. Correspondence between phenotypic resistance and restriction products of the CAPS marker JHI-Hv50k-2016-391380 (HindIII) on chromosome 6H in F_2 from a cross between resistant accessions and the susceptible cv. Tatum

Crosses of cv. Tatum with resistant accessions	Size of restriction fragment	Genotype	R	S	Calculated value χ^2	Probability *p*
k-18552	**536**	**GG (R)**	**4**	**0**	**34.693**	**< 0.001**
	536, 282, 254	**AG**	**14**	**0**		
	282, 254	**AA (S)**	**0**	**10**		
k-8877	**536**	**GG (R)**	**13**	**0**	**21.0115**	**< 0.05**
	536, 282, 254	**AG**	**6**	**13**		
	282, 254	**AA (S)**	**1**	**8**		
k-14936	536	GG (R)	6	3	1.54553	> 0.05
	536, 282, 254	AG	9	12		
	282, 254	AA (S)	3	4		
k-8829	**536**	**GG (R)**	**5**	**1**	**10.37037**	**< 0.05**
	536, 282, 254	**AG**	**0**	**6**		
	282, 254	**AA (S)**	**0**	**2**		
k-5900	**536**	**GG (R)**	**3**	**1**	**6.831019**	**< 0.05**
	536, 282, 254	**AG**	**5**	**10**		
	282, 254	**AA (S)**	**0**	**7**		
30341	536	GG (R)	2	2	0.31111	> 0.05
	536, 282, 254	AG	5	4		
	282, 254	AA (S)	2	3		

Note. R – resistance, S – susceptibility. The association between the resistance phenotype and the marker is significant at $p < 0.05$, highlighted in bold. The minimum table value of χ^2 at significance level α of 0.05 was 5.991 for all barley samples.

studied MMs. The absence/presence of MMs in susceptible plants of a particular hybrid combination is a different matter. For example, in the class of susceptible F_2 plants in the Tatum × k-8877 combination, the homozygotes of the susceptible parent for the 4H-924 RsaI marker were 100 %, while for the 4H-684 NruI marker, six out of ten plants were homozygous and two heterozygous for susceptibility. Similar results were obtained for the KASP markers: for all four markers, six out of ten susceptible plants were homozygous for the susceptible parent's allele, and three, heterozygous. In this cross, the four markers on chromosome 4H had a diagnostic efficiency of more than 0.7 (Table S12).

Susceptible plants predominated in the k-8829 × Tatum combination. The resistance phenotype split was 4 (R):7 (MR):54 (S), so only these four resistant plants and ten susceptible plants were included in the analysis. For the CAPS markers 6H380 HindIII and 4H-684NruI, as well as the KASP marker 4H-471, all heterozygous plants were associated with susceptibility, suggesting a recessive inheritance pattern. For the 4H-924RsaI marker, all susceptible plants had the genotype of the susceptible parent (Table S13).

In the k-5900 × Tatum combination, in addition to the proven significant correlation between the CAPS marker 6H380 HindIII and marker 4H-886 (F × Rin1), fragment analysis shows no obvious correspondence between the presence/absence of markers 6H-178 and 3H-56 in the group of susceptible plants (Table S14).

Discussion

Currently, 103 loci associated with juvenile and adult resistance to *Ptt* and a large number of MMs have been identified using GWAS technology and mapping in double haploid populations (Steffenson et al., 1996; Richter et al., 1998; Friesen et al., 2006; Manninen et al., 2006; Yun et al., 2006; Grewal et al., 2008, 2012; Cakir et al., 2011; Liu et al., 2011; Berger et al., 2013; König et al., 2013, 2014; O'Boyle et al., 2014; Afanasenko et al., 2015, 2022; Wang et al., 2015; Koladia et al., 2017; Richards et al., 2017; Wonneberger et al., 2017; Amezrou et al., 2018; Martin et al., 2018; Dinglasan et al., 2019; Novakazi et al., 2019; Rozanova et al., 2019; Rehman et al., 2025). However, there are only a few publications presenting the results of validation of *Ptt* resistance QTL markers identified in GWAS in a different genetic background (Grewal et al., 2010; Afanasenko et al., 2022).

Breeding barley for resistance to *Ptt* requires effective QTL markers controlling both qualitative and quantitative resistance. To validate the SNP markers of *Ptt* resistance loci on chromosomes 3H, 4H, and 6H identified in GWAS (Richards et al., 2017; Amezrou et al., 2018; Novakazi et al., 2019), barley accessions, the SNP haplotypes of which were associated with several *Ptt* resistance loci (Table 1), were selected. These accessions were resistant to a wide range of *Ptt* pathotypes in the juvenile phase (Table 2) and against a provocative background (late sowing) in the adult plant phase in the field (unpublished data).

Analysis of segregation for juvenile resistance in F_2 from crosses of these accessions with the susceptible cultivar Tatum indicated complex inheritance of the trait, confirming the GWAS results. A distinctive feature of resistance assessment in segregating populations to *Ptt*, as well as to other hemibiotrophic pathogens, is the presence of a group of plants with intermediate reactions (scores 5.0–5.9). In the presence of several QTLs in the parental components of the cross, intermediate plant reactions are due to the presence of recombinants with different numbers of genetic determinants of resistance and different gene interactions.

Several *Ptt* resistance loci are known on chromosome 4H in the following intervals: 1.13 cM (Grewal et al., 2008); 3.31 cM (Afanasenko et al., 2015; Wonneberger et al., 2017); 47.27–52.69 cM (Richards et al., 2017; Novakazi et al., 2019); 64.3 cM (Steffenson et al., 1996); 77.0 cM (Martin et al., 2018); 97.66 cM (Amezrou et al., 2018); 113.1 cM (Martin et al., 2018); 121–123 cM (König et al., 2014); 150–175 cM (Friesen et al., 2006).

In this study, we examined markers of a locus located on chromosome 4H in the position of 50.0–50.4 cM, which we previously identified as a result of GWAS (Novakazi et al., 2019), as well as markers of loci identified by other researchers in the range of 52–52.2 cM (Richards et al., 2017), 97.0–97.20 cM (Amezrou et al., 2018) and at the 2.0 cM position (Wonneberger et al., 2017). The choice of MM for studying co-segregation in segregating populations was based on the correlation of certain SNP haplotypes of markers identified in GWAS with the resistance phenotype. So, four "peak" markers associated with resistance to *Ptt* isolate No. 13 were identified on chromosome 4H in 98 barley accessions (average damage score of 3.54) (Table 9). The CCAT SNP haplotypes of these four markers were associated with resistance, while the GTGC SNP haplotypes of the same markers were associated with susceptibility (average damage score of 5.45) in 347 barley accessions (data kindly provided by F. Novakazi). However, among the 98 accessions, 16 with the CCAT haplotype were susceptible to the pathogen, and among the 347 accessions, 147 were resistant, although they had the GTGC haplotype. These data indicate that, despite the association of certain SNP marker haplotypes with resistance, random combinations of the same SNP haplotypes in susceptible accessions and vice versa are possible, which suggests the possibility of a false assumption about the presence of resistance-associated loci in certain accessions identified in GWAS.

For most of the studied resistance loci markers on chromosomes 3H, 4H, and 6H, no polymorphism was detected in the MMs between the parental components of the cross (resistant accession × susceptible cultivar Tatum).

Of the 10 markers and 28 different primer combinations of these markers on chromosome 4H, only three were polymorphic in fragment PCR analysis, two were polymorphic when used as CAPS markers, and one KASP marker was polymorphic between the parental accessions in only one combination. A significant association between the F_2 plant resistance trait and marker 4H-924 was found in the Tatum × k-8877 cross, as well as with marker 4H-886 (F × Rin1) in the Tatum × k-5900 cross. The obtained data indicate the possibility of using these markers in breeding if the k-8877 and k-5900 accessions are used as donors of resistance to *Ptt*.

In previous studies, determining the effectiveness of SNP markers of the resistance locus on chromosome 3H in the interval 46.29–54.3 cM using KASP genotyping revealed five markers that were 100–80 % effective in the double haploid population and in two segregating populations and were associated with resistance in the CLS, Morex, Fox cultivars, and accession k-21578 (Afanasenko et al., 2022). It was shown that the resistance locus on chromosome 3H contains at least two QTLs controlling resistance to *Ptt* in the intervals of 46.0–48.44 cM and 51.27–54.8 cM (Afanasenko et al., 2022). In this study, the same markers, located in the positions of 45.82–47.2 cM (3H-1) and 52.46–54.53 cM (3H-2), were used to examine segregating barley populations (Table 4). Of the seven primer–marker pairs studied on chromosome 3H, only one marker, JHI-Hv50k-2016-166356 (3H-56), detected polymorphism in five resistant accessions with the Tatum cultivar. However, only one cross, Tatum × k-8829, revealed a significant association between the marker and plant resistance. No correlation was found between these markers and the resistance phenotype using KASP genotyping.

Resistance in the studied accessions was also associated with several loci on chromosome 6H. Previously, using double haploid mapping populations on chromosome 6H in the interval of 52.00–55.03 cM in barley accessions CI9819, CI5791, and k-23874, a large RPt5 gene was identified that determines high-quality resistance to *Ptt* (Manninen et al., 2006; Potokina et al., 2010; Koladia et al., 2017). As a result of GWAS (Novakazi et al., 2019), resistance loci were also identified in this position, the markers of which were combined into four groups, depending on their location on the genetic and physical maps of barley (Table 4). Previously, using barley accessions CI9819, CI5791 and k-23874 as test genotypes, we demonstrated the effectiveness of two markers of resistance loci on chromosome 6H, which were used in this study: JHI-Hv50k-2016-391380 HindIII (6H-380) at position 52.20 (6H-2) and BOPA2_12_31178 (6H-178) at position 55.03 cM (6H-3) (unpublished data).

Table 9. Mean infection responses of barley accessions with defined SNP haplotypes of four markers on chromosome 4H after inoculation with isolate *P. teres* f. *teres* No. 13

JHI-Hv50k-2016-237471, 50.0 cM	JHI-Hv50k-2016-237684, 50.2 cM	JHI-Hv50k-2016-237839, 50.3 cM	JHI-Hv50k-2016-241935, 50.4 cM	Number of accessions	Mean infection response
C	C	A	C	4	3.66
C	C	A	T	98	3.54
G	T	G	C	347	5.45

A significant association ($p < 0.05$) between the JHI-Hv50k-2016-391380 HindIII (6H-380) marker and the *Ptt* resistance phenotype in F_2 plants was found in combinations from crossing the susceptible cv. Tatum with the k-5900 (Turkmenistan), k-8829 (Italy), k-8877 (Spain), and k-18552 (Australia) accessions. These data indicate the possibility of using these accessions as resistance donors and the CAPS marker JHI-Hv50k-2016-391380 HindIII in marker-assisted selection (MAS).

It is known that QTLs on chromosome 6H at the studied locus control high resistance in barley genotypes (Afanasenko et al., 1998; Manninen et al., 2006; Koladia et al., 2017). The presence of a highly significant association of the resistance phenotype of F_2 plants in four cross combinations involving the k-5900, k-8829, k-8877, and k-18552 accessions of the SNP haplotype of the 6H-380 HindIII marker masks the presence of other QTLs. However, in the class of susceptible plants in a given cross combination, the markers must correspond to the genotype of the susceptible parent. For example, the KASP markers 4H-471, 4H-347, and 4H-935 and the CAPS marker 4H-924 RsaI, in combination with k-8877, can be effective for culling susceptible plants. However, when using the entire plant accession, no significant correlation was found between the resistance phenotypes and the genotypes of these markers.

Conclusion

Therefore, the absence or presence of amplification products of a polymorphic marker on the parental components of a cross in resistant F_2 plants with polygenic inheritance does not prove that there is no correlation between the marker and the resistance trait, as the presence of a major resistance gene masks the expression of other QTLs.

New donors of resistance to *Ptt* were identified: accessions k-5900 (Turkmenistan), k-8829 (Italy), k-8877 (Spain) and k-18552 (Australia), in which the QTL on chromosome 6H is located at position 52.2 cM, 125,903,650 bp. Accessions k-8877 and k-5900 also have a QTL on chromosome 4H in the position of 50.00–50.99 cM, 57,098,155–63,065,507 bp, and accession k-8829 has a QTL on chromosome 3H at position 47.2 cM, 1,196,27,830 bp. Resistance donors and validated MMs with proven efficacy can be used in MAS to develop barley cultivars resistant to net blotch.

References

Afanasenko O., Jalli M., Pinnschmidt H., Filatova O., Platz G. Development of an international standard set of barley differential genotypes for *Pyrenophora teres* f. *teres*. *Plant Pathol*. 2009;58(4):665-676. doi 10.1111/j.1365-3059.2009.02062.x

Afanasenko O.S., Koziakov A.V., Hedlay P.E., Lashina N.M., Anisimova A.V., Manninen O., Jalli M., Potokina E.K. Mapping of the loci controlling the resistance to *Pyrenophora teres* f. *teres* and *Cochliobolus sativus* in two double haploid barley populations. *Russ J Genet Appl Res.* 2015;5(3):242-253. doi 10.1134/S2079059715030028

Afanasenko O., Rozanova I., Gofman A., Lashina N., Novakazi F., Mironenko N., Baranova O., Zubkovich A. Validation of molecular markers of barley net blotch resistance loci on chromosome 3H for marker-assisted selection. *Agriculture*. 2022;12(4):439. doi 10.3390/agriculture12040439

Amezrou R., Verma R.P.S., Chao S., Brueggeman R.S., Belgadi L., Arbaoui M., Rehman S., Gyawali S. Genome-wide association studies of net form of net blotch resistance at seedling and adult plant stages in spring barley collection. *Mol Breed*. 2018;38:58. doi 10.1007/s11032-018-0813-2

Cakir M., Gupta S., Li C., Hayden M., Mather D.E., Ablett G.A., Platz G.J., Broughton S., Chalmers K.J., Loughman R., Jones M.G.K., Lance R.C.M. Genetic mapping and QTL analysis of disease resistance traits in the barley population Baudin × AC Metcalfe. *Crop Pasture Sci*. 2011;62(2):152-161. doi 10.1071/cp10154

Clare S.J., Çelik Oğuz A., Effertz K., Sharma Poudel R., See D., Karakaya A., Brueggeman R.S. Genome-wide association mapping of *Pyrenophora teres* f. *maculata* and *Pyrenophora teres* f. *teres* resistance loci utilizing natural Turkish wild and landrace barley populations. *G3*. 2021;11(11):jkab269. doi 10.1093/g3journal/jkab269

Dinglasan E., Hickey L., Ziems L., Fowler R., Anisimova A., Baranova O., Lashina N., Afanasenko O. Genetic characterization of resistance to *Pyrenophora teres* f. *teres* in the International barley differential Canadian lake Shore. *Front Plant Sci*. 2019;10:326. doi 10.3389/fpls.2019.00326

Friesen T.L., Faris J.D., Lai Z., Steffenson B.J. Identification and chromosomal location of major genes for resistance to *Pyrenophora teres* in a doubled-haploid barley population. *Genome*. 2006;49(7):855-859. doi 10.1139/g06-024

Grewal T.S., Rossnagel B.G., Pozniak C.J., Scoles G.J. Mapping quantitative trait loci associated with barley net blotch resistance. *Theor Appl Genet*. 2008;116(4):529-539. doi 10.1007/s00122-007-0688-9

Grewal T.S., Rossnagel B.G., Scoles G.J. Mapping quantitative trait loci associated with spot blotch and net blotch resistance in a doubled-haploid barley population. *Mol Breed*. 2012;30(1):267-279. doi 10.1007/s11032-011-9616-4

Gupta S., Li C.D., Loughman R., Cakir M., Platz G., Westcott S., Bradley J., Broughton S., Lance R. Quantitative trait loci and epistatic interactions in barley conferring resistance to net type net blotch (*Pyrenophora teres* f. *teres*) isolates. *Plant Breed*. 2010;129(4):362-368. doi 10.1111/j.1439-0523.2009.01716.x

Jatayev S., Kurishbayev A., Zotova L., Khasanova G., Serikbay D., Zhubatkanov A., Botayeva M., Zhumalin A., Turbekova A., Soole K., Langridge P., Shavrukov Yu. Advantages of Amplifluor-like SNP markers over KASP in plant genotyping. *BMC Plant Biol*. 2017;17(2):254. doi 10.1186/s12870-017-1197-x

Koladia V.M., Faris J.D., Richards J.K., Brueggeman R.S., Chao S., Friesen T.L. Genetic analysis of net-form net blotch resistance in barley lines CIho 5791 and Tifang against a global collection of *P. teres* f. *teres* isolates. *Theor Appl Genet.* 2017;130(1):163-173. doi 10.1007/s00122-016-2801-4

König J., Perovic D., Kopahnke D., Ordon F. Development of an efficient method for assessing resistance to the net type of net blotch (*Pyrenophora teres* f. *teres*) in winter barley and mapping of quantitative trait loci for resistance. *Mol Breed*. 2013;32:641-650. doi 10.1007/s11032-013-9897-x

König J., Perovic D., Kopahnke D., Ordon F. Mapping seedling resistance to net form of net blotch (*Pyrenophora teres* f. *teres*) in barley using detached leaf assay. *Plant Breed*. 2014;133(3):356-365. doi 10.1111/pbr.12147

Lashina N.M., Mironenko N.V., Zubkovich A.A., Afanasenko O.S. Juvenile resistance of barley varieties and samples to net-, spot- and hybrid (net × spot) forms of *Pyrenophora teres*. *Mycol Phytopathol*. 2023;57(1):48-59. doi 10.31857/S0026364823010099 (in Russian)

Liu Z., Ellwood S.R., Oliver R.P., Friesen T.L. *Pyrenophora teres*: profile of an increasingly damaging barley pathogen. *Mol Plant Pathol.* 2011;12(1):1-19. doi 10.1111/j.1364-3703.2010.00649.x

Manninen O.M., Jalli M., Kalendar R., Schulman A., Afanasenko O., Robinson J. Mapping of major spot-type and net-type net-blotch resistance genes in the Ethiopian barley line CI 9819. *Genome*. 2006;49(12):1564-1571. doi 10.1139/g06-119

Martin A., Platz G.J., de Klerk D., Fowler R.A., Smit F., Potgieter F.G., Prins R. Identification and mapping of net form of net blotch re-

sistance in South African barley. *Mol Breed.* 2018;38(5):53. doi 10.1007/s11032-018-0814-1

Mascher M., Wicker T., Jenkins J., Plott C., Lux T., Koh C.S., Ens J., … Šimková H., Moscou M.J., Grimwood J., Schmutz J., Stein N. Long-read sequence assembly: A technical evaluation in barley. *Plant Cell*. 2021;33(6):1888-1906. doi 10.1093/plcell/koab077

Novakazi F., Afanasenko O., Anisimova A., Platz G.J., Snowdon R., Kovaleva O., Zubkovich A., Ordon F. Genetic analysis of a worldwide barley collection for resistance to net form of net blotch disease (*Pyrenophora teres* f. *teres*). *Theor Appl Genet.* 2019;132(9):2633-2650. doi 10.1007/s00122-019-03378-1

O'Boyle P.D., Brooks W.S., Barnett M.D., Berger G.L., Steffenson B.J., Stromberg E.L., Maroof M.A.S., Liu S.Y., Griffey C.A. Mapping net blotch resistance in 'Nomini' and CIho 2291 barley. *Crop Sci*. 2014;54(6):2596-2602. doi 10.2135/cropsci2013.08.0514

Pandis N. The chi-square test. *Am J Orthodontics Dentofacial Orthopaedics*. 2016;150(5):898-899. doi 10.1016/j.ajodo.2016.08.009

Potokina E.K., Hedley P., Afanasenko O.S., Lashina N.M., Anisimova A.V., Kozyakov A.V., Yalli M., Manninen O. Mapping of QTL (Quantitative Trait Loci) determining resistance to net blotch of barley. In: Technologies for creating and using varieties and hybrids with group and complex resistance to pests in plant protection. St. Petersburg, 2010;229-236 (in Russian)

Rehman S., Al-Jaboobi M., Verma R.P.S., Sanchez-Garcia M., Visioni A. Genome-wide association mapping of net form net blotch resistance in barley at seedling and adult plant stages. *Front Agron*. 2025;7:1525588. doi 10.3389/fagro.2025.1525588

Richards J.K., Friesen T.L., Brueggeman R.S. Association mapping utilizing diverse barley lines reveals net form net blotch seedling resistance/susceptibility loci. *Theor Appl Genet.* 2017;130(5):915-927. doi 10.1007/s00122-017-2860-1

Richter K., Schondelmaier J., Jung C. Mapping of quantitative trait loci affecting Drechslera teres resistance in barley with molecular markers. *Theor Appl Genet.* 1998;97:1225-34. doi 10.1007/s00122 0051014

Rozanova I.V., Lashina N.M., Mustafin Z.S., Gorobets S.A., Efimov V.M., Afanasenko O.S., Khlestkina E.K. SNPs associated with barley resistance to isolates of *Pyrenophora teres* f. *teres*. *BMC Genomics*. 2019;20:292. doi 10.1186/s12864-019-5623-3

Steffenson B.J., Hayes P.M., Kleinhofs A. Genetics of seedling and adult plant resistance to net blotch (*Pyrenophora teres* f. *teres*) and spot blotch (*Cochliobolus sativus*) in barley. *Theor Appl Genet.* 1996;92:552-558. doi 10.1007/bf00224557

Tekauz A. A numerical scale to classify reactions of barley to *Pyrenophora teres. Can J Plant Pathol*. 1985;7(2):181-183. doi 10.1080/ 07060668509501499

Wonneberger R., Ficke A., Lillemo M. Identification of quantitative trait loci associated with resistance to net form net blotch in a collection of Nordic barley germplasm. *Theor Appl Genet*. 2017;130: 2025-2043. doi 10.1007/s00122-017-2940-2

Yun S.J., Gyenis L., Bossolini E., Hayes P.M., Matus I., Smith K.P., Steffenson B.J., Tuberosa R., Muehlbauer G.J. Validation of quantitative trait loci for multiple disease resistance in barley using advanced backcross lines developed with a wild barley. *Crop Sci*. 2006;46(3):1179-1186. doi 10.2135/cropsci2005.08-0293

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

Received August 8, 2025. Revised September 18, 2025. Accepted September 18, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1246-1254

doi 10.18699/vjgb-25-134

Identification of *CtE1* gene nucleotide polymorphisms and development of SNP-based KASP markers in guar (*Cyamopsis tetragonoloba* (L.) Taub.)

L. Criollo Delgado [1*], D. Zewude [1*], D.S. Karzhaev [2], D.E. Polev [3], E.K. Potokina [1]

[1] Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia
[2] Institute of Forest and Environmental Management, Saint-Petersburg State Forest Technical University, St. Petersburg, Russia
[3] Laboratory of Metagenomics Research, Saint-Petersburg Pasteur Institute, St. Petersburg, Russia

e.potokina@skoltech.ru

Abstract. Guar (*Cyamopsis tetragonoloba* (L.) Taub), is an important short-day legume crop, whose cultivation is limited at high latitudes due its photoperiod sensitivity, that negatively impacts flowering and maturation of this industrial-oriented crop. In its close relative, soybean, the *E1* gene has been highly associated with the regulation of flowering time under long-day conditions. In this study we investigated the natural diversity of the *E1* homologue gene (*CtE1*) in a panel of 144 guar accessions. For this purpose, the *CtE1* gene was amplified and sequenced using Illumina. As a result, five novel SNPs were identified in the 5′-untranslated region, coding region, and 3′-untranslated region of the *CtE1* gene. One non-synonymous SNP was located in the coding region causing a conservative Arg→Lys substitution. Based on the identified SNP, five KASP markers linked to polymorphism in the target gene were developed and tested in the guar collection. No significant associations were detected between discovered SNPs and available data on variability in flowering time or vegetation period length in the cohort of 144 accessions. These findings suggest that natural variation of the *CtE1* gene in the studied germplasm collection has minimal effect on flowering or maturation. The limited functional allelic diversity observed in the *CtE1* gene of guar compared to the *E1* gene in soybean likely reflects differences in their evolutionary histories, domestication bottlenecks, and selection pressures.

Key words: guar; photoperiod; flowering time; *CtE1* gene; GT-seq genotyping; SNP; KASP markers

For citation: Criollo Delgado L., Zewude D., Karzhaev D.S., Polev D.E., Potokina E.K. Identification of *CtE1* gene nucleotide polymorphisms and development of SNP-based KASP markers in guar (*Cyamopsis tetragonoloba* (L.) Taub.). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1246-1254. doi 10.18699/vjgb-25-134

Funding. The work was supported by the Russian Science Foundation (grant No. 24-26-00073).

Выявление нуклеотидного полиморфизма гена *CtE1* и разработка KASP-маркеров для гуара (*Cyamopsis tetragonoloba* (L.) Taub.)

Л. Криолло Дельгадо [1*], Д. Зевуде [1*], Д.С. Каржаев [2], Д.Е. Полев [3], Е.К. Потокина [1]

[1] Сколковский институт науки и технологий (Сколтех), Москва, Россия
[2] Институт леса и природопользования, Санкт-Петербургский государственный лесотехнический университет, Санкт-Петербург, Россия
[3] Лаборатория метагеномных исследований, Санкт-Петербургский научно-исследовательский институт эпидемиологии и микробиологии им. Пастера, Санкт-Петербург, Россия

e.potokina@skoltech.ru

Аннотация. Гуар (*Cyamopsis tetragonoloba* (L.) Taub) – бобовое растение короткого дня, возделывание которого в высоких широтах ограничено из-за его чувствительности к длинному фотопериоду, что негативно влияет на цветение и созревание этой новой индустриально значимой сельскохозяйственной культуры. У ее близкого родственника – сои – ген *E1* вовлечен в регуляцию процессов цветения в условиях длинного дня. В нашем исследовании естественное разнообразие гена *CtE1* (гомолога *E1*) проанализировано на выборке из 144 образцов гуара. Для этой цели ген *CtE1* был амплифицирован и секвенирован с использованием Illumina, в результате чего идентифицировано пять новых SNP в 5′- и 3′-нетранслируемых областях, а также в кодирующей части гена *CtE1*. На основе выявленных SNP разработаны и протестированы пять KASP-маркеров, связанных с полиморфизмом целевого гена в изученной коллекции гуара. Один несинонимичный SNP был локализован в кодирующей области; этот полиморфизм приводит к консервативной аминокислотной замене Arg→Lys в кодируемом белке. Значимой связи между обнаруженными SNP и имеющимися данными об изменчивости сроков начала цветения или продолжительности вегетационного периода в выборке из 144 образцов не обнаружено. Полученные результаты свидетельствуют о том, что естественный нуклеотидный полиморфизм

* Both authors contributed equally to the work.

гена *CtE1*, представленный в изученной коллекции образцов гуара, не оказывает существенного влияния на цветение и созревание растений. Ограниченное функциональное аллельное разнообразие, выявленное в гене *CtE1* гуара по сравнению с аллельным разнообразием гена *E1* у сои, вероятно, отражает различия в их эволюционной истории и различные направления искусственного отбора в процессе селекции.
Ключевые слова: гуар; фотопериод; время цветения; ген *CtE1*; GT-seq генотипирование; SNP; KASP-маркеры

Introduction

Guar (*Cyamopsis tetragonoloba* (L.) Taub), is an industrial-oriented short-day legume crop mainly cultivated for the production of guar gum (galactomannan) – a compound present in the seed endosperm of guar. This polysaccharide forms a viscous gel in water, and due to its thickening properties is widely used in several industrial sectors including oil and gas industry, cosmetics and food production (Benakanahalli et al., 2021). Currently, India and Pakistan are the main manufacturers and exporters of guar gum in the world market. However, there is growing interest in guar gum in many countries, and in the past two decades, guar rightfully gained the status of an important economical crop worldwide (Verma et al., 2025).

The main limiting factor for guar cultivation in Russia is its photoperiod sensitivity, which affects the timing of flowering and maturation of guar plants (Grigoreva et al., 2021 a, b). For the closely related legume soybean, loci that influence flowering and maturation under long-day conditions have been the subject of in-depth study for decades (Cao et al., 2017; Han et al., 2019). As a result, different alleles of genes involved in the photoperiod response were discovered, which are now used in breeding programs to adapt soybean varieties to diverse geographic regions and farming systems (Liu et al., 2020).

Among the genes identified to date as related to soybean vegetation period, *E1* has been recognized as the most critical regulator of flowering time in soybean (Watanabe et al., 2012; Xia et al., 2012), and as a key selection locus in breeding programs (Xia, 2017). These characteristics made *E1* the first and most significant target for CRISPR-Cas mutagenesis, aimed at developing new soybean germplasm with broad adaptability across different latitudes (Han et al., 2019).

Recently, an ortholog of the soybean *E1* gene was identified in the guar genome, showing 80 % identity at the coding peptide level and a similar intron–exon structure (Criollo Delgado et al., 2025). Like the other members from *E1* family genes, *CtE1* encodes a protein containing a putative bipartite nuclear localization signal (NLS) and a DNA-binding B3-like domain. This suggests that the genetic pathways underlying the basic mechanisms of photoperiod response may be similar in soybean and guar, and therefore the selection of photoperiod-insensitive guar varieties may follow the same pathway as in soybean.

In soybean, the legume-specific *E1* gene suppresses flowering of plants under long-day (LD) conditions, thus, non-synonymous mutations in this gene result in a dysfunctional polypeptide, promoting flowering of plants in high latitudes (Xu et al., 2015). At least 5 misfunctional alleles were described for the *E1* locus in soybean: *e1-fs* (frame shift), *e1-as* (amino acid substitution), *e1-b3a* (mutation in B3 domain), *e1-re* (retrotransposon insertion), *e1-p* (have SNPs or InDels in the coding sequence or 5′ upstream), and *e1-nl* (null) allele has a 130 kb deletion which includes the entire *E1* gene (Liu et al., 2020). Development of functional markers for *E1* polymorphisms has made significant contributions to both germplasm evaluation and marker-assisted selection (MAS) of soybean. Specifically, Kompetitive Allele Specific PCR (KASP) markers developed for SNPs at the *E1-E4* loci, allowed to reveal the most advantageous allele combinations for soybean cultivars propagated in various regions of China (Liu et al., 2020). In this regard, it might be relevant to assess the level of polymorphism of the *CtE1* gene in guar, represented in the natural intraspecific diversity of this legume crop, in order to identify alleles as possible targets for selection.

In the present paper we have evaluated nucleotide variability of *CtE1* gene using the diversity panel of 144 guar accessions of different geographic origin. We developed KASP markers for all SNPs detected and estimated association between the revealed haplotypes and phenotypic performance of the guar varieties.

Materials and methods

Plant material. A diversity panel consisting of 144 guar accessions, encompassing early- and late flowering/maturing varieties and landraces originating from India, Pakistan, United States, were described earlier (Grigoreva et al., 2021 a). In the same paper the performance of these accessions under field conditions in Krasnodar region (45°02′55″ N) in 2017 and 2018 was evaluated. Here, we used the field evaluation data to search for a link between alleles of *CtE1* gene and variation of the agrobiological traits of guar plants from different accessions. Two traits most relevant to the putative function of *CtE1* gene were considered: (1) flowering time defined as the number of days from sowing to flowering, recorded when 50 % of the plants in the accession have produced flower buds, and (2) length of vegetation period, which was calculated as the number of days from sowing to maturation (50 % of plants per accessions had mature pods).

Isolation of DNA, amplification of PCR and Sanger sequencing. As a first step, Sanger sequencing of *CtE1* was performed on several plants with contrasting maturation times to assess the presence of polymorphisms within a small but diverse panel of genotypes. For Sanger sequencing genomic DNA from one plant per each accession was extracted from the 7-days seedlings following the protocol described by Ivanova et al. (2008). For further high-throughput genotyping using Illumina, a bulk DNA (5–7 plants) per accession was analyzed.

The extracted DNA was stored at −20 °C and subsequently assessed for quality and integrity using 1.5 % agarose gel electrophoresis. DNA concentration and purity were measured using a NanoDrop spectrophotometer (Desjardins, Conklin, 2010). For Sanger sequencing genomic DNA was subjected to PCR using a pair of primers designed for the *CtE1* predicted sequence (Criollo Delgado et al., 2025) to amplify the 5′ untranslated region (5′ UTR), coding region, and 3′ untranslated region (3′ UTR). Information on the primer sequences is present on Table 1 (primers E1-F and E1-R).

Table 1. Primer pairs used for amplification of the *CtE1* gene in guar. Primers 1, 2 were used for Sanger sequencing. Primers 3–8 were included for Illumina sequencing purposes; adapter sequences are marked in color: forward (yellow) and reverse (green)

No.	Name of primers	Sequence
1	E1-F	GAGCCTCCATTCTCATTTCACAAAG
2	E1-R	CGACCAATAACAGTGTTGGCATAG
3	Guar_E1_P1_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCACGACAAAGGTGAAATG
4	Guar_E1_P1_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCCTCCTCCTTTGTTCTTCTC
5	Guar_E1_P2_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGAAGAACAAAGGAGGAGGAG
6	Guar_E1_P2_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTGATTCCACTTCCCAATAAG
7	Guar_E1_P3_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGGAAGGAACGCCGATTAG
8	Guar_E1_P3_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGTGTTGGCATAGTGATAGAA

PCR amplification was performed using primers E1-F and E1-R to generate 803 bp product. The PCR reaction mix (25 µL) consisted of 1 µl genomic DNA, 1× PCR buffer, 3 mM MgCl2, 0.4 µM of each primer, 100 µM dNTPs, and 2.5 unit of TaqDNA polymerase. The initial denaturation was performed at 94 °C for 2 min; followed by 30 cycles of denaturation at 98 °C for 30 s, annealing at 54 °C for 1 min, and extension at 72 °C for 45 s; and with a final extension at 72 °C for 10 min. The PCR products were purified from PCR mix using QIAquick PCR Purification Kit (Qiagen). The purified samples were submitted to a commercial sequencing facility (Evrogen, Moscow) for further processing.

Primer design for Illumina sequencing. High-quality DNA samples of 144 guar accessions were used for PCR with specifically designed primers. Three pairs of primers were designed to amplify three overlapping regions of *CtE1* gene covering the same region as for Sanger sequencing, but overhangs were included in the primer's sequences to facilitate subsequent Illumina application. Table 1 lists the forward (Guar_E1_P1_F, Guar_E1_P2_F, Guar_E1_P3_F) and reverse (Guar_E1_P1_R, Guar_E1_P2_R, Guar_E1_P3_R) primer sequences, the overhangs are color-coded: yellow for the forward and green for the reverse primers. Primer design was performed using the Integrated DNA Technologies (IDT) online primer design tool (https://eu.idtdna.com/pages/tools/primerquest).

PCR amplification was carried out using the cycler C1000 Touch (Bio-Rad, USA) to target the genomic region of *CtE1* in the guar genome. Each primer pair was tested separately using genomic DNA as the template. The PCR mix (25 µL) contained 1× HF buffer, 0.4 µM of each primer, 200 µM dNTPs, and 1 unit of Phusion® High-Fidelity DNA Polymerase (NEB, USA). The thermal cycling conditions were as follows: initial denaturation at 98 °C for 2 min, followed by 30 cycles of denaturation at 98 °C for 30 s, annealing at 62 °C for 1 min, and extension at 72 °C for 45 s. A final extension step was performed at 72 °C for 10 min.

Library preparation, sample pooling and Illumina sequencing. The sequencing library was prepared for a Genotyping-in-Thousands by Sequencing (GT-SEQ) (Campbell et al., 2015) approach from the PCR-amplified products. All PCR products were first purified using ammonium acetate precipitation to eliminate unincorporated nucleotides, salts, and other impurities from the reaction mixture. The concentration of each cleaned DNA sample was then measured using a NanoDrop spectrophotometer (Desjardins, Conklin, 2010). Each sample had an initial volume of 20 µL. For pooling, 4 µL was taken from samples with a DNA concentration below 10 ng/µL, while 2 µL was taken from samples with a concentration above 10 ng/µL. Approximately equal concentrations were allowed, since Illumina sequencing provides excess coverage of the target locus for each sample. This compensates for variability in concentration without exact quantification, which is acceptable for amplicon sequencing at high coverage depths (e. g., 16S) (Kennedy et al., 2014).

The selected volumes were pooled together into a single 2 mL Eppendorf tube to create a composite library. The pooled library was then prepared for high-throughput sequencing with the Illumina MiSeq. Nextera XT DNA Library Prep Kit was used for a two-step PCR workflow. First PCR was performed with gene specific primers + overhangs (Table 1), the second PCR was performed to add Illumina adapters + indices. For the pooled library, only one dual Illumina index was used, which significantly reduced the cost of sequencing. Paired-end sequencing was employed with 2 × 250 bp read mode.

Bioinformatics pipeline for SNP detection. Quality assessment of the raw reads was performed using FastQC (Andrews, 2010) with default parameters to evaluate base quality, GC content, and potential adapter contamination. Subsequently, high-quality reads were aligned to the reference target sequence, specifically, the *CtE1* guar gene (Criollo Delgado et al., 2025), using the BWA-MEM algorithm. SNPs were then identified using a variant calling pipeline that involved SAMtools for alignment processing and BCFtools for SNP calling and variant filtration (Li, 2011; Danecek et al., 2021).

Development of KASP assays. The KASP primers genotyping assay design tool (https://primerdigital.com/tools/kasp.html) (Kalendar et al., 2022) was used to design KASP primers for detected SNPs. Two allele-specific primers were designed carrying unique tails: FAM (5′ GAAGGTGACCAAGTTCAT GCT 3′) and HEX (5′ GAAGGTCGGAGTCAACGGATT 3′)

Table 2. KASP primers for *CtE1* gene assessment: two allele-specific forward primers with unique 5′ tails (shown in bold) and targeted SNPs at their 3′ ends

Primers name	Sequence	SNP position
F_SNP1_FAM	**GAAGGTGACCAAGTTCATGCT**TTACATGCAAAGCAAATAACAATAGTTG	–89
F_SNP1_HEX	**GAAGGTCGGAGTCAACGGATT**TTTTACATGCAAAGCAAATAACAATAGTAG	
R_SNP1	GGTGGGAATTTTAGTTGTGATGAAAT	
F_SNP2_FAM	**GAAGGTGACCAAGTTCATGCT**AAAAACCCTTTCATCATTTCAATTTTAATGTA	–54
F_SNP2_HEX	**GAAGGTCGGAGTCAACGGATT**AAAACCCTTTCATCATTTCAATTTTAATGAA	
R_SNP2	TGTTCATGTTTGAAGTAGAAGAGATG	
F_SNP3_FAM	**GAAGGTGACCAAGTTCATGCT**GGTGATGATCACGCGAGA	+296
F_SNP3_HEX	**GAAGGTCGGAGTCAACGGATT**GGGTGATGATCACGCGAAA	
R_SNP3	CAAACTCTAATCGGCGTTCC	
F_SNP4_FAM	**GAAGGTGACCAAGTTCATGCT**CTTCTTAATTGGTATTCTCTTCACCTTT	*43
F_SNP4_HEX	**GAAGGTCGGAGTCAACGGATT**TCTTAATTGGTATTCTCTTCACCTCT	
R_SNP4	ACCAATAACAGTGTTGGCATAGT	
F_SNP5_FAM	**GAAGGTGACCAAGTTCATGCT**GGTATTCTCTTCACCTTTCAACTC	*49
F_SNP5_HEX	**GAAGGTCGGAGTCAACGGATT**GTATTCTCTTCACCTTTCAACCC	
R_SNP5	ACCAATAACAGTGTTGGCATAGT	

Note. SNP positions are numbered relative to the *CtE1* CDS: – upstream of start codon; + within CDS; * downstream of stop codon (den Dunnen et al., 2016).

respectively, with the targeted SNPs at the 3′end (penultimate nucleotide), and a common primer was designed to pair with both forward and reverse primers. KASP genotyping primers are provided in Table 2. SNP positions were numbered relative to the *CtE1* coding sequence (CDS), where positive numbers indicate positions within the CDS (with 1 corresponding to the A of the ATG start codon), negative numbers (−) indicate nucleotides upstream (5′) of the start codon, and asterisks (*) denote nucleotides downstream (3′) of the stop codon (den Dunnen et al., 2016).

The KASP assay was conducted in 8 μL PCR reaction volume comprising 2 μL of genomic DNA (5 ng/μL), 3 μL of 2×KASP-TF V5.0 2X Master Mix (LGC, Biosearch Technologies) and 0.2 μL of allele-specific primer mix, making the final concentration of forward primers in the reaction volume 0.05 mM each, and 0.10 mM of common reverse primer. PCR cycling was performed with QuantStudio 5 cycler (Thermo Fisher Scientific, USA) using the following protocol: pre-incubation 30 °C 30 s (Pre-Read stage fluorescence measurement), pre-denaturation at 95 for 10 min, followed by 10 touchdown cycles (95 °C for 15 s; touchdown from 62 °C to 55 °C with 1.5 °C decrease per cycle for 60 s), followed by 60 additional cycles (94 °C for 20 s; 55 °C for 60 s), 30 °C for 1 min (Post-Read stage fluorescence measurement).

Statistical analysis. Descriptive statistics and estimates of variance were done by using the R package '*agricolae*' (https://cran.r-project.org/package = agricolae) (de Mendiburu, 2023). To check the effect of allelic variants on flowering and maturation time traits, ANOVA was used. For each analysis of variance, we also evaluated normality of residuals distribution using the Shapiro–Wilk Test. When the assumptions for residuals normality were not satisfied, the Kruskal–Wallis rank sum test served as a robust non-parametric alternative.

Results

Identification of novel SNPs in the gene *CtE1*

Sanger sequencing of *CtE1* in eight guar varieties with contrasting maturation times confirmed the presence of polymorphisms in the sequence of the gene previously predicted *in silico*, and identified five SNPs (Table 3). Positions of the SNPs were determined relatively to the CDS of the *CtE1* gene (Criollo Delgado et al., 2025) and a reference sequence of 1846 bp encompassing the CDS and upstream/downstream regions of the *CtE1* gene, which was extracted from guar genome assembly Cte V1.0 (GCA_037177725.1). Two SNPs were identified in the upstream region of the *CtE1* gene, the first SNP was located in the 5′ untranslated region (5′ UTR) at position –89 relative to CDS (SNP1) and the second SNP was located in the 5′ UTR at position –54 (SNP2), where the nucleotide thymine (T) was substituted with adenine (A) in both cases. One non-synonymous SNP was found within the coding region at position 296 from the start codon (SNP3), showing a guanine (G) to adenine (A) substitution that causes an amino acid change from arginine to lysine. Additionally, two SNPs were detected in the 3′ untranslated region (3′ UTR) at positions *43 (SNP4) and *49 (SNP5), both involving a transition from thymine (T) to cytosine (C). The analyzed *CtE1* gene sequence spans 803 base pairs, covering the 5′ UTR, coding region, and 3′ UTR, and all 5 SNPs are shown in Figure 1. As shown in Table 3, out of the 8 varieties examined by Sanger sequencing, the only distinct *CtE1* haplotype was revealed in accession Cat.52580.

Illumina sequencing

To extend the *CtE1* genotyping to the entire guar collection of over 144 accessions we avoided the use of cost-consuming Sanger sequencing and instead applied a method previously

Table 3. Polymorphisms of the *CtE1* gene among 8 guar accessions showing variation of flowering and maturation time

VIR cat. No.	Variety/Origin	Flowering time 2017, days	Vegetation period 2017, days	Flowering time 2018, days	Vegetation period 2018, days	SNP1	SNP2	SNP3	SNP4	SNP5
94	–/India	39	94	30	78	AA	TT	GG	TT	TT
52586	Lewis/USA	35	94	30	76	TT	TT	GG	TT	TT
22	–/India	35	94	33	87	TT	TT	GG	TT	TT
52572	Vavilovskij 130/–	36	91	34	86	TT	TT	GG	TT	TT
52581	–/India	35	99	34	110	TT	TT	GG	TT	TT
52580	–/India	39	87	39	110	AA	TA	AA	CC	TC
–	–/India	nd	nd	nd	nd	TT	TT	GG	TT	TT
–	–/Pakistan	nd	nd	nd	nd	TT	TT	GG	TT	TT

Note. For the accessions the variety/origin is indicated if known. nd – not determined.

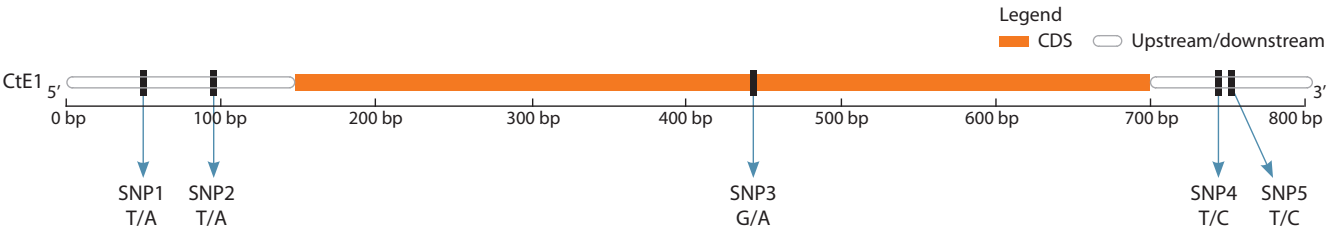


Fig. 1. Location of SNPs in the CDS and upstream/downstream regions of the *CtE1* gene revealed by Sanger sequencing among 8 guar cultivars differing in flowering/maturing time.

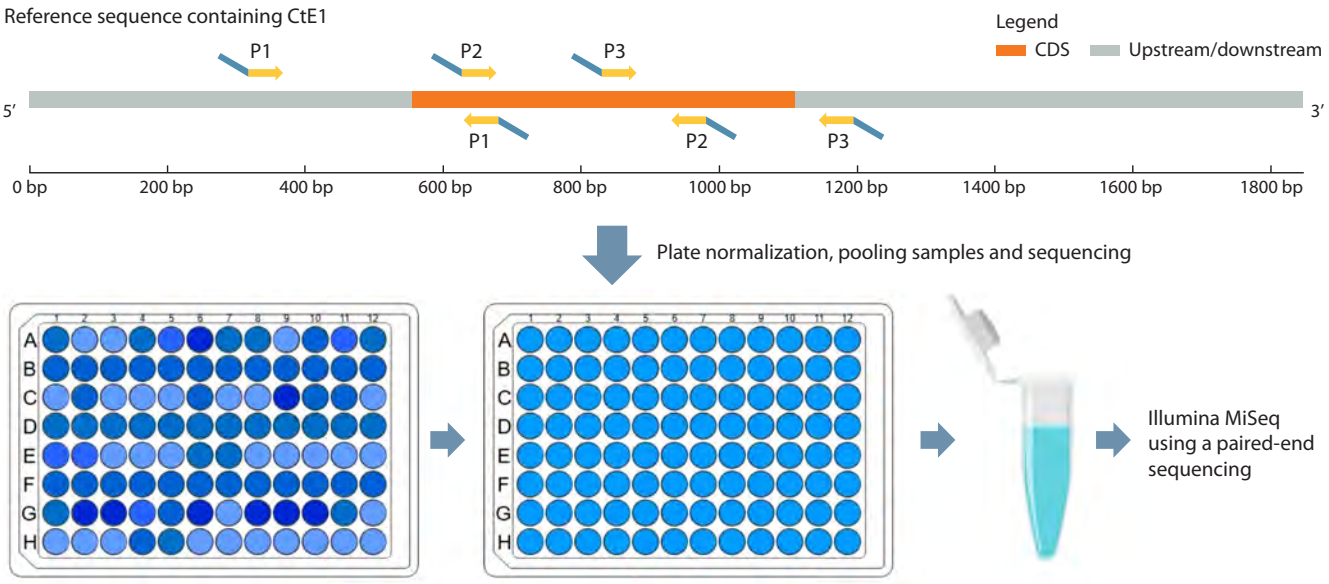


Fig. 2. Scheme of the GT-seq approach and distribution of primers for GT-seq in the CDS and upstream/downstream regions of the *CtE1* gene. By numbers primer pairs (forward and reverse) are indicated. The green part of the primers represents the target-specific sequence. Blue tails of the primers represent the Illumina sequencing adapters.

described as Illumina Genotyping in Thousands by Sequencing (GT-Seq) (Campbell et al., 2015). We created a pooled library containing multiplex PCR products of 3 regions spanning the *CtE1* gene to identify all possible polymorphisms in the target sequence in the collection of 144 accessions. Three pairs of primers with Illumina sequencing adapters were designed (Table 1) enabling all amplicons of all individuals to be pooled into a single sequencing library. Figure 2 shows the location of the primers in the sequenced region (847 bp) compared to a reference sequence of 1846 bp.

No barcoding of individual samples was performed, so when running the Illumina MiSeq, only one dual Illumina

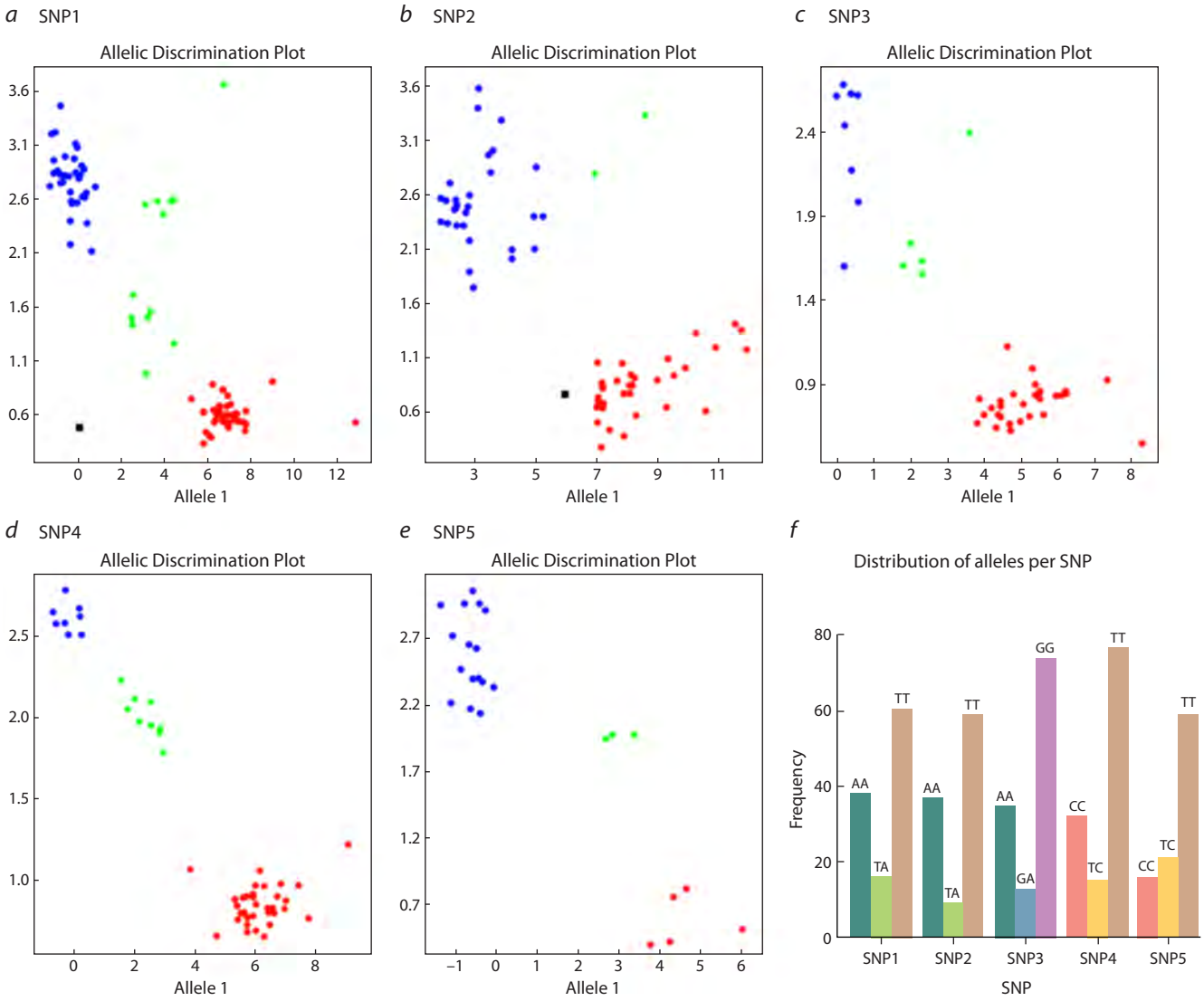


Fig. 3. Clustering of alleles of SNPs in the *CtE1* gene using KASP assays.
a–e, allelic discrimination plots of KASP markers located on the five SNP loci. SNP1–SNP5 correspond to these in Table 3. The clusters of accessions are represented on the scatter plot on the x-axis (Allele 1) and y-axis (Allele 2); *f*, distribution of alleles per SNP loci. The Figure does not reflect genotyping results for the entire collection: if some samples were incorrectly or incompletely genotyped, they were re-analyzed in additional runs.

index was used to barcode the entire library. As a result, 4,880,840 raw reads were obtained from Illumina and quality-checked using FastQC to evaluate base quality, GC content, and adapter contamination. 4,513,800 (92.48 %) of the reads were then successfully aligned to the *CtE1* reference guar genome assembly (Cte V1.0, GCA_037177725.1). With the data available, each of the three amplicons was covered by an average of 10,448 reads for each of the 144 guar accessions. As a result, the same 5 SNPs, that were discovered by Sanger sequencing of 8 accessions contrasting in flowering and maturing time, were again detected, and no additional polymorphism was found among 144 guar accessions.

The GT-seq analysis revealed a single missense mutation in the coding sequence of the *CtE1* gene within the examined intraspecific diversity of guar, resulting in an Arg→Lys amino acid substitution. However, unlike the loss-of-function mutations observed in soybean that lead to truncated or nonfunctional *E1* proteins, no such deleterious variants were detected. Nevertheless, an attempt was made to assess the field performance of guar plants carrying different *CtE1* alleles. To facilitate this, KASP assays were developed for the identified SNPs.

High-throughput KASP genotyping of polymorphisms in the *CtE1* gene

Five KASP markers linked to polymorphisms in the *CtE1* gene were developed based on the SNPs identified through the Illumina GT-seq approach and tested with 144 guar accessions. Each KASP marker enabled apparent clustering of accessions into three genotype classes (homozygous allele1, homozygous allele2, and heterozygous) (Fig. 3). The heterozygosity level estimated for SNPs in the *CtE1* gene in the studied collection of 144 guar accessions ranged from 0.086 to 0.218 (Fig. 3*f*), which is in line with the average heterozygosity level of 0.127 reported for soybean germplasm collections (Potapova et al., 2023).

Analysis of association between SNPs in the *CtE1* gene and flowering/maturation time of guar varieties

Since the guar diversity panel encompassing studied 144 guar accessions was evaluated in 2017 and 2018 in the field conditions of Krasnodar region (Grigoreva et al., 2021a), we use the opportunity to explore any possible correlation existing between the revealed nucleotide polymorphisms of *CtE1* and agronomic performance of guar varieties carrying different alleles of the gene.

Flowering time observed for the studied accessions varied from 32 to 45 days in 2017 and from 29 to 42 days in 2018. Notably, the correlation between flowering time of the guar collection in 2017 and 2018 was not statistically significant (r2 = 0.05, *p*-value > 0.05). Heritability of flowering time in guar estimated under stable natural conditions in India ranged from 52 % (Remzeena, Anitha, 2021) to 81 % (Choyal et al., 2022). However, guar propagation in the Krasnodar region often faces challenges, such as, for example, the spring drought of 2018, which resulted in damage to plant seedlings. Therefore, the observed year-to-year variations in the flowering time of guar genotypes may be due to differences in environmental factors, as well as the lack of standardized agricultural practices for this recently introduced legume crop.

Length of vegetation period varied respectively in the range 72–116 days and 72–110 days in two years. Here, the correlation between the overall vegetative period of guar accessions in 2017 and 2018 was low, but statistically significant (r2 = 0.19, *p*-value < 0.01). As with flowering time, this low correlation can presumably be explained by extreme drought conditions in spring 2018.

Given genotyping data for 5 SNPs in the *CtE1* gene for 144 guar accessions, we attempted to perform ANOVA for each SNP, considering the three genotypes (two alternative homozygous and heterozygous) as factor levels and the number of days from sowing to flowering (maturing) as the dependent variable.

Of all the markers we tested, only one SNP (SNP2) demonstrated an association with flowering time in guar in 2018, that approached statistical significance (ANOVA, $p = 0.052$). Heterozygous genotypes for SNP2 (TA) tended to flower slightly later than both homozygotes (TT and AA), a result that is difficult to explain from a biological perspective. Therefore, it can be concluded that the natural polymorphisms of the *CtE1* gene identified in the available collection of 144 guar genotypes do not exert a significant effect on flowering or maturation time.

Discussion

In the present study, we assessed natural allelic variation of the *CtE1* gene using a diversity panel of 144 guar accessions from different geographic origins. *CtE1* was previously identified as a homolog of *E1*, the major flowering time regulator in soybean; however, its genetic diversity and functional role in guar had not yet been reported. The *E1* gene plays a key role in the functional network of photoperiodic flowering regulation in soybean. Since the molecular identity of this gene was successfully elucidated in 2012 (Xia et al., 2012), numerous studies have underscored the significant impact that mutations in this gene can have on photoperiod sensitivity (Zhai et al., 2015; Han et al., 2019; Fang et al., 2024a, b; Gao et al., 2024).

During the adaptation of cultivated soybean northward to high latitudes under longer daylengths, the *E1* gene, like some other important flowering inhibitor genes (e. g., *E3*, *E4*, *Tof5*, *Tof11*, and *Tof12*), has accumulated sequence polymorphisms, which reduced photoperiod sensitivity to produce early flowering. Thus, the variation leading to early flowering was artificially selected, allowing cultivated soybean to adapt to high latitude areas (Lin et al., 2021). Several functional and non-functional/dysfunctional *E1* alleles (e. g., *E1*, *e1-as*, *e1-fs*, *e1-nl*) have been identified in soybean, which vary by geography/maturity group (Hou et al., 2023). However, not all of them equally contribute to flowering phenotype. For example, *e1fs* and *e1nl* are functionally deficient, leading to very early flowering and maturity, while *e1-as* is a weak mutant allele with an effect intermediate between that of the *E1* genotype and the functionally deficient alleles (Xia et al., 2012).

The high similarity in coding peptide sequences between the soybean *E1* gene and the guar *CtE1* gene, along with their comparable intron-exon structures, suggests that the genetic pathways governing the fundamental mechanisms of photoperiod response may be conserved across these two legume species (Criollo Delgado et al., 2025). This structural and sequence conservation implies that intraspecific genetic variation at the *CtE1* locus in guar could potentially contribute to variation in photoperiod sensitivity, similar to the functional allelic diversity observed at the *E1* locus in soybean.

However, within the natural allelic diversity of *CtE1* evaluated in this study, no clearly dysfunctional alleles were identified. Of the five SNPs discovered in the *CtE1* gene, only one (SNP3) was found in the coding region and resulted in a non-synonymous arginine-to-lysine substitution. This alteration is located at amino acid position 99, situated within the B3-like domain, which in guar spans amino acid residues 61–171 (Criollo Delgado et al., 2025). Among the soybean *E1* polymorphisms discovered to date, a similar *e1-b3a* mutation was found, which also occurs in the middle of the B3 domain of the *E1* gene. The *e1-b3a* represents 5bp (3 SNP and 2-bp deletion) mutation which leads to a frameshift causing a premature stop codon at the middle of the B3-like domain. As a result, the soybean *e1-b3a/e1-b3a* genotype flowered significantly earlier than *E1/E1* and *E1/e1-b3a* (Zhai et al., 2015). In contrast, the functional significance of the Arg→Lys amino acid substitution identified within the B3 domain of guar, remains uncertain, as arginine and lysine share similar physicochemical properties. This substitution is considered conservative and is therefore predicted to have a minimal effect on protein function (Betts, Russell, 2003; Ryan, Ó'Fágáin, 2007; Banayan et al., 2024). On the other hand, it has been reported that amino acid substitutions in the *E1* sequence also can lead to significant functional changes, if they occur in the region of bipartite Nuclear Localization Signal (NLS). For example, the point mutation from arginine to threonine at position 15 in the soybean *E1* gene (known as *e1-as* mutant) occurs at exactly the first basic domain of the bipartite NLS, leading to different subcellular localization of the resulting protein and affecting flowering phenotype (Xia et al., 2012).

Two SNPs (SNP1 and SNP2) were discovered in 5′UTR region of the *CtE1* gene. Similarly, mutations *e1-re* and *e1-p* were described at the 5′UTR region of the *E1* gene in soy-

bean. The *e1-re* allele is characterized by the insertion of a long interspersed nuclear element (LINE) located 148 bases upstream of the start codon, whereas the *e1-p* mutant exhibits sequence variation in the 5′ upstream region compared to *E1*. The effects of both alleles on flowering time in soybean have not been well studied (Tsubokura et al., 2014).

The limited functional allelic diversity observed in the *CtE1* gene of guar, compared to the *E1* gene in soybean, likely reflects differences in their evolutionary histories, domestication bottlenecks, and selection pressures. Soybean was domesticated in a region spanning 30–45°N in China and is now cultivated globally across a broad latitudinal range, from 53°N to 35°S (Lin et al., 2021). In contrast, guar was domesticated in India and Pakistan (Ravelombola et al., 2021), and to this day, these countries remain the primary centers of guar cultivation. This more geographically restricted domestication of guar and cultivation range may have resulted in reduced selection for photoperiodic adaptation and, consequently, lower allelic diversity at key flowering-time loci such as *CtE1*.

On other hand, it is still possible that within the intraspecific diversity of guar there exist genotypes carrying more severe mutations in the *CtE1* gene that can substantially impair its function; however, such genotypes were not present in the studied cohort of 144 accessions.

Furthermore, it has been reported that *E1* homologues in various legumes exhibit differing roles in flowering, highlighting functional diversification within the *E1* gene family (Zhang et al., 2016; Cao et al., 2017). For instance, in *Phaseolus vulgaris*, the *E1* homologue known as *PvE1L* acts as a flowering repressor, mirroring the function of the *E1* gene in soybean. Ectopic expression of *PvE1L* has been shown to delay flowering onset in soybean (Zhang et al., 2016). In contrast, *Medicago truncatula*'s *E1* homologue, *MtE1L*, does not influence flowering when ectopically expressed in soybean. This variation suggests that the functional roles of *E1* homologues in legumes may be linked to lineage specificity and genomic duplication events. This underscores the complexity of flowering regulation within the legume family.

The CRISPR/Cas9 system has recently emerged as a powerful tool for targeted genome editing and functional genomics research. In soybean, its application has enabled in-depth investigation of the *E1* gene's role in photoperiod regulation, through CRISPR/Cas9-mediated mutagenesis followed by phenotypic analysis of flowering time (Wan et al., 2022). A similar approach can be implemented in guar by generating CRISPR/Cas9-induced mutants with targeted alterations in the *CtE1* gene. This would allow for a direct functional assessment of *CtE1* and its role in regulating flowering time and photoperiod sensitivity in guar. In addition, the application of CRISPR/Cas-based mutagenesis could potentially benefit guar breeding programs not only by enabling the creation of *CtE1* mutants, but also by targeting other flowering-related genes homologous to the soybean *E* maturity genes, such as *CtE2–CtE4*. This approach could facilitate a detailed investigation of the genetic network regulating flowering time in guar. This may also facilitate the development of novel photoperiod-insensitive guar germplasm, analogous to soybean mutants that have expanded soybean cultivation into higher latitudes.

Conclusions

In this study, we characterized nucleotide variability of the *CtE1* gene, the guar ortholog of soybean *E1* gene, in a diverse panel of 144 guar accessions and identified five novel SNPs across the 5′ UTR, coding, and 3′ UTR regions. We developed KASP markers for these SNPs to provide a robust genotyping tool to explore *CtE1* haplotypes in larger germplasm collections. Genotyping of 144 guar samples for five *CtE1* SNPs revealed only one SNP in the coding part of the gene, causing an Arg→Lys substitution. Given the conservative nature of this amino acid substitution, its functional impact is likely limited. No significant associations were detected so far between discovered SNPs and available data on variability in flowering time or vegetation period length. Our findings indicate that natural variation in *CtE1* within the studied guar germplasm has little impact on flowering time or maturation. We hypothesize that the geographically restricted domestication and cultivation range of guar may have led to reduced selection pressure for photoperiodic adaptation, resulting in lower allelic diversity at key flowering-time loci such as *CtE1*.

References

Andrews S. FastQC. A Quality control tool for high throughput sequence data. Babraham bioinformatics, 2010. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Banayan N.E., Loughlin B.J., Singh S., Forouhar F., Lu G., Wong K., Neky M., Hunt H.S., Bateman L.B., Tamez A., Handelman S.K., Price W.N., Hunt J.F. Systematic enhancement of protein crystallization efficiency by bulk lysine-to-arginine (KR) substitution. *Protein Sci*. 2024;33(3):e4898. doi 10.1002/pro.4898

Benakanahalli N.K., Sridhara S., Ramesh N., Olivoto T., Sreekantappa G., Tamam N., Abdelbacki A.M.M., Elansary H.O., Abdelmohsen S.A.M. A framework for identification of stable genotypes basedon MTSI and MGDII indexes: An example in guar (*Cymopsis tetragonoloba* L.). *Agronomy*. 2021;11(6):1221. doi 10.3390/agronomy11061221

Betts M.J., Russell R.B. Amino acid properties and consequences of substitutions. In: Barnes M.R., Gray I.C. (Eds.) Bioinformatics for geneticists. John Wiley & Sons. 2003;289-316. doi 10.1002/0470867302.ch14

Campbell N.R., Harmon S.A., Narum S.R. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour*. 2015;15(4):855-867. doi 10.1111/1755-0998.12357

Cao D., Takeshima R., Zhao C., Liu B., Jun A., Kong F. Molecular mechanisms of flowering under long days and stem growth habit in soybean. *J Exp Bot*. 2017;68(8):1873-1884. doi 10.1093/jxb/erw394

Choyal P., Dewangan R., Nd R., Xaxa S., Seervi K.S., Seervi D. Genetic variability studies in cluster bean [*Cyamopsis tetragonoloba* (L.) Taub]. *Pharma Innovation J*. 2022;11(2):2364-2368

Criollo Delgado L., Zamalutdinov A., Potokina E. Identification of soybean *E1-E4* gene orthologs in the guar genome using comprehensive transcriptome assembly and annotation. *Front Biosci*. 2025;17(2):26548. doi 10.31083/fbs26548

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008. doi 10.1093/gigascience/giab008

de Mendiburu F. Agricolae: Statistical Procedures for Agricultural Research (Version 1.3-7). Computer software. 2023. Available: https://cran.r-project.org/web/packages/agricolae/index.html

den Dunnen J.T., Dalgleish R., Maglott D.R., Hart R.K., Greenblatt M.S., McGowan-Jordan J., Roux A.-F., Smith T., Antonarakis S.E., Taschner P.E.M. HGVS recommendations for the descrip-

tion of sequence variants: 2016 update. *Hum Mutat*. 2016;37(6): 564-569. doi 10.1002/humu.22981

Desjardins P., Conklin D. NanoDrop Microvolume quantitation of nucleic acids. *J Visualized Exp*. 2010;45:e25651. doi 10.3791/2565

Fang C., Du H., Wang L., Liu B., Kong F. Mechanisms underlying key agronomic traits and implications for molecular breeding in soybean. *J Genet Genomics*. 2024a;51(4):379-393. doi 10.1016/j.jgg.2023.09.004

Fang C., Sun Z., Li S., Su T., Wang L., Dong L., Li H., … Lin X., Zatybekov A., Liu B., Kong F., Lu S. Subfunctionalisation and self-repression of duplicated *E1* homologues finetunes soybean flowering and adaptation. *Nat Commun*. 2024b;15(1):6184. doi 10.1038/s41467-024-50623-3

Gao Y., Zhang Y., Ma C., Chen Y., Liu C., Wang Y., Wang S., Chen X. Editing the nuclear localization signals of *E1* and *E1Lb* enables the production of tropical soybean in temperate growing regions. *Plant Biotechnol J*. 2024;22(8):2145-2156. doi 10.1111/pbi.14335

Grigoreva E., Barbitoff Y., Changalidi A., Karzhaev D., Volkov V., Shadrina V., Safronycheva E., Ben C., Gentzbittel L., Potokina E. Development of SNP set for the marker-assisted selection of guar (*Cyamopsis tetragonoloba* (L.) Taub.) based on a custom reference genome assembly. *Plants*. 2021a;10(10):2063. doi 10.3390/plants10102063

Grigoreva E., Tkachenko A., Arkhimandritova S., Beatovic A., Ulianich P., Volkov V., Karzhaev D., Ben C., Gentzbittel L., Potokina E. Identification of key metabolic pathways and biomarkers underlying flowering time of guar (*Cyamopsis tetragonoloba* (L.) Taub.) via integrated transcriptome-metabolome analysis. *Genes*. 2021b; 12(7):952. doi 10.3390/genes12070952

Han J., Guo B., Guo Y., Zhang B., Wang X., Qiu L.-J. Creation of early flowering germplasm of soybean by CRISPR/Cas9 technology. *Front Plant Sci*. 2019;10:1446. doi 10.3389/fpls.2019.01446

Hou Z., Fang C., Liu B., Yang H., Kong F. Origin, variation, and selection of natural alleles controlling flowering and adaptation in wild and cultivated soybean. *Mol Breeding*. 2023;43(5):36. doi 10.1007/s11032-023-01382-4

Integrated DNA Technologies. IDT (n.d.). Retrieved December 16, 2024. Available: https://www.idtdna.com/page

Ivanova N.V., Fazekas A.J., Hebert P.D.N. Semi-automated, membrane-based protocol for DNA isolation from plants. *Plant Mol Biol Rep*. 2008;26(3):186-198. doi 10.1007/s11105-008-0029-4

Kalendar R., Shustov A.V., Akhmetollayev I., Kairov U. Designing allele-specific competitive-extension PCR-based assays for high-throughput genotyping and gene characterization. *Front Mol Biosci*. 2022;9:773956. doi 10.3389/fmolb.2022.773956

Kennedy K., Hall M.W., Lynch M.D.J., Moreno-Hagelsieb G., Neufeld J.D. Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol*. 2014;80(18):5717-5722. doi 10.1128/AEM.01451-14

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993. doi 10.1093/bioinformatics/btr509

Lin X., Liu B., Weller J.L., Abe J., Kong F. Molecular mechanisms for the photoperiodic regulation of flowering in soybean. *J Integr Plant Biol*. 2021;63(6):981-994. doi 10.1111/jipb.13021

Liu L., Song W., Wang L., Sun X., Qi Y., Wu T., Sun S., Jiang B., Wu C., Hou W., Ni Z., Han T. Allele combinations of maturity genes *E1-E4* affect adaptation of soybean to diverse geographic regions and farming systems in China. *PLoS One*. 2020;15(7):e0235397. doi 10.1371/journal.pone.0235397

Potapova N.A., Zlobin A.S., Perfil'ev R.N., Vasiliev G.V., Salina E.A., Tsepilov Y.A. Population structure and genetic diversity of the 175 soybean breeding lines and varieties cultivated in West Siberia and other regions of Russia. *Plants*. 2023;12(19):3490. doi 10.3390/plants12193490

Ravelombola W., Manley A., Adams C., Trostle C., Ale S., Shi A., Cason J. Genetic and genomic resources in guar: A review. *Euphytica*. 2021;217(11):199. doi 10.1007/s10681-021-02929-2

Remzeena A., Anitha P. Genetic variability, heritability and genetic advance in cluster bean [*Cyamopsis tetragonoloba* (L.) Taub.] genotypes. *Indian J Agric Res*. 2021. doi 10.18805/IJARe.A-5779

Ryan B.J., Ó'Fágáin C. Arginine-to-lysine substitutions influence recombinant horseradish peroxidase stability and immobilisation effectiveness. *BMC Biotechnology*. 2007;7(1):86. doi 10.1186/1472-6750-7-86

Tsubokura Y., Watanabe S., Xia Z., Kanamori H., Yamagata H., Kaga A., Katayose Y., Abe J., Ishimoto M., Harada K. Natural variation in the genes responsible for maturity loci *E1, E2, E3* and *E4* in soybean. *Ann Bot*. 2014;113(3):429-441. doi 10.1093/aob/mct269

Verma S., Dubey N., Dhugga K.S., Gill K.S., Randhawa G.S. Cluster bean: From garnering industrial importance to molecular research interventions for the improvement of commercially viable traits. *S Afr J Bot*. 2025;178:307-317. doi 10.1016/j.sajb.2025.01.022

Wan Z., Liu Y., Guo D., Fan R., Liu Y., Xu K., Zhu J., Quan L., Lu W., Bai X., Zhai H. CRISPR/Cas9-mediated targeted mutation of the *E1* decreases photoperiod sensitivity, alters stem growth habits, and decreases branch number in soybean. *Front Plant Sci*. 2022;13: 1066820. doi 10.3389/fpls.2022.1066820

Watanabe S., Harada K., Abe J. Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breed Sci*. 2012;61(5): 531-543. doi 10.1270/jsbbs.61.531

Xia Z. Research progress in whole-genome analysis and cloning of genes underlying important agronomic traits in soybean. *Chin Bull Bot*. 2017;52(2):148-158. doi 10.11983/CBB16087

Xia Z., Watanabe S., Yamada T., Tsubokura Y., Nakashima H., Zhai H., Anai T., Sato S., Yamazaki T., Lü S., Wu H., Tabata S., Harada K. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc Natl Acad Sci USA*. 2012;109(32):E2155-E2164. doi 10.1073/pnas.1117982109

Xu M., Yamagishi N., Zhao C., Takeshima R., Kasai M., Watanabe S., Kanazawa A., Yoshikawa N., Liu B., Yamada T., Abe J. The soybean-specific maturity gene *E1* family of floral repressors controls night-break responses through down-regulation of *FLOWERING LOCUS T* orthologs. *Plant Physiol*. 2015;168(4):1735-1746. doi 10.1104/pp.15.00763

Zhai H., Lü S., Wu H., Zhang Y., Zhang X., Yang J., Wang Y., Yang G., Qiu H., Cui T., Xia Z. Diurnal expression pattern, allelic variation, and association analysis reveal functional features of the *E1* gene in control of photoperiodic flowering in soybean. *PLoS One*. 2015; 10(8):e0135909. doi 10.1371/journal.pone.0135909

Zhang X., Zhai H., Wang Y., Tian X., Zhang Y., Wu H., Lü S., Yang G., Li Y., Wang L., Hu B., Bu Q., Xia Z. Functional conservation and diversification of the soybean maturity gene *E1* and its homologs in legumes. *Sci Rep*. 2016;6(1):29548. doi 10.1038/srep29548

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

Received August 18, 2025. Revised October 24, 2025. Accepted October 29, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1255-1266

doi 10.18699/vjgb-25-130

Using a wheat line with wild emmer genetic material to improve modern *Triticum aestivum* L. varieties by a complex of economically useful traits

O.A. Orlovskaya [1], K.K. Yatsevich [1], L.V. Milko [1], N.M. Kaznina [2], N.I. Dubovets [1], A.V. Kilchevsky [1]

[1] Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus
[2] Institute of Biology of the Karelian Research Centre of the Russian Academy of Sciences, Petrozavodsk, Russia
O.Orlovskaya@igc.by

Abstract. Wild emmer *Triticum dicoccoides* samples have a high content of protein and microelements in their grain, but when crossed with common wheat varieties, undesirable properties of a wild relative (low yield, spike fragility and difficult threshing) can be transmitted to the hybrid along with valuable traits. The possibility of improving economically useful traits of modern common wheat varieties using a wheat line with wild emmer genetic material (l29), combining high cytological stability with improved nutritional value and productivity, was studied. The F_4–F_5 hybrids obtained as a result of crossing in the forward and reverse directions of four common spring wheat varieties with l29 were studied. A C-banding technique and genotyping with SSR markers were used to determine the introgression fragments of *T. dicoccoides* genetic material. Cytological stability was assessed based on the study of chromosome behavior in microsporogenesis. The grain content of macro- (K, P, Ca and Mg) and microelements (Zn, Fe, Cu and Mn) was established by atomic emission spectrometry with inductively coupled plasma; the grain quality indices were measured on an Infra LUM FT-12 analyzer. The C-banding and microsatellite analysis data indicate a high frequency of alien genetic material introgression in the genome of hybrid forms. All variants of the l29 introgression of wild emmer material (1BL, 2BS, 3B, 5B and 6AL) were identified among the progeny of eight crossing combinations. The recombinant chromosome 3B was found in all hybrid combinations. The hybrids were characterized by a high level of cytological stability (the meiotic index was 90.0–98.0 %). The effectiveness of using a wheat line with *T. dicoccoides* genetic material to enhance modern varieties in terms of the content of protein, gluten and mineral composition of grain without reducing productivity was shown. Secondary introgression hybrids, exceeding the initial varieties by a set of grain quality characteristics and not inferior to them in terms of basic productivity indicators, were obtained.

Key words: common wheat *Triticum aestivum* L.; wild emmer *Triticum dicoccoides*; wheat introgressive lines; C-banding; SSR analysis; microsporogenesis; grain quality; productivity

For citation: Orlovskaya O.A., Yatsevich K.K., Milko L.V., Kaznina N.M., Dubovets N.I., Kilchevsky A.V. Using a wheat line with wild emmer genetic material to improve modern *Triticum aestivum* L. varieties by a complex of economically useful traits. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2025;29(8):1255-1266. doi 10.18699/vjgb-25-130

Funding. The work was supported by the Belarusian Republican Foundation for Fundamental Research (grant No. B 20P-240) and project 2.10.2 of the State Program for Fundamental Research "Biotechnology-2."

Использование линии пшеницы с генетическим материалом дикой полбы для улучшения современных сортов *Triticum aestivum* L. по комплексу хозяйственно полезных признаков

О.А. Орловская [1], К.К. Яцевич [1], Л.В. Милько [1], Н.М. Казнина [2], Н.И. Дубовец [1], А.В. Кильчевский [1]

[1] Институт генетики и цитологии Национальной академии наук Беларуси, Минск, Беларусь
[2] Институт биологии Карельского научного центра Российской академии наук, Петрозаводск, Россия
O.Orlovskaya@igc.by

Аннотация. Образцы дикой полбы *Triticum dicoccoides* обладают высоким содержанием белка и микроэлементов в зерне, но при скрещивании с сортами мягкой пшеницы гибридам наряду с ценными признаками могут передаваться и нежелательные свойства дикорастущего сородича (низкая урожайность, ломкость колоса, трудный обмолот). В рамках данного исследования изучена возможность улучшения

современных сортов мягкой пшеницы по комплексу хозяйственно полезных признаков посредством использования линии пшеницы с генетическим материалом дикой полбы (л29), сочетающей высокую цитологическую стабильность с улучшенной питательной ценностью и продуктивностью. Исследовали гибриды F_4–F_5, полученные от скрещивания в прямом и обратном направлениях четырех сортов яровой мягкой пшеницы с л29. Фрагменты интрогрессии генетического материала *T. dicoccoides* выявляли с применением метода дифференциального окрашивания хромосом по Гимза (С-бэндинг) и генотипирования маркерами SSR. Оценку цитологической стабильности проводили на основе изучения поведения хромосом в микроспорогенезе. Содержание макро- (K, P, Ca, Mg) и микроэлементов (Zn, Fe, Cu, Mn) в зерне определяли методом атомно-эмиссионной спектрометрии с индуктивно связанной плазмой; показатели качества зерна – на анализаторе «Инфра ЛЮМ ФТ-12». Данные С-бэндинга и микросателлитного анализа свидетельствуют о высокой частоте включения чужеродного генетического материала в геном гибридных форм. Среди потомства восьми комбинаций скрещивания выявлены все характерные для л29 варианты интрогрессий материала полбы (1BL, 2BS, 3B, 5B, 6AL), причем рекомбинантная хромосома 3B обнаружена во всех гибридных комбинациях. Для гибридов характерен высокий уровень цитологической стабильности (мейотический индекс составил 90.0–98.0 %). Показана эффективность использования линии пшеницы с включением генетического материала *T. dicoccoides* для улучшения современных сортов по содержанию белка, клейковины и минерального состава зерна без снижения продуктивности. Выделены вторичные интрогрессивные гибриды, превосходящие родительские сорта по комплексу признаков качества зерна в оба года исследования и не уступающие им по основным показателям продуктивности.

Ключевые слова: мягкая пшеница *Triticum aestivum* L.; дикая полба *Triticum dicoccoides*; интрогрессивные линии; С-бэндинг; SSR-анализ; микроспорогенез; качество зерна; продуктивность

Introduction

Wheat, one of the most widely grown cereal crops across the globe, is a major source of nutrients in a human diet. At the same time, intensive breeding aimed at increasing productivity has led to a significant erosion of the wheat gene pool in terms of grain nutritional value, resulting in a low protein, mineral and vitamin content of grain in modern varieties (Shewry et al., 2016; Marcos-Barbero et al., 2021). Research over the past two decades has shown that wild relatives and landraces are characterized by a higher biological value of grain than cultivated varieties (Heidari et al., 2016; Goel et al., 2018; Arora et al., 2019; Zeibig et al., 2022). Also, related species of common wheat have higher variability in terms of economically valuable traits (Cakmak et al., 2000; Akcura, Kokten, 2017), which is also valuable for breeding. In this regard, great hopes are pinned on distant hybridization to solve the problem of nutrient deficiency in wheat grain. At present, there are examples of wheat lines with an increased mineral and protein content in grain being developed by incorporating genetic material from various related species into their genome (Tiwari et al., 2010; Savin et al., 2018; Liu et al., 2021).

As a rule, wheat genotypes with enhanced grain quality have lower productivity; therefore, one of the priority areas of breeding is the development of varieties that combine both high yield and good grain quality. This study examined the possibility of improving common wheat varieties in relation to a complex of economically useful traits using a line containing wild emmer genetic material (*Triticum dicoccoides*). Previously, the chromosomal localization of the genetic material of tetraploid and hexaploid *Triticum* species samples was established in the introgressive wheat lines we had developed based on C-banding data and genotyping results using microsatellite (SSR) and single nucleotide polymorphism (SNP) markers (Orlovskaya et al., 2020; Orlovskaya et al., 2023b).

Analysis of the main quality indicators (protein content, content and quality of gluten, vitreousness and thousand grain weight) and grain mineral composition (K, P, Ca, Mg, Zn, Fe, Cu and Mn content) made it possible to identify lines with alien genetic material, exceeding original varieties by studied traits, that are of interest for common wheat breeding (Orlovskaya et al., 2023a). Line 29 of the combination Rassvet × *T. dicoccoides* k-5199, combining high nutritional value and productivity, is included in crossing with modern varieties to enhance the traits of common wheat valuable for breeding.

Materials and methods

The F_4–F_5 hybrids obtained by the direct and reverse crossing of common spring wheat modern varieties of Belarusian selection Darya, Toma, Laska and Lyubava with the introgressive line 29 of the Rassvet × T. dicoccoides k-5199 (l29) combination (eight crossing combinations in total) were used as research material. The plants were grown in the experimental fields of the Institute of Genetics and Cytology, NAS of Belarus, in 2021–2022 (Minsk, the Republic of Belarus, 53.924256° north latitude and 27.695015° east longitude) on sod-podzolic sandy loam soil.

Mineral fertilizers were applied in the following doses: nitrogen – 80 kg of active substance per 1 ha (kg active substance/ha); phosphorus – 70 kg active substance/ha; and potash – 90 kg active substance/ha. During harvesting, the following traits were taken into account: the number of productive shoots per plant; the length and number of spikelets and grains of the main spike; grain weight of the main spike and plant; and thousand grain weight. To assess the traits, 15 plants of each genotype were randomly selected.

To analyze the genomic structure of the hybrid material, a variant of the Giemsa method of differential chromosome staining (C-banding) was used (Badaeva et al., 1994), which allows recognizing individual A, B, D chromosomes of genomes in the karyotype, as well as chromosome aberrations involving regions with diagnostic C-blocks. The stained preparations were analyzed using the Amplival microscope (Carl Zeiss, Jena) with an Apochromat 100x objective and 1.32 MI aperture. The selected metaphase plates were photographed using a LeicaDC 300 digital video camera (Leica

Camera AG, Germany). The obtained images were processed using the Adobe Photoshop CC 2017 graphic editor (Adobe Systems, USA).

DNA was isolated from the seedlings of five plants for each genotype using the GeneJET Plant Genomic DNA Purification Kit (Thermo Fisher Scientific, Lithuania) according to the manufacturer's protocol. Genotyping of hybrids and parental forms was performed using SSR markers (WMC, GWM) mapped in the hexaploid wheat genome (Somers et al., 2004). The previously conducted genotyping of l29 using SSR and SNP markers showed the presence of wild emmer genetic material in chromosomes 1BL, 2BS, 3B, 5B and 6AL (Orlovskaya et al., 2023b). In this regard, we used markers designated only for these chromosomes: *Xgwm18*, *Xgwm374.2*, *Xgwm403*, *Xgwm274*, *Xgwm268*, *Xgwm11*, *Xgwm131*, *Xgwm498* for 1B; *Xgwm210*, *Xgwm614*, *Xgwm257*, *Xgwm410*, *Xgwm630*, *Xgwm148*, *Xgwm429*, *Xgwm319* – 2BS; *Xgwm389*, *Xgwm493*, *Xgwm533.2*, *Xgwm566*, *Xgwm285*, *Xgwm108*, *Xgwm107*, *Xgwm 264* – 3B; *Xgwm234*, *Xgwm159*, *Xgwm544*, *Xgwm67*, *Xgwm499*, *Xgwm554*, *Xgwm271*, *Xgwm408*, *Xgwm604*, *Xwmc99* – 5B; *Xgwm427*, *Xwmc621*, *Xwmc254* and *Xgwm169* – 6AL.

The conditions for carrying out the polymerase chain reaction (PCR) are described in the work of M.S. Röder et al. (1998). The PCR fragments' separation was performed on an automatic sequencer ABI PRISM 3500 (Applied Biosystems, USA). The size of fragments was calculated using the computer program Gene Mapper (version 5.0) developed by Applied Biosystems, USA. Putative chromosomal localization was determined based on consensus wheat chromosome maps for SSR markers (Somers et al., 2004). MapChart 2.32 software was used to visualize chromosome maps.

Microsporogenesis was studied on temporary squash preparations using the generally accepted method (Pausheva, 1988). For each combination of crossing and initial forms, 30 metaphase I plates and 50–80 plates of the following meiotic stages were analyzed: anaphases I and II, metaphase II and tetrads. The preparations were studied on an Amplival microscope (Carl Zeiss, Jena) with an Apochromat 100x objective and aperture 1.32 MI.

The content of macro- (K, P, Ca and Mg) and microelements (Zn, Fe, Cu and Mn) in grain was determined at the Center for Analytical and Spectral Measurements of the State Scientific Institution "B.I. Stepanov Institute of Physics of the NAS of Belarus" on IRIS Intrepid II XDL DUO (the atomic emission spectrometer). For each sample, the analysis was carried out in two biological replicates; measurements were repeated 10 times for each sample. The total protein and gluten content in grain and the quality of gluten were identified using an Infra LUM FT-12 infrared analyzer (Lumex, RF) according to GOST ISO 12099-2017. To assess the quality of gluten, the GDI (the Gluten Deformation Index, GOST 13586.1-68) indicator was used. According to the GDI values, strong (45–77 conventional units, quality group I), satisfactory weak (78–102 conventional units, quality group II) and unsatisfactory weak (more than 102 conventional units, quality group III) gluten is distinguished.

The experiment results were analyzed using descriptive statistics and ANOVA methods in the software packages Statistica 10.0 (StatSoft, USA) and MS Excel. The differences between the groups were assessed using ANOVA and Fisher's LSD criterion (Fisher's least significant difference).

Results

Analysis of *T. dicoccoides* genetic material introgression in the genome of hybrid forms

For the chromosomal identification of alien chromatin in the common wheat genome, cytological (C-banding) and molecular (SSR analysis) methods were used. The karyotype of l29, which served as the genetic material donor of wild emmer, was studied in detail, as we described earlier (Orlovskaya et al., 2023b). The line contains fragments of *T. dicoccoides* genetic material in the proximal region of the long arm of chromosome 1B and in the distal regions of the short arm of chromosome 2B and the long arm of 5B, as well as a pair of 3B chromosomes of emmer, which substituted the corresponding homologs of the variety Rassvet. The presence of polymorphism determined by the differential staining pattern of these four chromosomes between l29 and the Toma, Lyubava wheat varieties included in the experiment made it possible to trace the process of transferring alien chromatin to the karyotype of hybrid forms based on these varieties. The progeny from the crossing of Toma × l29 and l29 × Toma was heterogeneous in its chromosomal composition. In the direct combination of crossing (Toma × l29), chromosome 1B with a fragment of emmer chromatin in the L-arm and chromosome 3B of emmer were detected, with the latter being present in both disomic and monosomic states (one chromosome from the wheat variety and the other from emmer) (Fig. 1).

In all the plants obtained in the reverse crossing combination (l29 × Toma), a pair of 5B chromosomes with a fragment of *T. dicoccoides* genetic material in the L-arm and chromosome 3B of emmer, which in half of the analyzed plants was present in a disomic state and, in the other half, in a monosomic state, were detected (Fig. 2*a*).

In addition to the above chromosomes, some plants had a pair of 2B chromosomes with a fragment of emmer genetic material (Fig. 2*b*). The absence of variations in the chromosomal composition was characteristic of the progeny of hybrids resulting from the cross Lyubava × l29: all plants contained chromosomes 2B and 5B introduced from l29 with fragments of emmer genetic material in their karyotypes (Fig. 2*c*). Plants in the progeny of the l29 × Lyubava combination contained emmer chromatin introgression similar to the direct crossing combination. It should be noted here that identifying a possible substitution of 3B chromosomes of the variety Lyubava in the karyotypes of hybrids with corresponding emmer homologs using C-banding was not possible due to the similarity of their differential staining patterns.

Due to the absence of polymorphism, we were not able to estimate the frequency of emmer chromatin inclusion in the karyotypes of hybrids obtained with the involvement of Darya and Laska varieties using C-banding. SSR markers were used to identify introgressed fragments in those hybrids, as well as to detect alien material in chromosomes with a small number of diagnostic heterochromatic blocks. Out of 38 SSR markers used, 23 markers revealed polymorphism between the parent varieties and l29. A molecular analysis made it possible to reveal recombination events in all analyzed hybrids. For the

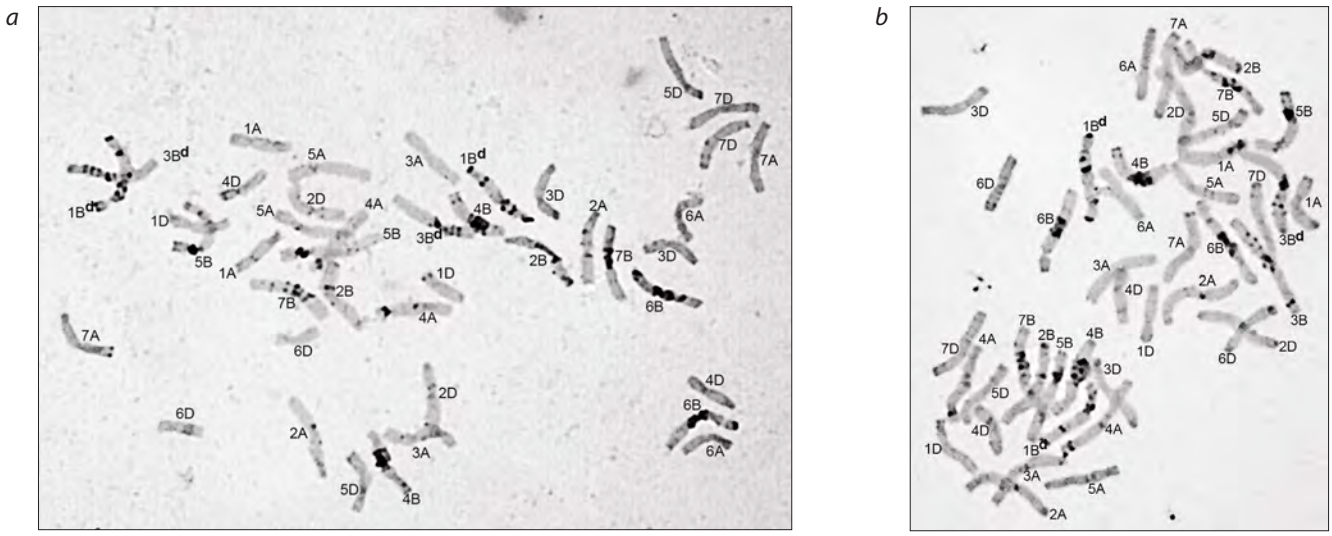


Fig. 1. Plant karyotypes in the progeny from the cross Toma × l29.

a – with a pair of 1B chromosomes with a fragment of *T. dicoccoides* genetic material and a pair of 3B chromosomes of emmer; *b* – with a pair of 1B chromosomes with a fragment of *T. dicoccoides* genetic material and a heteromorphic pair of 3B/3B^{d} chromosomes. The chromosomes with *T. dicoccoides* genetic material introduced from line 29 are designated by the superscript letter "d".

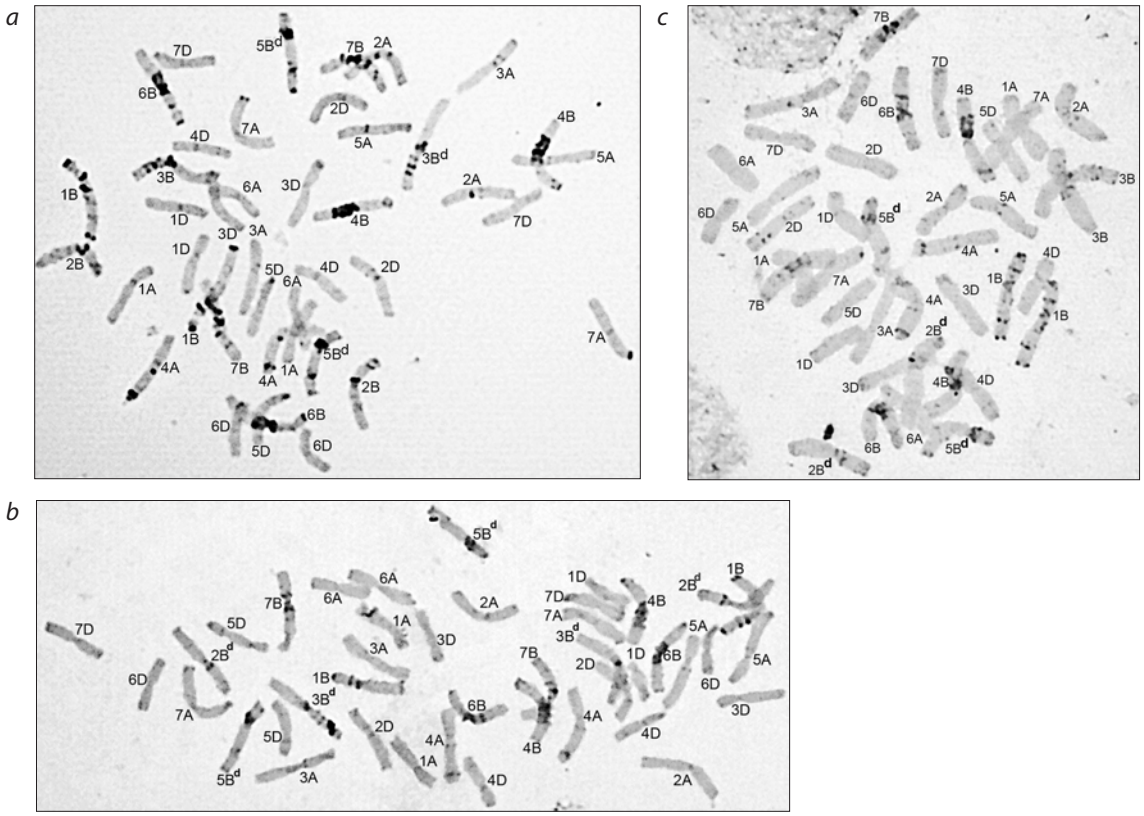


Fig. 2. The karyotypes of plants with different variants of introgression of *T. dicoccoides* genetic material.

a and *b* – hybrid l29 × Toma; *c* – hybrid Lyubava × l29. Chromosomes with *T. dicoccoides* genetic material introduced from line 29 are designated by the superscript letter "d".

hybrid forms based on the Toma and Lyubava varieties, microsatellite analysis confirmed the C-banding result and also revealed *T. dicoccoides* genetic material in other chromosomes (Table 1, Fig. 3).

Thus, in the hybrids Toma × l29 and l29 × Toma, a fragment of alien genetic material was found in the long arm of chromosome 6A, which could not be identified using C-banding due to the small number of heterochromatic blocks

Table 1. The chromosomal localization of *T. dicoccoides* genetic material in wheat hybrids according to C-banding and SSR analysis data

Crossing combination	C-banding	SSR-analysis
Line 29 Rassvet × *T. dicoccoides* k-5199	1BL, 2BS, 3B, 5BL	1BL, 2BS, 3B, 5B, 6AL
Darya × l29	–	3B, 5BL, 6AL
Darya × l29	1BL, 3B	1BL, 3B, 6AL
Laska × l29	–	2BS, 3B
Lyubava × l29	2BS, 5BL	2BS, 3B, 5BL
l29 × Darya	–	2BS, 3B, 5B, 6AL
l29 × Toma	2BS, 3B, 5BL	2BS, 3B, 5BL, 6AL
l29 × Laska	–	2BS, 3B
l29 × Lyubava	2BS, 5BL	2BS, 3BS, 5B

Note. "–" – identification of *T. dicoccoides* genetic material using C-banding was not performed.

in A-genome chromosomes. In the hybrids resulting from the Lyubava variety, in addition to the insertion of alien chromatin in chromosomes 2B and 5B, emmer fragments were found in chromosome 3B, which were not detected due to the similarity of the differential staining patterns of 3B chromosomes of Lyubava and l29. For the hybrids obtained with the involvement of the variety Darya, the inclusion of alien material was found in chromosomes 3B, 5B and 6A, and in the reverse crossing combination, there was also a fragment of emmer in the short arm of chromosome 2B (Fig. 3). The fewest number of introgression fragments among the studied material was found in hybrids resulting from the variety Laska (Table 1, Fig. 3).

The data obtained indicate a high frequency of introgression of *T. dicoccoides* genetic material in the genome of hybrid forms. Among the analyzed progeny of eight crossing combinations, all variants of emmer material introgression characteristic of l29 were identified, and in most cases, they were present in both homologs, which indicates the imminent completion of the karyotype stabilization process. The highest frequency of introgression of *T. dicoccoides* genetic material in the common wheat genome was found for chromosome 3B, and the lowest, for chromosome 1B (in eight and one hybrid combinations out of eight, respectively).

Analysis of cytological stability

Our previous analysis of chromosome behavior at various stages of microsporogenesis in common wheat lines with alien genetic material showed that l29 is one of the most stable genotypes among introgressive lines. The level of chromosome pairing at the stage of metaphase I in l29 was high: the number of chromosomes constituting bivalents was 99.76 %. The meiotic index (an important indicator of the normal course of the entire meiosis) of this genotype was maximum among the lines with the introgression of wild emmer genetic material – 93.0 % (Orlovskaya et al., 2023b). The high cytological stability of introgressive line 29 makes it possible to assume that all stages of meiosis in the hybrids developed with its involvement will also proceed without significant disturbances.

Metaphase I analysis in pollen mother cells (PMC) in the studied wheat genotypes showed a high level of bivalent chromosome pairing in both parental varieties and hybrid material. The number of chromosomes included in the bivalents exceeded 99 % (Table 2).

In the studied material, cells with only two univalents were detected, and their frequency of occurrence was insignificant (3.33–6.67 %). The high level of chromosome pairing in metaphase I ensured an insignificant number of disturbances at the subsequent stages of meiosis (Fig. 4). Moreover, as a rule, there were fewer disturbances in the second division (anaphase II) than in the first. In anaphase II, a decrease in the number of PMC with bridges (1.25–6.25 %) and asynchronous division (1.43–16.25 %) could be noted compared to the previous phases of meiosis: 2.5–14.28 and 3.64–18.64 %, respectively. At the final stage, the percentage of normal tetrads in all the studied material exceeded 90 (Fig. 4). The main disturbance at that stage was the presence of micronuclei in the tetrads, the number of which varied from 1 to 4; however, tetrads with 1 and 2 micronuclei were most often formed. PMC with 4 micronuclei were characteristic of only three hybrids (l29 × Laska; Laska × l29; and Darya × l29), and their frequency of occurrence was only 0.91–1.11 % of the total number of analyzed cells.

The number of pollen mother cells without aberrations was, as a rule, slightly higher in the hybrids of the reverse crossing combinations, but at the final stage, no significant differences were found between the hybrids of different crossing directions (Fig. 4). The maximum meiotic index (98 %) among the hybrid material was noted in the combination Lyubava × l29.

Thus, F_4 hybrids resulting from the crossing of modern varieties of common spring wheat with the introgressive line 29 Rassvet × *T. dicoccoides* k-5199 at all the studied meiotic stages showed a high level of stability comparable to that of parental varieties, which ensures the successful reproduction of the developed hybrid material.

The mineral content of grain

An increase in the mineral content of grain of major agricultural crops is a pressing issue, since micronutrient deficiency (the so-called "hidden hunger") produces a significant impact on human health and well-being (Gupta et al., 2021). In this regard, we assessed the mineral content of grain of second-

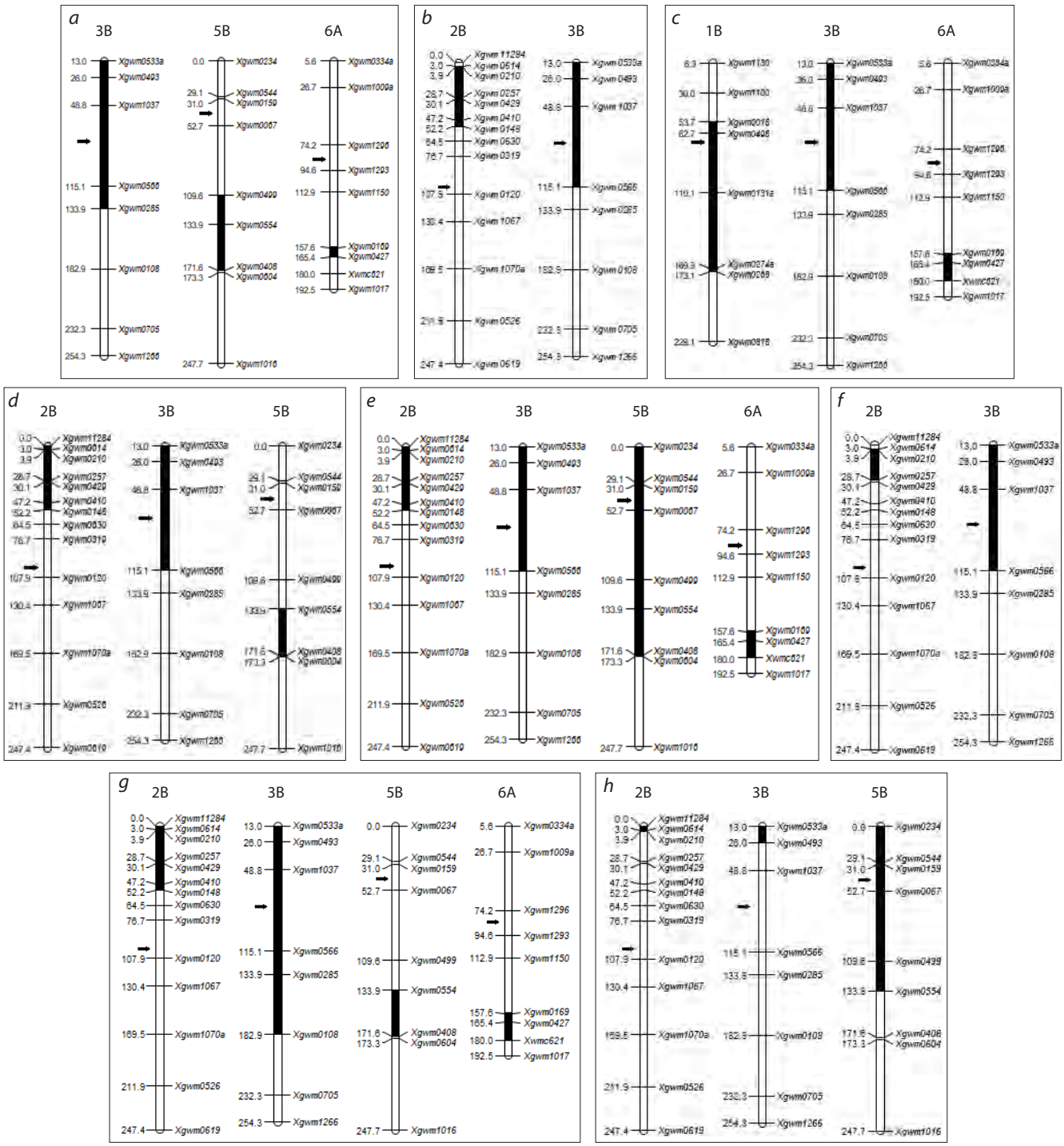


Fig. 3. A schematic illustration of *T. dicoccoides* genome fragments in hybrids.

a – Darya × l29; *b* – Laska × l29; *c* – Toma × l29; *d* – Lyubava × l29; *e* – l29 × Darya; *f* – l29 × Laska; *g* – l29 × Toma; *h* – l29 × Lyubava. The order of microsatellite markers corresponds to the genetic maps of *T. aestivum* chromosomes (Somers et al., 2004). Dark blocks indicate the length of the introgression fragments. To the left of the chromosome, the distance between the markers in cM is demonstrated; the centromere position is indicated by an arrow.

ary introgressive common wheat hybrids in comparison with the parental forms. It was found that over two years the micronutrient content in l29 was on average as follows: Cu – 3.1; Fe – 54.4; Mn – 25.1; and Zn – 30.9 mg/kg (Fig. 5), which turned out to be significantly higher than in the group of varieties where the values were 1.9, 45.4, 20.6 and 27.4 mg/kg, respectively.

In both years of study, this introgressive line surpassed all the studied varieties in the content of Cu, Fe and Mn, but it was slightly inferior to the variety Lyubava in the content of Zn (Fig. 5). On average, for 2021–2022, most hybrids accumulated more Cu, Fe and Mn than the original variety (a statistically significant excess in the Fe content was found for the genotypes Toma × l29; Laska × l29; l29 × Lyubava;

Table 2. Chromosome behavior in metaphase I of meiosis of secondary introgressive F_4 common wheat hybrids and their parental forms

Genotype	Bivalents, items				Univalents, items
	Closed	Opened	In total	%	
Darya	20.60 ± 0.12	0.40 ± 0.12	21.00 ± 0.00	100	–
Toma	19.87 ± 0.24	1.07 ± 0.24	20.93 ± 0.05	99.67	0.14 ± 0.09
Laska	20.65 ± 0.13	0.3 ± 0.13	20.95 ± 0.05	99.76	0.10 ± 0.10
Lyubava	19.3 ± 0.21	1.47 ± 0.21	20.93 ± 0.07	99.97	0.14 ± 0.13
Darya × l29	20.07 ± 0.20	0.90 ± 0.19	20.97 ± 0.03	99.86	0.06 ± 0.07
Toma × l29	20.40 ± 0.16	0.60 ± 0.16	21.00 ± 0.00	100	0
Laska × l29	19.23 ± 0.66	1.03 ± 0.18	20.93 ± 0.05	99.67	0.14 ± 0.09
Lyubava × l29	20.27 ± 0.16	0.73 ± 0.16	21.00 ± 0.00	100	0
l29 × Darya	19.97 ± 0.19	1.03 ± 0.19	21.00 ± 0.00	100	0
l29 × Toma	20.37 ± 0.15	0.60 ± 0.14	20.97 ± 0.03	99.86	0.06 ± 0.07
l29 × Laska	20.17 ± 0.21	0.80 ± 0.21	20.97 ± 0.03	99.86	0.06 ± 0.07
l29 × Lyubava	20.33 ± 0.15	0.60 ± 0.15	20.93 ± 0.05	99.67	0.14 ± 0.09

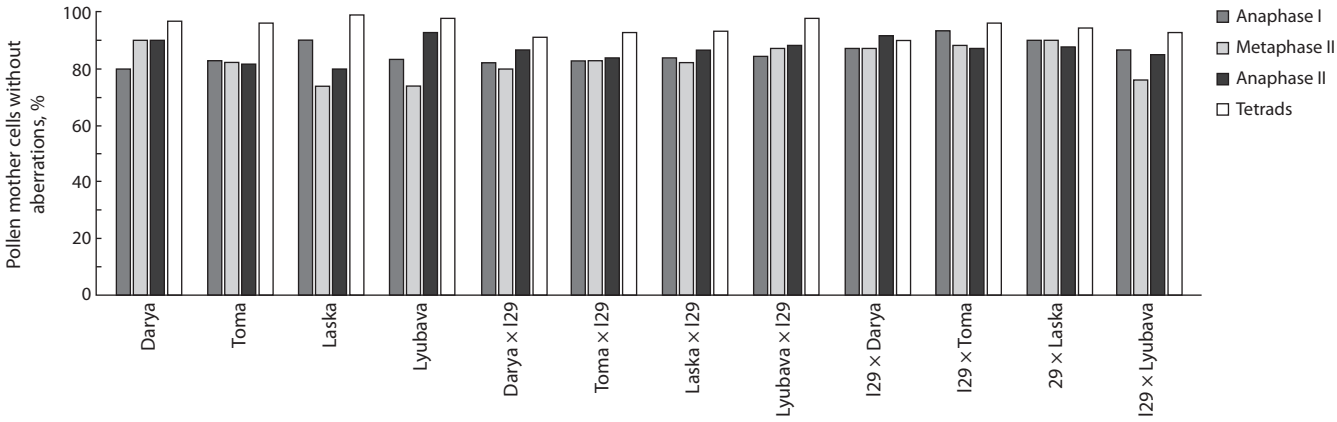


Fig. 4. Pollen mother cells without aberrations (%) at different meiotic stages of secondary introgressive F_4 common wheat hybrids and parental varieties.

and in the Mn content, for Toma × l29; l29 × Darya and l29 × Toma). At the same time, in 2022, the content of all the studied microelements in the hybrids was, as a rule, significantly higher than that of the original variety. In 2021, such tendency was noted for Fe content in the genotypes based on the variety Darya; for Mn content, in those based on the varieties Darya and Toma; for Cu content – genotypes Toma × l29 and Laska × l29. For the hybrids obtained with the involvement of varieties Darya and Toma, a high amount of Zn was also shown, significantly exceeding that in the parent variety (Fig. 5).

The highest level of Cu and Fe accumulation was observed in the hybrids Darya × l29 and l29 × Darya. In particular, the average Cu content in the grain of these hybrids over two years was 2.7 and 2.6 mg/kg, respectively, and the iron content was 50.3 and 49.7 mg/kg, respectively. The highest Mn content was observed in the hybrids Toma × l29 and Lyubava × l29 (24.0 and 23.6 mg/kg, respectively), and the highest Zn content was found in the hybrids l29 × Darya and l29 × Toma (30.7 and 29.2 mg/kg, respectively). The hybrids exceeded all the studied genotypes in the amount of Cu, Fe and Mn in grain, except for l29; and in Zn content – except for l29 and Lyubava.

As for macroelements, l29 was inferior to common wheat varieties (478.5 mg/kg) in Ca content (on average 369.0 mg/kg over two years); the hybrids did not reach the level of the parent variety, either (Table S1)[1]. As for macroelements, l29 slightly exceeded the varieties. Thus, on average over two years, the content of K, Mg and P in l29 was 5,207.8; 1,370.0 and 4,505.5 mg/kg, respectively; and in the group of varieties, 5,083.7; 1,364.9 and 4,137.8 mg/kg. In both years of study, a significant advantage over the original variety was revealed only for the hybrids resulting from the variety Laska in terms of K content and for the l29 × Daria hybrid – by P content. A reliable excess over the variety in terms of Mg content was observed only in 2022 in reverse crossing combinations (except for l29 × Lyubava).

[1] Supplementary Tables S1–S3 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Orl_Engl_29_8.pdf

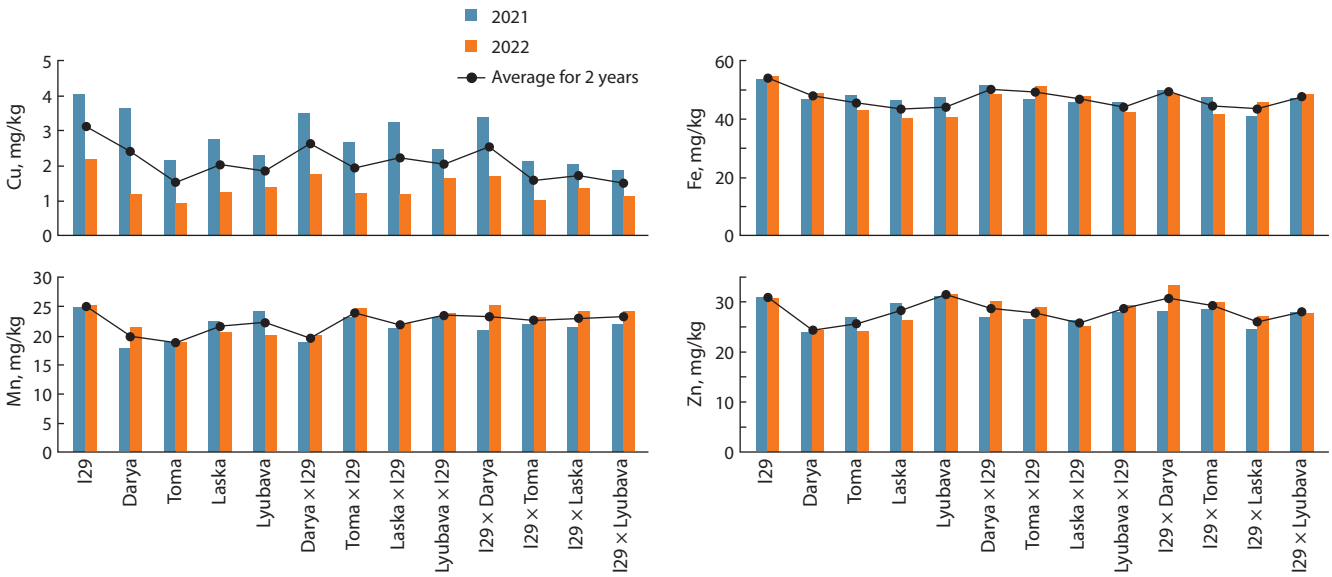


Fig. 5. The content of microelements in the grain of secondary common spring wheat hybrids and their parental forms in 2021 and 2022; average.

Main grain quality indicators

When assessing grain quality, such indicators as protein and gluten content and gluten quality are of great importance, since they determine the nutritional and baking capacity of wheat grain. High-protein genotypes are of particular interest to the breeders across the globe, since a high role of protein in the formation of wheat grain quality has been established (Tanin et al., 2022).

In both years of our study, the level of protein and gluten accumulation in l29 was higher compared to the modern varieties of the Belarusian selection. Thus, the protein content in l29 in 2021 and 2022 was 18.3 and 16.5 %, respectively; and in the varieties, 17.1–17.6 and 14.9–16.1 %, respectively. The gluten content in l29 in 2021 and 2022 was 43.0 and 28.1 %, respectively; in the varieties, 30.7–32.7 and 23.6–27.8 %, respectively. Most hybrids in both years of study were characterized by a higher protein and gluten content in grain compared to the parent variety (Fig. 6). At that, the hybrids l29 × Darya and l29 × Toma surpassed all genotypes in the protein content. The amount of gluten in those hybrids was inferior only to the genotype l29.

An excess of the studied hybrids over the original variety by the trait "gluten quality" in both years was found for the combinations based on the variety Darya. In 2022, such tendency was typical for most hybrid genotypes (Fig. 6). Over two years on average, the hybrids Lyubava × l29 (85.3 GDI conventional units) and Laska × l29 (85.5 GDI conventional units) had the lowest GDI values among all hybrid material and were slightly inferior in gluten quality only to the variety Laska (84.8 GDI conventional units).

Productivity indicators

Currently, wheat varieties that combine high grain quality and productivity indicators are of particular value. We have evaluated the obtained hybrids by the main quantitative traits (Tables S2, S3). On average, over two years, l29 exceeded the studied varieties by all the studied quantitative traits, but statistically significant superiority was noted only for the "spike length" and "grain weight per plant" indicators. In particular years, the introgressive line was characterized by significantly higher productive tillering (2022); grain weight per spike (2021); and thousand grain weight (2021). On average, for the "hybrids" group, higher values were noted for all the traits (except for thousand grain weight) compared to the "varieties" group, while a significant advantage was found for spike productivity traits (Table 3).

The main spike makes a great contribution to the overall plant productivity. The spike length, the number of spikelets and grains in the main spike in the studied hybrids with a high level of confidence exceeded the original variety or were at its level. In terms of grain productivity of a spike, the hybrids Laska × l29 and Lyubava × l29 can be distinguished, which surpassed not only the parent varieties, but also other hybrids in the grain number and weight in the main spike (Table S3). By the traits "grain weight per plant" and "thousand grain weight", no significant differences were found between the hybrids and the parent varieties: the indicators were within 3.2–4.3 and 33.8–41.0 g for varieties and 3.4–4.5 and 33.1–38.2 g for hybrids, respectively.

Discussion

It is known that *T. dicoccoides* samples have a number of economically useful traits: high protein and microelement content in grain, and resistance to biotic and abiotic factors (Chatzav et al., 2010; Lucas et al., 2017; Mohammadi et al., 2021). However, when crossing common wheat varieties with wild emmer, hybrids, along with valuable traits, can also inherit undesirable properties of their wild relative (low productivity, spike fragility and hard threshing) (Özkan et al., 2011). In addition, during interspecific hybridization, introgression that affects the functioning of the main genes of chromosome synapsis, which leads to a significant decrease in the meiotic index and a long formative process (cleavage according to morphological and economically useful traits), may occur.

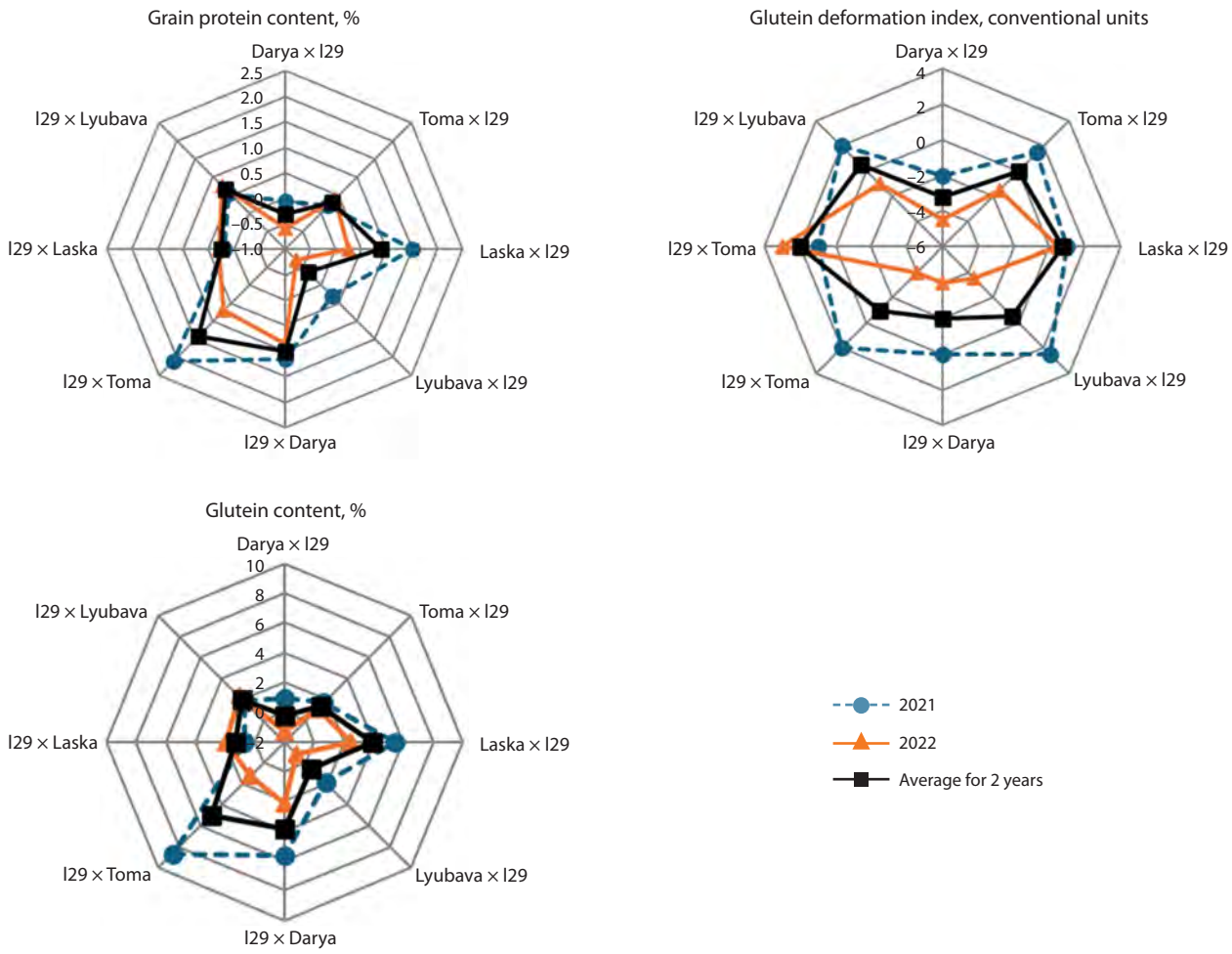


Fig. 6. The difference between the mean values of the grain quality traits of wheat hybrids and their parent varieties in 2021, 2022 and the average for two years.

Table 3. Quantitative trait values in wheat genotype groups for two years (X ± SEM)

Trait	Line 29	Varieties	Hybrids
Productive tillering, items	3.5 ± 0.2	3.2 ± 0.1	3.3 ± 0.1
Spike length, cm	10.3 ± 0.2*	9.8 ± 0.0	10.3 ± 0.1*
Number of spikelets per spike, items	19.6 ± 0.3	18.9 ± 0.2	19.8 ± 0.1*
Number of grains per spike, items	46.6 ± 1.2	42.2 ± 1.2	46.4 ± 0.8*
Grain weight per spike, g	1.9 ± 0.1	1.6 ± 0.1	1.7 ± 0.1
Grain weight per plant, g	4.7 ± 0.5*	3.7 ± 0.2	4.1 ± 0.1
Thousand grain weight, g	39.9 ± 1.1	36.6 ± 1.3	35.4 ± 0.9

* Statistically significant differences compared to the "varieties" group, where $p < 0.05$; X – mean trait values; and SEM – the standard error of the mean.

Cytogenetically stable forms with the inclusion of alien genetic material that has a positive effect on economically important traits are of interest for genetic and selection research. It is believed that when replacing a chromosome arm, a negative effect resulting from introgression is weaker than in the case where an entire wheat chromosome is replaced with an alien chromosome (Millet et al., 2013). Line 29 of the Rassvet × *T. dicoccoides* combination we have obtained is of high cytological stability, despite the presence of fragments of an alien genome in five chromosomes, including chromosomes 3B and 5B, where the main genes regulating the meiosis process in wheat are localized (Bhullar et al., 2014; Darrier et al., 2022). The value of this line is that, along with a high meiotic index and improved nutritional properties, l29 is at the level of modern common wheat varieties by the main quantitative indicators.

The inclusion of alien genetic material in the genome of all hybrids obtained from the crossing of varieties with l29 was

established. In the progeny developed with the involvement of varieties Darya, Toma, Laska and Lyubava, the number and combination of variants of introduced emmer fragments are different, which indicates the influence exerted on the process of fixation of alien genetic material in the hybrid genome of a wheat variety, or more precisely, the genotypic environment of the hybrid form at its development stage. At that, all hybrids were already cytologically stable in the fourth generation, which ensures the successful reproduction of produced hybrid material.

Our study demonstrated the efficiency of improving modern wheat varieties in terms of the protein, gluten and mineral content of grain using an introgressive wheat line as a source of wild emmer genetic material. The hybrids that in both years of study significantly surpassed the parental varieties in terms of the accumulation of the following elements were obtained: Cu – Toma × l29, l29 × Darya; Mn – Toma × l29, l29 × Darya, l29 × Toma; Zn – Darya × l29, l29 × Darya, l29 × Toma; K – Laska × l29, l29 × Laska; P – l29 × Darya; protein and gluten – Toma × l29, Laska × l29; and all reverse crossing combinations.

Moreover, all hybrid genotypes under study were not inferior to the varieties by the main quantitative traits and sometimes surpassed them: l29 × Toma – by the length of the main spike; Toma × l29 and l29 × Lyubava – by the number of spikelets in the spike; Lyubava × l29, l29 × Toma and l29 × Lyubava – by the number of grains in the spike. Comparison of secondary introgressive hybrids with the standard of common spring wheat (variety Lyubava) showed that throughout the entire period of study, Darya × l29 statistically significantly exceeded the standard in the content of Cu, Fe and K; l29 × Darya – Cu, Fe, K, P, protein and gluten; l29 × Toma and l29 × Lyubava – protein and gluten; Laska × l29 and l29 × Laska – Ca. In other traits, no significant differences were found between the standard variety and the hybrids. Only the number of grains per spike in the variety Lyubava was significantly smaller compared to the hybrids (except for Darya × l29 and l29 × Lyubava).

At present, increased attention is given to the search for genes that affect the content of micro- and macroelements in wheat grain. The largest number of studies in this area is related to zinc and iron as the most important microelements in the formation of plant productivity. For example, a number of loci associated with the content of these microelements were identified using QTL (quantitative trait loci) mapping and GWAS (genome wide association study) (Hao et al., 2014; Velu et al., 2018). At the CIMMYT breeding center, the use of GWAS to analyze the genetics of Zn accumulation in grain based on the material of 330 lines of common wheat allowed identifying 39 marker-trait associations. Two QTL regions with a large effect on the studied trait were found on chromosomes 2B and 7B (Velu et al., 2018). In the work of J. Tong et al. (2020), based on the information about the genes involved in Zn and Fe homeostasis in model plants (arabidopsis and rice), 254 orthologs were identified in wheat. The genes were found on all wheat chromosomes, with their largest number on the second (23 %), fifth (15 %) and third (14 %) homologous groups of chromosomes (Tong et al., 2020).

In our study, the maximum level of Zn accumulation in grain was observed in the l29 × Darya and l29 × Toma hybrids with the fragments of wild emmer genetic material in chromosomes 2BS, 3B, 5B and 6AL. At that, the quantitative trait loci associated with Zn and Fe content in wheat grain (described in the work of J. Tong et al. (2020)) coincide in their localization with the regions of introgression of alien genetic material in chromosomes 2B, 3B and 5B in these hybrids. It can be noted that these genotypes also had a high protein and gluten content of grain exceeding the parent varieties and other hybrids. High rates by these traits were also found in the hybrid Toma × l29 with the inclusion of alien material in chromosomes 1BL, 3B and 6AL.

In the works of foreign scientists, there are data on the role of various chromosomes in regulating the process of protein accumulation in wheat grain (Liu et al., 2019; Alemu et al., 2021). For example, the loci associated with protein and gluten content in the grains of plants grown under different environmental conditions were found on chromosome 5B (Gonzalez-Hernandez et al., 2004; Giancaspro et al., 2019; Alemu et al., 2021). Genetic regions controlling protein content in wheat were also identified on chromosomes 3A and 3B (Kartseva et al., 2023), as well as 2B and 6A (Muqaddasi, et al., 2020). The review by C. Paina and P.L. Gregersen (2023) provides data on the presence of 325 QTL on all the wheat chromosomes involved in the regulation of protein accumulation in grain. Some of these loci exert a negative effect on yield. However, a study of wheat lines obtained from the crossing with wild emmer (Liu et al., 2019; Fatiukha et al., 2020) did not show any significant association between the protein content and the thousand grain weight. At present, there are data on the identification of regions of the wheat genome that produce a positive effect on the protein accumulation in grain without affecting yield, as well as regions with a favorable effect on both indicators (Thorwarth et al., 2019; Ruan et al., 2021), which indicates the possibility of overcoming negative correlation between the two economically valuable traits and developing genotypes with a high grain protein content without reducing productivity.

Conclusion

The efficiency of using a wheat line with the inclusion of *T. dicoccoides* genetic material to enhance modern varieties in terms of the protein content, gluten and mineral composition of grain without reducing productivity is demonstrated. C-banding and microsatellite analysis data indicate a high frequency of wild emmer genetic material introgression in the genome of hybrid forms obtained from the crossing of common spring wheat varieties Darya, Toma, Laska and Lyubava of the Belarusian selection with the introgressive line 29 Rassvet × *T. dicoccoides* k-5199 (four direct and four reverse crossing combinations).

Among the progeny of eight crossing combinations analyzed, all introgression variants of alien genetic material characteristic of line 29 were revealed, and in most cases, they are present in both homologs, which indicates the stabilization of the karyotype. All F_4 hybrids are characterized by a high level of cytological stability (the meiotic index was 90.0–98.0 %). Secondary introgressive hybrids surpassing parent varieties by a complex of grain quality traits in both years were identified: l29 × Darya (protein, gluten, Cu, Mn, Zn and P content); l29 × Toma (protein, gluten, Mn and Zn

content); Laska × l29 and l29 × Laska (protein, gluten and K content); and Toma × l29 (protein, gluten, Cu and Mn content). In addition, these hybrids are not inferior to modern common wheat varieties in terms of the main productivity indicators, which increases their value for breeding.

References

Akcura M., Kokten K. Variations in grain mineral concentrations of Turkish wheat landraces germplasm. *Qual Assur Saf Crops Foods*. 2017;9(2):153-159. doi 10.3920/QAS2016.0886

Alemu A., El Baouchi A., El Hanafi S., Kehel Z., Eddakhir K., Tadesse W. Genetic analysis of grain protein content and dough quality traits in elite spring bread wheat (*Triticum aestivum*) lines through association study. *J Cereal Sci*. 2021;100:103214. doi 10.1016/j.jcs.2021.103214

Arora S., Cheema J., Poland J., Uauy C., Chhuneja P. Genome-wide association mapping of grain micronutrients concentration in *Aegilops tauschii*. *Front Plant Sci*. 2019;10:54. doi 10.3389/fpls.2019.00054

Badaeva E.D., Badaev N.S., Gill D.S., Filatenko A.A. Intraspecific karyotype divergence in *Triticum araraticum* (*Poaceae*). *Plant Syst Evol*. 1994;192(1):117-145. doi 10.1007/BF00985912

Bhullar R., Nagarajan R., Bennypaul H., Sidhu G.K., Sidhu G., Rustgi S., von Wettstein D., Gill K.S. Silencing of a metaphase I-specific gene results in a phenotype similar to that of the pairing homeologous 1 (*Ph1*) gene mutations. *Proc Natl Acad Sci USA*. 2014;111(39):14187-14192. doi 10.1073/pnas.1416241111

Cakmak I., Ozkan H., Braun H.J., Welch R.M., Romheld V. Zinc and iron concentration in seeds of wild, primitive and modern wheats. *Food Nutr Bull*. 2000;21(4):401-403. doi 10.1177/15648265000 2100411

Chatzav M., Peleg Z., Ozturk L., Yazici A., Fahima T., Cakmak I., Saranga Y. Genetic diversity for grain nutrients in wild emmer wheat: potential for wheat improvement. *Ann Bot*. 2010;105(7):1211-1220. doi 10.1093/aob/mcq024

Darrier B., Colas I., Rimbert H., Choulet F., Bazile J., Sortais A., Jenczewski E., Sourdille P. Location and identification on chromosome 3B of bread wheat of genes affecting chiasma number. *Plants*. 2022;11(17):2281. doi 10.3390/plants11172281

Fatiukha A., Filler N., Lupo I., Lidzbarsky G., Klymiuk V., Korol A.B., Pozniak C., Fahima T., Krugman T. Grain protein content and thousand kernel weight QTLs identified in a durum × wild emmer wheat mapping population tested in five environments. *Theor Appl Genet*. 2020;133(1):119-131. doi 10.1007/s00122-019-03444-8

Giancaspro A., Giove S.L., Zacheo S.A., Blanco A., Gadaleta A. Genetic variation for protein content and yield-related traits in a durum population derived from an inter-specific cross between hexaploid and tetraploid wheat cultivars. *Front Plant Sci*. 2019;10:1509. doi 10.3389/fpls.2019.01509

Goel S., Singh B., Gwal S., Jaat R.S., Singh N.K. Variability in Fe and Zn content among Indian wheat landraces for improved nutritional quality. *Indian J Genet Plant Breed*. 2018;78(4):426-432. doi 10.31742/IJGPB.78.4.4

Gonzalez-Hernandez J., Elias E., Kianian S. Mapping genes for grain protein concentration and grain yield on chromosome 5B of *Triticum turgidum* (L.) var. *dicoccoides*. *Euphytica*. 2004;139:217-225. doi 10.1007/s10681-004-3157-4

Gupta P.K., Balyan H.S., Sharma S., Kumar R. Biofortification and bioavailability of Zn, Fe and Se in wheat: present status and future prospects. *Theor Appl Genet*. 2021;134(1):1-35. doi 10.1007/s00122-020-03709-7

Hao Y., Velu G., Peña R.J., Singh S., Singh R.P. Genetic loci associated with high grain zinc concentration and pleiotropic effect on kernel weight in wheat (*Triticum aestivum* L.). *Mol Breed*. 2014;34(4):1893-1902. doi 10.1007/s11032-014-0147-7

Heidari B., Padash S., Dadkhodaie A. Variations in micronutrients, bread quality and agronomic traits of wheat landrace varieties and commercial cultivars. *Aust J Crop Sci*. 2016;10(3):377-384. doi 10.21475/ajcs.2016.10.03.p7231

Kartseva T., Alqudah A.M., Aleksandrov V., Alomari D.Z., Doneva D., Arif M.A.R., Börner A., Misheva S. Nutritional genomic approach for improving grain protein content in wheat. *Foods*. 2023;12(7):1399. doi 10.3390/foods12071399

Liu J., Huang L., Wang C., Liu Y., Yan Z., Wang Z., Xiang L., Zhong X., Gong F., Zheng Y., Liu D., Wu B. Genome-wide association study reveals novel genomic regions associated with high grain protein content in wheat lines derived from wild emmer wheat. *Front Plant Sci*. 2019;10:464. doi 10.3389/fpls.2019.00464

Liu J., Huang L., Li T., Liu Y., Yan Z., Tang G., Zheng Y., Liu D., Wu B. Genome-wide association study for grain micronutrient concentrations in wheat advanced lines derived from wild emmer. *Front Plant Sci*. 2021;12:651283. doi 10.3389/fpls.2021.651283

Lucas S.J., Salantur A., Yazar S., Budak H. High-throughput SNP genotyping of modern and wild emmer wheat for yield and root morphology using a combined association and linkage analysis. *Funct Integr Genomics*. 2017;17(6):667-685. doi 10.1007/s10142-017-0563-y

Marcos-Barbero E.L., Pérez P., Martínez-Carrasco R., Arellano J.B., Morcuende R. Genotypic variability on grain yield and grain nutritional quality characteristics of wheat grown under elevated CO_2 and high temperature. *Plants*. 2021;10(6):1043. doi 10.3390/plants 10061043

Millet E., Rong J.K., Qualset C.O., McGuire P.E., Bernard M., Sourdille P., Feldman M. Production of chromosome-arm substitution lines of wild emmer in common wheat. *Euphytica*. 2013;190(1):1-17. doi 10.1007/s10681-012-0725-x

Mohammadi M., Mirlohi A., Majidi M.M., Soleimani Kartalaei E. Emmer wheat as a source for trait improvement in durum wheat: a study of general and specific combining ability. *Euphytica*. 2021;217:64. doi 10.1007/s10681-021-02796-x

Muqaddasi Q.H., Brassac J., Ebmeyer E., Kollers S., Korzun V., Argillier O., Stiewe G., Plieske J., Ganal M.W., Röder M.S. Prospects of GWAS and predictive breeding for European winter wheat's grain protein content, grain starch content, and grain hardness. *Sci Rep*. 2020;10(1):12541. doi 10.1038/s41598-020-69381-5

Orlovskaya O., Dubovets N., Solovey L., Leonova I. Molecular cytological analysis of alien introgressions in common wheat lines derived from the cross of *Triticum aestivum* with *T. kiharae*. *BMC Plant Biol*. 2020;20(Suppl. 1):201. doi 10.1186/s12870-020-02407-2

Orlovskaya O.A., Vakula S.I., Khotyleva L.V., Kilchevsky A.V. Mineral composition of bread wheat lines with introgressions of alien genetic material. *Trudy po Prikladnoy Botanike, Genetike i Selektsii = Proceedings on Applied Botany, Genetics and Breeding*. 2023a;184(1):42-52. doi 10.30901/2227-8834-2023-1-42-52 (in Russian)

Orlovskaya O.A., Leonova I.N., Solovey L.A., Dubovets N.I. Molecular cytological analysis of alien introgres sions in common wheat lines created by crossing of *Triticum aestivum* with *T. dicoccoides* and *T. dicoccum*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023b;27(6):553-564. doi 10.18699/VJGB-23-67

Özkan H., Willcox G., Graner A., Salamini F., Kilian B. Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genet Resour Crop Evol*. 2011;58:11-53. doi 10.1007/s10722-010-9581-5

Paina C., Gregersen P.L. Recent advances in the genetics underlying wheat grain protein content and grain protein deviation in hexaploid wheat. *Plant Biol*. 2023;25(5):661-670. doi 10.1111/plb.13550

Pausheva Z.P. Practicum on Plant Cytology. Moscow, 1988 (in Russian)

Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P., Ganal M.W. A microsatellite map of wheat. *Genetics*. 1998;149:2007-2023. doi 10.1093/genetics/149.4.2007

Ruan Y., Yu B., Knox R.E., Zhang W., Singh A.K., Cuthbert R., Fobert P., DePauw R., Berraies S., Sharpe A., Fu B.X., Sangha J. Conditional mapping identified quantitative trait loci for grain protein concentration expressing independently of grain yield in Canadian durum wheat. *Front Plant Sci*. 2021;12:642955. doi 10.3389/fpls.2021.642955

Savin T.V., Abugaliyeva A.I., Cakmak I., Kozhakhmetov K. Mineral composition of wild relatives and introgressive forms in wheat selection. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2018;22(1):88-96. doi 10.18699/VJ18.335 (in Russian)

Shewry P.R., Pellny T.K., Lovegrove A. Is modern wheat bad for health? *Nat Plants*. 2016;2(7):16097. doi 10.1038/nplants.2016.97

Somers D.J., Isaac P., Edwards K. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor Appl Genet*. 2004;109(6):1105-1114. doi 10.1007/s00122-004-1740-7

Tanin M.J., Sharma A., Saini D.K., Singh S., Kashyap L., Srivastava P., Mavi G.S., Kaur S., Kumar V., Kumar V., Grover G., Chhuneja P., Sohu V.S. Ascertaining yield and grain protein content stability in wheat genotypes having the *Gpc-B1* gene using univariate, multivariate, and correlation analysis. *Front Genet*. 2022;13:1001904. doi 10.3389/fgene.2022.1001904

Thorwarth P., Liu G., Ebmeyer E., Schacht J., Schachschneider R., Kazman E., Reif J.C., Würschum T., Longin C. Dissecting the genetics underlying the relationship between protein content and grain yield in a large hybrid wheat population. *Theor Appl Genet*. 2019;132(2):489-500. doi 10.1007/s00122-018-3236-x

Tiwari V.K., Rawat N., Neelam K., Kumar S., Randhawa G.S., Dhaliwal H.S. Substitutions of 2S and 7U chromosomes of *Aegilops kotschyi* in wheat enhance grain iron and zinc concentration. *Theor Appl Genet*. 2010;121(2):259-269. doi 10.1007/s00122-010-1307-8

Tong J., Sun M., Wang Y., Zhang Y., Rasheed A., Li M., Xia X., He Z., Hao Y. Dissection of molecular processes and genetic architecture underlying iron and zinc homeostasis for biofortification: from model plants to common wheat. *Int J Mol Sci*. 2020;21(23):9280. doi 10.3390/ijms21239280

Velu G., Singh R.P., Crespo-Herrera L., Juliana P., Dreisigacker S., Valluru R., Stangoulis J., … Balasubramaniam A., Chatrath R., Gupta V., Singh G.P., Joshi A.K. Genetic dissection of grain zinc concentration in spring wheat for mainstreaming biofortification in CIMMYT wheat breeding. *Sci Rep*. 2018;8(1):13526. doi 10.1038/s41598-018-31951-z

Zeibig F., Kilian B., Frei M. The grain quality of wheat wild relatives in the evolutionary context. *Theor Appl Genet*. 2022;135(11):4029-4048. doi 10.1007/s00122-021-04013-8

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

Received March 31, 2025. Revised June 24, 2025. Accepted June 25, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1267-1276

doi 10.18699/vjgb-25-131

Flax transposons: unraveling their impact on domestication and agronomic trait variation

M.A. Duk [2], V.A. Stanin[1], A.A. Kanapin [1], A.A. Samsonova [1], T.A. Rozhmina[3], M.G. Samsonova [1]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
[2] Ioffe Institute of the Russian Academy of Sciences, St. Petersburg, Russia
[3] Federal Research Center for Bast Fiber Crops, Torzhok, Russia
m.g.samsonova@gmail.com

Abstract. Flax is an important agricultural crop with multifunctional uses. Diversified breeding for oil content in seeds and fiber in stems has led to the emergence of two morphotypes – fiber flax and oilseed flax. Previously, using single nucleotide polymorphisms (SNPs), we characterized the genetic diversity of 306 flax samples from the collection of the Russian Federal Research Center for Bast Crops. However, larger structural variations, such as mobile genetic elements, also play a significant role in shaping agronomically important plant traits and can be used for further flax improvement. Here, we used the same flax collection to predict sites of new transposon insertions and to assess the role of such insertions in the formation of agronomically important traits, as well as in the process of flax domestication. We discovered 588,480 new transposon insertion sites not present in the reference flax genome (NCBI assembly ASM22429v2), the majority of which were attributed to retrotransposons of the *Copia* and *Gypsy* superfamilies, while among DNA transposons, insertion sites of the *MULE-MuDR*, *hAT*, and *CMC-EnSpm* superfamilies were most common. Unlike SNPs, which were significantly more numerous in oilseed flax than in fiber flax, we did not find such a substantial difference in the number of insertions of different transposon families per plant among samples of different morphotypes. Analysis of genomic regions affected by recent breeding efforts revealed a total of 61 candidate regions, of which 18 regions overlapped with QTLs associated with important agronomic traits. Interestingly, 5 regions of reduced genetic diversity in kryazhs and cultivars compared to landraces were also identified as regions of reduced diversity when using single nucleotide polymorphisms as markers. A genome-wide association study (GWAS) identified 50 TE insertions associated with different phenotypic traits, with many associations confirmed by multiple models or detected in data from multiple years. Thus, transposon insertion sites are an important source of genetic diversity in flax, alongside single nucleotide polymorphisms, making them suitable for further crop improvement in breeding.

Key words: flax; *Linum usitatissimum*; transposons; GWAS; genetic diversity; breeding

For citation: Duk M.A., Stanin V.A., Kanapin A.A., Samsonova A.A., Rozhmina T.A., Samsonova M.G. Flax transposons: unraveling their impact on domestication and agronomic trait variation. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1267-1276. doi 10.18699/vjgb-25-131

Funding. The work was supported by the RSF grant No. 23-16-00037.

Транспозоны льна: роль в генетическом разнообразии, окультуривании и детерминации хозяйственно ценных признаков

М.А. Дук [2], В.А. Станин[1], А.А. Канапин [1], А.А. Самсонова [1], Т.А. Рожмина[3], М.Г. Самсонова [1]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия
[2] Физико-технический институт им. А.Ф. Иоффе Российской академии наук, Санкт-Петербург, Россия
[3] Федеральный научный центр лубяных культур, Торжок, Россия
m.g.samsonova@gmail.com

Аннотация. Лен – важная сельскохозяйственная культура многофункционального использования. Разнонаправленная селекция на содержание масла в семенах и волокна в стеблях привела к возникновению двух морфотипов – льна-долгунца и льна масличного. Ранее, используя однонуклеотидные полиморфизмы, мы охарактеризовали генетическое разнообразие у 306 образцов льна из коллекции российского Федерального научного центра лубяных культур. Вместе с тем более крупные структурные вариации также играют существенную роль в формировании агрономически важных признаков растений и могут быть применены для дальнейшего улучшения льна. Здесь мы задействовали ту же коллекцию льна, чтобы предсказать сайты

новых инсерций транспозонов и оценить роль таких инсерций в формировании агрономически важных признаков, а также в процессе окультуривания льна. Мы обнаружили 588 480 новых сайтов инсерций транспозонов, не содержащихся в референсном геноме льна (сборка NCBI ASM22429v2), из которых большая часть приходилась на ретротранспозоны суперсемейств *Copia* и *Gypsy*, а среди ДНК транспозонов чаще всего встречались сайты инсерции суперсемейств *MULE-MuDR*, *hAT* и *CMC-EnSpm*. В отличие от однонуклеотидных полиморфизмов, которых было значительно больше у льна масличного, чем у долгунцов, мы не обнаружили столь существенной разницы в числе инсерций разных семейств транспозонов на одно растение у образцов разного морфотипа. Анализ геномных областей, затронутых недавними селекционными усилиями, выявил в общей сложности 61 район-кандидат, из которых 18 районов пересекались с QTL, ассоциированными с важными агрономическими признаками. Интересно, что пять участков уменьшения генетического разнообразия у культурных сортов и кряжей при сравнении их со староместными сортами также были идентифицированы как участки уменьшения разнообразия при использовании в качестве маркеров однонуклеотидных полиморфизмов. При полногеномном поиске ассоциаций найдено 50 инсерций ТЕ, ассоциированных с разными фенотипическими признаками, причем многие ассоциации подтверждаются несколькими моделями или обнаруживаются в данных по нескольким годам. Таким образом, сайты инсерции транспозонов – важный источник генетического разнообразия у льна наряду с однонуклеотидными полиморфизмами, что позволяет использовать их для дальнейшего улучшения культуры при селекции.
Ключевые слова: лен; *Linum usitatissimum*; транспозоны; GWAS; генетическое разнообразие; селекция

Introduction

Flax is an important agricultural crop grown for both fiber and oil, used in many areas such as the production of varnishes and paints, linoleum, composites, and the textile and food industries (Goudenhooft et al., 2019). Long-term breeding of flax for oil content in seeds and fiber in stems has led to the appearance of two morphotypes – fiber flax and oilseed flax. Fiber flax is characterized by less branching, greater stem length and plant height, while oilseed flax is characterized by a larger number of seeds, and hence a greater number of inflorescences, with a shorter main stem length. In the late 19th and early 20th centuries, Russia was the main supplier of high-quality flax fiber, obtained from Russian heritage landraces, also known as "kryazh" (plural: kryazhs) resulting from folk selection. Kryazhs and Russian landraces made a decisive contribution to the gene pool of modern flax cultivars (Helbaek, 1959; Duk et al., 2021).

Previously, using single nucleotide polymorphisms, we characterized the genetic diversity of 306 flax samples from the collection of the Russian Federal Research Center for Bast Crops (FRC BC). We found significant differentiation between oilseed and fiber flax populations and identified genomic regions affected by modern breeding (Kanapin et al., 2022; Duk et al., 2025).

However, larger structural variations, such as transposon insertions (TEs), also play a significant role in shaping agronomically important plant traits and can be used for further improvement of the flax cultivars. It is known that TEs constitute a large part of plant genomes (Quesneville, 2020), and their insertions can lead to genome rearrangements, gene silencing, and rewiring of gene networks (Bourque et al., 2018), and can also be a source for the emergence of new genes and non-coding RNAs (Pulido, Casacuberta, 2023).

TEs are conventionally divided into two classes. Class I includes retrotransposons, which increase their copy number in the genome by insertion via an RNA intermediate (Mhiri et al., 2022), resulting in long terminal repeats potentially constituting up to 90 % of a plant's genome. Class II includes DNA transposons, which operate on a "cut-and-paste" principle, moving around the genome and usually not increasing their copy number. The highest TE activity is observed during periods of stress (Schrader, Schmitz, 2019). Most often, new insertions have a negative effect and are removed from the population; however, sometimes they can promote plant adaptation to unfavorable environmental conditions (Niu et al., 2019) and, because of this, be preserved in the population as a result of positive selection.

In this work, we used the same collection of 306 flax samples from the FRC BC (Duk et al., 2021; Kanapin et al., 2022) to predict sites of new TE insertions in groups of samples of different morphotypes and breeding status and to compare the distribution patterns of TE insertion sites and polymorphic sites across the genome. We also assessed the role of TE insertions in the formation of agronomically important traits and in the process of flax domestication.

Materials and methods

A total of 306 flax samples from the Federal Research Center for Bast Crops (FRC BC, Torzhok, Russia) collection were used in this study. The panel comprised 182 fiber flax and 120 oilseed flax varieties. The oilseed group included 99 intermediate, 16 crown, and five large-seeded accessions; the morphotype of the four remaining accessions was undetermined. Based on breeding status, the accessions were categorized as follows: 230 cultivated varieties (including 141 cultivars and 89 breeding lines), 40 landraces, and 31 kryazhs.

Genomic DNA was extracted from leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Netherlands). Library preparation and sequencing were performed at BGI on an Illumina platform, generating 150 bp paired-end reads. This yielded approximately 7.63 billion raw reads, totaling 1,143.850625 Gb of data. The average genome coverage was 9.3×, corresponding to 3.7 Gb per sample.

TE insertion sites were predicted using the TEMP2 software package (Yu T. et al., 2021) in insertion2 mode to identify non-reference insertions. Consensus TE sequences for the search were generated *de novo* with the REPET package's TEdenovo module (Flutre et al., 2011). To address inherent imprecision in insertion coordinates, we clustered insertions of the same TE family that were within 200 bp of each other

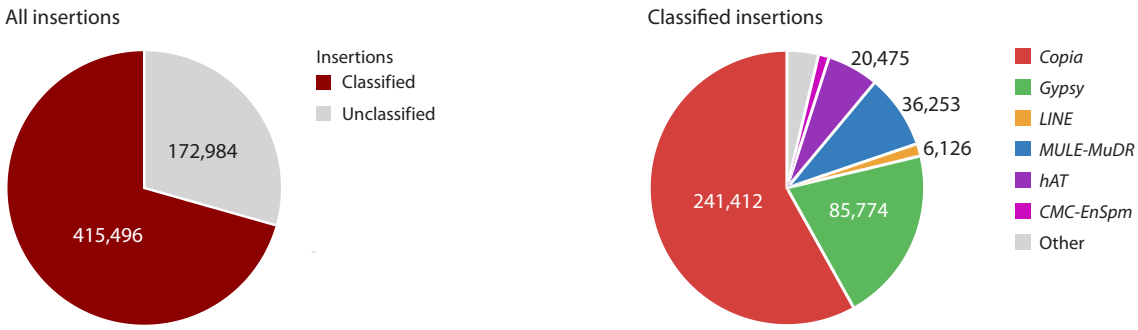


Fig. 1. Total number of new TE insertions and the number of insertions of individual TE superfamilies in the collection samples.

(twice the sequencing fragment length) and assigned them a unified coordinate at the center of the cluster. The final calls were converted to VCF format, and key insertions were visually validated using the Integrative Genomics Viewer (IGV).

For population and GWAS analyses, we further consolidated insertion calls to minimize false positives. Insertions across accessions located within a 50 bp window were merged into a single locus, with the position defined as the window's midpoint. This 50 bp threshold was empirically chosen as it maximized the number of insertions with a minor allele frequency (MAF) >5 %, thereby increasing statistical power while reducing the likelihood of spurious associations.

The genetic structure in the dataset was evaluated using the Principal Component Analysis (PCA) as well as the ADMIXTURE software v.1.3.0 (Alexander et al., 2009). The analyses were performed for K values ranging from 2 to 5. The phylogenetic tree was constructed with the ape package and drawn using the ggtree R package (Yu G., 2020). To assess genetic diversity due to TE insertions and calculate the fixation index (Fst), the VCFtools tool (Danecek et al., 2011) was used, with a window size of 200 kb. The window size was chosen in accordance with the average size of regions found as "hotspots" in the PrimatR package for R (https://github.com/daewoooo/primatR).

To identify possible genomic regions that underwent selection during breeding, we compared cultivars with the corresponding kryazhs and landraces. In each comparison, we calculated Fst and ROD = $1 - \pi_1/\pi_2$ statistics, where π is the genetic diversity of the corresponding sample group. Genomic regions with high population differences between the two groups (highest Fst values, top 5 % of the entire genome and top 2.5 % of ROD values) were considered as possible regions with traces of selection.

Values of 13 phenotypic traits measured in plants grown on the experimental fields of FRC BC in 2019 (one cultivation), 2020 (two cultivations with a two-week shift), and 2021 (one cultivation) were used (Kanapin et al., 2022): DSI – fusarium wilt severity index, EFL – elementary fiber length, FC – fiber content, FW – fiber weight, IL – inflorescence length, NI – number of internodes, Oil – oil content in seeds, PH – plant height, Nsed – number of seeds per plant, STI – stem tapering index, TL – technical stem length, TW – weight of the technical part of the plant, Tswgt – 1,000-seed weight. The genome-wide association analysis was performed using Blink, FarmCPU, SUPER, MLMM, MLM, GLM models in the GAPIT package (Wang, Zhang, 2020) with a threshold FDR rate 0.9. To link associations with genes, the genome annotation provided by the S. Cloutier group (You, Cloutier, 2020) was used. To calculate the effect of the insertion on the trait, the trait values in samples with and without the insertion were compared, with reliability confirmed by the Mann–Whitney statistical test (Mann, Whitney, 1947).

Results

Composition of the flax mobilome

We identified a total of 588,480 new transposable element (TE) insertions across 306 flax samples, 172,984 (29.4 %) of which could not be classified (Fig. 1). Among the classified insertions, the *Copia* superfamily was predominant, representing 41 % of all insertions (58 % of classified insertions). The *Gypsy* superfamily was the next most abundant, comprising 15 % of all insertions (20 % of classified ones). Retrotransposons of the *LINE* superfamily constituted 1 % of insertions. Among Class II DNA transposons, the *Mutator* (*MULE-MuDR*) superfamily was the most common (6 % of all insertions), followed by the *hAT* (4.3 %) and *CMC-EnSpm* (1.3 % of classified insertions) superfamilies.

Population structure analysis using PCA indicated that genetic differentiation between fiber and oilseed morphotypes was primarily driven by insertions from the *Copia* and *Gypsy* superfamilies, as well as the *hAT-Ac* family, as these were the only markers for which the first principal component separated the two groups (Fig. S1)[1].

Genomic landscape of the flax mobilome

The genomic distribution of TE insertions relative to genes varied significantly among superfamilies (Fig. 2, Table S1). Overall, 22 % of all insertions were located within or in close proximity (<2 kb) to genes. A strong bias for intergenic regions was observed for the *Gypsy* (85 %) and *CMC-EnSpm* (74 %) superfamilies, with the majority of their insertions located far (>2 kb) from genes. In contrast, approximately half of all *LINE* and *hAT* insertions were found near or within genes. The *Copia* superfamily showed a pronounced preference for genic regions, inserting into genes 1.4 times more frequently than the overall average. Furthermore, *Gypsy* elements that did land within genes were 1.3 times more likely to be in introns and three times less likely to be in exons compared to the general TE population. Exonic insertions were exceptionally

[1] Supplementary Figures S1–S4 and Tables S1–S5 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Duk_Engl_29_8.pdf

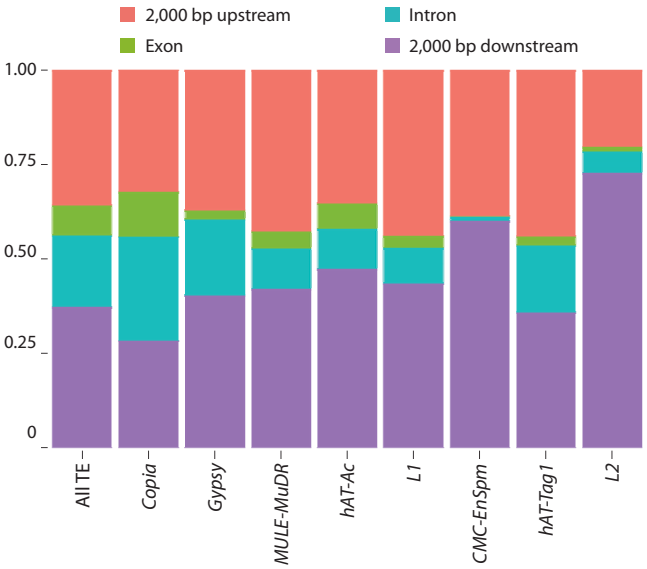


Fig. 2. Location of TE insertion sites relative to genes.

rare for the *CMC-EnSpm* and *L2* transposons. Across all superfamilies, intragenic insertions were predominantly intronic (71 %), with only 29 % located in exons.

Using TE insertion data for population structure and relatedness analyses

Principal Component Analysis revealed limited population structure based on TE insertions. While some differentiation between fiber and oilseed morphotypes was visible along the second and third principal components, no clear grouping by breeding status was observed (Fig. 3*a*, *b*). ADMIXTURE analysis indicated that the optimal number of genetic populations (K) was two, based on cross-validation error (Fig. 3*c*). However, the error for K = 3 was only marginally higher. At K > 2, a distinct genetic component (shown in green in Fig. 3*c*) became apparent specifically within oilseed flax accessions. No discernible differences in admixture patterns were associated with breeding status.

Phylogenetic reconstruction supported the population structure, grouping the accessions into three distinct clades (Fig. 4).

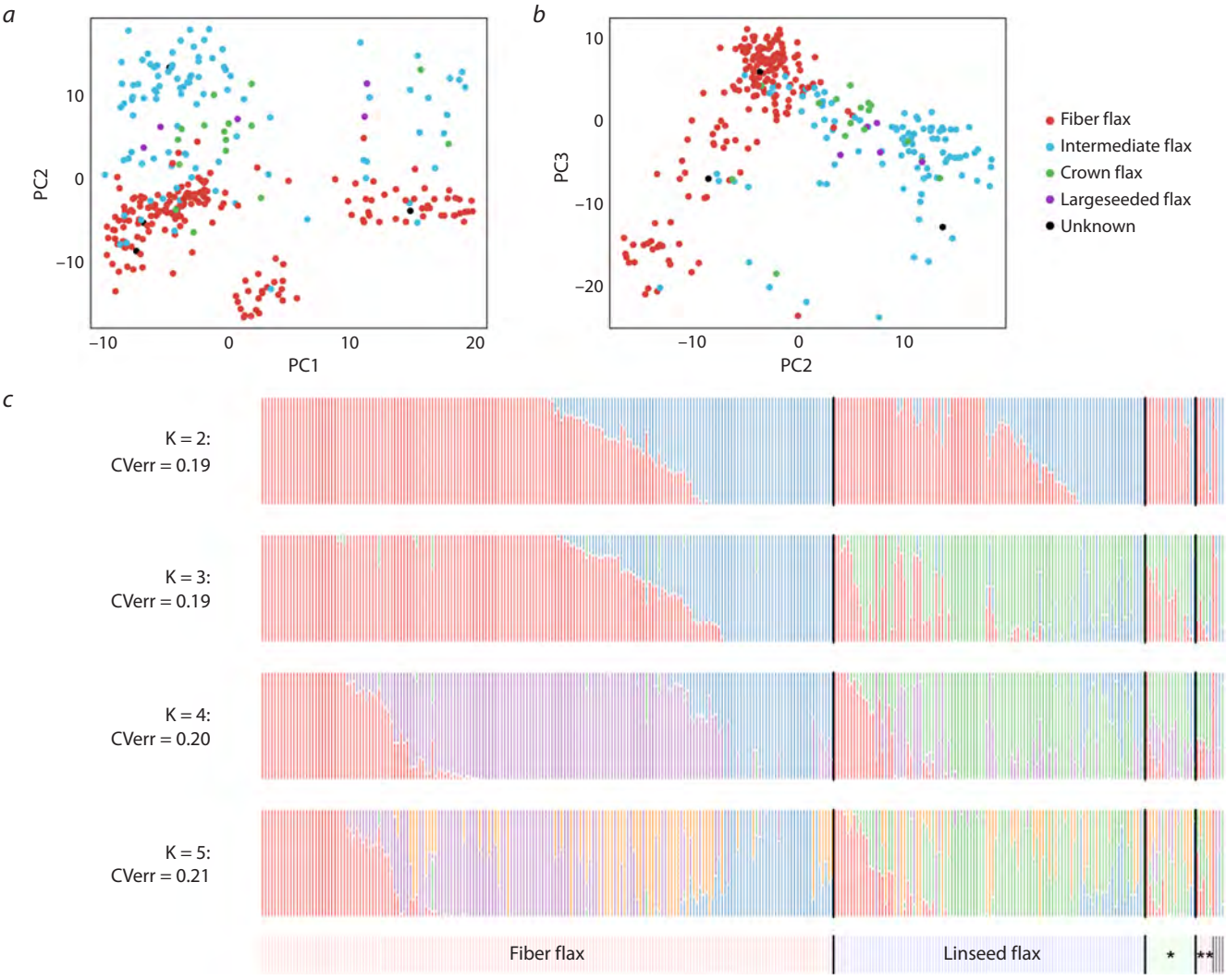


Fig. 3. Population structure of flax accessions.

a, b – visualization of flax accession clustering using PCA; *c* – ADMIXTURE graphs for K = 2–5 populations. CVerr – cross-validation error value during analysis; * crown; ** large-seeded accessions.

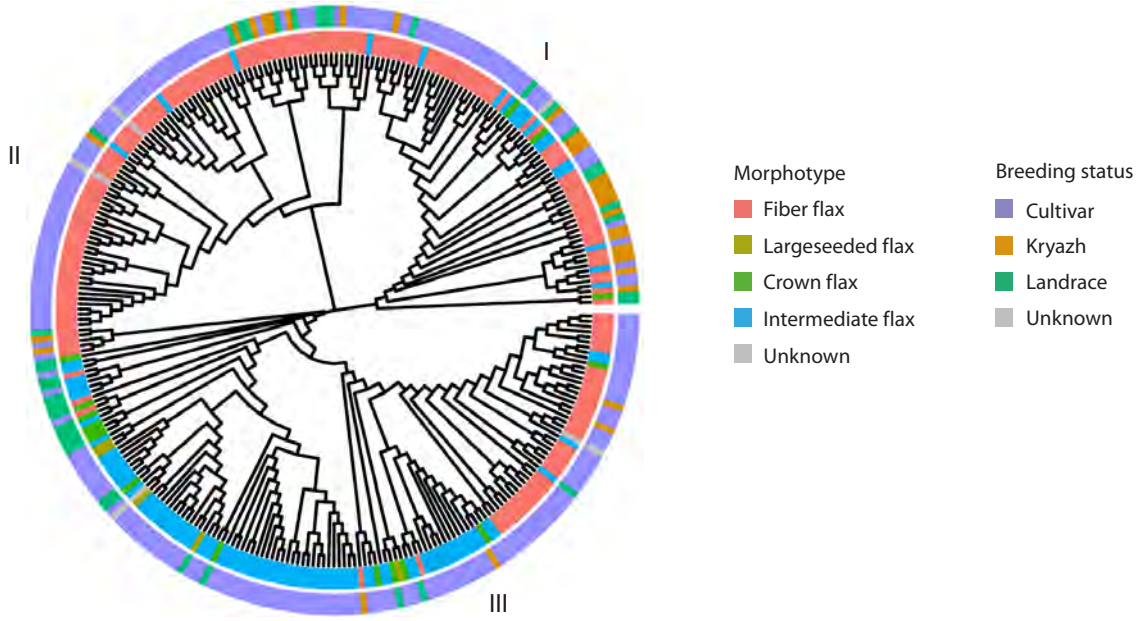


Fig. 4. NJ tree colored by morphotype and breeding status.

Clades I and II were composed almost exclusively of fiber flax, while Clade III contained nearly all oilseed flax samples.

Kryazhs were predominantly found in Clade I, confirming their shared genetic background (Duk et al., 2021). Each clade was significantly enriched for specific TE insertions. Clade I was characterized by ten unique TE sequences, primarily from the *Copia* (five) and *Gypsy* (two) superfamilies, with single sequences from *RC/Helitron* and *hAT-Ac*. Clades II and III were enriched with two and three *Copia* superfamily insertions, respectively (Table S2).

Identifying selective sweeps and analysis of agronomic traits using transposon insertions

To uncover genomic signatures of selection, we scanned for regions exhibiting significantly reduced diversity in specific comparisons: a) fiber flax cultivars *vs.* kryazhs or landraces; b) landraces *vs.* kryazhs; and c) oilseed cultivars *vs.* oilseed landraces. We also compared genetic diversity between fiber and oilseed cultivars to identify regions associated with their divergent agronomic traits. This analysis identified 61 candidate selective sweep regions (Fig. 5), with nine regions detected in multiple comparisons. Notably, 18 of these candidate regions overlap with known quantitative trait loci (QTLs) for important agronomic traits (You, Cloutier, 2020) (Table S3), linking these signatures of selection to specific phenotypic outcomes.

Comparative genomic analysis revealed distinct selective sweep patterns between flax morphotypes and breeding groups, with many overlapping known quantitative trait loci (QTLs). The comparison between oilseed and fiber flax revealed only one selective sweep signal in fiber flax, which co-localizes with the oil content QTL QOIL-Lu6.4 on chromosome 6 (Table S3). Cultivated varieties of both morphotypes showed reduced diversity (as compared to landraces) in regions of chromosome 12 overlapping with QTLs QIOD-Lu12.3, QLIN-Lu12.3, QLIO-Lu12.3, associated with iodine content, linoleic and linolenic acid content, respectively.

In fiber flax cultivars, selective sweep signals were observed 1) in a region of chromosome 12, overlapping with QTL uq.C12-1, 2) in a region of chromosome 3 overlapping with Lu3-25559600, 3) in a region of chromosome 8 overlapping with QOLE-Lu8.1, 4) in a region of chromosome 9 overlapping with QSTE-Lu9.2, and 5) in regions of chromosome 6 overlapping with Lu2564 and QOIL-Lu6. These QTLs are associated with plant height and stem length, seed mucilage content, oleic and stearic acid content, and oil content, respectively.

In kryazhs, selective sweep signals were detected in 1) regions of chromosome 7 overlapping with QPM-crc-LG7 and QPAL-Lu7.3, which are associated with powdery mildew incidence and palmitic acid content, respectively; 2) in regions of chromosome 2 overlapping with scaffold43-1111162 and QOIL-Lu2.1, which are associated with 1,000-seed weight and oil content, respectively, and 3) in regions of chromosome 3 overlapping with QLio-LG3.1, QLin-LG3.1, Marker4371 and scaffold156-76129, for which association with linolenic and linoleic acid content, plant height, and number of bolls, respectively, has been shown.

In oilseed flax cultivars, reduced diversity was also observed in regions of chromosome 8 overlapping with scaffold635-4397 and QOLE-Lu8.1, which are associated with the number of branches and oleic acid content, respectively; conversely, increased diversity compared to landraces was observed in regions of chromosome 7 overlapping with QLIN-Lu7.2, QLIO-Lu7.2, QPAL-Lu7.3 and QIOD-Lu7.2, which are associated with linolenic, linoleic, palmitic acid, and iodine content, respectively.

To further elucidate the genetic basis of agronomic traits, we conducted a GWAS utilizing transposon insertions as molecular markers. We discovered 50 TE insertions signifi-

Fig. 5. Genomic regions under selection. Circos diagram showing genome-wide selective sweep loci identified in different comparisons: fiber flax *vs.* oilseed flax (F/O), fiber flax cultivars *vs.* landraces (F/L), fiber flax cultivars *vs.* kryazhs (F/K), kryazhs *vs.* fiber flax landraces (K/L), oilseed flax cultivars *vs.* landraces (O/L). QTL – QTLs published in (You, Cloutier, 2020).

cantly associated with traits such as Fusarium wilt resistance, productivity, and fiber content. Many of these associations were robust, being confirmed by multiple models or repeated across growing seasons. A ~20 % subset exhibited pleiotropic effects, associating with several traits simultaneously (see the Table and Table S4). The potential functional importance of these insertions is underscored by the finding that four reside within known QTLs and two are located in genomic regions with significantly reduced diversity (Fig. 5).

Specifically, 15 associations were supported by multiple models, and 12 were linked to multiple traits or years. Six widely distributed insertions (found in >50 accessions) were selected for experimental validation, with the results detailed in the Figures S2–S4.

Discussion

A detailed characterization of flax genetic diversity is of paramount importance for its long-term and sustainable production and diversification, as well as for the overall success of its breeding programs. Previously, using SNPs, we characterized the genetic diversity of the core flax collection (306 samples) of the Federal Research Center for Bast Crops (FRC BC) (Duk et al., 2021; Kanapin et al., 2022). This collection, one of the best in the world, includes flax varieties from Eurasia with a significant proportion of local varieties. Recently, thanks to advances in bioinformatics and the improvement of sequencing technologies, other sources of genomic diversity, including TEs and structural variation, have become available for analysis, which can play a significant role in shaping agronomically important plant traits and can be used for further improvement of agricultural crops.

Analysis of new TE insertion sites performed in this work showed that, along with SNPs, TEs are an important source of genetic diversity in flax. The predominant insertion sites, as in many other agricultural plants (Stanin et al., 2025), are retrotransposon insertions, and among DNA transposons, insertions of the *MULE-MuD*R, *hAT*, and *CMC-EnSpm* superfamilies are most common (Fig. 1).

In contrast to SNP variation, which was greater in oilseed flax and landraces than in fiber flax, the number of TE family insertions did not differ substantially between morphotypes. Despite this, TE insertion patterns still corroborate the close relationship and common origin of the kryazhs, similarly to SNPs. Furthermore, the insertion profiles of the *Copia*, *Gypsy*, and *hAT-Ac* superfamilies clearly distinguish oilseed from fiber flax varieties (Fig. S1).

The genomic location of TEs relative to genes significantly influences gene expression and can lead to diverse phenotypic changes. In flax, most *Gypsy* and *CMC-EnSpm* insertions are located in intergenic regions. Furthermore, among the TEs that have inserted into genes, only 29 % are found in exons, suggesting that insertions in these coding regions are preferentially purged by natural selection (Fig. 2).

Analysis of genomic regions impacted by recent breeding (Table S3) revealed a striking disparity in selection signals between oilseed and fiber flax cultivars, with 9 and 32 regions identified, respectively. Notably, chromosomes 2, 5, 6, 9–11, and 15 showed no signals of selective improvement in oilseed flax (Fig. 5). Despite the fewer regions in oilseed flax, two of them overlap with known QTLs for fatty acid synthesis. We also identified 10 genomic regions showing divergent selection between the two morphotypes. A comparison of fiber flax cultivars and kryazhs revealed numerous selective sweep regions, but these showed little overlap. This likely reflects differing breeding objectives for modern fiber flax, driven by new industrial uses and climate change. This hypothesis is supported by the detection of 13 regions with reduced diversity in modern fiber cultivars compared to kryazhs.

Interestingly, five regions showing reduced genetic diversity in cultivars and kryazhs compared to landraces are also identified as regions of the diversity reduction when SNPs are used as markers (Table S5). Among these, noteworthy is the region of reduced diversity in cultivated oil flax varieties within Chr4_12400001–12600000, which contains the gene *Lus10036915* associated with pathogen defense, as well as the region Chr8_22400001–22600000, which overlaps with QTLs scaffold_6354_3971 and QOLELu8.1, associated with branch number and oleic acid content, respectively.

The most robust association identified in our GWAS was for the TE insertion Chr4_14756320, which was significantly linked to 22 distinct trait-year combinations using all analytical models (see the Table). This variant is a *hAT-Ac* family insertion situated ~2 kb upstream of *Lus10041548*, a gene with an *AGAMOUS*-like ortholog function that controls meristem determinacy and floral development (Yanofsky et al., 1990). The insertion's association with increased inflorescence length but decreased fiber quality suggests it may inhibit fiber cell initiation in the meristem. Its genomic position overlaps with three key QTLs: QYLD-Lu4.1 (seed yield), QPLH-Lu4.3 (plant height), and QDTM-Lu4.1 (early maturity).

TE insertions associated with phenotypic traits and located less than 3,500 bp from genes with known function

Coordinate of the TE insertion (superfamily)	Trait (effect)***	Insertion localization relative to the closest gene	Closest gene (ortholog)	Annotation
Chr1_3002494 (*Copia*)	Oil (+)	Intron	*Lus10042426* (*AT5G10770.1*)	Eukaryotic aspartyl protease family protein
Chr1_10250548 (–)	EFL (–) TL (–)		*Lus10022657* (*AT3G04380.1*)	SET domain-containing protein lysine methyltransferase family protein
Chr1_11870656 (*hAT-Tag1*)	TL (–) NI (–) PH (–)		*Lus10009386* (*AT1G08230.2*)	Transmembrane amino acid transporter family protein
Chr1_13735350 (*Copia*; *MULE-MuDR*)	TL (–)		*Lus10008435* (*AT5G37290.1*)	ARM repeat superfamily protein
Chr1_20792291 (–)	TW (–)		*Lus10015841* (*AT3G62870.1*)	Ribosomal L7Ae/L30e/S12e/Gadd45 family protein
Chr1_28316405 (*hAT-Ac*)	Tswgt (+)	1,309 upstream	*Lus10018915* (*AT1G71890.1*)	Major facilitator superfamily protein
Chr2_2089015 (*Copia*)	Tswgt (+)	Exon	*Lus10016394* (*AT3G08030.1*)	Protein of unknown function, DUF642
Chr3_522289 (*Copia*)	TW (+)	1,318 upstream	*Lus10013432* (*AT5G23130.1*)	Peptidoglycan-binding LysM domain-containing protein
Chr3_19617078 (*Gypsy*)	DSI (–)	Intron	*Lus10008232* (*AT5G46050.1*)	Peptide transporter 3
Chr3_20336100 (*Copia*)	STI (–)	1,882 downstream	*Lus10007194* (*AT4G01690.1*)	Flavin containing amine oxidoreductase family
Chr3_24121740 (*Copia*)	FC (+)	Exon	*Lus10017112* (*AT1G48130.1*)	1-cysteine peroxiredoxin 1
Chr3_24426254 (*MULE-MuDR*), *M/C	FC (–)	190 upstream	*Lus10017063* (*AT5G49480.1*)	Ca^{2+}-binding protein 1
Chr4_11952218 (*Copia*)	DSI (–)	853 downstream	*Lus10036789* (*AT1G13980.1*)	sec7 domain-containing protein
Chr4_14756320 (*hAT-Ac*)	EFL (–), FW (–), IL (+), NI (–), PH (–), TL (–), FC (–), TW (–)	2,059 upstream	*Lus10041548* (*AT3G61120.1*)	AGAMOUS-like 13, **QYLD-Lu4.1 (YLD), QPLH-Lu4.3 (PLH), QDTM-Lu4.1 (DTM)
Chr4_14994752 (*Copia*)	TL (–), NI (–)	Intron	*Lus10041602* (*AT3G52150.1*)	RNA-binding (RRM/RBD/RNP motifs) family protein, **QPLH-Lu4.3 (PLH), QDTM-Lu4.1 (DTM)
Chr6_1356894 (*Copia*)	TL (–)		*Lus10033725* (*AT1G12470.1*)	Zinc ion binding
Chr6_11075470 (*Copia*)	TL (+)	2,456 upstream	*Lus10002212* (*AT5G01300.1*)	PEBP (phosphatidyle-thanolamine-binding protein) family protein
Chr6_11780716 (*hAT-Ac*)	Tswgt (+)	1,006 upstream	*Lus10017263* (*AT4G36040.1*)	Chaperone DnaJ-domain superfamily protein
Chr6_17837550 (*Copia*)	TW (+)	400 upstream	*Lus10024282* (*AT1G58250.1*)	Golgi-body localisation protein domain; RNA pol II promoter Fmp27 protein domain
Chr7_15829004 (*Copia*)	FW (–)	709 upstream	*Lus10038370* (*AT4G20050.1*)	Pectin lyase-like superfamily protein **QIOD-Lu7.2 (IOD), QLIN-Lu7.2 (LIN), QLIO-Lu7.2 (LIO)
Chr8_4775776 (–)	EFL (–), TW (+), PH (–), TL (–)	34 upstream	*Lus10023944* (*AT4G38380.1*)	MATE efflux family protein
Chr8_12565858 (*hAT-Ac*)	NI (+), PH (+)	1,199 downstream	*Lus10025285* (*AT2G37370.1*)	Centrosomal protein

Table (end)

Coordinate of the TE insertion (superfamily)	Trait (effect)***	Insertion localization relative to the closest gene	Closest gene (ortholog)	Annotation
Chr8_16467884 (*MULE-MuDR*)	TW (–), FW (–)	972 downstream	*Lus10004540* (*AT1G29850.2*)	Double-stranded DNA-binding family protein
Chr8_16923232 (*Copia*), *Д/М	FC (+)	Intron	*Lus10000673* (*AT1G32450.1*)	Nitrate transporter 1.5
Chr8_18366764 (*Copia*)	PH (+), NI (+)	Exon	*Lus10007810* (*AT5G36930.2*)	Disease resistance protein (TIR-NBS-LRR class) family
Chr8_22233522 (*Copia*)	PH (+), NI (+)	2,260 upstream	*Lus10039488* (*AT5G43700.1*)	AUX/IAA transcriptional regulator family protein **QOLE-Lu8.1 (OLE)
Chr11_14960430 (*Copia*)	TW (–), TL (–), NI (–)	3,421 downstream	*Lus10008566* (*AT1G55200.1*)	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain
Chr12_11155523 (*MULE-MuDR*)	TW (–)	373 upstream	*Lus10030109* (*AT4G02530.1*)	Chloroplast thylakoid lumen protein
Chr13_1437052 (*Copia*)	EFL (–)	Exon	*Lus10010648* (*AT3G03770.1*)	Leucine-rich repeat protein; kinase family protein
Chr13_2213745 (*Copia*)	Oil (+)	271 upstream	*Lus10026978* (*AT2G26110.1*)	Protein of unknown function (DUF761)
Chr13_3391752 (*Copia*)	PH (–), NI (–), TW (–)	1,071 downstream	*Lus10026032* (*AT4G36670.1*)	Major facilitator superfamily protein
Chr13_6346942 (*hAT-Ac*)	FW (–)	253 upstream	*Lus10002084* (*AT5G64040.1*)	Photosystem I reaction center subunit PSI-N, chloroplast, putative/PSI-N, putative (PSAN)
Chr13_11641862 (*Copia*)	EFL (–)	972 downstream	*Lus10009758* (*AT4G17650.1*)	Polyketide cyclase/dehydrase and lipid transport protein
Chr13_14372739 (*Copia*)	PH (–)	780 upstream	*Lus10034660* (*AT4G29680.1*)	Alkaline-phosphatase-like family protein
Chr14_9124491 (*Copia*)	DSI (–)	153 upstream	*Lus10005526* (*AT2G37690.1*)	Phosphoribosylaminoimidazole carboxylase, putative/AIR carboxylase, putative
Chr14_9853868 (*hAT-Ac*)	Nsed (+)	984 downstream	*Lus10000815* (*AT2G27410.1*)	Domain of unknown function (DUF313)
Chr14_16743646 (*Copia*)	FW (–)	466 downstream	*Lus10017981* (*AT1G14590.1*)	Nucleotide-diphospho-sugar transferase family protein

* Insertion falls within a genomic region under selection identified in different comparisons: fiber flax *vs*. oilseed flax (F/O), oilseed flax cultivars *vs*. landraces (O/L). ** Insertion also falls within a QTL published in (You, Cloutier, 2020). TE insertion coordinate is the midpoint of a 50 bp window. *** Effect of the TE insertion on the trait: (+) – positive, (–) – negative.

We also noted that the same phenotypic effect could be caused by different TE families inserting at identical sites, such as Chr1_13735350 (*Copia/MULE-MuDR*) and Chr13_2207778 (*MULE-MuDR/hAT-Tag1*).

Contrary to expectation, several exonic insertions appear to enhance trait performance. For instance, *Copia* insertions in the exon of *Lus10016394*, *Lus10017112*, and *Lus10007810* were associated with positive effects, indicating that these genes likely function as suppressors of their respective traits (see the Table).

We identified pleiotropic TE insertions affecting multiple plant architecture traits. An intronic *hAT-T*ag1 insertion in *Lus10009386* (Chr1_11870656), a gene involved in amino acid transport (Meyer et al., 2006), negatively impacted technical stem length, plant height, and internode number. A comparable negative effect on stem and internode development was caused by an intronic *Copia* insertion in *Lus10041602* (Chr4_14994752), a gene encoding a photosynthetic apparatus component with a role in germination stress response (Xu et al., 2013); this locus also overlaps with QTLs for plant height and early maturity (see the Table).

Conversely, a *Copia* insertion upstream of *Lus10039488* (Chr8_22233522) enhanced plant height and internode number. This gene modulates early auxin responses (Liscum, Reed, 2002), and the insertion lies within the QOLE-Lu8.1 QTL associated with oleic acid.

The TE insertion at Chr1_10250548 located in an intron of the *Lus10022657* gene, whose ortholog contributes to the

transcriptional suppression of pseudogenes and transposons (Veiseth et al., 2011), had a negative effect on technical stem length and fiber length (see the Table).

Some insertions can also have different effects on different traits. For example, the insertion Chr8_4775776, located upstream of *Lus10023944*, had a negative effect on fiber length but a positive effect on the weight of the technical part. Such an influence is more preferable when it comes to oil flax, and the insertion was likely preserved during the selection process. The ortholog of this gene belongs to the MATE protein family, which is involved in protection from toxins and the synthesis of beneficial compounds (Takanashi et al., 2014).

The statistical robustness of several associations was confirmed by their discovery with multiple analytical models. A notable example is a *Copia* insertion at Chr1_3002494, located within an intron of the *Lus10042426* gene, which was associated with increased oil content. The ortholog of this gene is implicated in plant immunity (Breitenbach et al., 2014). Other robust associations include intronic *Copia* insertions at Chr6_1356894 and Chr8_16923232. The corresponding genes are involved in root development metabolism (Takemoto et al., 2018) and nitrate transport (Lin et al., 2008), respectively. Furthermore, the insertion at Chr8_16923232 is located within a selective sweep region that differentiates oilseed flax from fiber flax (see the Table).

From an applied perspective, TE insertions that confer advantageous traits can be harnessed in breeding. A compelling case is the association between reduced disease severity (DSI) and two insertions: a *Gypsy* element in the 3′-flanking region of *Lus10008232* (implicated in seed stress resistance and pathogen defense) (Karim et al., 2005) and a *Copia* element within the auxin transport and cell wall organization gene *Lus10036789* (Geldner et al., 2003). These variants provide direct targets for marker-assisted selection to enhance disease resistance.

Conclusion

The genomic landscape of transposon insertions in flax is non-uniform, revealing patterns consistent with the divergent selection pressures applied to different morphotypes. These insertions contribute substantially to phenotypic diversity and environmental adaptation. Consequently, transposons serve as a crucial source of molecular markers that, together with single nucleotide polymorphisms, can be harnessed to select for desired characteristics in breeding programs.

References

Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664. doi 10.1101/gr.094052.109

Bourque G., Burns K.H., Gehring M., Gorbunova V., Seluanov A., Hammell M., Imbeault M., Izsvák Z., Levin H.L., Macfarlan T.S., Mager D.L., Feschotte C. Ten things you should know about transposable elements. *Genome Biol*. 2018;19(1):199. doi 10.1186/s13059-018-1577-z

Breitenbach H.H., Wenig M., Wittek F., Jordá L., Maldonado-Alconada A.M., Sarioglu H., Colby T., Knappe C., Bichlmeier M., Pabst E., Mackey D., Parker J.E., Vlot A.C. Contrasting roles of the apoplastic aspartyl protease APOPLASTIC, *ENHANCED DISEASE SUSCEPTIBILITY1*-DEPENDENT1 and LEGUME LECTIN-LIKE PROTEIN1 in Arabidopsis systemic acquired resistance. *Plant Physiol*. 2014;165(2):791-809. doi 10.1104/pp.114.239665

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158. doi 10.1093/bioinformatics/btr330

Duk M., Kanapin A., Rozhmina T., Bankin M., Surkova S., Samsonova A., Samsonova M. The genetic landscape of fiber flax. *Front Plant Sci*. 2021;12:764612. doi 10.3389/fpls.2021.764612

Duk M.A., Kanapin A.A., Samsonova A.A., Bankin M.P., Samsonova M.G. The IIIVmrMLM method uncovers new genetic variants associated with resistance to Fusarium wilt in flax. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(3):380-391. doi 10.18699/vjgb-25-41

Flutre T., Duprat E., Feuillet C., Quesneville H. Considering transposable element diversification in *de novo* annotation approaches. *PloS One*. 2011;6(1):e16526. doi 10.1371/journal.pone.0016526

Geldner N., Anders N., Wolters H., Keicher J., Kornberger W., Muller P., Delbarre A., Ueda T., Nakano A, Jürgens G. The *Arabidopsis* GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth. *Cell*. 2003;112(2):219-230. doi 10.1016/s0092-8674(03)00003-5

Goudenhooft C., Bourmaud A., Baley C. Flax (*Linum usitatissimum* L.) fibers for composite reinforcement: exploring the link between plant growth, cell walls development, and fiber properties. *Front Plant Sci*. 2019;10:411. doi 10.3389/fpls.2019.00411

Helbaek H. Domestication of food plants in the Old World. Joint efforts by botanists and archeologists illuminate the obscure history of plant domestication. *Science*. 1959;130(3372):365-372. doi 10.1126/science.130.3372.365

Kanapin A., Rozhmina T., Bankin M., Surkova S., Duk M., Osyagina E., Samsonova M. Genetic determinants of fiber-associated traits in flax identified by Omics data integration. *Int J Mol Sci*. 2022;23(23):14536. doi 10.3390/ijms232314536

Karim S., Lundh D., Holmström K.O., Mandal A., Pirhonen M. Structural and functional characterization of AtPTR3, a stress-induced peptide transporter of *Arabidopsis*. *J Mol Model*. 2005;11(3):226-236. doi 10.1007/s00894-005-0257-6

Lin S.H., Kuo H.F., Canivenc G., Lin C.S., Lepetit M., Hsu P.K., Tillard P., Lin H.L., Wang Y.Y., Tsai C.B., Gojon A., Tsay Y.F. Mutation of the *Arabidopsis NRT1.5* nitrate transporter causes defective root-to-shoot nitrate transport. *Plant Cell*. 2008;20(9):2514-2528. doi 10.1105/tpc.108.060244

Liscum E., Reed J.W. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol Biol*. 2002;49(3-4):387-400. doi 10.1023/A:1015255030047

Meyer A., Eskandari S., Grallath S., Rentsch D. AtGAT1, a high affinity transporter for γ-aminobutyric acid in *Arabidopsis thaliana*. *J Biol Chem*. 2006;281(11):7197-7204. doi 10.1074/jbc.M510766200

Mhiri C., Borges F., Grandbastien M.-A. Specificities and dynamics of transposable elements in land plants. *Biology*. 2022;11(4):488. doi 10.3390/biology11040488

Niu X.M., Xu Y.C., Li Z.W., Bian Y.T., Hou X.H., Chen J.F., Zou Y.P., Jiang J., Wu Q., Ge S., Balasubramanian S., Guo Y.L. Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc Natl Acad Sci USA*. 2019;116(14):6908-6913. doi 10.1073/pnas.1811498116

Pulido M., Casacuberta J.M. Transposable element evolution in plant genome ecosystems. *Curr Opin Plant Biol*. 2023;75:102418. doi 10.1016/j.pbi.2023.102418

Quesneville H. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. *Mob DNA*. 2020;11:28. doi 10.1186/s13100-020-00223-x

Schrader L., Schmitz J. The impact of transposable elements in adaptive evolution. *Mol Ecol*. 2019;28(6):1537-1549. doi 10.1111/mec.14794

Stanin V.A., Duk M.A., Kanapin A.A., Samsonova A.A., Surkova S.Y., Samsonova M.G. Chickpea diversity driven by transposon insertion

polymorphism. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2025;29(1):61-71. doi 10.18699/vjgb-25-08

Takemoto K., Ebine K., Askani J.C., Krüger F., Gonzalez Z.A., Ito E., Goh T., Schumacher K., Nakano A., Ueda T. Distinct sets of tethering complexes, SNARE complexes, and Rab GTPases mediate membrane fusion at the vacuole in Arabidopsis. *Proc Natl Acad Sci USA.* 2018;115(10):E2457-E2466. doi 10.1073/pnas.1717839115

Veiseth S.V., Rahman M.A., Yap K.L., Fischer A., Egge-Jacobsen W., Reuter G., Zhou M.M., Aalen R.B., Thorstensen T. The SUVR4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis. *PLoS Genet.* 2011;7(3):e1001325. doi 10.1371/journal.pgen.1001325

Wang J., Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genom Proteom Bioinform.* 2021;19(4):629-640. doi 10.1016/j.gpb.2021.08.005

Xu T., Lee K., Gu L., Kim J.I., Kang H. Functional characterization of a plastid-specific ribosomal protein PSRP2 in *Arabidopsis thaliana* under abiotic stress conditions. *Plant Physiol Biochem.* 2013; 73:405-411. doi 10.1016/j.plaphy.2013.10.027

Yanofsky M.F., Ma H., Bowman J.L., Drews G.N., Feldmann K.A., Meyerowitz E.M. The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature.* 1990; 346(6279):35-39. doi 10.1038/346035a0

You F., Cloutier S. Mapping quantitative trait loci onto chromosome-scale pseudomolecules in flax. *Methods Protoc.* 2020;3(2):28. doi 10.3390/mps3020028

Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinform.* 2020;69(1):e96. doi 10.1002/cpbi.96

Yu T., Huang X., Dou S., Tang X., Luo S., Theurkauf W.E., Lu J., Weng Z. A benchmark and an algorithm for detecting germline transposon insertions and measuring *de novo* transposon insertion frequencies. *Nucleic Acids Res.* 2021;49(8):e44. doi 10.1093/nar/gkab010

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

Received July 30, 2025. Revised September 24, 2025. Accepted September 25, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1277-1287

doi 10.18699/vjgb-25-132

The influence of allelic variants of the *Vrn-A1* gene on the duration of the vegetation period in *Triticum dicoccoides*

G.Yu. Chepurnov [1, 2], Z. Chen [2, 3], A.G. Blinov [2], V.M. Efimov [2, 3], N.P. Goncharov [2, 3]

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

chepurnovgy@bionet.nsc.ru

Abstract. The duration of the vegetation period (DVP) is an important agronomic trait in cereal. The main influence on it in wheat is exerted by *Vrn* genes, which determine the growth habit (spring *vs.* winter) and DVP. In the present study, 137 wild emmer *Triticum dicoccoides* (Körn. ex Aschers. et Graebn.) Schweinf. accessions were evaluated according to the growth habit trait, among which 39 spring ones were identified. The nucleotide sequences of the promoter region of the *Vrn-A1* gene were established in the spring accessions by sequencing. Five allelic variants of *Vrn-A1* genes previously found in *T. dicoccoides* were identified, namely *Vrn-A1b.1*, *Vrn-A1b.2*, *Vrn-A1b.4*, *Vrn-A1d*, *Vrn-A1u*. Three spring accessions PI355457, PI190919, PI560817 simultaneously contained two alleles of the *Vrn-A1* gene: *Vrn-A1d* and previously undescribed functional allelic variant designated by the authors as *Vrn-A1b.8*. The promoter region of this allele had several deletions relative to the intact variant. One of such deletions covered 8 bp of the VRN box. In a single experiment, under controlled greenhouse conditions, the relationship between the allelic variants of the *Vrn-A1* gene and the duration of the vegetation period of the *T. dicoccoides*' spring accessions was studied using the 2B-PLS (Two-Block Partial Least Squares) analysis. The correlation coefficient (r) between these traits was 0.534. The correlation coefficient between the duration of the vegetation period of wild emmer plants and the regions of origin of the studied accessions was also calculated (r = 0.478). It was shown that accessions with identical alleles of the *Vrn-A1* gene and originating from the same region can differ significantly from each other in the duration of the vegetation period. The presence of phenotypic differences with the same allelic composition of the *Vrn-A1* gene indicates the contribution of other hereditary factors localized in the genomes of these accessions, which determines their value as new donors of genetic resources that contribute to the expansion of the biodiversity of common and durum wheat commercial cultivars.

Key words: wild emmer; *Triticum dicoccoides*; spring growth habit; *Vrn-A1* gene; alleles; duration of vegetation period; 2B-PLS analysis

For citation: Chepurnov G.Yu., Chen Z., Blinov A.G., Efimov V.M., Goncharov N.P. The influence of allelic variants of the *Vrn-A1* gene on the duration of the vegetation period in *Triticum dicoccoides*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1277-1287. doi 10.18699/vjgb-25-132

Funding. The study was carried out with the financial support of the state budget project FWNR-2025-0009.

Acknowledgements. The plants were grown at the Center for Plant Reproduction of ICG SB RAS within the budgetary project FWNR-2022-0017. Sequencing was completed at the SB RAS Genomics Core Facility (Novosibirsk, Russia, URL: http://www.niboch.nsc.ru/doku.php/corefacility, accessed on April 2, 2025). The authors are grateful to Doctor of Sciences (Biology) A.B. Shcherban (Institute of Cytology and Genetics SB RAS) for providing additional information on the *Triticum dicoccoides* accessions studied in his investigation.

Влияние аллельных вариантов гена *Vrn-A1* на длину вегетационного периода у *Triticum dicoccoides*

Г.Ю. Чепурнов [1, 2], Ч. Чэнь [2, 3], А.Г. Блинов [2], В.М. Ефимов [2, 3], Н.П. Гончаров [2, 3]

[1] Сибирский научно-исследовательский институт растениеводства и селекции – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, р.п. Краснообск, Новосибирская область, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

chepurnovgy@bionet.nsc.ru

Аннотация. Длина (продолжительность) вегетационного периода является одним из ключевых хозяйственно важных признаков зерновых культур. У пшениц основное влияние на него оказывают гены *Vrn*, контролирующие тип (яровость *vs*. озимость) и скорость развития растений. В настоящей работе по признаку «тип развития» изучено 137 образцов дикой полбы *Triticum dicoccoides* (Körn. ex Aschers. et Graebn.) Schweinf., среди которых выяв-

лено 39 яровых. Методом секвенирования у яровых образцов установлены нуклеотидные последовательности промоторной области гена *Vrn-A1*. Идентифицировано пять аллельных вариантов: *Vrn-A1b.1*, *Vrn-A1b.2*, *Vrn-A1b.4*, *Vrn-A1d*, *Vrn-A1u*. Три яровых образца, PI355457, PI190919 и PI560817 *T. dicoccoides*, содержали одновременно два аллеля гена *Vrn-A1*: *Vrn-A1d* и не описанный ранее аллельный вариант, обозначенный авторами как *Vrn-A1b.8*. Промоторная область этого аллеля имела несколько делеций относительно интактного варианта, одна из которых захватывала 8 п.н. VRN-бокса. В контролируемых условиях теплицы у образцов была определена длина вегетационного периода и методом 2B-PLS анализа (two-block partial least squares analysis) оценена связь между аллельными вариантами гена *Vrn-A1* и длиной вегетационного периода яровых образцов *T. dicoccoides*. Коэффициент корреляции (r) между этими признаками составил 0.534. Был рассчитан также коэффициент корреляции между длиной вегетационного периода растений и регионами происхождения образцов (r = 0.478). Показано, что образцы, имеющие одинаковые аллели гена *Vrn-A1* и происходящие из одного региона, могут значительно отличаться между собой по длине вегетационного периода. Наличие фенотипических различий при одинаковом аллельном составе гена *Vrn-A1* указывает на вклад иных наследственных факторов, локализованных в геномах этих образцов, что обусловливает их ценность в качестве новых источников генетических ресурсов, способствующих расширению биоразнообразия сортов возделываемых видов пшениц.

Ключевые слова: дикая полба; *Triticum dicoccoides*; яровой тип развития; ген *Vrn-A1*; аллели; длина вегетационного периода; 2B-PLS анализ

Introduction

Wild emmer *Triticum dicoccoides* (Körn. ex Aschers. et Graebn.) Schweinf. is one of the hypothetic ancestors of cultivated tetraploid (2*n* = 4*x* = 28) wheat (Rivera et al., 2025). Its range covers the Fertile Crescent (Southwest Asia) and extends from Israel, Jordan, Lebanon, Syria, southern Turkey and northern Iraq to southwestern Iran (Özkan et al., 2011; Lack, van Slageren, 2020), where emmer *T. dicoccum* Schrank ex Schübl. was domesticated (Novoselskaya-Dragovich et al., 2025). Due to this extensive areal, *T. dicoccoides* retains polymorphism and has significant potential to improve modern cultivated wheat species (Kato et al., 1997; Nevo, 2001; Dong et al., 2010; Feng et al., 2017). Allelic variants of *T. dicoccoides* genes determining agronomically important traits have often been used to improve the resistance of durum and bread wheat plants to various diseases such as ear fusarium (Soresi et al., 2017, 2021), yellow rust (Sela et al., 2014; Zhang H. et al., 2016), powdery mildew (Xue et al., 2012; Ouyang et al., 2014; Liang et al., 2015; Saidou et al., 2015; Qiu et al., 2021), and others. In addition to the introgression of immunity-related genes, wild emmer is widely used to improve other traits in cultivated wheat species, such as increased adaptability due to the transfer of dominant *Vrn* (response to vernalization) genes and their alleles responsible for the formation of the growth habit (spring *vs*. winter) and the duration of the vegetation period (Strejčková et al., 2023).

The growth habit is the most important trait that determines wide adaptability of wheat plants to various climatic conditions (Law, Worland, 1997). Winter-type wheat requires prolonged exposure to low positive temperatures (typically ≥50 days of vernalization) for transition from vegetative to reproductive development (Kiss et al., 2025). This mechanism causes a delay in the vegetative phase of plants, preventing damage to floral meristems by low temperatures. Spring wheat delays the transition from vegetative to reproductive development during a single growing season without vernalization (Distelfeld et al., 2009a). It has been more than once shown that the *Vrn* genes, which control the growth and development characteristics (duration of ontogenesis) of wheat plants, determine not only the growth habit (spring/winter), but also the duration of development phases (Efremova, Chumanova, 2023), i. e. they control the duration of the life cycle from germination to ripening and, as a result, affect early flowering and yield (Flood, Halloran, 1986; Goncharov, 1998; Distelfeld et al., 2009a; Kamran et al., 2014; Shcherban et al., 2015a; Afshari-Behbahanizadeh et al., 2024; Plotnikov et al., 2024; etc.). In addition to these genes, the duration of the vegetation period in wheat is also affected by other genes, such as *Ppd* (response to photoperiod), which determine the sensitivity of plants to photoperiod, and *Eps* (earliness *per se*), which determine the earliness without the influence of external signals (Distelfeld et al., 2009a; Kamran et al., 2014; Würschum et al., 2018). It is noted that the *Vrn* gene system accounts for up to 75 % of variability for this trait, while the other two systems – for about 25 % (Stelmakh, 1998). The significant influence of *Vrn* genes on phenology (particularly flowering time regulation) has motivated extensive research into these loci. By now, studies have characterized these genes' genomic structure and chromosomal localization, while also elucidating their interactions with other genes controlling developmental timing (Yan et al., 2003, 2004b, 2006; Fu et al., 2005; Distelfeld et al., 2009b; Chen A., Dubcovsky, 2012; Shcherban et al., 2012a, b, 2013, 2015a; Chen F. et al., 2013; Kippes et al., 2014–2016; Shcherban, Salina, 2017).

Mutations of three genes, *Vrn-1*, *Vrn-2* and *Vrn-3*, cause the spring growth habit in many species of the genus *Triticum* L. (Goncharov, 2004a, b; Yan et al., 2004a, b, 2006; Golovnina et al., 2010; Shcherban, Salina, 2017). In common wheat (*Triticum aestivum* L.) (Kippes et al., 2014, 2015) and *T. sphaerococcum* Perc. (Goncharov, Shitova, 1999), the fourth *Vrn* gene, *Vrn-D4*, has been described. The expression of *Vrn-1* serves as the primary molecular trigger initiating the inflorescence development cascade (Yan et al., 2003; Trevaskis et al., 2007). The *Vrn-1* gene encodes MADS-box transcription factors (Yan et al., 2004a; Trevaskis et al., 2007), which reduce the expression of *Vrn-2* genes and induce the expression of *Vrn-3* genes, which function as florigen (Dubcinsky et al., 2006; Yan et al., 2006; Hemming et al., 2008). It has been shown that the spring growth habit in hexaploid (2*n* = 6*x* = 42) wheat can be determined by mutations in the *Vrn-1*, *Vrn-D4* and *Vrn-3* genes, which cause their expression without low temperature (vernalization) (Yan et al., 2003, 2004b, 2006; Fu et al., 2005; Chen A., Dubtsovsky, 2012; Shcherban et al., 2012a, b, 2013, 2015a; Kippes et al., 2014, 2015; Shcherban, Salina, 2017),

or due to a decrease in the number of zinc finger domains and CCTs that form the *Vrn-2* codes, or form a cyclic composition of domain structures (Distelfeld et al., 2009b; Chen F. et al., 2013; Kippes et al., 2016).

The spring growth habit in *T. dicoccoides* is inherited in a dominant manner (Goncharov, 1998). In this species, allelic variants of the gene determining the spring growth habit are described only in the VRN-1 locus (Yan et al., 2004a; Shcherban et al., 2015b; Konopatskaya et al., 2016; Muterko et al., 2016; etc.). To date, seven such alleles are known, four of which contain deletions of different lengths in the promoter region (*Vrn-A1b.2*, *Vrn-A1b.7*, *Vrn-A1f* and *Vrn-A1d*); two alleles have the structure of these elements in the same region (*Vrn-A1a.3*) and a deletion in the first intron (*Vrn-A1c*); one allele differs from the intact sequence by 29 nucleotide substitutions, one deletion and one SNP insertion in the promoter region (*Vrn-B1dic*) (Yan et al., 2004a; Shcherban et al., 2015b; Konopatskaya et al., 2016; Muterko et al., 2016).

All of the above-mentioned allelic variants of *Vrn-1* genes were previously detected in a study of 92 spring and winter accessions of *T. dicoccoides* (Yan et al., 2004a; Shcherban et al., 2015b; Konopatskaia et al., 2016; Muterko et al., 2016). However, these studies cover only a portion of the wild emmer accessions available in collections. According to the GRIN NPGS report, based on the results of 2001 trials, 792 *T. dicoccoides* accessions were sown at the USDA research station in Idaho. 292 of them were classified as spring or facultative forms (URL: https://npgsweb.ars-grin.gov/gringlobal/method?id=491608, accessed April 2, 2025). However, unlike bread wheat, the studies published to date do not provide information on the effect of the identified allelic variants of *Vrn-1* genes on the change in the duration of the growing season of spring *T. dicoccoides* plants.

The present study has two main objectives: (i) sequencing and analysis of the promoter region of the *Vrn-A1* gene, including VRN-box and GArG-box, in 39 previously unstudied spring accessions of *T. dicoccoides*, (ii) assessment of the associative relationship between the allelic variants of the *Vrn-A1* gene and the duration of the growing season in spring accessions of *T. dicoccoides* under controlled conditions.

Material and methods

Plant material, growing conditions, assessment of the growth habit and duration of the vegetation period. The plant material for the study was 137 *T. dicoccoides* accessions of various ecological and geographical origins, of which 39 accessions with a spring growth habit were identified and taken for further study (Table 1, Fig. 1). Progeny seeds were obtained from each accession to assess the growth habit (spring *vs*. winter) and heading time. The plants were planted as 5-day-old seedlings (10 per accession) in a hydroponic greenhouse of the Institute of Cytology and Genetics SB RAS without preliminary vernalization. The plants were grown at a temperature of 23–25 °C, under long-day (16 h) conditions, at standard humidity. The number of days from planting to heading was recorded for each plant individually. Based on the data obtained, the average value of this feature for each accession was estimated.

Total DNA isolation, PCR amplification, and nucleotide sequence analysis of the *Vrn-A1* gene promoter. Total DNA was isolated using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol.

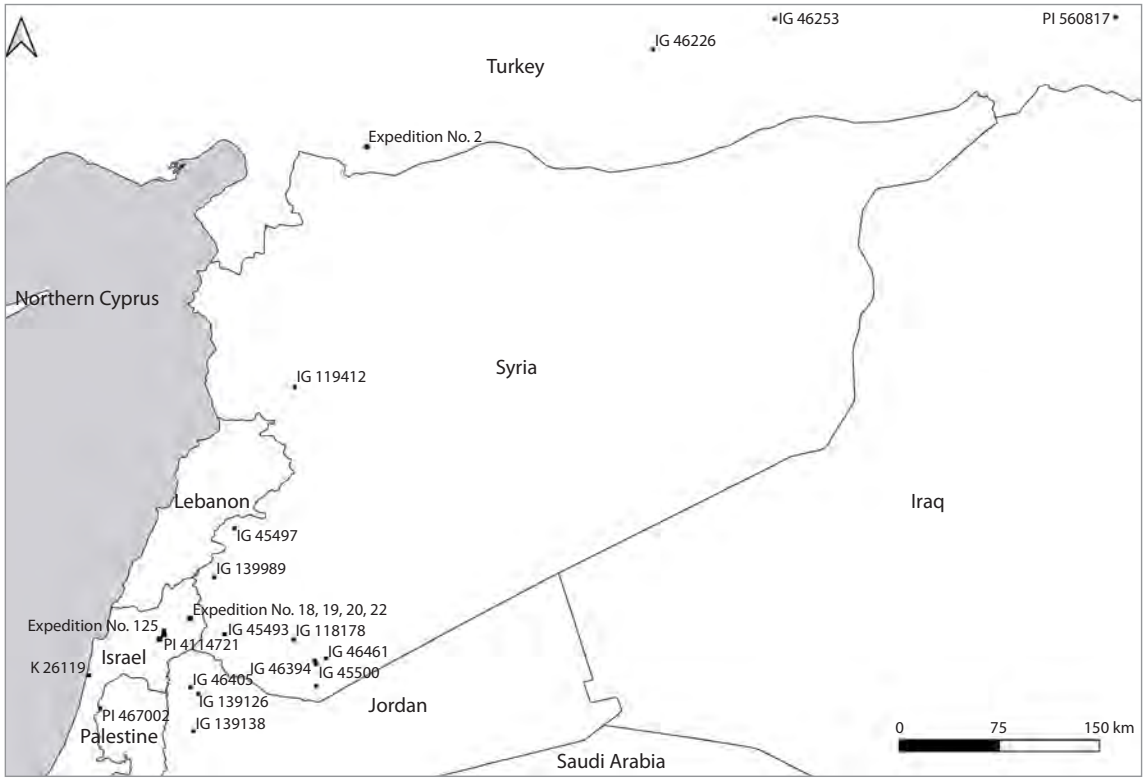


Fig. 1. Map of the collection sites of *T. dicoccoides* accessions studied in this work.

Table 1. *Triticum dicoccoides* accessions used in the study, their collection sites, time to heading, and *Vrn-A1* allelic variants

No.	Accession name	Collection site	Heading time (day), mean ± standard deviation	*Vrn-A1* allelic variant
1	IG 46405	Jordan	93 ± 10.1	*Vrn-A1u*
2	IG 45491	Syria	83 ± 1.7	
3	IG 139138	Jordan	83 ± 5.1	
4	IG 45493	Syria	69 ± 2.4	
5	IG 46394		83 ± 1.8	
6	IG 45500		69 ± 9.4	
7	IG 46226	Turkey	73 ± 2.9	
8	Expedition No. 2		89 ± 7.7	
9	IG 46461	Syria	69 ± 1.3	
10	IG 45497		73 ± 4.3	
11	PI 415151	Israel	86 ± 7.5	*Vrn-A1d*
12	PI 414721		83 ± 6.9	
13	IG 139126	Jordan	62 ± 7.8	
14	PI 467016	Israel	83 ± 1.9	
15	PI 467002		44 ± 2.8	
16	k-26119		46 ± 2.1	
17	PI 467027		56 ± 8.6	
18	DC 91	Unknown	76 ± 7.4	
19	PI 346783	Hungary	67 ± 6.9	
20	PI 355457	Germany	62 ± 8.7	*Vrn-A1d/Vrn-A1b.7*
21	PI 190919	Belgium	56 ± 1.6	
22	PI 560817	Turkey	62 ± 1.5	
23	Expedition No. 20	Israel	54 ± 4.3	*Vrn-A1d*
24	Expedition No. 8		45 ± 5.4	
25	Expedition No. 19		46 ± 7.4	
26	Expedition No. 18		54 ± 5.8	
27	Expedition No. 22		72 ± 5.3	
28	Expedition No. 125		72 ± 3.3	
29	k-5199	Israel	41 ± 9.4	*Vrn-A1b.1*
30	IG 45495	Syria	93 ± 7.2	
31	Unknown	Belarus	38 ± 2.4	*Vrn-A1b.2*
32	PI 352366	Germany	38 ± 3.6	
33	k-41965	Israel	48 ± 8.2	
34	IG 46253	Turkey	34 ± 6.4	
35	PI 30989	Israel	93 ± 3.6	*Vrn-A1b.4*
36	IG 139989	Syria	86 ± 9.7	
37	IG 119412		80 ± 10.2	
38	IG 118178		93 ± 7.8	
39	Expedition No. 116	Israel	86 ± 3.6	

50–100 mg of freshly collected leaves from each sample were used for DNA extraction. The quantity and quality of the isolated DNA were determined using a NanoDrop2000 spectrophotometer (Thermo Scientific, Waltham, USA) and electrophoretic separation in 1 % agarose gel containing ethidium bromide (0.5 mg/ml) in 1xTAE. Polymerase chain reactions (PCR) were performed in 20-μl reactions containing 10 mM Tris-HCl (pH 8.9), 1 mM $(NH_4)_2SO_4$, 4 mM $MgCl_2$, 200 μM of each dNTP, 0.5 μM of each primer, 1 unit of Taq DNA polymerase, and 0.1 μg of genomic DNA. Primers VRN1AF (5′-GAAAGGAAAAATTCTGCTCG-3′) and VRN1-INT1R (5′-GCAGGAAATCGAAATCGAAG-3′) were

used to probe the target region of the promoter region. The primers amplified a 688 bp fragment (in the case of the intact allelic variant) from position –439 bp upstream of the start codon to 246 bp upstream of the start codon. The PCR program included an initial denaturation step for 5 min at 94 °C and 33 amplification cycles consisting of 30 sec of denaturation at 94 °C, 40 sec of annealing at 52 °C and 1 min of elongation at 72 °C. PCR products were separated by agarose gel electrophoresis and purified using the QIAquick Gel Extraction Kit (QIAGEN). PCR products were cloned into the pAL2-T vector using the Quick-TA kit protocol (Eurogen, Moscow, Russia). Sequencing reactions were performed using 200 ng of product and the BigDye Terminator v3.1 sequencing kit (Thermo Scientific, USA) on an ABI 3130XL genetic analyzer (Applied Biosystems, Waltham, MA, USA) at the Genomics Center of the Siberian Branch of the Russian Academy of Sciences (URL: http://www.niboch.nsc.ru/doku.php corefacility, accessed April 2, 2025).

Statistical analysis of data. For statistical processing, quantitative and qualitative characteristics of the samples were used. The analysis of allelic variants of the *Vrn-A1* gene and their relationship with the heading date (duration of the vegetation period) was carried out taking into account previously published data obtained under identical growing conditions (hydroponic system, 16-hour photoperiod) (Shcherban et al., 2015b; Konopatskaia et al., 2016; Muterko et al., 2016). The data are presented as a "samples–features" matrix. Each object is described by three features: the heading date of the sample (quantitative), the allelic variant of the *Vrn-A1* gene (qualitative), and the region of accession collection sites (qualitative). The two-block partial least squares method (2B-PLS) was applied to each pair of blocks (Rohlf, Corti, 2000). Then, the correlations between the obtained bicomponents were calculated. Calculations and visualization of the obtained results were carried out in the software package for statistical analysis Statistica 12.6 (StatSoft).

Results

Study of growth habit and duration of the vegetation period in the studied *T. dicoccoides* accessions

To study the growth habit (spring/winter type) of plants, 137 accessions of wild emmer *T. dicoccoides* were planted at a hydroponic greenhouse at the Institute of Cytology and Genetics SB RAS. Among the studied accessions, 12 (IG 45495, IG 45491, IG 46394, PI 414721, IG 45500, PI 355457, PI 560817, k-41965, k-26119, PI 467002, k-5199, PI 352366) had been previously characterized as spring ones (URL: https://npgsweb.ars-grin.gov/gringlobal/method?id=491608; URL: https://www.genesys-pgr.org; URL: https://grs.icarda.org, accessed April 2, 2025). These were used as spring controls. Among the controls, plants of accession IG 45495 (Syria) were the latest heading. Of the 125 previously unstudied accessions, 98 did not do transition to reproductive development and remained at the tillering stage, while the remaining 27 accessions formed spikes no later than the late spring control IG 45495. Considering that all seedlings were planted without vernalization, we classified 98 accessions that failed to head as winter types, and 27 accessions as spring ones. In subsequent studies of the nucleotide sequences of *Vrn-A1* gene alleles, only the 27 identified spring accessions and 12 spring controls were studied.

For all spring accessions, the duration of the period from seedling planting to heading (in days) was recorded. The plants exhibited substantial variation in the duration of the vegetation period; the earliest-maturing accession, IG 46253 (Turkey), reached heading in 34 days, while the latest-maturing one, IG 45495, took 93 days (Table 1). The obtained data were used to calculate the correlation between the allelic variants of the *Vrn-A1* gene and the duration of the vegetation period in plants.

Analysis of the nucleotide sequences of the promoter region of the *Vrn-A1* gene

The studied *T. dicoccoides* accessions revealed six distinct variants of the *Vrn-A1* gene promoter. Five variants corresponded to alleles previously described in this species. Ten accessions carried the *Vrn-A1u* allele; two, *Vrn-A1b.1*; four, *Vrn-A1b.2*; five, *Vrn-A1b.4*; and eighteen, *Vrn-A1d* (Table 1). Among 39 analyzed sequences, none contained additional SNPs or other mutations in the VRN-box and GArG-box regions compared to previously described variants (Fig. 2). Three accessions (PI 355457, PI 190919, PI 560817) were of particular interest as they simultaneously contained two *Vrn-A1* gene promoters. One sequence was identical to the *Vrn-A1d* allelic variant. The other sequence contained three deletions relative to the intact promoter variant: a 32 bp deletion located between –234 bp and –201 bp upstream of the start codon; a 19 bp deletion between –159 bp and –139 bp upstream; and a 1 bp deletion at –138 bp upstream. The deletion located farthest from the start codon encompassed 8 bp of the VRN-box, while the remaining sequence of this site contained a T to C substitution at position –197 bp upstream. This allelic variant, discovered in the present study, was designated as *Vrn-A1b.8*. All sequences have been deposited in GenBank (accession numbers PV699347–PV699388).

2B-PLS analysis

For statistical analysis of the obtained results, we used three data blocks describing the accessions: heading time (quantitative trait), *Vrn-A1* gene allelic variants (the number of binary traits equals the number of alleles), and region of origin (the number of binary traits equals the number of origin regions). For each pair of blocks (heading–allele, heading–region of origin, allele–region of origin), we applied the two-block partial least squares method (2B-PLS) (Rohlf, Corti, 2000). The data on *Vrn-A1* gene allelic variants and heading times of accessions obtained in the present study were supplemented with similar data from previously published studies where plants were grown under identical conditions (hydroponic greenhouse, 16-hour photoperiod) (Shcherban et al., 2015b; Konopatskaia et al., 2016; Muterko et al., 2016).

During the analysis, we also considered information about the geographic origin (collection site) of accessions to evaluate the influence of different *Vrn-1* gene alleles on plant heading time, contributing to wheat adaptation to environments. For accessions containing two *Vrn-A1* alleles, we treated them as a separate variant, recording the alleles present in the accession as separated by a slash (e. g., *Vrn-A1d*/*Vrn-A1f*). We only considered the first pair of axes (designated as uAx1 and

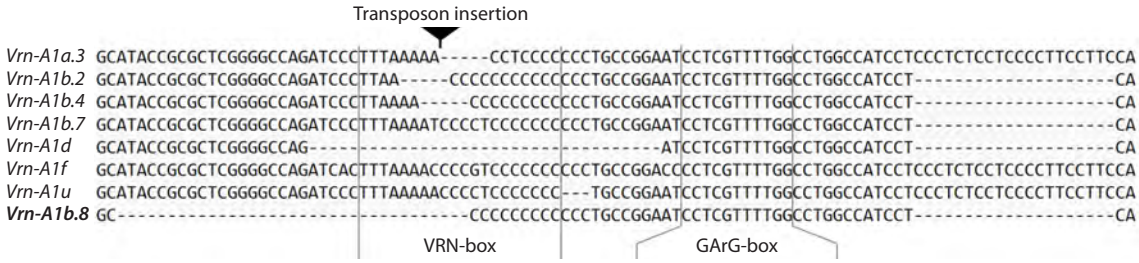


Fig. 2. Alignment of *Vrn-A1* gene promoter sequences found in 39 spring *T. dicoccoides*. The newly identified sequence is shown in bold.

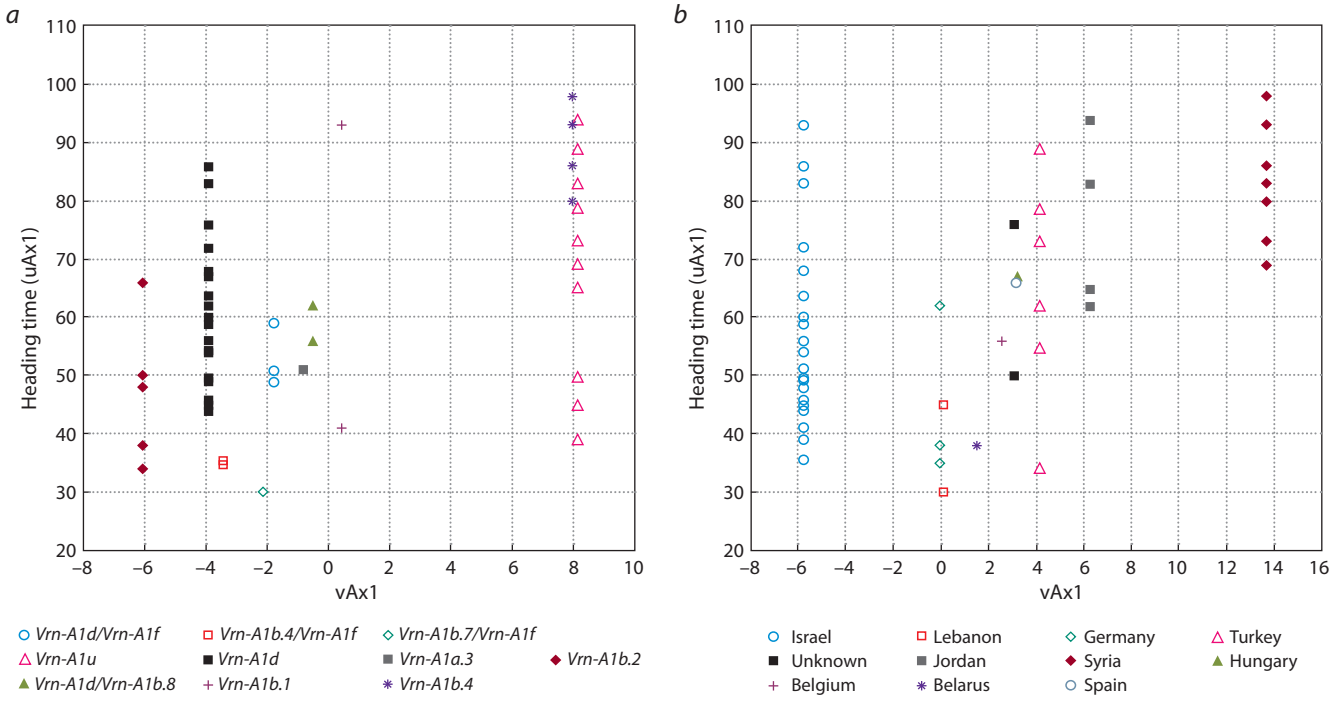


Fig. 3. Configuration of samples on the planes of the first bicomponents.
a – vAx1 calculated based on allelic variants of the *Vrn-A1* gene, *b* – vAx1 calculated for regional samples.

vAx1) showing the highest covariance (Fig. 3). When a block contained only one trait ("heading time"), it constituted the sole (first) bicomponent of that block (uAx1).

When applying 2B-PLS analysis to assess the relationship between the duration of the vegetation period and allelic variants of the *Vrn-A1* gene, we derived one axis (uAx1) from the "heading time" trait and another axis (vAx1) from the binary "accession–allele" matrix. The correlation coefficient (r) between the axes was 0.53 (moderate association), p-value = 3.88×10^{-6} (Table 2). No specific allelic variant of the *Vrn-A1* gene showed a correlation coefficient with axis uAx1 ≥0.5. The only allele with a close value (0.45) was *Vrn-A1b.4*, while other allelic variants showed correlations <0.3 with the "heading time" trait. The strongest associations with axis vAx1 were observed for alleles *Vrn-A1d* (r = –0.61) and *Vrn-A1u* (r = 0.8).The opposite signs of the correlation coefficients for these two allelic variants suggest their opposing effects on the trait.

Using the same method, we analyzed the influence of genetic adaptation mechanisms by deriving axis vAx1 from the "accession–origin region" matrix. The second axis (uAx1), as in the previous case, corresponded to the "heading time" trait. The correlation coefficient between the axes showed a weaker association than in the analysis of the relationship between vegetation period duration and *Vrn-A1* alleles (r = 0.47, p-value = 4.92×10^{-5}) (Table 3). Accessions from Syria showed a correlation coefficient with plant heading time of r = 0.46, while accessions from other regions demonstrated insignificant associations with this trait (r < 0.3). The strongest associations with axis vAx1 were observed for accessions from Israel and Syria (r = –0.864 and r = 0.812, respectively). The difference in signs of the correlation coefficients indicates opposing effects of different genetic adaptation mechanisms on the duration of the vegetation period.

Discussion

Combinations of allelic variants of *Vrn-1* genes significantly influence the agronomically important trait of "duration of the vegetation period" in cultivated wheat species (Flood, Halloran, 1986; Goncharov, 1998; Kato et al., 1998; Distelfeld

Table 2. Correlation matrix (×1000) between *Vrn-A1* allelic variants and plant heading time

Trait	uAx1	vAx1	Heading time	*Vrn-A1a.3*	*Vrn-A1b.1*	*Vrn-A1b.2*	*Vrn-A1b.4*	*Vrn-A1b.4/ Vrn-A1f*	*Vrn-A1b.7/ Vrn-A1f*	*Vrn-A1d*	*Vrn-A1d/ Vrn-A1f*	*Vrn-A1d/ Vrn-A1b.8*	*Vrn-A1u*
uAx1	–	534	–	–75	55	–295	450	–274	–229	–109	–111	–21	290
vAx1	534	–	534	–19	14	–347	415	–110	–48	–611	–71	–21	802
Heading time	–	534	–	–75	55	–295	450	–274	–229	–109	–111	–21	290
Vrn-A1a.3	–75	–19	–75	–	–22	–39	–36	–22	–15	–106	–27	–27	–67
Vrn-A1b.1	55	14	55	–22	–	–56	–51	–31	–22	–152	–39	–39	–96
Vrn-A1b.2	–295	–347	–295	–39	–56	–	–91	–56	–39	–271	–69	–69	–171
Vrn-A1b.4	450	415	450	–36	–51	–91	–	–51	–36	–246	–62	–62	–155
Vrn-A1b.4/Vrn-A1f	–274	–110	–274	–22	–31	–56	–51	–	–22	–152	–39	–39	–96
Vrn-A1b.7/Vrn-A1f	–229	–48	–229	–15	–22	–39	–36	–22	–	–106	–27	–27	–67
Vrn-A1d	–109	–611	–109	–106	–152	–271	–246	–152	–106	–	–187	–187	–466
Vrn-A1d/Vrn-A1f	–111	–71	–111	–27	–39	–69	–62	–39	–27	–187	–	–48	–118
Vrn-A1d/Vrn-A1b.8	–21	–21	–21	–27	–39	–69	–62	–39	–27	–187	–48	–	–118
Vrn-A1u	290	802	290	–67	–96	–171	–155	–96	–67	–466	–118	–118	–

Note. Color here and in Table 3: light red, light green – $p < 0.001$, red, green – $p < 10^{-4}$.

Table 3. Matrix of relationships (×1000) between the region of origin or habitat of accession and the time of plant heading

Trait	uAx1	vAx1	Heading time	Belgium	Belarus	Germany	Hungary	Israel	Jordan	Lebanon	Spain	Syria	Turkey	Unknown
uAx1	–	478	–	–41	–171	–213	39	–265	213	–249	31	460	65	14
vAx1	478	–	478	44	26	–1	56	–864	221	2	55	812	182	75
Heading time	–	478	–	–41	–171	–213	39	–265	213	–249	31	460	65	14
Belgium	–41	44	–41	–	–15	–27	–15	–132	–32	–22	–15	–52	–39	–22
Belarus	–171	26	–171	–15	–	–27	–15	–132	–32	–22	–15	–52	–39	–22
Germany	–213	–1	–213	–27	–27	–	–27	–232	–55	–39	–27	–92	–69	–39
Hungary	39	56	39	–15	–15	–27	–	–132	–32	–22	–15	–52	–39	–22
Israel	–265	–864	–265	–132	–132	–232	–132	–	–270	–188	–132	–449	–336	–188
Jordan	213	221	213	–32	–32	–55	–32	–270	–	–45	–32	–107	–80	–45
Lebanon	–249	2	–249	–22	–22	–39	–22	–188	–45	–	–22	–75	–56	–31
Spain	31	55	31	–15	–15	–27	–15	–132	–32	–22	–	–52	–39	–22
Syria	460	812	460	–52	–52	–92	–52	–449	–107	–75	–52	–	–134	–75
Turkey	65	182	65	–39	–39	–69	–39	–336	–80	–56	–39	–134	–	–56
Unknown	14	75	14	–22	–22	–39	–22	–188	–45	–31	–22	–75	–56	–

Note. For Germany, Hungary, Belarus, and Belgium, these indicate not natural habitats, but GenBank deposit locations.

et al., 2009a; Kamran et al., 2014; Shcherban et al., 2015a; Smolenskaya et al., 2022; Smolenskaya, Goncharov, 2023; Afshari-Behbahanizadeh et al., 2024; Plotnikov et al., 2024; Goncharov et al., 2025). Studying the allelic composition of these genes in wild species and the relationship between different alleles and vegetation period duration is crucial for expanding diversity and enhancing plasticity of cultivated species, as well as their adaptation to specific growing conditions in different regions. Populations of *T. dicoccoides* contain spring accessions where mutant variants of *Vrn-1* genes emerged independently from those in widely cultivated *T. aestivum* and *T. durum* Desf. (Shcherban et al., 2015b; Konopatskaia et al., 2016; Muterko et al., 2016). Introgression of wild emmer alleles into cultivated species would allow to expand their polymorphism and enable finer tuning of heading times in commercial cultivars.

In this study, we examined the growth habit (spring *vs*. winter) in 137 *T. dicoccoides* accessions and analyzed the promoter region of the *Vrn-A1* gene in 39 identified spring accessions. Thirty-six of them contained five allelic variants (*Vrn-A1b.1*, *Vrn-A1b.2*, *Vrn-A1b.4*, *Vrn-A1d*, *Vrn-A1u*) previously described in *T. dicoccoides* (Yan et al., 2004a; Shcherban et al., 2015b; Konopatskaia et al., 2016; Muterko et al., 2016). The presence of the *Vrn-A1b.1*, *Vrn-A1b.2* or *Vrn-A1d* alleles in wild emmer genomes has been shown to be a determining factor for spring growth habit (Yan et al., 2004a; Shcherban et al., 2015b; Konopatskaia et al., 2016; Muterko et al., 2016). Three of the 39 spring *T. dicoccoides* accessions simultaneously carried two different *Vrn-A1* variants. One allele sequence matched the previously described *Vrn-A1d* promoter variant, while the other contained deletions differing from known variants. The National Center for Biotechnology Information (NCBI) database contains no nucleotide sequences identical to this allelic variant. The closest match was *Vrn-A1b.7*, from which our newly identified variant differed by a 32 bp deletion located between –234 bp and –201 bp upstream of the start codon, and a T to C substitution at –197 bp upstream. We designated this novel variant as *Vrn-A1b.8*. The deletion in the *Vrn-A1b.8* promoter encompassed 8 bp of the VRN-box region.

Searching for *Vrn-A1* promoter sequences with similar VRN-box deletions revealed the closest match to be the *Vrn-A1o* allele (Zhang B. et al., 2023), which carries a larger (10 bp) deletion from the 5′ end of VRN-box. The *Vrn-A1o* allele has been shown to confer spring growth habit in common wheat (Zhang B. et al., 2023). This suggests that *Vrn-A1b.8* may also determine spring growth habit. However, the presence of the dominant *Vrn-A1d* allele, which induces spring growth habit even as a single copy (Golovnina et al., 2010), in these accessions currently prevents definitive assessment of the *Vrn-A1b.8* effect on spring type. The presence of two dominant alleles in one accession could result from plant material heterogeneity or copy number variation (CNV) due to locus duplication. Fixation of two different *Vrn-1* alleles in *T. dicoccoides* genomes has been previously demonstrated (Konopatskaia et al., 2016). Moreover, this phenomenon has been observed in several other tetraploid wheat species (Golovnina et al., 2010; Chhuneja et al., 2015). Therefore, CNV in the genomes of these three spring *T. dicoccoides* accessions with two *Vrn-A1* promoter copies appears to be a more probable explanation.

Following the determination of nucleotide sequences of the *Vrn-A1* gene promoter in this study, we evaluated the relationship between its allelic variants and the vegetation period duration in wild emmer. For our analysis, we supplemented the data with heading time values from several other studies where promoter allelic variants had been precisely identified through nucleotide sequencing (Shcherban et al., 2015b; Konopatskaia et al., 2016; Muterko et al., 2016). Combining data from different investigations, even under similar conditions, may introduce certain biases, although excluding portions of observations seemed unjustified as it would substantially reduce our accession sample (bulk). Considering the specific nature of our data, we selected the 2B-PLS method for statistical analysis due to its advantages over traditional approaches. Classical methods such as ANOVA and multiple regression require strict assumptions, including normal distribution, and demonstrate high sensitivity to multicollinearity and outliers. The use of quadratic criteria in these methods can lead to biased estimates, particularly with small sample sizes, increasing the probability of Type I errors. In contrast, the 2B-PLS method offers greater robustness through its use of latent variables, resulting in reduced sensitivity to outliers and multicollinearity. These characteristics make our chosen method particularly suitable for analyzing biological data characterized by statistical noise and complex factor interactions, which is especially important given our study's specific features, including limited sample sizes.

Since copy number variation (CNV) of the dominant *Vrn-A1* gene affects the duration from emergence to heading (Grogan et al., 2016), we considered the presence of two different alleles of dominant *Vrn* genes in a single accession as a distinct variant. We established correlation coefficients of $r = 0.534$ and $r = 0.478$ for the relationships "heading time × allelic variants" and "heading time × regions of origin", respectively. While these coefficients allowed assessment of parameter relationships, their values preclude definitive conclusions about whether *Vrn-A1* allelic variants or region-specific genetic factors predominantly influence the vegetation period duration trait in *T. dicoccoides*.

Previous studies have repeatedly demonstrated significant effects of specific dominant *Vrn-A1* alleles on maturation timing in common wheat (Royo et al., 2020; Qiu et al., 2021; Chumanova, Efremova, 2024). Our results suggest that analogous effects of this gene's allelic variants in wild emmer *T. dicoccoides* are less pronounced. We acknowledge that our experiments were conducted exclusively under controlled greenhouse conditions without replicates and with a limited number of accessions. These methodological features impose certain limitations on result interpretation. Nevertheless, despite these limitations, our data demonstrate weak associations between the studied parameters. Similar experimental approaches – particularly testing under controlled conditions without replicates – have been employed in previously published studies evaluating plant heading times (Kippes et al., 2014, 2015; Palomino, Cabrera, 2023). Despite simplified designs, these authors confirmed phenotypic differences between compared groups, supporting the validity of such approaches. Additionally, we must acknowledge that our study analyzed only the *Vrn-A1* promoter region. While this is the most variable region in *T. dicoccoides*, combinations of the *Vrn-A1* with *Vrn-B1* alleles, as well as *Ppd-1* allele combinations, may influence vegetation period duration. Genotypes characteristic of specific wild emmer collection regions showed lower correlation with heading times than *Vrn-A1* allelic variants, suggesting minimal influence of geographic origin on plant heading times.

The information obtained in this study could be valuable for breeding spring bread and durum wheats, particularly considering that *T. dicoccoides* is actively used as a genetic donor for these species (Badaeva et al., 2024). Furthermore, studies have demonstrated that *Aegilops tauschii* Coss. (syn. *Ae. squarrosa* L.) accessions (Takumi et al., 2011; Chepurnov et al., 2023), similar to *T. dicoccoides* (Table 1), exhibit significant polymorphism in the duration of the vegetation period trait. Therefore, hybridization of *Ae. tauschii* with spring

T. dicoccoides accessions could facilitate the production of artificial hexaploid ($2n = 6x = 42$) amphidiploids, which may serve as a promising platform for successful introgression of novel dominant *Vrn* gene allelic variants determining vegetation period duration from these species into bread wheat.

Conclusion

This study identified a novel *Vrn-A1b.8* allele in spring *T. dicoccoides* accessions. We detected a significant association (p-value = 3.88×10^{-6}) between allelic variants of the dominant *Vrn-A1* gene and vegetation period duration, as well as an association (p-value = 4.92×10^{-5}) between this parameter and the geographic origins (collection sites) of the wild emmer accessions. The research revealed that some *T. dicoccoides* accessions sharing identical *Vrn-A1* alleles and originating from the same eco-geographical region show substantial variation in duration of the vegetation period. The observed phenotypic variability for this trait despite identical *Vrn-A1* allelic composition suggests the involvement of additional genetic determinants controlling this characteristic in these accessions. These findings highlight the potential value of wild emmer accessions as genetic resources (donors) for expanding the genetic diversity of commercial bread and durum wheat varieties..

References

Afshari-Behbahanizadeh S., Puglisi D., Esposito S., De Vita P. Allelic variations in vernalization (*Vrn*) genes in *Triticum* spp. *Genes*. 2024; 15:251. doi 10.3390/genes15020251

Badaeva E.D., Davoyan R.O., Tereshchenko N.A., Lyalina E.V., Zoshchuk S.A., Goncharov N.P. Cytogenetic features of intergeneric amphydiploids and genome-substituted forms of wheat. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(7):716-730. doi 10.18699/VJGB-24-80

Chen A., Dubcovsky J. Wheat TILLING mutants show that the vernalization gene *Vrn1* down-regulates the flowering repressor *Vrn2* in leaves but is not essential for flowering. *PLoS Genet.* 2012;8: e1003134. doi 10.1371/journal.pgen.1003134

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

Chepurnov G., Ovchinnikova E., Blinov A., Chikida N., Belousova M., Goncharov N.P. Analysis of the structural organization and expression of the *Vrn-D1* gene controlling growth habit (spring vs. winter) in *Aegilops tauschii* Coss. *Plants.* 2023;12:3596. doi 10.3390/plants12203596

Chhuneja P., Arora J.K., Kaur P., Kaur S., Singh K. Characterization of wild emmer wheat *Triticum dicoccoides* germplasm for vernalization alleles. *J Plant Biochem Biotechnol.* 2015;24:249-253. doi 10.1007/s13562-014-0281-7

Chumanova E., Efremova T. Marker-assisted development of wheat lines of the winter cultivar Bezostaya 1 and the effects of interaction between alleles of *Vrn-A1L* and *Vrn-B1* loci on heading time. *Cereal Res Commum.* 2024;52(4):1287-1298. doi 10.1007/s42976-023-00478-z

Distelfeld A., Li C., Dubcovsky J. Regulation of flowering in temperate cereals. *Curr Opin Plant Biol.* 2009a;12:178-184. doi 10.1016/j.pbi.2008.12.010

Distelfeld A., Tranquilli G., Li C., Yan L., Dubcovsky J. Genetic and molecular characterization of the *Vrn2* loci in tetraploid wheat. *Plant Physiol.* 2009b;149:245-257. doi 10.1104/pp.108.129353

Dong P., Wei Y.-M., Chen G.-Y., Li W., Wang J.-R., Nevo E., Zheng Y.-L. Sequence-related amplified polymorphism (*SRAP*) of wild emmer wheat (*Triticum dicoccoides*) in Israel and its ecological association. *Biochem Syst Ecol.* 2010;38(1):1-11. doi 10.1016/j.bse.2009.12.015

Dubcovsky J., Loukoianov A., Fu D., Valarik M., Sanchez A., Yan L. Effect of photoperiod on the regulation of wheat vernalization genes *Vrn1* and *Vrn2*. *Plant Mol Biol.* 2006;60:469-480. doi 10.1007/s11103-005-4814-2

Efremova T.T., Chumanova E.V. Stages of growth and development of wheat and their importance in the formation of productivity elements. *Pisma v Vavilovskii Zhurnal Genetiki i Selektsii = Letters to Vavilov Journal of Genetics and Breeding.* 2023;9(2):54-80. doi 10.18699/LettersVJ-2023-9-09 (in Russian)

Feng K., Nie X., Cui L., Deng P., Wang M., Song W. Genome-wide identification and characterization of salinity stress-responsive miRNAs in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*). *Genes.* 2017;8:156. doi 10.3390/genes8060156

Flood R.G., Halloran G.M. Genetics and physiology of vernalization response in wheat. *Adv Agron*. 1986;39:87-125. doi 10.1016/S0065-2113(08)60466-6

Fu D., Szűcs P., Yan L., Helguera M., Skinner J.S., Von Zitzewitz J., Hayes P.M., Dubcovsky J. Large deletions within the first intron in *Vrn1* are associated with spring growth habit in barley and wheat. *Mol Genet Genomics*. 2005;273:54-65. doi 10.1007/s00438-004-1095-4

Golovnina K.A., Kondratenko E., Blinov A.G., Goncharov N.P. Molecular characterization of vernalization loci *Vrn1* in wild and cultivated wheats. *BMC Plant Biol*. 2010;10:168. doi 10.1186/1471-2229-10-168

Goncharov N.P. Genetic resources of wheat related species: the *Vrn* genes controlling growth habit (spring *vs.* winter). *Euphytica*. 1998; 100:371-376. doi 10.1023/A:1018323600077

Goncharov N.P. Response to vernalization in wheat: its quantitative or qualitative nature. *Cereal Res Commun*. 2004a;32:323-330. doi 10.1007/BF03543317

Goncharov N.P. Genetics of growth habit (spring vs. winter) in tetraploid wheats: production and analysis of near-isogenic lines. *Hereditas*. 2004b;130:125-130. doi 10.1111/j.1601-5223.1999.00125.x

Goncharov N.P. Shitova I.P. The inheritance of growth habit in old local varieties and landraces of hexaploid wheat. *Russ J Genet*. 1999;35(4):386-392

Goncharov N.P., Plotnikov K.O., Chepurnov G.Y., Kamenev I.A., Nemtsev B.F., Smolenskaya S.E., Blinov A.G. Global warming and the genetics of wheat type and rate of development. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Biologiya = Tomsk State University Journal of Biology*. 2025;69:68-75. doi 10.17223/19988591/69/8 (in Russian)

Grogan S.M., Brown-Guedira G., Haley S.D., McMaster G.S., Reid S.D., Smith J., Byrne P.F. Allelic variation in developmental genes and effects on winter wheat heading date in the U.S. great plains. *PLoS One*. 2016;11:e0152852. doi 10.1371/journal.pone.0152852

Hemming M.N., Peacock W.J., Dennis E.S., Trevaskis B. Low-temperature and daylength cues are integrated to regulate *FLOWERING LOCUS T* in barley. *Plant Physiol*. 2008;147:355-366. doi 10.1104/pp.108.116418

Kamran A., Iqbal M., Spaner D. Flowering time in wheat (*Triticum aestivum* L.): a key factor for global adaptability. *Euphytica*. 2014; 197:1-26. doi 10.1007/s10681-014-1075-7

Kato K., Mori Y., Beiles A., Nevo E. Geographical variation in heading traits in wild emmer wheat, *Triticum dicoccoides*. I. Variation in vernalization response and ecological differentiation. *Theor Appl Genet*. 1997;95:546-552. doi 10.1007/s001220050595

Kato K., Tanizoe C., Beiles A., Nevo E. Geographical variation in heading traits in wild emmer wheat, *Triticum dicoccoides*. II. Variation in heading date and adaptation to diverse ecogeographical conditions. *Hereditas*. 1998;128(1):33-39. doi 10.1111/j.1601-5223.1998.00033.x

Kippes N., Zhu J., Chen A., Vanzetti L., Lukaszewski A., Nishida H., Kato K., Dvorak J., Dubcovsky J. Fine mapping and epistatic inter-

actions of the vernalization gene *Vrn-D4* in hexaploid wheat. *Mol Gen Genomics*. 2014;289:47-62. doi 10.1007/s00438-013-0788-y

Kippes N., Debernardi J.M., Vasquez-Gross H.A., Akpinar B.A., Budak H., Kato K., Chao S., Akhunov E., Dubcovsky J. Identification of the *Vernalization 4* gene reveals the origin of spring growth habit in ancient wheats from South Asia. *Proc Natl Acad Sci USA*. 2015;112(39):E5401-E5410. doi 10.1073/pnas.151488311

Kippes N., Chen A., Zhang X., Lukaszewski A.J., Dubcovsky J. Development and characterization of a spring hexaploid wheat line with no functional *Vrn2* genes. *Theor Appl Genet*. 2016;129:1417-1428. doi 10.1007/s00122-016-2713-3

Kiss T., Horváth Á.D., Cseh A., Berki Z., Balla K., Karsai I. Molecular genetic regulation of the vegetative–generative transition in wheat from an environmental perspective. *Ann Bot*. 2025;135(4):605-628. doi 10.1093/aob/mcae174

Konopatskaia I., Vavilova V., Kondratenko E.Ya., Blinov A., Goncharov N.P. *Vrn1* genes variability in tetraploid wheat species with a spring growth habit. *BMC Plant Biol*. 2016;16:244. doi 10.1186/s12870-016-0924-z

Lack H.W., Van Slageren M. The discovery, typification and rediscovery of wild emmer wheat, *Triticum turgidum* subsp. *dicoccoides* (Poaceae). *Willdenowia*. 2020;50(2):207-216. doi 10.3372/wi.50.50206

Law C.N., Worland A.J. Genetic analysis of some flowering time and adaptive traits in wheat. *New Phytol*. 1997;137(1):19-28. doi 10.1046/j.1469-8137.1997.00814.x

Liang Y., Zhang D.-Y., Ouyang S., Xie J., Wu Q., Wang Z., Cui Y., … Dvorak J., Huo N., Sun Q., Gu Y.-Q., Liu Z. Dynamic evolution of resistance gene analogs in the orthologous genomic regions of powdery mildew resistance gene *MlIW170* in *Triticum dicoccoides* and *Aegilops tauschii*. *Theor Appl Genet*. 2015;128:1617-1629. doi 10.1007/s00122-015-2536-7

Muterko A., Kalendar R., Salina E. Novel alleles of the *VERNALIZATION1* genes in wheat are associated with modulation of DNA curvature and flexibility in the promoter region. *BMC Plant Biol*. 2016;16(Suppl.1):9. doi 10.1186/s12870-015-0691-2

Nevo E. Genetic resources of wild emmer, *Triticum dicoccoides*, for wheat improvement in the third millennium. *Isr J Plant Sci*. 2001;49(1):77-92. doi 10.1560/XJQN-9T4H-VTL3-CDXU

Novoselskaya-Dragovich A.Yu., Fisenko A.V., Konovalov F.A., Lyapunova O.A., Kudryavtsev A.M., Badaeva E.D., Goncharov N.P. Assessment of genetic diversity of tetraploid wheat *Triticum dicoccum* Schrank ex Schübl. based on long terminal repeat (LTR) retrotransposon integration sites and C-banding analysis. *Genet Resour Crop Evol*. 2025;72:8819-8839. doi 10.1007/s10722-025-02487-9

Ouyang S., Zhang D., Han J., Zhao X., Cui Y., Song W., Huo N., … Ling H.-Q., Luo M., Gu Y., Sun Q., Liu Z. Fine physical and genetic mapping of powdery mildew resistance gene *MlIW172* originating from wild emmer (*Triticum dicoccoides*). *PLoS One*. 2014;9:e100160. doi 10.1371/journal.pone.0100160

Özkan H., Willcox G., Graner A., Salamini F., Kilian B. Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genet Resour Crop Evol*. 2011;58:11-53. doi 10.1007/s10722-010-9581-5

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;24:16041. doi 10.3390/ijms242216041

Plotnikov K.O., Klimenko A.I., Ovchinnikova E.S., Lashin S.A., Goncharov N.P. Analysis of the effects of the *Vrn1* and *Ppd-1* alleles on adaptive and agronomic traits in common wheat (*Triticum aestivum* L.). *Plants*. 2024;13:1453. doi 10.3390/plants13111453

Qiu L., Liu N., Wang H., Shi X., Li F., Zhang Q., Wang W., Guo W., Hu Z., Li H., Ma J., Sun Q., Xie C. Fine mapping of a powdery mildew resistance gene *MlIW39* derived from wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*). *Theor Appl Genet*. 2021;134:2469-2479. doi 10.1007/s00122-021-03836-9

Rivera D., Ferrer-Gallego P., Obón C., Alcaraz F., Laguna E., Goncharov N.P., Kislev M. Fossil or non-fossil: a case study in the archaeological wheat *Triticum parvicoccum* (Poaceae: Triticeae). *Genes*. 2025;16(3):274. doi 10.3390/genes16030274

Rohlf F.J., Corti M. Use of two-block partial least-squares to study covariation in shape. *Syst Biol*. 2000;49:740-753. doi 10.1080/106351500750049806

Royo C., Dreisigacker S., Soriano J.M., Lopes M.S., Ammar K., Villegas D. Allelic variation at the vernalization response (*Vrn-1*) and photoperiod sensitivity (*Ppd-1*) genes and their association with the development of durum wheat landraces and modern cultivars. *Front Plant Sci*. 2020;11:838. doi 10.3389/fpls.2020.00838

Saidou M., Wang C., Alam Md.A., Chen C., Ji W. Genetic analysis of powdery mildew resistance gene using SSR markers in common wheat originated from wild emmer (*Triticum dicoccoides* Thell). *Turkish J Field Crops*. 2015;21:10. doi 10.17557/tjfc.83589

Sela H., Ezrati S., Ben-Yehuda P., Manisterski J., Akhunov E., Dvorak J., Breiman A., Korol A. Linkage disequilibrium and association analysis of stripe rust resistance in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) population in Israel. *Theor Appl Genet*. 2014;127:2453-2463. doi 10.1007/s00122-014-2389-5

Shcherban A., Salina E.A. Evolution of *Vrn1* homoeologous loci in allopolyploids of *Triticum* and their diploid precursors. *BMC Plant Biol*. 2017;17:188. doi 10.1186/s12870-017-1129-9

Shcherban A., Emtseva M., Efremova T. Molecular genetical characterization of vernalization genes *Vrn-A1*, *Vrn-B1* and *Vrn-D1* in spring wheat germplasm from Russia and adjacent regions. *Cereal Res Commun*. 2012a;40:351-361. doi 10.1556/CRC.40.2012.3.4

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

Shcherban A., Khlestkina E., Efremova T., Salina E.A. The effect of two differentially expressed wheat *Vrn-B1* alleles on the heading time is associated with structural variation in the first intron. *Genetica*. 2013;141(4-6):133-141. doi 10.1007/s10709-013-9712-y

Shcherban A., Börner A., Salina E.A. Effect of *Vrn-1* and *Ppd D1* genes on heading time in European bread wheat cultivars. *Plant Breed*. 2015a;134(1):49-55. doi 10.1111/pbr.12223

Shcherban A., Strygina K., Salina E.A. *Vrn-1* gene-associated prerequisites of spring growth habit in wild tetraploid wheat *T. dicoccoides* and the diploid A genome species. *BMC Plant Biol*. 2015b;15:94. doi 10.1186/s12870-015-0473-x

Smolenskaya S.E., Goncharov N.P. Allelic diversity of the *Vrn* genes and the control of growth habit and earliness in wheat. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(8):933-946. doi 10.18699/VJGB-23-108

Smolenskaya S.E., Efimov V.M., Kruchinina Y.V., Nemtsev B.F., Chepurnov G.Y., Ovchinnikova E.Y., Belan I.A., Zuev E.V., Chenxi Zhou, Piskarev V.V., Goncharov N.P. Earliness and morphotypes of common wheat cultivars of Western and Eastern Siberia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2022;26(7):662-674. doi 10.18699/VJGB-22-81

Soresi D., Zappacosta D., Garayalde A., Irigoyen I., Basualdo J., Carrera A. A valuable QTL for Fusarium head blight resistance from *Triticum turgidum* L. ssp. *dicoccoides* has a stable expression in durum wheat cultivars. *Cereal Res Commun*. 2017;45:234-247. doi 10.1556/0806.45.2017.007

Soresi D., Bagnaresi P., Crescente J.M., Díaz M., Cattivelli L., Vanzetti L., Carrera A. Genetic characterization of a fusarium head blight resistance QTL from *Triticum turgidum* ssp. *dicoccoides*. *Plant Mol Biol Rep*. 2021;39:710-726. doi 10.1007/s11105-020-01277-0

Stelmakh A.F. Genetic systems regulating flowering response in wheat. *Euphytica*. 1998;100(1-3):359-369. doi 10.1023/A:1018374116006

Strejčková B., Mazzucotelli E., Čegan R., Milec Z., Brus J., Çakır E., Mastrangelo A.M., Özkan H., Šafář J. Wild emmer wheat, the pro-

genitor of modern bread wheat, exhibits great diversity in the *Vernalization1* gene. *Front Plant Sci*. 2023;13:1106164. doi 10.3389/fpls.2022.1106164

Takumi S., Koyama K., Fujiwara K., Kobayashi F. Identification of a large deletion in the first intron of the *Vrn-D1* locus, associated with loss of vernalization requirement in wild wheat progenitor *Aegilops tauschii* Coss. *Genes Genet Syst*. 2011;86:183-195. doi 10.1266/ggs.86.183

Trevaskis B., Hemming M.N., Dennis E.S., Peacock W.J. The molecular basis of vernalization-induced flowering in cereals. *Trends Plant Sci*. 2007;12:352-357. doi 10.1016/j.tplants.2007.06.010

Würschum T., Langer S.M., Longin C.F.H., Tucker M.R., Leiser W.L. A three component system incorporating *Ppd-D1*, copy number variation at *Ppd-B1*, and numerous small effect quantitative trait loci facilitates adaptation of heading time in winter wheat cultivars of worldwide origin. *Plant Cell Environ*. 2018;41:1407-1416. doi 10.1111/pce.13167

Xue F., Ji W., Wang C., Zhang H., Yang B. High-density mapping and marker development for the powdery mildew resistance gene *PmAS846* derived from wild emmer wheat (*Triticum turgidum* var. *dicoccoides*). *Theor Appl Genet*. 2012;124:1549-1560. doi 10.1007/s00122-012-1809-7

Yan L., Loukoianov A., Tranquilli G., Helguera M., Fahima T., Dubcovsky J. Positional cloning of the wheat vernalization gene *Vrn1*. *Proc Natl Acad Sci USA*. 2003;100:6263-6268. doi 10.1073/pnas.0937399100

Yan L., Helguera M., Kato K., Fukuyama S., Sherman J., Dubcovsky J. Allelic variation at the *Vrn1* promoter region in polyploid wheat. *Theor Appl Genet*. 2004a;109:1677-1686. doi 10.1007/s00122-004-1796-4

Yan L., Loukoianov A., Blechl A., Tranquilli G., Ramakrishna W., San-Miguel P., Bennetzen J.L., Echenique V., Dubcovsky J. The wheat *Vrn2* gene is a flowering repressor down-regulated by vernalization. *Science*. 2004b;303:1640-1644. doi 10.1126/science.1094305

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;103:19581-19586. doi 10.1073/pnas.060714210

Zhang B., Guo Y., Fan Q., Li R., Chen D., Zhang X. Characterization and distribution of novel alleles of the vernalization gene *Vrn-A1* in Chinese wheat (*Triticum aestivum* L.) cultivars. *Crop J*. 2023;11:852-862. doi 10.1016/j.cj.2022.10.002

Zhang H., Zhang L., Wang C., Wang Y., Zhou X., Lv S., Liu X., Kang Z., Ji W. Molecular mapping and marker development for the *Triticum dicoccoides* – derived stripe rust resistance gene *YrSM139-1B* in bread wheat cv. Shaanmai 139. *Theor Appl Genet*. 2016;129:369-376. doi 10.1007/s00122-015-2633-7

Conflict of interest. The authors declare no conflict of interest.
Received April 3, 2025. Revised July 22, 2025. Accepted July 23, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1288-1294

doi 10.18699/vjgb-25-135

Frame-based mathematical models – a tool for the study of molecular genetic systems

F.V. Kazantsev [1, 2, 3], S.A. Lashin [1, 2, 3], Yu.G. Matushkin [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
kazfdr@bionet.nsc.ru

Abstract. This paper reviews existing approaches for reconstructing frame-based mathematical models of molecular genetic systems from the level of genetic synthesis to models of metabolic networks. A frame-based mathematical model is a model in which the following terms are specified: formal structure, type of mathematical model for a particular biochemical process, reactants and their roles. Typically, such models are generated automatically on the basis of description of biological processes in terms of domain-specific languages. For molecular genetic systems, these languages use constructions familiar to a wide range of biologists in the form of a list of biochemical reactions. They rely on the concepts of elementary subsystems, where complex models are assembled from small block units ("frames"). In this paper, we have shown an example with the generation of a classical repressilator model consisting of three genes that mutually inhibit each other's synthesis. We have given it in three different versions of the graphic standard, its characteristic mathematical interpretation and variants of its numerical calculation. We have shown that even at the level of frame models it is possible to identify qualitatively new behaviour of the model through the introduction of just one gene into the model structure. This change provides a way to control the modes of behaviour of the model through changing the concentrations of reactants. The frame-based approach opens the way to generate models of cells, tissues, organs, organisms and communities through frame-based model generation tools that specify structure, roles of modelled reactants using domain-specific languages and graphical methods of model specification.

Key words: Hill functions; mathematical modelling; gene networks; frame-based models; domain-specific languages

For citation: Kazantsev F.V., Lashin S.A., Matushkin Yu.G. Frame-based mathematical models – a tool for the study of molecular genetic systems. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1288-1294. doi 10.18699/vjgb-25-135

Funding. This work was supported by the Ministry of Science and Higher Education of the Russian Federation (the Federal Scientific-technical programme for genetic technologies development for 2019–2030, agreement No. 075-15-2025-516).

Фреймовые математические модели – инструмент исследования молекулярно-генетических систем

Ф.В. Казанцев [1, 2, 3], С.А. Лашин [1, 2, 3], Ю.Г. Матушкин [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
kazfdr@bionet.nsc.ru

Аннотация. Представлен обзор подходов реконструкции фреймовых математических моделей молекулярно-генетических систем от уровня генетического синтеза до метаболических сетей. Фреймовые математические модели – это модели, в которых заданы: формальная структура, математический формализм под конкретный биохимический процесс, реагенты этого процесса и их роль в нем. Обычно такие модели созданы в автоматическом режиме на базе описания биологического процесса в терминах языков предметной области. Для молекулярно-генетических систем такие языки используют конструкции, привычные для широкого круга исследователей-биологов, например список биохимических реакций. Они основываются на концепции элементарных подсистем, где комплексная модель собрана из небольших блоков – «фреймов». В настоящей работе мы показали пример с генерацией классической модели репрессилятора, состоящей из трех генов, продукты синтеза которых последовательно ингибируют биосинтез друг друга. Для этого примера мы привели три версии графической нотации описания структуры модели, их типичную математическую интерпретацию

и варианты вычислительных экспериментов. Показали, что даже на уровне фреймовых моделей возможно идентифицировать качественно новое поведение благодаря добавлению в структуру модели всего одного гена. Такая модификация предоставляет пути контроля режимами функционирования модели через изменение концентраций ее реагентов. Подход, основанный на фреймах, открывает пути генерации моделей клеток, тканей, органов, организмов и сообществ с помощью программных инструментов, которые формализуют структуру модели и роль ее реагентов, используя как предметно-ориентированные языки, так и графические методы спецификации моделей.
Ключевые слова: функции Хилла; математическое моделирование; генные сети; фреймовые модели; предметно-ориентированные языки спецификации модели

Introduction

In the era of accumulation of large genetic data, the question of high-throughput analysis of these data using methods of mathematical and computer modelling has arisen. The last 30 years of scientific experience have prepared the theoretical basis for computational analysis tasks. The mechanisms of biochemical catalysis were studied, the rates of biochemical processes were estimated, and the scenarios of transcriptional synthesis and the influence of external factors were examined (Alon, 2007; Wittig et al., 2018; Kolmykov et al., 2021; Vorontsov et al., 2024; Rigden, Fernández, 2025). A mathematical foundation has been prepared, which, together with the development of experimental and computational technologies, has set the trend for the transition from small models of enzymatic kinetics to full-genome models of bacterial (Karr et al., 2014) and animal cells (Norsigian et al., 2019). Existing approaches rely on the concepts of elementary subsystems, where complex models are assembled from small block units. The rejection of "monolithic" models in a single mathematical formalism (e. g. the ODE system) in favour of representing and storing models as a set of model units has set the direction for the integration of multiple tools through standards. These model units carry structural information about the role of each reactant and the mathematical interpretation of the biological process (Malik-Sheriff et al., 2019). This tremendous amount of knowledge allows to pass through ordering of information to the formalization of models, that is, the ability to propose a formal structure and type of mathematical model for a specific biochemical process with known reactants – a *frame model*. If we can offer a formal algorithm for translating knowledge into a model, then the way of setting the initial configuration can be left as usual for scientists – through a domain-specific language to describe the model.

What are frame models?

Frame-based mathematical models in the field of molecular genetic systems (MGS) modelling are most often understood as models that have been made using domain-specific model description languages (DSL) and tools for model generation on their basis. For MGS, such languages use constructions familiar to a wide range of biologists in the form of a list of biochemical reactions, for example (Shapiro et al., 2003, 2013; Hoops et al., 2006). This way allows one to set the reactants of biochemical processes and their roles. It is the information about the role of reactants that is the necessary element for algorithms of frame models construction, crucial parameters for formal generation rules. Moreover, the mathematical formalism may differ depending on the problem to be solved. These can be likened to models in the form of ODE system, discrete models or based on Boolean logic (Beal et al., 2011; Galdzicki et al., 2014).

In this paper, we review approaches for describing models that consider only the levels of transcription, translation, enzymatic synthesis, signaling networks and metabolic pathways. This limitation reveals the structure and hierarchical arrangement of such models, from the basic processes of synthesis to the arrangement of everything into a system of interacting metabolic pathways. Within the framework of building such MGS models, the bottom-up approach is natural – when one moves from models of simple processes to their combination, obtaining a synergistic/emergent effect (Kolodkin et al., 2012, 2013). This process of increasing the model complexity resembles a design based on the principle of staking "domino tiles" by the rule of reactants overlapping.

In addition to domain specific languages, the process of designing frame models can be started from a structural model/schema/graph of interacting entities with specified roles of participants. Such a graphical way of specifying the reactants and their roles for modelling is more illustrative and allows the use of additional analysis tools in case of working with a big amount of data. These graphs are the input markup for the frame model generation stage. In general, it is possible to unambiguously switch from representing the role of reactants from a list of biochemical reactions to a graph representation and vice versa. There are several standards for the presentation of structural information: SBML (Hucka et al., 2015), SBGN (Moodie et al., 2015), SBOL (Galdzicki et al., 2014).

Basic theorems underlying the approaches

Some theoretical issues concerning the integration of frame models are worth clarifying first. The simple frame models of MGS are built on the basis of chemical kinetics equations representing various ways of the kinetic mass action law. The construction is carried out within the paradigm of local independence of functional properties of elementary subsystems from their compartmental localization within the original system. If a particular structural model or reactions scheme is available, the instantaneous concentration rate of any substance is equal to the sum of the local concentration rates of that substance for each reaction in which that substance participates as a substrate or product. The theoretical basis for this is Korzukhin's theorem, which is crucial for modelling chemical kinetics and states: "For any set of non-negative curves given on a finite time interval and any given accuracy, there exists such (there may be more than one) biochemical scheme composed only of bimolecular and monomolecular reactions that the mathematical model constructed from this biochemical scheme approximates the given set of curves

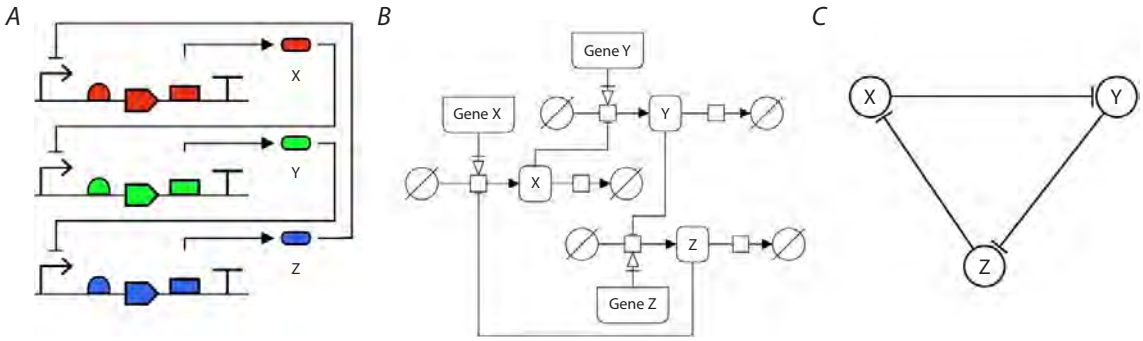


Fig. 1. Graphical representation of the repressilator model in different standards: *A*, SBOL notation describes in detail the arrangement/structure of a DNA molecule and the functional relationship between the elements. *B*, SBGN notation carries more information about the processes that compose the model. *C*, Notation of protein-protein regulatory interactions, specific for describing Hypothetical Gene Networks (omits many details of the mechanisms).

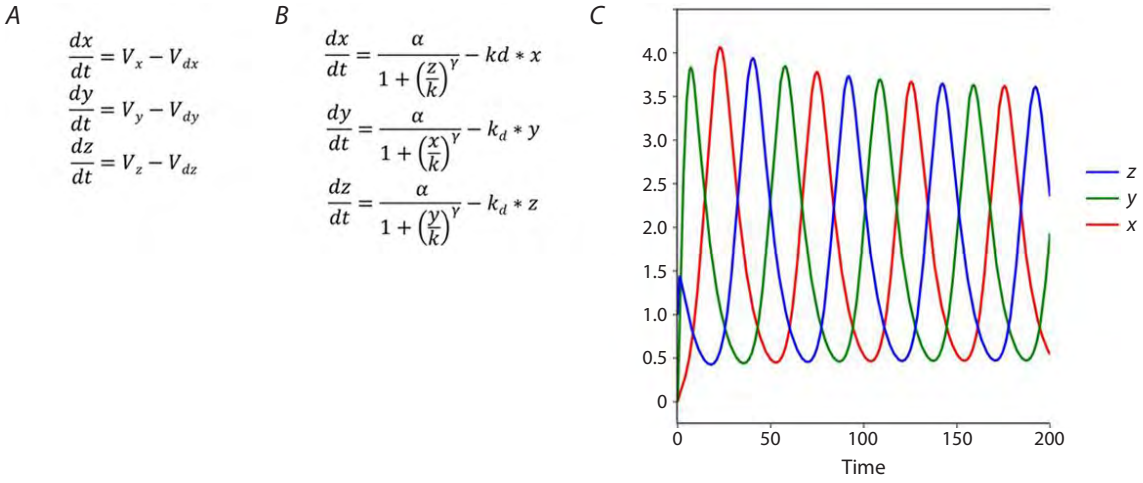


Fig. 2. Generalized view of the frame mathematical model and characteristic plots of concentration variation with time obtained for the structure from Figure 1.

A, Model representation as combination of elementary subsystems. V_i – as synthesis processes, V_{di} – as dissipation processes. *B*, The model ODE system description. *C*, The dynamics is obtained with parameters $\gamma > 2$, $\alpha = 1$, $k = 0.5$, $k_d = 0.1$ and initial point $x_0 = y_0 = 0$, $z_0 = 1.0$.

with a given accuracy" (Zhabotinsky, Zaikin, 1973). The extension of these ideas is formulated in the framework of the generalised chemical kinetic modelling method (GCKMM), proposed by Vitaly Likhoshvay (Likhoshvai et al., 2001). Some examples of this approach are presented below.

Frame model examples

Processes of genetic synthesis – transcription, translation

When designing frame models at the level of genetic synthesis, it is necessary to set the process structure and roles of the reactants in the process. These reactants are genes and transcription factors. An interesting approach to describe the genetic level is the SBOL approach (Galdzicki et al., 2014) (The Synthetic Biology Open Language, sbolstandard.org). It allows describing the structure of a DNA molecule with the location of genes, binding sites of transcription factors, regulatory relations of synthesis products from genes and some other properties. It is possible to describe them both in text form using a domain-specific modelling language or in a graph form using a special graphical editor (Der et al., 2017; Cox et al., 2018). There is a set of tools for working with this standard and tools for graphical interpretation of such models (Fig. 1*A*).

For further presentation a demonstration of a well-known three genes model is given, each of them inhibits synthesis from a neighboring gene – a repressilator (Elowitz, Leibler, 2000). Figure 1 shows a possible graphical interpretation of such a model under various types of representation.

Once the structure of relations between molecules is given, it is possible to build a "frame mathematical model" describing the dynamics of the process. The role of the element allows one to understand at a glance the contribution of the selected subsystem and its negative/positive effect in the overall system of equations (Fig. 2*A*). For this purpose, we can use approaches that are described in such tools as SBMLSqeezer (Dräger et al., 2015), MGSgenerator (Kazantsev et al., 2009), MicroTranscriptMod (Lakhova et al., 2022). These tools, given the role of each of the reactants in the process, generate equations for the rate of product synthesis (Fig. 2*B*). The more details on the role of the reactants are described, the more precisely

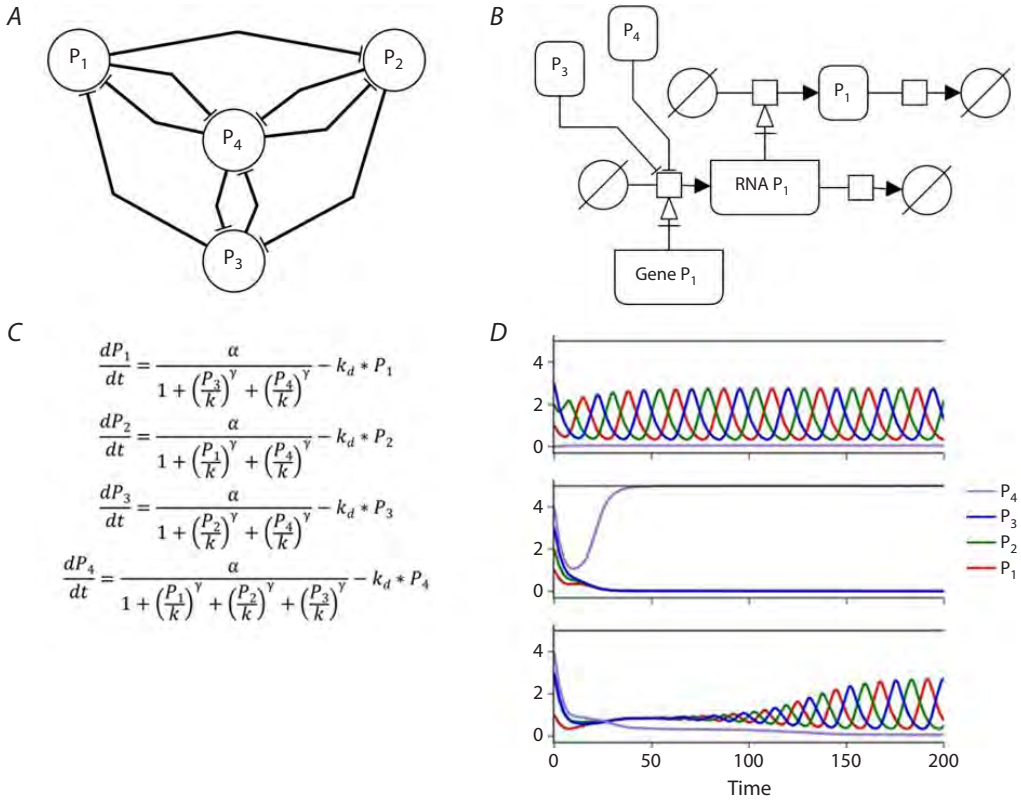


Fig. 3. A four-node model of hypothetical gene network (HGS) and its characteristic behavior. *A*, Structural model, where arcs define the conditions of biosynthesis inhibition. *B*, An extended description of the processes behind each of the HGS nodes in SBGN standard. *C*, Model equations that correspond to the structure. *D*, Characteristic plots of concentration *vs* time obtained for the presented structure from Figure *A*.

The dynamics are obtained at parameters $\gamma = 3$, $\alpha = 1$, kb = 0.5, $k_d = 0.2$ and starting points [P_1, P_2, P_3, P_4]: (1) [1.0, 2.0, 3.0, 0.0]; (2) [1.0, 2.0, 3.0, 4.0]; (3) [1.0, 2.8, 3.0, 4.0].

their behavior can be specified in the model. At the stage of computational experiments, all such tools combine elementary subsystems (separate processes) into one unifying structure, the one common model. The typical view of the mathematical model and its dynamics for each of the structures presented in Figure 1 will look as it is shown in Figure 2.

Hypothetical gene networks

A model does not often require excessive detail. There is a class of models where the relationship of genes and their synthesis products with other genes is modelled. The structural model is represented by a unipartite graph, where each node represents both the process of transcription of the coding part of a gene and translation of its mRNA (or synthesis of its protein) (Fig. 1*C*, Fig. 3). A graph node is considered as a unified transcription-translation process. Directed arcs (arrows) in such a graph define an inhibiting or activating effect on another node (on itself is also allowed). This class of models is named Hypothetical Gene Network (HGN) (Likhoshvai et al., 2003).

Hypothetical gene networks with cyclic inhibitory effects of reactants (which are specified with the relationships: "protein P_i inhibits the synthesis of product P_j from gene g_j", see Figure 3) are exhaustively described in (Fadeev, Likhoshvai, 2003). Each edge in a graph representation of such models affects the generalized transcription/translation process of the node to which it is directed. Moreover, when generating ODE models for these graphs, a third process – decay of the synthesis product – is added to the mentioned processes. A node in such graphs is understood as follows: "An RNA molecule r_i is synthesised from the gene g_i, from which a P_i protein is synthesised and this protein is degraded/dissociated over time". Figure 3*A* shows an example of an HGS model of four nodes and nine edges specifying the conditions of biosynthesis inhibition. The structure is obtained by inserting one additional node into the model shown in Figure 1. The additional node inhibits the others, and they in turn inhibit it. The resulting mathematical model is presented in Figure 3*C*. While the model presented in Figure 1 has one unstable steady-state condition and stable oscillatory behaviour under parameters presented in Figure 2, the introduction of an additional node allows the behaviour of the model to change depending on the concentration of reactants (Figure 3*D*): at low initial concentrations of P_4 the system oscillates stably as the initial model, and at sufficiently high concentrations of P_4 it enters the stationary regime. Moreover, regime switching can be controlled by small changes in concentrations. Changing the regulation mechanism to non-competitive does not change the regimes of the model behavior. The model found in Supplementary Materials[1] as a Copasi file is a model version with the non-competitive regulation mechanism in it.

[1] Supplementary Materials are available at:
https://vavilovj-icg.ru/download/pict-2025-29/appx50.zip

Studying such a class of models and forming a knowledge base of their dynamics allows us to identify possible behavior at the level of structural models of target biological systems without performing calculations.

Enzymatic reaction

For enzymatic synthesis processes, the key aspect is the presence of an enzyme, which catalyzes the process but is not consumed in the course of the reaction. To reconstruct a model of enzymatic reactions, the reaction mechanisms, the order of interaction of molecular players of the reaction with the enzyme, the steps of transformation and release of the product should be taken into account. Once one has an assumption about the mechanisms of the enzyme's relationships with substrates and products, a suitable form of representing these interactions as a graph may be suggested. There are a number of works in this area. One can use ready-made solutions for building frame models of enzymatic reactions on graphs (King, Altman, 1956; Cornish-Bowden, 1977) (www.biokin.com/tools/king-altman/). In addition, the Copasi tool has a set of predefined frame-based mathematical models for enzymatic systems (Hoops et al., 2006). These models may not only be used as examples in a case study, but also be valuable in developing and analyzing a model within the Copasi toolkit: design a set of elementary subsystems (model structure); give them a mathematical law of velocity based on frame models; get a ready system of equations; perform computational experiments in both continuous and discrete stochastic formalisms; perform computational analyses of the model to fit the parameters to the experimental data and test the robustness of the model to variation of the parameters.

SBMLsqeezer can serve as an independent source of frame models (Dräger et al., 2015). It is both a database of ready-made model variants and a tool that can match a well-annotated structure in the form of an SBML model to a suitable model variant. This tool can be embedded into the CellDesigner application (Funahashi et al., 2008). There is a ready set of equations adapted to experimental data for both enzymatic reactions and transcription-translation processes in the bacterium *E. coli* (Kazantsev et al., 2018). These models may serve as a training sample because they contain accompanying information about the data items on which the models were built.

Metabolic pathways

Frame models of metabolic pathways can also be derived from structural information in the form of graphs. It is possible to build a model through descriptions of reactions in tabular form with COBRApy (Ebrahim et al., 2013) and BIOUML (Kolpakov et al., 2022). These tools allow the construction of a whole-genome model in terms of flux balance modelling (Orth et al., 2010). But if one needs to work within continuous models, the Path2Models project (Büchel et al., 2013) may be used, in which 140,000 frame models were generated based on structural models from the KEGG database (Kanehisa, 2000). These models are available at the biomodels.net resource (Malik-Sheriff et al., 2019). This kind of automation in model building is also available as part of the Cellerator package for the Mathematica modelling environment (Shapiro et al., 2003, 2013).

Signalling pathways

Signalling pathways require different approaches. Within such pathways, it is necessary to take into account the change of states of one molecule and/or the formation of molecular complexes, the change of conformational states, the consideration of active centres of molecules, etc. For automating the generation of these types models, the BioNetGen resource is being developed (bionetgen.org) (Harris et al., 2016). The key feature is that a series of allowed states of molecules is described, their active centres, and the rules of interaction through the active centres. These data are specified within a domain specific language. The visualization of this kind of relationship is specified within the "SBGN:entity relations" standard (sbgn.github.io). In order to try these models, one can use the VScode development environment module (code.visualstudio.com). BioNetGen algorithms build reaction chains themselves (structural models) and propose frame models for them in the widely used SBML standard. Then one can run a series of computational experiments using both discrete methods and continuous methods in several specialized computational tools that support SBML models as input data.

Designing and depositing of the model

When designing models using automation and autogeneration, one faces the problem of identifying the right entity in the lists of variables and parameters. If the model is formatted as a monolithic system of differential equations within the particular syntax in one of the engineering modelling environments, one has to map each of the x_i of the model to the proper biology entity through reading the accompanying publications. At best, authors will name variables as short acronyms for proteins or metabolites. On the other hand, it is more difficult to come up with some general rule for naming parameters correctly. It was the transition to the representation of MGS models as a set of elementary subsystems corresponding to an independent biochemical process that allowed to solve most of the mentioned issues (Miller et al., 2010; Hucka et al., 2015). In this concept, a model is not a set of equations, but an instruction on how to assemble a target model in the target mathematical formalism from tens, hundreds and thousands of pieces of elementary subsystems distributed over compartments and perform a series of computational experiments with it. In order to end up with a development-friendly model, it is better to follow a series of recommendations for the design of such elementary blocks using the systems biology ontology (SBO) (Courtot et al., 2011). This ontology allows to associate the rate equations of processes and the parameters of these equations with the meanings that were given in classical studies.

The problem of model annotation is well highlighted in a publication on the model reproducibility crisis (Tiwari et al., 2021), where the authors showed that 51 % of mathematical models published on the largest online resource (biomodels.net) are not reproducible, for various reasons. It is the frame models that partially solve the issue of both *repeatability* of computational experiments and *reproducibility* of modelling results, as the relevant toolkits contain references to the formalism, to the methods used and correctly describe the parameters with the use of ontologies. All of those questions are studied in depth due to community efforts. One of such

communities is “co.mbine.org”, an initiative to coordinate the development of various standards and formats for computational models in systems biology.

Discussion

Advancement of technology has given impulse to the processes of development of artificial languages for describing models within scientific fields. We have reviewed existing solutions for designing frame-based mathematical models of molecular genetic systems. For each of them there are specific tools for representing models and performing computational experiments. The publication (Tiwari et al., 2021) proposed metrics for evaluating the resulting models in terms of readiness for reuse in new research. If one follows the proposed guidelines for incorporating annotations into a model that can be made available to modelling tools, it will enhance the possibilities of automating model processing. Model automation is an interesting route with the goal of being able to integrate off-the-shelf subsystems into comprehensive cellular, intercellular and organ level models. More and more software libraries for engineering simulation environments are becoming available where molecular genetic systems modelling approaches can already be used. Even in questions of designing industrial samples of bacterial synthesis, one comes to embrace standardization for the subsequent automation of processes. This applies to the issues of model development, their integration into the production biotechnological cycle and monitoring with updating of knowledge bases (Herold et al., 2017).

Whereas in the 2000s there were trends towards developing proprietary solutions that included a “friendly user interface”, now the trends tend towards the use of highly specialized core software for each of the stages and/or the use of specialized libraries via API calls: VScode as an editor for BioNetGen DSL; yEd or Cytoscape as tools for displaying model structure; Copasi as a general-purpose tool for computational experiments, etc.

Data analysis is also performed by general-purpose statistical processing libraries or off-the-shelf tools (dashboard) that only need to load data. Now a necessary skill for work in systems biology is proficiency in Python/R/Bash scripting languages for building pipelines and linking data between function calls of specific libraries.

References

Alon U. Network motifs: theory and experimental approaches. *Nat Rev Genet.* 2007;8(6):450-461. doi 10.1038/nrg2102

Beal J., Lu T., Weiss R. Automatic compilation from high-level biologically-oriented programming language to genetic regulatory networks. *PLoS One*. 2011;6(8):e22490. doi 10.1371/journal.pone.0022490

Büchel F., Rodriguez N., Swainston N., Wrzodek C., Czauderna T., Keller R., Mittag F., … Saez-Rodriguez J., Schreiber F., Laibe C., Dräger A., Le Novère N. Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol.* 2013;7:116. doi 10.1186/1752-0509-7-116

Cornish-Bowden A. An automatic method for deriving steady-state rate equations. *Biochem J.* 1977;165(1):55-59. doi 10.1042/bj1650055

Courtot M., Juty N., Knüpfer C., Waltemath D., Zhukova A., Dräger A., Dumontier M., … Wilkinson D.J., Wimalaratne S., Laibe C., Hucka M., Le Novère N. Controlled vocabularies and semantics in systems biology. *Mol Syst Biol.* 2011;7(1):543. doi 10.1038/msb.2011.77

Cox R.S., Madsen C., McLaughlin J., Nguyen T., Roehner N., Bartley B., Bhatia S., … Voigt C.A., Zundel Z., Myers C., Beal J., Wipat A. Synthetic Biology Open Language Visual (SBOL Visual). Version 2.0. *J Integr Bioinform.* 2018;15(1):20170074. doi 10.1515/jib-2017-0074

Der B.S., Glassey E., Bartley B.A., Enghuus C., Goodman D.B., Gordon D.B., Voigt C.A., Gorochowski T.E. DNAplotlib: programmable visualization of genetic designs and associated data. *ACS Synth Biol.* 2017;6(7):1115-1119. doi 10.1021/acssynbio.6b00252

Dräger A., Zielinski D.C., Keller R., Rall M., Eichner J., Palsson B.O., Zell A. SBMLsqueezer 2: context-sensitive creation of kinetic equations in biochemical networks. *BMC Syst Biol.* 2015;9(1):68. doi 10.1186/s12918-015-0212-9

Ebrahim A., Lerman J.A., Palsson B.O., Hyduke D.R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol.* 2013;7(1):74. doi 10.1186/1752-0509-7-74

Elowitz M.B., Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000;403(6767):335-338. doi 10.1038/35002125

Fadeev S.I., Likhoshvai V.A. On hypothetical gene networks. *Sib Zh Ind Mat.* 2003;6(3):134-153

Funahashi A., Matsuoka Y., Jouraku A., Morohashi M., Kikuchi N., Kitano H. CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proc IEEE*. 2008;96(8):1254-1265. doi 10.1109/JPROC.2008.925458

Galdzicki M., Clancy K.P., Oberortner E., Pocock M., Quinn J.Y., Rodriguez C.A., Roehner N., … Villalobos A., Wipat A., Gennari J.H., Myers C.J., Sauro H.M. The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat Biotechnol.* 2014;32(6):545-550. doi 10.1038/nbt.2891

Harris L.A., Hogg J.S., Tapia J.-J., Sekar J.A.P., Gupta S., Korsunsky I., Arora A., Barua D., Sheehan R.P., Faeder J.R. BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics*. 2016;32(21):3366-3368. doi 10.1093/bioinformatics/btw469

Herold S., Krämer D., Violet N., King R. Rapid process synthesis supported by a unified modular software framework. *Eng Life Sci.* 2017;17(11):1202-1214. doi 10.1002/elsc.201600020

Hoops S., Sahle S., Gauges R., Lee C., Pahle J., Simus N., Singhal M., Xu L., Mendes P., Kummer U. COPASI – a COmplex PAthway SImulator. *Bioinformatics*. 2006;22(24):3067-3074. doi 10.1093/bioinformatics/btl485

Hucka M., Bergmann F.T., Hoops S., Keating S.M., Sahle S., Schaff J.C., Smith L.P., Wilkinson D.J. The Systems Biology Markup Language (SBML): language specification for level 3 version 1 core. *J Integr Bioinform.* 2015;12(2):382-549. doi 10.2390/biecoll-jib-2015-266

Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):27-30. doi 10.1093/nar/28.1.27

Karr J.R., Phillips N.C., Covert M.W. WholeCellSimDB: a hybrid relational/HDF database for whole-cell model predictions. *Database* (*Oxford*). 2014;2014:bau095. doi 10.1093/database/bau095

Kazantsev F.V., Akberdin I.R., Bezmaternykh K.D., Likhoshvai V.A. The tool for automatic generation of gene networks mathematical models. *Vestnik VOGiS.* 2009;13(1):163-169

Kazantsev F.V., Akberdin I.R., Lashin S.A., Ree N.A., Timonov V.S., Ratushnyi A.V., Khlebodarova T.M., Likhoshvai V.A. MAMMOTh: a new database for curated mathematical models of biomolecular systems. *J Bioinform Comput Biol.* 2018;16(01):1740010. doi 10.1142/S0219720017400108

King E.L., Altman C. A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *J Phys Chem*. 1956;60(10):1375-1378. doi 10.1021/j150544a010

Kolmykov S., Yevshin I., Kulyashov M., Sharipov R., Kondrakhin Y., Makeev V.J., Kulakovskiy I.V., Kel A., Kolpakov F. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res.* 2021;49(D1):D104-D111. doi 10.1093/nar/gkaa1057

Kolodkin A., Simeonidis E., Balling R., Westerhoff H.V. Understanding complexity in neurodegenerative diseases: *in silico* reconstruc-

tion of emergence. *Front Physiol.* 2012;3:291. doi 10.3389/fphys.2012.00291

Kolodkin A., Simeonidis E., Westerhoff H.V. Computing life: add *logos* to biology and *bios* to physics. *Prog Biophys Mol Biol.* 2013;111(2-3):69-74. doi 10.1016/j.pbiomolbio.2012.10.003

Kolpakov F., Akberdin I., Kiselev I., Kolmykov S., Kondrakhin Y., Kulyashov M., Kutumova E., Pintus S., Ryabova A., Sharipov R., Yevshin I., Zhatchenko S., Kel A. BioUML – towards a universal research platform. *Nucleic Acids Res.* 2022;50(W1):W124-W131. doi 10.1093/nar/gkac286

Lakhova T.N., Kazantsev F.V., Mukhin A.M., Kolchanov N.A., Matushkin Y.G., Lashin S.A. Algorithm for the reconstruction of mathematical frame models of bacterial transcription regulation. *Mathematics*. 2022;10(23):4480. doi 10.3390/math10234480

Likhoshvai V.A., Matushkin Y.G., Ratushnyi A.V., Anako E.A., Ignatieva E.V., Podkolodnaya O.A. Generalized chemokinetic method for gene network simulation. *Mol Biol.* 2001;35(6):919-925. doi 10.1023/A:1013254822486

Likhoshvai V.A., Matushkin Y.G., Fadeev S.I. Problems in the theory of the functioning of genetic networks. *Sib Zh Ind Mat.* 2003;6(2):64-80

Malik-Sheriff R.S., Glont M., Nguyen T.V.N., Tiwari K., Roberts M.G., Xavier A., Vu M.T., … Park Y.M., Buso N., Rodriguez N., Hucka M., Hermjakob H. BioModels – 15 years of sharing computational models in life science. *Nucleic Acids Res.* 2019;48(D1):D407-D415. doi 10.1093/nar/gkz1055

Miller A.K., Marsh J., Reeve A., Garny A., Britten R., Halstead M., Cooper J., Nickerson D.P., Nielsen P.F. An overview of the CellML API and its implementation. *BMC Bioinformatics*. 2010;11:178. doi 10.1186/1471-2105-11-178

Moodie S., Le Novère N., Demir E., Mi H., Villéger A. Systems Biology Graphical Notation: Process Description language level 1 version 1.3. *J Integr Bioinform.* 2015;12(2):263. doi 10.1515/jib-2015-263

Norsigian C.J., Pusarla N., McConn J.L., Yurkovich J.T., Dräger A., Palsson B.O., King Z. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* 2019;48(D1):D402-D406. doi 10.1093/nar/gkz1054

Orth J.D., Thiele I., Palsson B.Ø. What is flux balance analysis? *Nat Biotechnol.* 2010;28(3):245-248. doi 10.1038/nbt.1614

Rigden D.J., Fernández X.M. The 2025 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res.* 2025;53(D1):D1-D9. doi 10.1093/nar/gkae1220

Shapiro B.E., Levchenko A., Meyerowitz E.M., Wold B.J., Mjolsness E.D. Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*. 2003;19(5):677-678. doi 10.1093/bioinformatics/btg042

Shapiro B.E., Meyerowitz E.M., Mjolsness E. Using cellzilla for plant growth simulations at the cellular level. *Front Plant Sci.* 2013;4:408. doi 10.3389/fpls.2013.00408

Tiwari K., Kananathan S., Roberts M.G., Meyer J.P., Sharif Shohan M.U., Xavier A., Maire M., … Ng S., Nguyen T.V.N., Glont M., Hermjakob H., Malik-Sheriff R.S. Reproducibility in systems biology modelling. *Mol Syst Biol.* 2021;17(2):e9982. doi 10.15252/msb.20209982

Vorontsov I.E., Eliseeva I.A., Zinkevich A., Nikonov M., Abramov S., Boytsov A., Kamenets V., … Medvedeva Y.A., Jolma A., Kolpakov F., Makeev V.J., Kulakovskiy I.V. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res.* 2024;52(D1):D154-D163. doi 10.1093/nar/gkad1077

Wittig U., Rey M., Weidemann A., Kania R., Müller W. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Res.* 2018;46(D1):D656-D660. doi 10.1093/nar/gkx1065

Zhabotinsky A.M., Zaikin A.N. Autowave processes in a distributed chemical system. *J Theor Biol.* 1973;40(1):45-61. doi 10.1016/0022-5193(73)90164-1

Conflict of interest. The authors declare no conflict of interest. The authors have no relevant financial or non-financial interests to disclose.
Received June 17, 2025. Revised August 12, 2025. Accepted August 22, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1295-1303

doi 10.18699/vjgb-25-136

Computational prediction of the interaction network between long non-coding RNAs and microRNAs in maize based on the transcriptome of the *fuzzy tassel* mutant line

J. Yan[1], A.Yu. Pronozin[2, 3], D.A. Afonnikov[1, 2, 3]

[1] Novosibirsk State University, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
t.yan5@g.nsu.ru

Abstract. Long non-coding RNAs (lncRNAs) play an important role in the regulation of gene expression, including interactions with microRNAs (miRNAs), acting as molecular "sponges". Bioinformatics methods are generally used to predict such interactions. To refine computational predictions, additional evidence based on the co-expression of miRNAs and lncRNAs can be incorporated. In the present study, we investigated potential interactions between lncRNAs and miRNAs in the maize mutant line *fuzzy tassel* (*fzt*), which is characterized by reduced expression of certain miRNAs due to a mutation in the *Dicer-like1* (*DCL1*) gene in shoot and tassel tissues. Transcriptome assembly was performed based on RNA-seq data from maize shoot and tassel tissues of control and mutant lines, with data obtained from the NCBI SRA archive. In the shoot, 10 lncRNAs with significantly altered expression levels between control and mutant groups were identified, 9 of which were upregulated in the mutant plants. In the tassel, 34 differentially expressed lncRNAs were identified, with 20 showing increased expression in the mutant line. For lncRNAs with increased expression and miRNAs with decreased expression in the mutant line, potential interactions were predicted using the machine learning algorithm PmliPred. The IntaRNA program was used to confirm possible complementary binding for the identified miRNA–lncRNA pairs, which enabled the construction of competing endogenous RNA (ceRNA) networks. Structural analysis of these networks revealed that certain lncRNAs are capable of binding multiple miRNAs simultaneously, supporting their regulatory role as "sponges" for miRNAs. The results obtained deepen our understanding of post-transcriptional regulation in maize and open new perspectives for breeding strategies aimed at improving stress tolerance and crop productivity.

Key words: lncRNA; miRNA; gene regulation; maize; *fuzzy tassel* (*fzt*); *DCL1*; bioinformatics; RNA interaction; competing endogenous RNA (ceRNA)

For citation: Yan J., Pronozin A.Yu., Afonnikov D.A. Computational prediction of the interaction network between long non-coding RNAs and microRNAs in maize based on the transcriptome of the *fuzzy tassel* mutant line. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1295-1303. doi 10.18699/vjgb-25-136

Funding. This work was supported by the budgetary project No. FWNR-2022-0020.

Acknowledgements. Data processing was carried out using the computational resources of the "Bioinformatics" Center for Collective Use at the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences.

Компьютерное предсказание сети взаимодействий длинных некодирующих РНК и микроРНК кукурузы на основе транскриптома мутантной линии *fuzzy tassel*

Ц. Янь[1], А.Ю. Пронозин[2, 3], Д.А. Афонников[1, 2, 3]

[1] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
t.yan5@g.nsu.ru

Аннотация. Длинные некодирующие РНК (днРНК) играют важную роль в регуляции экспрессии генов, включая взаимодействия с микроРНК (миРНК), выполняя функцию молекулярных «губок». Для предсказания таких взаимодействий, как правило, применяются методы биоинформатики. Для уточнения предсказаний компьютерных программ можно использовать дополнительные данные на основе коэкспрессии миРНК и днРНК. В настоящей работе исследуются потенциальные взаимодействия между днРНК и миРНК у мутантной

линии кукурузы *fuzzy tassel* (*fzt*), характеризующейся сниженной экспрессией некоторых миРНК вследствие мутации в гене *Dicer-like1* (*DCL1*) в тканях побега и соцветия. Проведена сборка транскриптомов на основе данных RNA-seq побега и соцветия кукурузы контрольной и мутантной линий. Данные были взяты из архива SRA NCBI. Для побега было идентифицировано десять днРНК, достоверно изменяющих свой уровень экспрессии между контрольной и мутантной группами, девять из них повышают экспрессию у мутантных растений. Для соцветия идентифицировано 34 дифференциально экспрессирующихся днРНК (20 с повышенным уровнем экспрессии у мутантных линий). Для днРНК с повышенным уровнем собственной экспрессии и миРНК с пониженным уровнем экспрессии в мутантных линиях были предсказаны потенциальные взаимодействия с помощью алгоритма машинного обучения PmliPred. С использованием программы IntaRNA подтверждена возможность комплементарного связывания для выявленных пар миРНК–днРНК, что позволило построить конкурирующие эндогенные РНК-сети. Анализ структуры этих сетей показал, что отдельные днРНК способны связывать несколько миРНК одновременно, подтверждая их регуляторную функцию в качестве «губок» для миРНК. Полученные результаты углубляют понимание посттранскрипционной регуляции у кукурузы и открывают перспективы для селекционных разработок, направленных на повышение стрессоустойчивости и продуктивности растений.

Ключевые слова: днРНК; миРНК; регуляция генов; кукуруза; мутация *fzt*; *DCL1*; биоинформатика; взаимодействие РНК; конкурирующие эндогенные РНК

Introduction

In recent years, the rapid development of next-generation high-throughput sequencing technologies has enabled the identification of tens of thousands of non-protein-coding transcripts (Sheng et al., 2023). Initially, these sequences were considered transcriptional noise. However, subsequent studies have revealed that approximately 75 % of cellular transcripts lack protein-coding potential, yet they actively participate in the regulation of gene expression (Wang L., Wang J.W., 2015). Non-coding RNAs (ncRNAs) are generally classified into housekeeping and regulatory types. Regulatory ncRNAs can be further divided into small and long non-coding RNAs based on their transcript length (Li R. et al., 2016). To date, the biological functions of small ncRNAs, particularly microRNAs (miRNAs), have been extensively studied; they are capable of repressing mRNA expression at both transcriptional and post-transcriptional levels. In contrast, the functions of long non-coding RNAs (lncRNAs) remain poorly understood, especially in plants.

Recent studies have revealed that lncRNAs and miRNAs engage in complex interactions that play crucial roles in numerous biological processes. Several mechanisms underlying these interactions have been identified (Pronozin, Afonnikov, 2025). For example, lncRNAs can function as molecular “sponges”, binding complementarily to miRNAs and thereby preventing their interaction with target mRNAs. Such interactions contribute to the regulation of plant growth, development, tissue differentiation, and stress responses. However, due to the limited scale of experimental studies, bioinformatic approaches are increasingly needed to identify these interactions (Sheng et al., 2023).

To date, the PmliPred method has been developed to identify interactions between lncRNAs and miRNAs (Kang et al., 2020). This method is based on deep learning for predicting molecular interactions. Information on potential miRNA–lncRNA interactions can be valuable for modeling regulatory networks involved in gene expression. Furthermore, the obtained results can serve as a basis for subsequent functional experiments and may have practical applications in breeding programs. It should also be noted that potential miRNA–lncRNA interactions can be inferred from co-expression analyses (He et al., 2020).

The present study aims to identify interactions between lncRNAs and miRNAs in maize using bioinformatic approaches, taking into account co-expression data of miRNAs and lncRNAs. The *fuzzy tassel* (*fzt*) mutant line of maize, which exhibits disrupted miRNA biogenesis due to a mutation in the *Dicer-like1* (*DCL1*) gene, a key player in the processing of miRNA precursors, was used as a model for this study (Thompson et al., 2014). Impaired *DCL1* function leads to reduced levels of several mature miRNAs, which in turn causes an imbalance in regulatory interactions and, consequently, in the expression of miRNAs and their target mRNAs (Thompson et al., 2014). We hypothesize that the decreased concentration of miRNAs may reduce the formation of duplexes with lncRNAs that act as molecular “sponges”. In this scenario, the degradation rate of lncRNA “sponges” would decrease, leading to an increase in their abundance. Thus, similar to mRNAs exhibiting elevated expression in the *fzt* maize line (Thompson et al., 2014), lncRNAs with increased levels in this line may serve as targets of these miRNAs. The results obtained from this study are expected to enhance our understanding of post-transcriptional regulation in plants and may inform the development of novel breeding strategies aimed at improving stress tolerance and crop productivity (Zhang L. et al., 2009; Sun Q. et al., 2013).

Materials and methods

Transcriptome data. In this study, RNA-seq data were obtained from the open NCBI Sequence Read Archive (SRA) database (accession numbers GSM1277448–GSM1277461, see the Table) (Thompson et al., 2014). The samples were divided into two groups: control and mutant. The mutant lines contained a deletion in the *Dicer-like1* (*DCL1*) gene, which plays a key role in the processing of miRNA precursors. Gene expression was assessed separately for whole seedling and tassel tissues, including both long RNAs and miRNAs.

As shown previously (Thompson et al., 2014), expression of 22 miRNAs was significantly reduced in the seedling (miR398b-5p, miR408a-b-3p, miR408b-5p, miR394a-b-5p, miR167c-3p*, miR156a-3p*, miR167b-3p*, miR319b,d-5p*, miR169i-k-5p, miR167a-d-5p, miR168b-3p*, miR168a-3p*, miR156d-f-g-3p*, miR398a-b-3p, miR528a-b-3p, miR156e-3p*, miR397a-b-5p, miR159a-5p, miR2118b, miR399e,i-j-3p,

RNA-seq libraries of maize (*Zea mays*) obtained from seedling and tassel tissues of control plants and the *fuzzy tassel* (*fzt*) mutant line

SRA identifier	Library name	RNA source	Plant type
GSM1277448	A619_mRNA_1	Seedling	Control
GSM1277449	A619_mRNA_2		
GSM1277450	A619_mRNA_4		
GSM1277451	A619_mRNA_5		
GSM1277452	fzt_mRNA_1		Mutant
GSM1277453	fzt_mRNA_2		
GSM1277454	fzt_mRNA_4		
GSM1277455	fzt_mRNA_5		
GSM1277456	1Tm_Nsib	Tassel	Control
GSM1277457	2Tm_Nsib		
GSM1277458	3Tm_Nsib		
GSM1277459	1Tm_fzt		Mutant
GSM1277460	2Tm_fzt		
GSM1277461	3Tm_fzt		

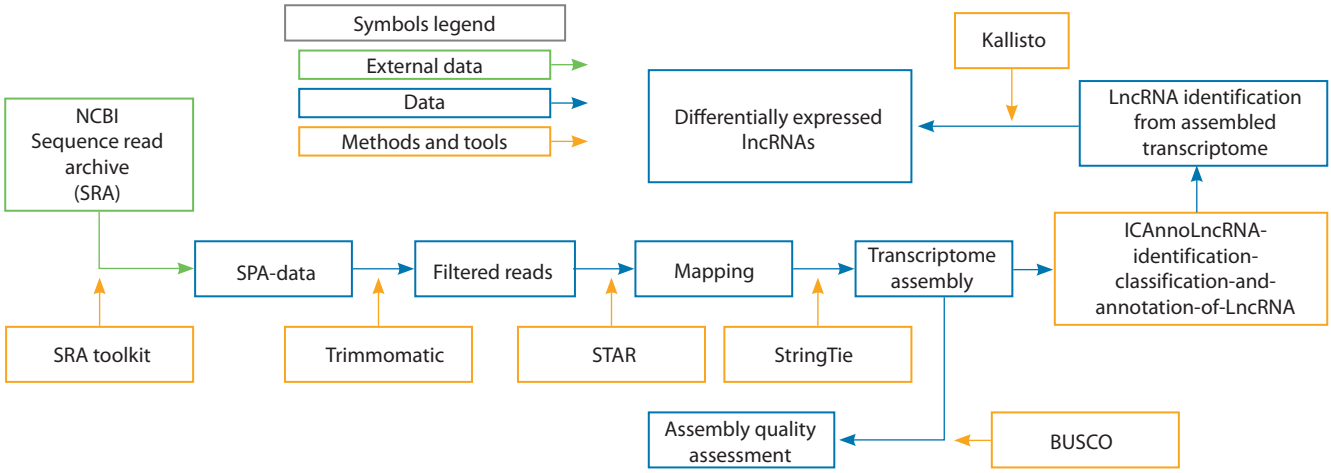


Fig. 1. Workflow of the bioinformatics pipeline for maize transcriptome assembly.

Green rectangles represent the description of external data sources; blue rectangles indicate library data and intermediate results, and orange rectangles denote software tools used in the analysis.

miR160a-e,g-5p, miR398a-5p*) and 14 miRNAs in the tassel (miR167d-3p*, miR167a-d-5p, miR172e, miR408a-b-3p, miR398b-5p*, miR394a-b-5p, miR167c-3p*, miR398a-b-3p, miR319a-d-3p, miR159a-b,f,j-k-3p, miR528a-b-5p, miR160a-e,g-5p, miR166j-k,n-3p, miR159a-5p*).

The reference genome of maize (*Zea mays*) version 5 (Zm-B73-REFERENCE-NAM-5.0) was used in this study, downloaded along with its annotation from the Ensembl Plants database (Bolser et al., 2016).

MiRNA sequences were obtained from miRBase version 22.1 (https://www.mirbase.org/).

Bioinformatics analysis. This study consisted of two main blocks of bioinformatics analysis: transcriptome assembly followed by the differential expression analysis of lncRNAs; prediction of miRNA–lncRNA interactions using deep learning–based approaches. A detailed description of each analytical step is provided below.

Transcriptome assembly and analysis of maize. Transcriptome assembly (Fig. 1) included the following steps: data preprocessing, transcriptome assembly, identification and annotation of lncRNAs, and quantification of transcript expression levels.

Read filtering was performed using Trimmomatic (Bolger et al., 2014) with the following parameters: removal of adapter sequences, elimination of short reads shorter than 36 nucleotides, and quality-based trimming of low-quality reads. After preprocessing, the filtered reads were aligned to the *Z. mays* reference genome using STAR (Dobin et al., 2013). Based on

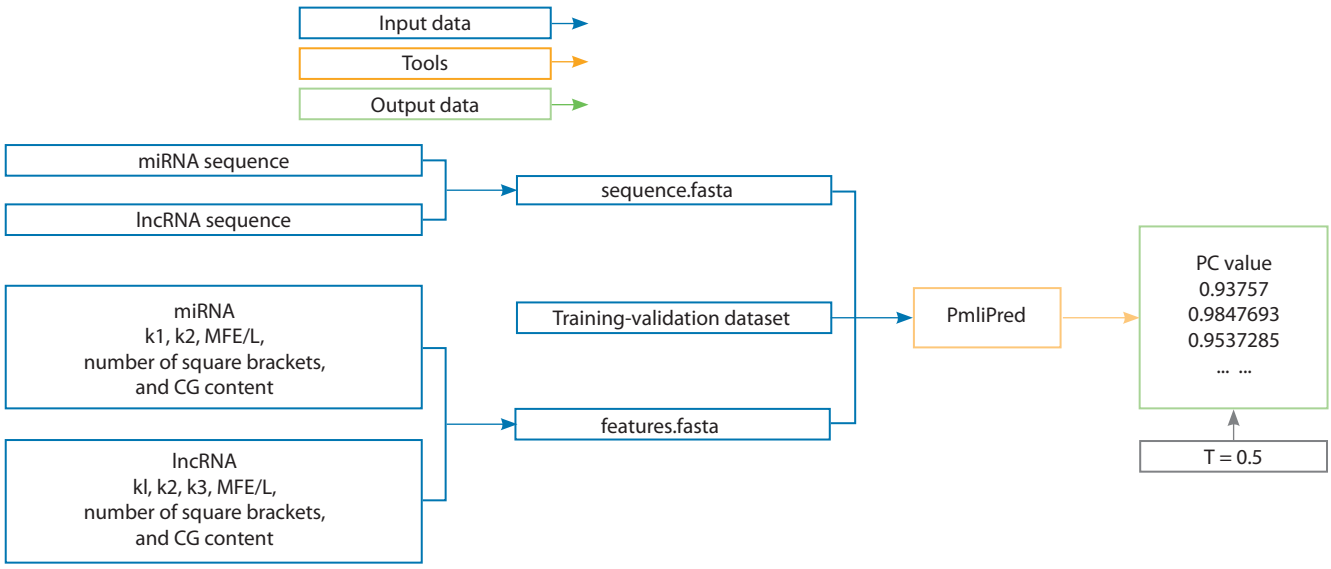


Fig. 2. Workflow of predicting interactions between lncRNAs and miRNAs using the PmliPred model.

Blue rectangles represent the input data, green rectangles indicate the output results, and orange rectangles denote the software tools used in the analysis. The threshold value of confidence probability (T = 0.5) is shown.

the alignment results, transcriptome assembly was conducted using StringTie (Pertea et al., 2015). The completeness and quality of the assembled transcriptome were evaluated with BUSCO (Simão et al., 2015). Identification and annotation of lncRNAs were performed using ICAnnoLncRNA (Pronozin, Afonnikov et al., 2023). Expression levels of identified lncRNAs and other transcripts were quantified using Kallisto (Bray et al., 2016).

Differential expression analysis of lncRNAs in maize. Differential expression analysis of lncRNAs was performed separately for shoot and inflorescence tissues by comparing wild-type (control) and mutant (*fzt*) maize lines. Statistical analysis was conducted using the DESeq2 and edgeR packages (Robinson et al., 2010; Love et al., 2014). Transcripts were considered significantly differentially expressed at a p-value < 0.05, adjusted for multiple testing.

For the differentially expressed lncRNAs, heatmaps of normalized expression values were generated to visualize expression patterns across biological replicates and to confirm the consistency of expression changes between the control and mutant groups.

Analysis of interactions between miRNAs and lncRNAs. Interactions between miRNA and lncRNA molecules were predicted using the PmliPred method (Kang et al., 2020), which involves several consecutive analytical stages (Fig. 2). At the first stage, input data were prepared, including nucleotide sequences of miRNAs and lncRNAs that exhibited downregulated miRNA expression and upregulated lncRNA expression in mutant plants compared with the control. The input to the program also included quantitative sequence features extracted by the built-in algorithms of the model, as well as the training dataset provided with the software package (Kang et al., 2020). For miRNAs, the following features were used: k-mer frequencies (k = 1, k = 2), minimum free energy normalized by length (MFE/L), number of paired nucleotides in the secondary structure, and GC content ratio. For lncRNAs, an additional feature representing k-mer frequencies (k = 3) was extracted.

The processed data were analyzed using the PmliPred program to estimate the interaction probability between miRNA–lncRNA pairs (output parameter PC, confidence probability). A miRNA–lncRNA pair was considered to have a potential interaction when the PC value was ≥ 0.5. The results were presented in a table containing probability scores, which reflected the predicted strength of interaction between miRNA and lncRNA molecules.

Analysis and visualization of interactions between miRNAs and lncRNAs. The obtained miRNA–lncRNA pairs were divided into two groups based on their interaction parameters: lncRNAs with increased expression levels in the mutant line (test group) and lncRNAs with decreased expression levels (control group). Both groups of lncRNAs were compared with all miRNAs showing reduced expression levels, as reported by Thompson et al. (2014) (see section "Transcriptomic data"). As a threshold for selecting potential interactions in the test group, the maximum value of the PC parameter calculated by the PmliPred program for the control group was used. If for a given miRNA–lncRNA pair from the test group, the PC parameter exceeded any of the PC values from the control group, such miRNA–lncRNA pairs were considered to interact.

The sequences of the selected miRNAs and lncRNAs were uploaded into the IntaRNA program (Mann et al., 2017) for the identification and visualization of base-pairing interactions. Among all predicted interactions, only those pairs were retained, in which the number of unpaired nucleotides within the interaction region of the two molecules was fewer than 4, and the length of the interaction region exceeded 16 nucleotides.

Such interactions between lncRNAs and miRNAs have important biological significance. lncRNAs can function as competing endogenous RNAs (ceRNAs), or "sponges", by binding to miRNAs and thereby preventing them from inter-

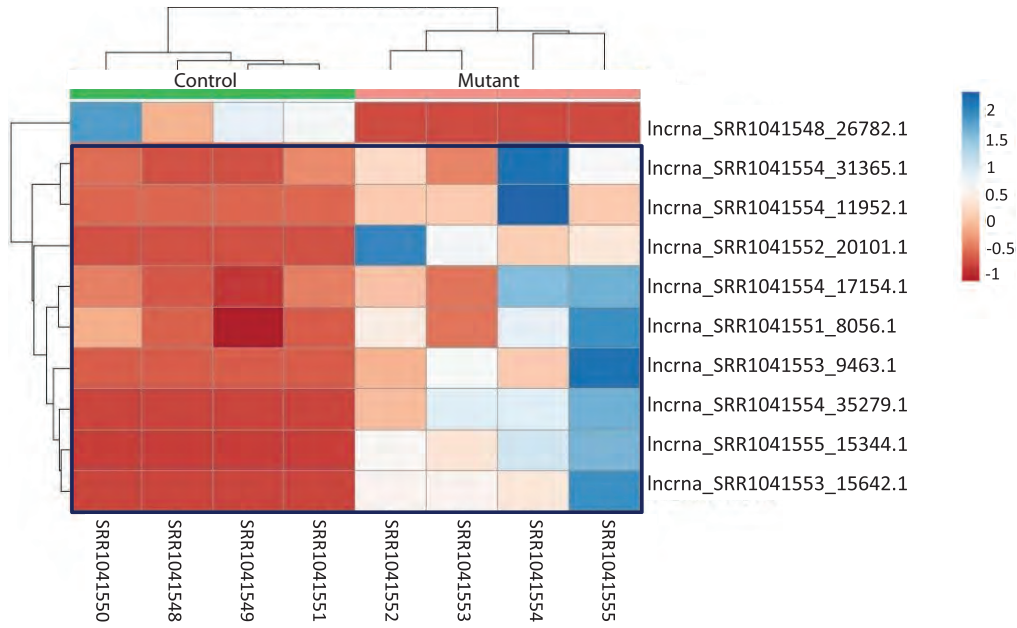


Fig. 3. Heatmap of differentially expressed lncRNAs in seedling tissue.

Here and in Fig. 4: the color scale on the right represents normalized expression levels, with blue indicating high expression, and red indicating low expression. Cells corresponding to lncRNAs with increased expression in the mutant line are highlighted with a blue frame.

acting with their mRNA targets. This mechanism contributes to the regulation of gene expression involved in plant growth, development, and stress responses (Pronozin, Afonnikov, 2025).

Results

Transcriptome assembly

As a result of the transcriptomic analysis of *Z. mays*, covering both seedling and tassel stages for control and mutant (*fzt*) lines, high-quality raw data were obtained. The average percentage of uniquely mapped reads during alignment using STAR (Dobin et al., 2013) was 84.73 %, while only 3.10 % of reads remained unmapped. For the aligned reads, the average mismatch rate per nucleotide was 0.76 %, indicating high sequencing accuracy and the reliability of the data for subsequent analyses.

The transcriptome assemblies generated using StringTie (Pertea et al., 2015) were evaluated with the BUSCO tool (Simão et al., 2015). In all 14 libraries, the proportion of complete BUSCO groups exceeded 95 %, reaching a maximum of 98.8 % (252 out of 255 expected orthologs detected in library SRR1041561). These metrics indicate the completeness and high quality of the obtained assemblies, confirming their suitability for subsequent expression analysis and the identification of noncoding RNAs.

Differential expression of lncRNAs between control and mutant *Z. mays* samples

In seedling tissue, 10 lncRNAs were identified as significantly differentially expressed between the control and mutant groups (Table S1)[1]. Among these, nine lncRNAs showed increased expression in the mutants, suggesting that they may serve as targets for miRNAs and participate in post-transcriptional regulatory mechanisms. These transcripts were subsequently considered as candidate miRNA targets in further analyses.

The heatmap (Fig. 3) illustrates systematic differences in the expression of these lncRNAs across the analyzed transcriptomic libraries. For 9 out of the 10 lncRNAs, expression levels were higher in the mutant plants compared with the control.

In tassel tissue, the number of differentially expressed lncRNAs was considerably higher, with a total of 34 lncRNAs identified (Table S2). Among these, 20 lncRNAs exhibited increased expression in the mutant line. Notably, pronounced differences in transcription levels were observed for several lncRNAs that displayed strong tissue-specific expression patterns unique to the tassel.

The heatmap of lncRNA expression in tassel tissue (Fig. 4) also illustrates systematic differences across the analyzed transcriptomic libraries. lncRNAs with decreased and increased expression levels in the mutant plants formed two clearly distinct clusters.

Overall, the identified lncRNAs represent a prioritized set for subsequent analysis of interactions with miRNAs and for further functional annotation.

Assessment of the accuracy of miRNA–lncRNA interaction predictions

The evaluation of the model's ability to distinguish lncRNAs from the test group (with increased expression in mutants) from those in the control group (with decreased expression) is presented in Fig 5.

In seedling tissue, the interaction scores for the test lncRNAs (with increased expression in mutants) were shifted above 0.5, suggesting a potential ability of these transcripts to

[1] Supplementary Tables S1 and S2 are available at: https://vavilovj-icg.ru/download/pict-2025-29/appx51.pdf

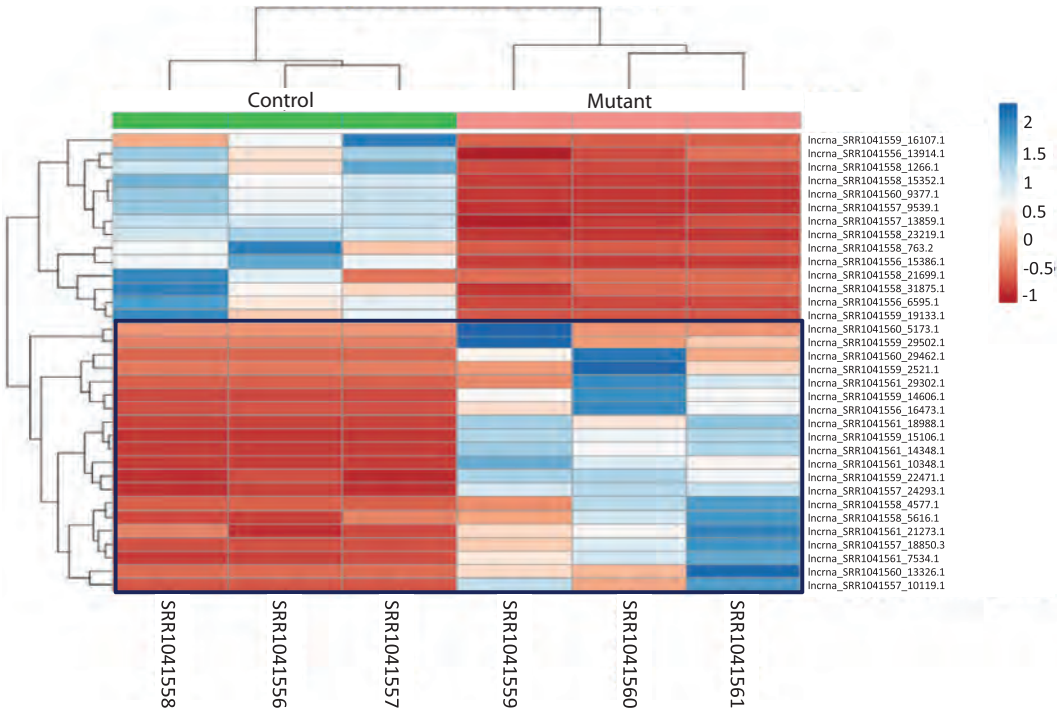


Fig. 4. Heatmap of differentially expressed lncRNAs in tassel tissue.

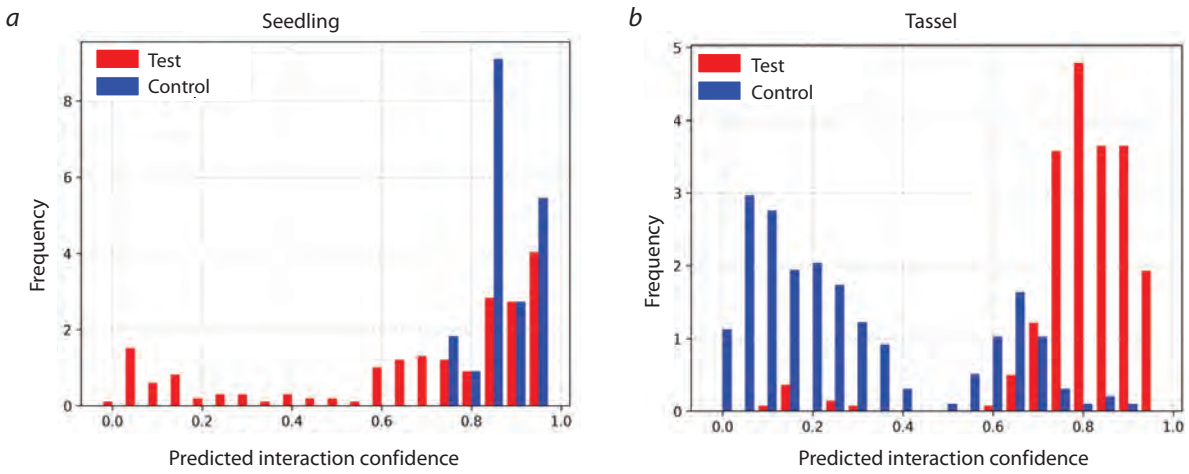


Fig. 5. Distribution of predicted miRNA–lncRNA interaction scores in seedlings and tassel tissues.

a – seedlings: interaction scores for the test group (lncRNAs with increased expression) are shown in red, and for the control group (lncRNAs with decreased expression), in blue; *b* – tassel: similarly, red represents the test group, and blue, the control group (lncRNAs with decreased expression). The *X*-axis represents the predicted interaction confidence (PC) calculated by PmliPred, and the *Y*-axis indicates the number of miRNA–lncRNA pairs analyzed.

bind miRNAs. However, it should be noted that the control group contained only a single lncRNA with a high predicted score. Because the control in this experiment consisted of only one lncRNA (with decreased expression in mutants), it was difficult to accurately assess the precision and discriminatory power of the PmliPred model.

In tassel tissue, the differences between the groups were even more pronounced: interaction scores for the test lncRNAs were predominantly above 0.5, whereas the control lncRNAs displayed a distribution shifted below 0.5. This behavior of the model indicates its ability to effectively distinguish biological classes based on the predicted miRNA–lncRNA interaction parameters.

Thus, the PmliPred model demonstrated high discriminatory power and can be used for the preliminary selection of lncRNAs potentially involved in interactions with miRNAs.

miRNA–lncRNA interaction networks in maize tissues

The results obtained using the miRNA–lncRNA interaction prediction tool IntaRNA are shown in Fig. 6. For example,

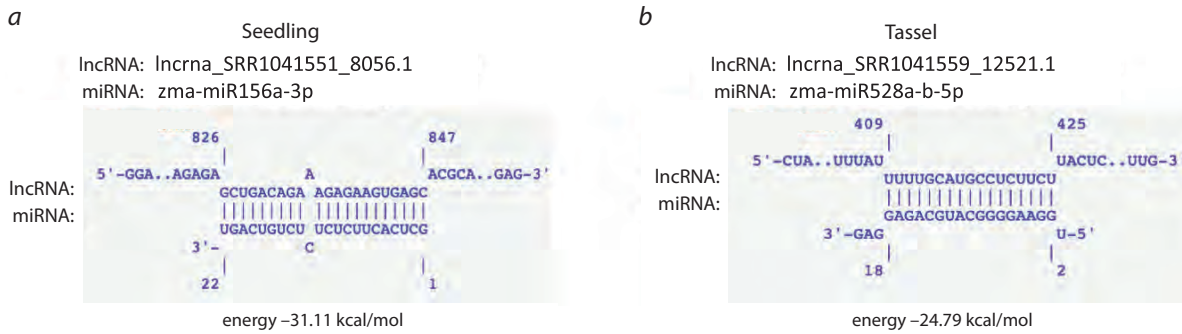


Fig. 6. Examples of miRNA–lncRNA interactions predicted using IntaRNA.

a – seedlings: zma-miR156a-3p and lncRNA_mapped_SRR1041551_8056.1; *b* – tassel: zma-miR528a-b-5p and lncRNA_mapped_SRR1041559_12521.1. Regions of base pairing and the interaction structures are shown, calculated based on minimum free energy (kcal/mol).

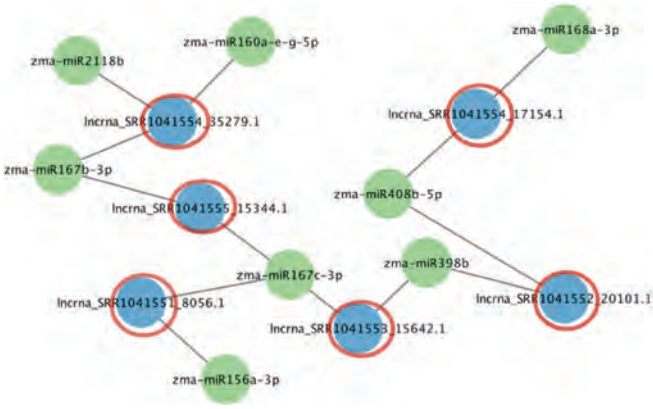


Fig. 7. miRNA–lncRNA interaction networks in maize seedling tissue.

Here and in Fig. 8: green circular nodes represent miRNAs, blue circular nodes represent lncRNAs, and red circles indicate lncRNAs that are potentially functioning as "sponges".

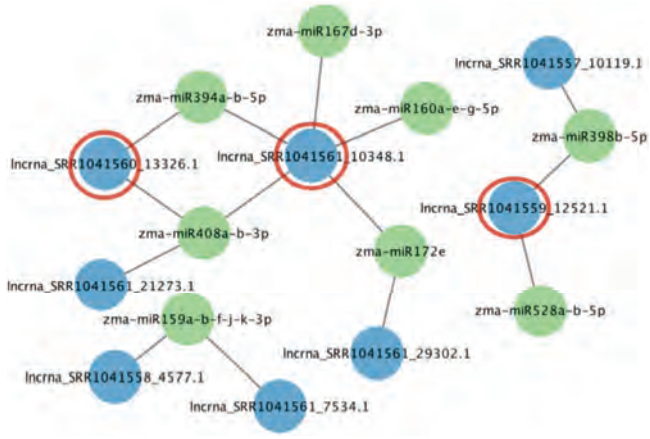


Fig. 8. miRNA–lncRNA interaction networks in maize tassel tissue.

two RNA pairs clearly formed stable and extensive regions of complementary binding. In total, 13 reliable miRNA–lncRNA pairs were identified in seedling tissue, and 14 pairs, in tassel tissue. These data confirm that the selected lncRNAs not only exhibit increased expression in the mutants but also possess a high potential for specific interactions with miRNAs, the expression of which is reduced in the mutants. This makes them justified candidates as miRNA targets.

Visualization of the identified interactions using Cytoscape (Figs. 7 and 8) revealed that some lncRNAs are potentially capable of binding multiple miRNAs simultaneously. For example, lncrna_SRR1041561_10348.1 (in tassel) interacts with five different miRNAs: zma-miR160a-e-g-5p, zma-miR167d-3p, zma-miR394a-b-5p, zma-miR408a-b-3p, and zma-miR172e, suggesting its potential role as a "sponge" within a ceRNA mechanism. Another example is lncrna_SRR1041554_35279.1 in seedling tissue, which interacts with zma-miR2118b, zma-miR160a-e-g-5p, and zma-miR167b-3p.

Discussion

The analysis revealed that 9 out of 14 identified lncRNAs are potentially capable of interacting with multiple miRNAs simultaneously, suggesting their possible role as competing endogenous RNAs (ceRNAs), molecular "sponges" that bind miRNAs and prevent their interaction with target mRNAs. Through this mechanism, lncRNAs can indirectly regulate the expression of various genes involved in key biological processes.

Among the predicted miRNA partners are well-characterized regulators of plant growth, development, and stress responses (Jones-Rhoades et al., 2006; Sunkar et al., 2012):

- miR156 regulates the transition from the juvenile to the adult phase, flowering, leaf morphogenesis, and branching by suppressing *SPL* genes (Preston et al., 2013; Wang H., Wang H., 2015);
- miR167 and miR160 regulate the auxin signaling pathway by suppressing *ARF* genes, thereby influencing root formation, leaf, flower, and seed development, as well as somatic embryogenesis (Caruana et al., 2020; Barrera-Rojas et al., 2021; Wang Y. et al., 2020);
- miR168 participates in maintaining the stable level of the AGO1 protein, a central component of the RNA interference (RNAi) machinery, thereby regulating the entire miRNA pathway (Martínez de Alba et al., 2011; Li W. et al., 2012);

- miR172 regulates the onset of flowering and organogenesis by repressing *AP2*-type transcription factor genes (Ripoll et al., 2015; Zhang B. et al., 2015);
- miR2118 activates the biogenesis of phased small interfering RNAs (phasiRNAs), playing a critical role in plant immunity and anther development (Canto-Pastor et al., 2019; Jiang P. et al., 2020);
- miR398 and miR408 provide antioxidant protection by regulating the levels of superoxide dismutases and metal-binding proteins, and they also respond to a wide range of abiotic stresses (Jiang A. et al., 2021; Zou et al., 2021; Gao et al., 2022);
- miR394 influences leaf morphogenesis, fruit development, and meristem activity (Song et al., 2015; Sun P. et al., 2017);
- miR528 is involved in redox homeostasis, resistance to viral infections, salt stress response, and regulation of lignification (Wu et al., 2017; Sun Q. et al., 2018).

Functional annotation of the interacting miRNAs indicates that most of them are involved not only in the development of plant morphological structures, but also in the complex regulatory networks controlling responses to biotic and abiotic stresses.

Moreover, the identified ceRNA networks confirm that post-transcriptional regulation in plants is mediated through finely coordinated interactions between non-coding and coding RNAs. The presence of lncRNAs capable of binding multiple regulatory miRNAs suggests the existence of potential hubs of regulatory cross-talk within RNA networks, which represents a particularly promising target for functional validation.

The obtained results emphasize the importance of a systems-level approach to transcriptomic data analysis, as such strategies enable the identification of hidden layers of gene regulation and promising molecular targets. Furthermore, these findings may serve as a theoretical foundation for the development of new agronomically valuable maize varieties with enhanced stress tolerance and improved adaptive traits.

Conclusion

In this study, a comprehensive analysis of *Z. mays* transcriptomic data was conducted to identify potential interactions between miRNAs and lncRNAs. Based on the results of differential expression analysis comparing control and mutant samples, lncRNAs and miRNAs with potential interactions were identified.

The PmliPred model, based on machine learning approaches, was applied to predict potential miRNA–lncRNA pairs. Subsequent structural analysis using IntaRNA confirmed the presence of stable complementary binding sites between the selected molecules, indicating high reliability of the predicted interactions.

Based on the selected interaction pairs, competing endogenous RNA (ceRNA) networks were constructed, demonstrating that individual lncRNAs are capable of binding multiple miRNAs simultaneously. This supports the hypothesis that they participate in post-transcriptional regulatory mechanisms as miRNA "sponges", capable of modulating the activity of regulatory molecules and influencing the expression of target genes.

Additionally, key interactions were visualized using Cytoscape, allowing a clear representation of the structure and potential functional significance of the identified regulatory connections. The results confirm the role of lncRNAs as important components of plant regulatory networks and provide a foundation for further functional studies.

References

Barrera-Rojas C.H., Otoni W.C., Nogueira F.T.S. Shaping the root system: the interplay between miRNA regulatory hubs and phytohormones. *J Exp Bot.* 2021;72(20):6822-6835. doi 10.1093/jxb/erab299

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

Bolser D.M., Staines D.M., Perry E., Kersey P.J. Ensembl Plants: integrating tools for visualizing, mining, and analyzing plant genomic data. In: van Dijk A. (Ed.) Plant Genomics Databases. Methods in Molecular Biology. Vol. 1533. New York: Humana Press, 2017; 1-31. doi 10.1007/978-1-4939-6658-5_1

Bray N., Pimentel H., Melsted P., Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525-527. doi 10.1038/nbt.3519

Canto-Pastor A., Santos B., Valli A., Summers W., Schornack S., Baulcombe D. Enhanced resistance to bacterial and oomycete pathogens by short tandem target mimic RNAs in tomato. *Proc Natl Acad Sci USA.* 2019;116(7):2755-2760. doi 10.1073/pnas.181438011

Caruana J.C., Dhar N., Raina R. Overexpression of *Arabidopsis microRNA167* induces salicylic acid-dependent defense against *Pseudomonas syringae* through the regulation of its targets *ARF6* and *ARF8*. *Plant Direct.* 2020;4(9):e00270. doi 10.1002/pld3.270

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. doi 10.1093/bioinformatics/bts635

Gao Y., Feng B., Gao C., Zhang M., Zhang X., Han X., Liu P., Wang B., Li Z. The evolution and functional roles of *miR408* and its targets in plants. *Int J Mol Sci.* 2022;23(1):530. doi 10.3390/ijms23010530

He X., Guo S., Wang Y., Wang L., Shu S., Sun J. Systematic identification and analysis of heat-stress-responsive lncRNAs, circRNAs and miRNAs with associated co-expression and ceRNA networks in cucumber (*Cucumis sativus* L.). *Physiol Plant.* 2020;168(3):736-754. doi 10.1111/ppl.12997

Jiang A., Guo Z., Pan J., Yu B., Chen D., Li Y. The PIF1-miR408-PLANTACYANIN repression cascade regulates light-dependent seed germination. *Plant Cell.* 2021;33(5):1506-1529. doi 10.1093/plcell/koab060

Jiang P., Lian B., Liu C., Fu Z., Shen Y., Cheng Z., Ding Y. 21-nt phasiRNAs direct target mRNA cleavage in rice male germ cells. *Nat Commun.* 2020;11(1):5191. doi 10.1038/s41467-020-19034-y

Jones-Rhoades M.W., Bartel D.P., Bartel B. MicroRNAs and their regulatory roles in plants. *Annual Rev Plant Biol.* 2006;57:19-53. doi 10.1146/annurev.arplant.57.032905.105218

Kang Q., Jun M., Jun C., Zhang Y., Wang W. PmliPred: a method based on hybrid model and fuzzy decision for plant miRNA–lncRNA interaction prediction. *Bioinformatics.* 2020;36(10):2986-2992. doi 10.1093/bioinformatics/btaa074

Li R., Yang Y.F., Li R., Zhu H.L., Luo Y.B. Functions of long non-coding RNA and its interacting mechanisms. *Chin Bull Life Sci.* 2016;28(6):703-711. doi 10.13376/j.cbls/2016090

Li W., Cui X., Meng Z., Huang X., Xie Q., Wu H., Jin H., Zhang D., Liang W. Transcriptional regulation of Arabidopsis *MIR168a* and *ARGONAUTE1* homeostasis in abscisic acid and abiotic stress responses. *Plant Physiol.* 2012;158(3):1279-1292. doi 10.1104/pp.111.188789

Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. doi 10.1186/s13059-014-0550-8

Mann M., Wright P.R., Backofen R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* 2017;45(W1):W435-W439. doi 10.1093/nar/gkx279

Martínez de Alba A.E., Jauvion V., Mallory A.C., Bouteiller N., Vaucheret H. The miRNA pathway limits AGO1 availability during siRNA-mediated PTGS defense against exogenous RNA. *Nucleic Acids Res.* 2011;39(21):9339-9344. doi 10.1093/nar/gkr590

Pertea M., Pertea G.M., Antonescu C.M., Chang T.C., Mendell J.T., Salzberg S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290-295. doi 10.1038/nbt.3122

Preston J.C., Hileman L.C. Functional evolution in the plant *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (*SPL*) gene family. *Front Plant Sci.* 2013;4:80. doi 10.3389/fpls.2013.00080

Pronozin A.Y., Afonnikov D.A. ICAnnoLncRNA: a snakemake pipeline for a long noncoding RNA search and annotation in transcriptomic sequences. *Genes.* 2023;14(7):1331. doi 10.3390/genes14071331

Pronozin A.Y., Afonnikov D.A. The role of long noncoding RNAs in plants. *Russ J Genet.* 2025;61:1-18. doi 10.1134/S1022795424701345

Ripoll J.J., Bailey L.J., Mai Q.A., Wu S.L., Hon C.T., Chapman E.J., Ditta G.S., Estelle M., Yanofsky M.F. microRNA regulation of fruit growth. *Nat Plants.* 2015;1(4):15036. doi 10.1038/nplants.2015.36

Robinson M.D., McCarthy D.J., Smyth G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-140. doi 10.1093/bioinformatics/btp616

Sheng N., Huang L., Gao L., Zhao W., Zhang Y., Wang X. A survey of computational methods and databases for lncRNA-miRNA interaction prediction. *IEEE/ACM Trans Comput Biol Bioinform.* 2023;20(5):2810-2826. doi 10.1109/TCBB.2023.3264254

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210-3212. doi 10.1093/bioinformatics/btv351

Song J.B., Shu X.X., Shen Q., Guo C.Y., Jiang J., Xie L.H., Liu Y.F., Yang Z.M. Altered fruit and seed development of transgenic rapeseed (*Brassica napus*) over-expressing microRNA394. *PLoS One.* 2015;10(5):e0125427. doi 10.1371/journal.pone.0125427

Sun P., Cheng C., Lin Y., Zhu Q., Lin J., Jin Y., Yuan H., Lin Y. Combined small RNA and degradome sequencing reveals complex microRNA regulation of catechin biosynthesis in tea (*Camellia sinensis*). *PLoS One.* 2017;12(2):e0171173. doi 10.1371/journal.pone.0171173

Sun Q., Csorba T., Skourti-Stathaki K., Proudfoot N.J., Dean C. R-loop stabilization represses antisense transcription at the *Arabidopsis FLC* locus. *Science.* 2013;340(6132):619-621. doi 10.1126/science.1234848

Sun Q., Liu X., Yang J., Wang Y., Zhang Y., Li J., Wang N., Li H., Wang J. MicroRNA528 affects lodging resistance of maize by regulating lignin biosynthesis under nitrogen-luxury conditions. *Mol Plant.* 2018;11(6):806-814. doi 10.1016/j.molp.2018.03.013

Sunkar R., Li Y.F., Jagadeeswaran G. Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 2012;17(4):196-203. doi 10.1016/j.tplants.2012.01.010

Thompson B.E., Basham C., Hammond R., Sidorenko L., Becker M.G., Neuffer M.G., Meeley R.B., Timmermans M.C.P., Chandler V.L., Hake S. The dicer-like1 homolog fuzzy tassel is required for the regulation of meristem determinacy in the inflorescence and vegetative growth in maize. *Plant Cell.* 2014;26(12):4702-4717. doi 10.1105/tpc.114.132670

Wang H., Wang H. The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Mol Plant.* 2015;8(5):677-688. doi 10.1016/j.molp.2015.01.008

Wang L., Wang J.W. Coding function for non-coding RNA in plants-insights from miRNA encoded peptide (miPEP). *Sci China Life Sci.* 2015;58(5):503-505. doi 10.1007/s11427-015-4854-z

Wang Y., Liu W., Wang X., Yang R., Wu Z., Wang H., Wang L., Hu Z., Guo S., Zhang H., Lin J., Fu C. MiR156 regulates anthocyanin biosynthesis through *SPL* targets and other microRNAs in poplar. *Hortic Res.* 2020;7:118. doi 10.1038/s41438-020-00341-w

Wu J., Yang R., Yang Z., Yao S., Zhao S., Wang Y., Li P., … Zhou X., Chu C., Qi Y., Cao X., Li Y. ROS accumulation and antiviral defence control by microRNA528 in rice. *Nat Plants.* 2017;3:16203. doi 10.1038/nplants.2016.203

Zhang B., Wang L., Zeng L., Zhang C., Ma H. *Arabidopsis* TOE proteins convey a photoperiodic signal to antagonize CONSTANS and regulate flowering time. *Genes Dev.* 2015;29(9):975-987. doi 10.1101/gad.251520.114

Zhang L., Chia J.M., Kumari S., Stein J.C., Liu Z., Narechania A., Maher C.A., Guill K., McMullen M.D., Ware D. A genome-wide characterization of microRNA genes in maize. *PLoS Genet.* 2009;5(11):e1000716. doi 10.1371/journal.pgen.1000716

Zou H., Guo X., Yang R., Wang D., Liu K., Gan L., Yuan Y., Qi J., Wang Y. MiR408-*SmLAC3* module participates in salvianolic acid B synthesis in *Salvia miltiorrhiza*. *Int J Mol Sci.* 2021;22(14):7541. doi 10.3390/ijms22147541

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

Received July 24, 2025. Revised October 6, 2025. Accepted October 7, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1304-1317

doi 10.18699/vjgb-25-137

Study of insecticidal and fungicidal potential of endophytic bacteria of wheat, soybean and rapeseed by bioinformatic analysis methods

T.N. Lakhova [1, 2], A.I. Klimenko [1], G.V. Vasiliev [1], E.Yu. Gyrnets [3], A.M. Asaturova [3], S.A. Lashin [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
[3] Federal State Budgetary Scientific Institution "Federal Research Center of Biological Plant Protection", Krasnodar, Russia
tlakhova@bionet.nsc.ru

Abstract. Endophytic bacteria play a key role in agricultural ecosystems, as they can affect the availability of various compounds, crop yield and growth, and provide resistance to diseases and pests. Therefore, the study of endophytes of agriculturally important crop plants is a promising task in the field of biological plant protection. Understanding the mechanisms of interaction between endophytic bacteria and plants will allow the use of these microorganisms as bioagents in the future and thus reduce dependence on chemical pesticides. In this paper, samples obtained from the leaves and/or roots of wheat, rapeseed and soybean are considered. Whole-genome sequencing of the isolates was performed. Using an analytical pipeline, the genomes of 15 strains of endophyte bacteria of cultivated plants were assembled and characterized. Their insecticidal and fungicidal potential was analyzed. Gene repertoire analysis performed with GenAPI showed a high degree of correspondence between the gene repertoires of strain BZR 585 against *Alcaligenes phenolicus*, BZR 762 and BZR 278 against *Alcaligenes* sp., BZR 588 and BZR 201P against *Paenochrobactrum pullorum*. All strains, with the exception of BZR 162, BZR 588 and BZR 201P, were found to contain genes encoding proteins with fungicidal activity, such as iturins, fengycins and surfactins. All strains also contained genes encoding proteins with insecticidal activity, namely GroEL, Spp1Aa1, Spp1Aa2, Vpb1Ab1, Vpb4Ca1, HldE, mycosubtilin, fengycin and bacillomycin. The obtained genomic data are confirmed by the results of previous experimental studies: high insecticidal activity of a number of strains (BZR 1159, BZR 936, BZR 920, etc.) against *Galleria mellonella*, *Tenebrio molitor* and *Cydia pomonella*, as well as fungicidal properties against *Fusarium*, *Alternaria*, *Trichothecium*, was demonstrated. This shows the practical significance of the identified genetic determinants for the creation of new biocontrol agents.
Key words: bioinformatics; comparative genomics; endophytes; biocontrol

For citation: Lakhova T.N., Klimenko A.I., Vasiliev G.V., Gyrnets E.Yu., Asaturova A.M., Lashin S.A. Study of insecticidal and fungicidal potential of endophytic bacteria of wheat, soybean and rapeseed by bioinformatic analysis methods. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1304-1317. doi 10.18699/vjgb-25-137

Funding. This research was funded by Russian Science Foundation grant No. 23-16-00260, https://rscf.ru/en/project/23-16-00260/. Genome sequencing was performed at the Center for Genomic Research, ICG SB RAS. High-performance computing was performed using resources of the "Bioinformatics" Joint Computational Center ICG SB RAS supported by the budget project FWNR-2022-0020.

Исследование инсектицидного и фунгицидного потенциала бактерий эндофитов пшеницы, сои и рапса методами биоинформатического анализа

Т.Н. Лахова [1, 2], А.И. Клименко [1], Г.В. Васильев [1], Е.Ю. Гырнец [3], А.М. Асатурова [3], С.А. Лашин [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
[3] Федеральное государственное бюджетное научное учреждение «Федеральный научный центр биологической защиты растений», Краснодар, Россия
tlakhova@bionet.nsc.ru

Аннотация. Эндофитные бактерии способны влиять на доступность различных соединений, урожайность и рост сельскохозяйственных растений, а также обеспечивать устойчивость к болезням и вредителям. Поэтому исследование эндофитов сельскохозяйственно значимых культурных растений является перспективной задачей в области биологической защиты растений. В данной работе рассмотрены изоляты штаммов бактерий, полученные из листьев и/или корней пшеницы, рапса и сои. Было произведено полногеномное секвенирование

изолятов. С помощью аналитического конвейера собраны и охарактеризованы геномы 15 штаммов бактерий-эндофитов культурных растений, проанализирован их инсектицидный и фунгицидный потенциал. Анализ генного репертуара с помощью программы GenAPI показал высокую степень соответствия между генным репертуаром штамма BZR 585 относительно *Alcaligenes phenolicus*, BZR 762 и BZR 278 относительно *Alcaligenes* sp., BZR 588 и BZR 201P относительно *Paenochrobactrum pullorum*. Во всех штаммах, за исключением BZR 162, BZR 588 и BZR 201P, найдены гены, кодирующие белки, обладающие фунгицидной активностью, такие как итурины, фенгицины и сурфактины. Также во всех штаммах найдены гены, кодирующие белки с инсектицидной активностью, а именно: GroEL, Spp1Aa1, Spp1Aa2, Vpb1Ab1, Vpb4Ca1, HldE, фенгицин, микосубтилин и бацилломицин. Полученные геномные данные подтверждены экспериментальными испытаниями: ранее показана высокая инсектицидная активность штаммов BZR 1159, BZR 936, BZR 920 и др. против *Galleria mellonella*, *Tenebrio molitor* и *Cydia pomonella*, а также фунгицидные свойства против *Fusarium*, *Alternaria*, *Trichothecium*. Это демонстрирует практическую значимость выявленных генетических детерминант для создания новых агентов биоконтроля.
Ключевые слова: биоинформатика; сравнительная геномика; эндофиты; биоконтроль

Introduction

Endophytic bacteria are components of the plant microbiome that can colonize roots, stems, or leaves, where they obtain stable sources of nutrients and, in turn, increase plant resistance to both biotic and abiotic stresses. These symbionts have growth-stimulating activity due to nitrogen fixation, phosphorus mobilization, and phytohormone synthesis, and also produce a wide range of metabolites, hydrolytic enzymes, and volatile compounds involved in the biological control of pests and diseases of agricultural crops. In addition, endophytes compete for ecological niches and induce systemic resistance in plants, creating a multi-level defense that can be comparable in effectiveness to chemical pesticides, but is safe for humans and the environment (Hamane et al., 2023; Ali et al., 2024). The study of endophytes in agriculturally important crops opens up prospects for the search for biocontrol agents for various pests and pathogens (insects, fungi, and others).

Endophytic bacteria are actively being studied using bioinformatics methods and comparative genomics approaches (Wang Z. et al., 2023; Zhang M. et al., 2023; Pellegrinetti et al., 2024; Yang K. et al., 2024; Dai et al., 2025). Sequencing their complete genomes provides a large array of information for identifying genes and their products that have insecticidal and fungicidal potential in terms of solving biocontrol problems. In addition, functional annotation of genes of interest can be performed based on well-characterized related strains of microorganisms, such as *Bacillus velezensis*, *B. thuringiensis*, and others.

For example, in the rhizosphere of plants, *B. velezensis* creates a favorable nutritional and physicochemical environment for root microbiota by forming a biofilm, which promotes plant growth and protection against phytopathogens, both through the secretion of antimicrobial compounds and through the formation of phytoimmune potential in plants. The study (Rabbee et al., 2019) summarized information on strain-specific gene clusters of *B. velezensis* associated with the biosynthesis of secondary metabolites, which play an important role in both suppressing pathogens and stimulating plant growth. For *B. velezensis* BRI3 strains and related lines, comparative analysis confirmed the preservation of the "core" gene clusters of lipopeptide synthetases (iturin, fengycin, and surfactin), with the simultaneous appearance of unique type III polyketide synthases, which allows these strains to exhibit broad fungicidal potential *in vitro* (Liu et al., 2024).

Pangenomic studies of *Burkholderia* bacteria, including a comparison of 18 endophytic and pathogenic strains, revealed the loss of classical virulence determinants and enrichment of antimicrobial compound synthesis gene clusters in endophytic symbionts. These changes are a sign of adaptation to an intracellular lifestyle and serve as an indirect marker of biological control potential (Liu et al., 2024).

For the representative of Pseudomonadota, *Ochrobactrum quorumnocens* A44 is capable of disrupting quorum sensing (QS) in Gram-negative bacteria by inactivating N-acyl-homoserine lactones (AHLs) and protecting plant tissues from soft rot pathogens, the virulence of which is regulated by QS. For this strain, isolated from the potato rhizosphere, and six related type strains of the genus *Ochrobactrum*, comparative genomics showed that the core genome contains 50–66 % of genes, and the variable part for each genome accounts for 8 to 15 % (Krzyżanowska et al., 2019).

The entomopathogenic strain *B. thuringiensis* ser. *israelensis*, a well-known source of δ-endotoxins (*cry*, *cyt*), plays a key role (occupies a special place) in research. The complete sequence of the 127-kb pBtoxis plasmid showed that the cluster of genes encoding toxins remains stable, while regulatory and mobile elements are actively reorganized, ensuring horizontal transfer of insecticidal protein genes and expanding the adaptive potential of the strain (Berry et al., 2002; Bolotin et al., 2017).

At the same time, microorganisms that establish symbiotic relationships with their hosts were found to contain GroEL proteins (an ATP-dependent molecular chaperone that is present in all forms of life and is one of the most conservative proteins in living organisms), which acted as toxins (Horwich et al., 2007; Shi et al., 2012; Kupper et al., 2014). Its homologue, XnGroEL, has been described in the *Xenorhabdus nematophila*, which retains its folding function but acquires the ability to bind to the insect's chitin cuticle and suppress the host's immune responses (Horwich et al., 2007).

Thus, the creation of databases on proteins associated with the insecticidal and fungicidal properties of endophytic bacteria opens up broad opportunities for bioinformatic analysis and *in silico* screening of bacteria with protective properties that are significant for agriculture. For example, the BPPRC database (Panneerselvam et al., 2022) summarizes information on some proteins and peptides with insecticidal activity.

In this paper, we present assemblies and annotations of the genomes of 15 strains of endophytic bacteria isolated from

various organs of wheat, soybean, and rapeseed plants. Annotation was performed relative to close bacterial reference genomes. We also demonstrated the presence of proteins with fungicidal and insecticidal properties in the studied strains.

Materials and methods

Strains of endophytic bacteria. In this research, we used the scientific equipment "Technological line for obtaining microbiological plant protection products of a new generation" (http://ckp-rf.ru/usu/671367). The objects of the study were bacterial strains from the bioresource collection of the Federal State Budgetary Scientific Institution Federal Research Center for Biological Resources "State Collection of Entomophagous Acarids and Microorganisms" (https://fncbzr.ru/brk-i-unu/unique-installation-1/). The strains under study were isolated from the roots and leaves of wheat, soybeans, and rapeseed. Samples were collected in four districts of the Krasnodar Territory (Krylovsky, Vyselkovsky, Pavlovsky, and Krasnodar). The general characteristics of the microbial strains under study are presented in Supplementary Material 1[1].

Sequencing. Bacterial DNA was isolated from individual colonies grown on agarized medium in Petri dishes using the D-Cells kit (Biolabmix, Russia) according to the method for Gram-negative bacteria. The colony was transferred to a 1.5 ml tube and resuspended in 150 µl of PBS buffer. After adding 20 µl of proteinase K and 150 µl of lysis buffer, the cells were incubated for 10 min at 56 °C. After adding 500 µl of LB buffer, the lysate was applied to the column and centrifuged for 30 seconds at 12,000g. The column was sequentially washed with 500 µl of WB1 and WB2 buffers, followed by centrifugation. DNA was eluted from the column using 60 µl of EB buffer. Ultrasonic DNA fragmentation was performed on a Covaris M220 sonicator (Shelton, USA) in a volume of 50 µl using a protocol optimized for average fragment lengths of 300 bp. DNA concentration was measured using a Qubit4 fluorometer with the DNA HS kit (Thermo Fisher Scientific, USA). Genomic libraries were prepared using the KAPA Hyper Prep kit (Roche, Switzerland) with KAPA double indices according to the manufacturer's instructions. Fifty nanograms of fragmented DNA were used for the experiment, with nine cycles of final PCR. The quality and molarity of the obtained libraries were calculated after analysis on a BA2100 bioanalyzer (Agilent, USA) and measurement of concentrations on a Qubit4 fluorometer. After normalization, the resulting libraries were pooled to a concentration of 4 nM/µl. Sequencing was performed on a GenolabM device (GeneMind, China) using 2×150 bp paired-end reads with the GenolabM V2.0 kit (FCM 300) according to the manufacturer's protocol.

Bioinformatic analysis. The genome sequencing results were quality checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), trimming was performed using fastp (Chen et al., 2018; Chen, 2023). Filtering for possible contamination by human reads (Hg38) was performed using BWA MEM (Li, 2013) and synthetic sequences (Univec (https://ftp.ncbi.nlm.nih.gov/pub/UniVec/, accessed April 2025) and blastn).

The taxonomic composition of the sequenced material was analyzed using the MetaPhlan4 tool (Blanco-Míguez et al., 2023). The initial taxonomic identification of the draft primary genome assemblies was performed by tetracorrelation search using the JSpeciesWS web service with the GenomesDB database (Richter et al., 2016). The draft primary assemblies were obtained in the same way as described below, except that Ragout scaffolding (Kolmogorov et al., 2014) and Pilon correction (Walker et al., 2014), were not used for the draft primary assemblies, and the results of the tetra-correlation search during primary taxonomic identification were used as reference genomes. The GFinisher assembler accepts multiple assemblies as input data (Guizelini et al., 2016). The authors of GFinisher note that the first assembly is the main one, and the rest are additional. In our case, depending on the quality control of all assemblies, the main assembly in GFinisher was one of the following: bins, the MinYS assembly (Guyomar et al., 2020), or the Spades assembly (Bankevich et al., 2012; Prjibelski et al., 2020) after mapping to the reference.

Next, taxonomic identity was refined in order to select the closest reference genomes for corrected bacterial genome assembly using a hybrid (*de novo* + reference-guided) cascade approach based on a set of closely related references (Fig. 1). To search for refined reference genomes, average nucleotide identity (ANI) analysis was performed using fastani (Jain et al., 2018) within the genus established during the initial taxonomic screening. Draft primary assemblies were compared by ANI with all publicly available genome assemblies within the same genus deposited in NCBI Genbank, which were downloaded in batch processing using ncbi-genome-download (https://zenodo.org/records/8192486, accessed April 2025).

Three draft assemblies were obtained. The first option was a *de novo* assembly of Spades. The assembly was binned using MaxBin2 (Wu et al., 2016). The resulting bins were filtered by completeness. If this parameter was <87 %, such bins were not considered further. The second assembly variant was a reference-guided assembly using MinYS. A closely related reference genome and filtered reads were fed into the input. Based on the results of the MinYS assembly, the main reference was selected from the closest reference genomes and used in the third assembly option and in the final assembly with correction. The third assembly option was performed by the Spades assembler based on mapped reads to the selected reference genome.

All draft assembly variants were scaffolded using Ragout, which can utilize information about a number of the closest reference genomes. Configuration files for each Ragout assembly variant were generated using a custom Python 3 script. The resulting draft assemblies at the contig and scaffold level were merged into a single assembly using the GFinisher program. Depending on the intermediate quality control results, the main assembly in GFinisher was either the Ragout assemblies or the bins obtained in the first draft assembly variant. GFinisher can output two assemblies, which were submitted for correction in Pilon. At each stage, the quality of the assemblies was controlled using QUAST (Gurevich et al., 2013). Based on the results of quality control, which was carried out after correction, the final genome assemblies were selected.

Genome annotation was performed using the prokka software pipeline (Seemann, 2014), additionally configured to

[1] Supplementary Materials 1–6 are available at:
https://vavilovj-icg.ru/download/pict-2025-29/appx52.zip

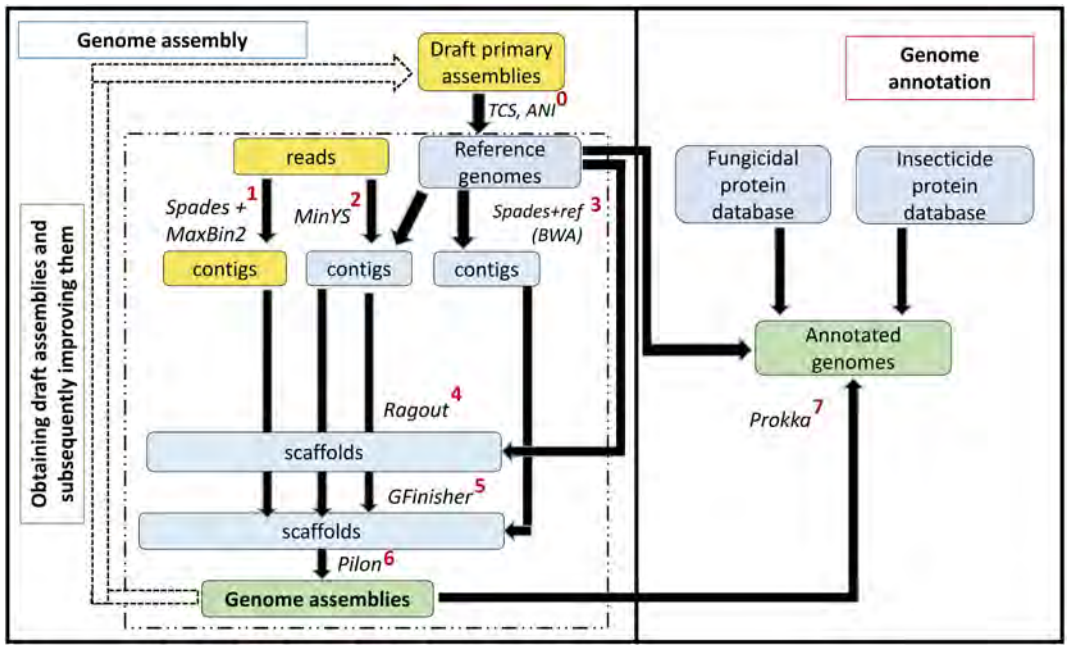


Fig. 1. Analytical pipeline for assembling bacterial genomes and annotating them.

Yellow blocks indicate data obtained at the preliminary stage of primary genome assembly, blue blocks indicate data obtained during genome assembly refinement, and green blocks indicate target results. Arrows indicate data interdependencies, and labels next to the arrows indicate the tools and methods used for data analysis or conversion. 0 – tetracorrelation analysis (JSpeciesWS and GenomesDB) and average nucleotide identity analysis (fastani) to establish the closest (and best annotated) reference genomes from GenBank; 1 – *de novo* assembly using Spades followed by binning in MaxBin2 (bin selection filter – completeness > 87 %); 2 – reference-guided assembly in MinYS; 3 – assembly using Spades of reads pre-mapped to the reference using BWA; 4 – scaffolding of previously obtained draft genomes (contigs) in Ragout, using information about the closest reference genome; 5 – genome polishing in GFinisher, combining the results of previously obtained draft assemblies into one common assembly; 6 – final correction in Pilon; 7 – annotation of final assemblies in Prokka, taking into account the most complete known genome annotations of the corresponding species/genus, as well as specially prepared databases of amino acid sequences of fungicidal and insecticidal proteins, the latter including BPPRC.

take into account the databases of protein sequences with insecticidal and fungicidal properties prepared within the framework of this study, as well as the annotation of refined reference genomes.

For each genome, a separate annotation database was created, consisting of the following parts: annotated protein sequences from the closest reference genome, protein sequences with insecticidal and fungicidal properties collected in this study (described below), and the BPPRC database (https://www.bpprc-db.org/, accessed June 2025).

To search for insecticidal proteins in the genomes under study, a database of protein sequences was compiled manually. It included the amino acid sequences of Cry and Vip proteins. The sequences were selected from the UniprotDB (Bateman et al., 2021) and NCBI Protein (https://www.ncbi.nlm.nih.gov/protein/, accessed October 2024) databases by genus name and protein name/function. In the complete genomic sequences of bacteria from the GenBank card, the protein product was found and the corresponding protein sequence was downloaded. Then, a search was performed using blastp on the formed database. The results were filtered by a threshold for two parameters: identity > 50 % and e-value < 0.001, where identity corresponds to the percentage of matching amino acids, and e-value corresponds to the statistical significance of the results. The BPPRC database, containing protein sequences of bacterial pesticidal proteins downloaded from the Bacterial Pesticidal Protein Database project, was also added to this variant.

Similarly, a database of protein sequences was compiled manually for certain proteins with fungicidal activity: iturins, fengycins, and surfactins. Using blastp, the results of alignment of proteins encoded in the genomes of the analyzed strains to the sequences from the compiled database were obtained. Initial filtering of the results was performed according to two parameters: identity > 50 % and e-value < 0.001.

Figure 1 shows a graphical diagram of the corresponding analytical pipeline. Venn diagrams for comparing the gene repertoire of the analyzed strains with that of the reference genomes were constructed using the Draw Venn Diagram web service (http://bioinformatics.psb.ugent.be/webtools/Venn/, accessed May 2025).

Subsequent analysis of the alignment results of the collected genomes based on insecticidal and fungicidal proteins was performed by visualizing information about the proteins found in the studied strains in the form of a heat map based on the iScore value (see Equation 1), which represents the percentage of identity weighted by the proportion of the aligned region to the total length of the reference protein. This representation allows us to take into account not only the percentage of identity of the local alignment found, but also the extent to which this alignment covers the length of the original protein.

The formula for calculating iScore is as follows:

$$\text{iScore} = \text{identity} \cdot \frac{\text{la}}{\text{ls}}, \quad (1)$$

where identity is the percentage of matching amino acids, la is the alignment length, and ls is the initial length of the amino acid sequence of the protein from the collected database.

Gene repertoire analysis was performed using GenAPI (Gabrielaite, Marvig, 2020).

Multiple alignment of concatenated amino acid sequences was performed using the GTDB-Tk tool (Chaumeil et al., 2022) based on a search for 120 bacterial marker genes. Multiple alignment was used to construct a phylogenetic tree in PhyML (Guindon et al., 2010). In PhyML, the default support level for internal branches is estimated using a Bayesian test. The iTOL web service (Letunic, Bork, 2024) (https://itol.embl.de/, accessed May 2025) was used to visualize the tree.

Program versions and launch parameters are listed in Supplementary Material 2.

Results

Taxonomic identification

A preliminary assessment of the taxonomic composition of sequenced genetic material using Metaphlan revealed heterogeneity in most of the analyzed samples to varying degrees (Table 1). Since not all reads for all samples belonged to a single organism, it was decided to perform binning of the initial draft genome assemblies to isolate the genomic fragments most fully represented in the microorganism sample. Next, taxonomic identification was performed for the bins using tetranucleotide frequency analysis and correlation coefficients (tetra-correlation search, TCS).

The taxonomic identity of the assembled genomes (see subsection Assembly and annotation) was refined during the analysis of average nucleotide identity by comparison with publicly available reference genomes of the same genus in NCBI Genbank (see Supplementary Material 3). The results of the identification are summarized in Figure 2 and Table 2.

Table 1. Results of the initial analysis of the taxonomic composition of the samples studied and subsequent binning

Sample	Species by Metaphlan, share of reads %	Bin number, bin completeness %	Species according to TCS
BZR 1159	*Brevundimonas naejangsanensis*, 100	1) 001 – 38.3 2) 002 – 57.9	1) *Brevundimonas naejangsanensis* DSM 23858 2) *Brevundimonas naejangsanensis* DSM 23858
BZR 162	*Ochrobactrum quorumnocens*, 78.9 Unclassified, 21.1	2) 002 – 87.9	2) *Ochrobactrum quorumnocens* A44
BZR 201	*Leucobacter aridicollis*, 56.1 Unclassified, 43.5 *Paenochrobactrum gallinarii*, 0.4	1) 001 – 95.3 2) 002 – 99.1	1) *Leucobacter aridicollis* L-9 2) *Paenochrobactrum glaciei* JCM 15115
BZR 206	*Leucobacter aridicollis*, 90.5 Unclassified, 9.5	1) 001 – 95.3 2) 002 – 75.7	1) *Leucobacter aridicollis* DSM 17380 2) *Leucobacter chinensis* NC76-1
BZR 278	*Alcaligenes faecalis*, 92.7 *Brevundimonas naejangsanensis* 0.4 Unclassified, 7.3	1) 001 – 100	1) *Alcaligenes nematophilus* A-TC2
BZR 466	*Leucobacter aridicollis*, 69.8 *Brevundimonas naejangsanensis*, 19.3 Unclassified, 11.5	1) 001 – 95.3 2) 002 – 95.3	1) *Leucobacter aridicollis* DSM 17380 2) *Brevundimonas naejangsanensis* DSM 23858
BZR 585	*Alcaligenes faecalis*, 91.1 Unclassified, 8.9	1) 001 – 100	1) *Alcaligenes nematophilus* A-TC2
BZR 588	Unclassified, 95.8 *Paenochrobactrum gallinarii* SGB85919, 4.2	1) 001 – 99.1 2) 002 – 94.4 3) 003 – 84.1	1) *Paenochrobactrum glaciei* JCM 15115 2) *Leucobacter chinensis* NC76-1 3) *Leucobacter komagatae* DSM 8803
BZR 635	Unclassified, 99.9 *Mesorhizobium hungaricum* SGB11031, 0.1	1) 001 - 38.3 2) 002 – 57.9	1) *Leucobacter* sp. G161 2) *Leucobacter* sp. G161 (*Leucobacter komagatae*)
BZR 736	*Bacillus cereus*, 87 Unclassified, 13	1) 002 – 99.1	1) *Bacillus cereus* BAG1O-3
BZR 762	*Alcaligenes faecalis*, 92.1 Unclassified, 7.9	1) 001 – 64.5 2) 002 – 35.5	1) *Alcaligenes nematophilus* A-TC2 2) *Alcaligenes nematophilus* A-TC2
BZR 920	*Bacillus velezensis*, 100	1) 001 – 51.4 2) 002 – 46.7	1) *Bacillus amyloliquefaciens* Bs006 2) *Bacillus velezensis* OB9
BZR 926	*Achromobacter* sp. MYb9, 47.9 Unclassified, 52.1	1) 001 – 74.8 2) 002 – 26.2	1) *Achromobacter marplatensis* B2 2) *Achromobacter marplatensis* HLE
BZR 936	*Bacillus velezensis*, 100	1) 001 – 45.8 2) 002 – 54.2	1) *Bacillus amyloliquefaciens* EGD-AQ14 2) *Bacillus velezensis* OB9

Note. The correspondence of the bin numbers in the third and fourth columns has been preserved.

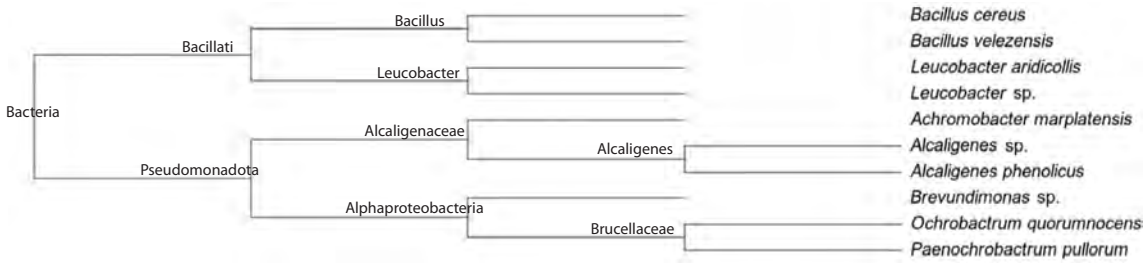


Fig. 2. General taxonomic tree of species represented in samples and selected for further analysis.

Table 2. Results of final taxonomic identification with indication of the NCBI identifier of the closest reference genome

No.	Strain	Taxonomic identity	Nearest related strain	Nearest reference genome
1	BZR 588	*Paenochrobactrum pullorum*	LMG 28095	GCF_041929845.1
2	BZR 1159	*Brevundimonas* sp.	LPMIX5	GCF_003576505.1
3	BZR 920	*Bacillus velezensis*	X-Bio_1	GCF_014230015.1
4	BZR 936	*Bacillus velezensis*	X-Bio_1	GCF_014230015.1
5	BZR 206	*Leucobacter aridicollis*	JUb111	GCF_024806925.1
6	BZR 926	*Achromobacter marplatensis*	LMG 26219	GCF_902859635.1
7	BZR 466	*Leucobacter aridicollis*	JUb111	GCF_024806925.1
8	BZR 736	*Bacillus cereus*	CMCC P0011	GCF_001635955.1
9	BZR 635	*Leucobacter* sp.	G161	GCF_001482305.1
10	BZR 762	*Alcaligenes* sp.	YSL9	GCF_030515105.1
11	BZR 278	*Alcaligenes* sp.	YSL9	GCF_030515105.1
12	BZR 201	*Leucobacter aridicollis*	JUb111	GCF_024806925.1
13	BZR 201P*	*Paenochrobactrum pullorum*	LMG 28095	GCF_041929845.1
14	BZR 162	*Ochrobactrum quorumnocens*	A44	GCF_002278035.1
15	BZR 585	*Alcaligenes phenolicus*	JC896	GCF_040430305.1

Note. Since two genomes were isolated from sample BZR 201, the second strain belonging to *Paenochrobactrum pullorum* from sample BZR 201 is designated BZR 201P (* in the second column).

As a result, genomes were assembled and 15 bacterial strains were taxonomically identified. Some of the samples had readings that could be attributed to different bacteria, which may indicate contamination of the samples prior to sequencing or that the culture was mixed.

For example, since *Paenochrobactrum pullorum* is found in samples taken from wheat roots (BZR 588) and winter rapeseed roots (BZR 201P), we conclude that this is not accidental contamination, but a normal representative of the rhizosphere of flowering plants. Thus, in the study (Hussain et al., 2025), this species was used as a part of a community of bacteria that affect phosphorus availability, yield, and wheat growth.

Assembly and annotation

As a result of the hybrid (*de novo* + reference-guided) approach, assemblies of 15 bacterial genomes were obtained. The main characteristics of the final assemblies obtained after QUAST quality control are presented in Table 3. Additionally, for comparison, the table shows such reference characteristics as assembly length (reference genome sequence length), CG composition of the reference genome, and reference-dependent indicators: the proportion of the reference represented in the studied genome (Genome fraction) and the duplication ratio (Duplication ratio).

However, the approach presented was not suitable for samples BZR 635 and BZR 926. Although closely related references were found for these samples, the Genome fraction indicator was low during assembly (4.274 and 1.511 %, respectively), which made it impossible to perform successful scaffolding in Ragout. Therefore, for BZR 635 and BZR 926, the assembly scheme consisted of the following steps. The primary assembly was a *de novo* Spades assembly without binning (the bins did not pass the set threshold). Rough assemblies were performed using MinYS and Spades with mapping to the reference. The steps with Ragout and Pilon were

Table 3. Characteristics of the final genome assembly and reference genome metrics

No.	Strain	Reference assembly length	CG content of reference, %	Genome fraction, %	Duplication ratio	Number of contigs	The largest contig, bp	Total length of assembly, bp	N50	L50	CG content, %
1	BZR 1159	3144717	66.41	87.420	1.000	239	82719	2772387	19497	39	67.17
2	BZR 162	4593578	53.41	88.176	1.002	60	487021	4079562	225616	6	53.23
3	BZR 201	3459535	67.32	93.609	1.000	10	783893	3424442	486624	3	67.45
4	BZR 201P	3664841	48.45	89.886	1.001	53	510935	3452129	153809	7	48.47
5	BZR 206	3459535	67.32	93.600	1.000	13	1015594	3387765	424194	3	67.51
6	BZR 278	4098772	56.49	95.093	1.000	12	1448095	4159417	406907	3	56.45
7	BZR 466	3459535	67.32	93.743	1.005	15	562204	3459663	424178	4	67.37
8	BZR 585	4121371	56.49	94.610	1.000	7	1900810	4183792	871829	2	56.42
9	BZR 588	3664841	48.45	85.527	1.029	60	412949	3238581	144220	8	48.54
10	BZR 635	3554188	65.32	4.274	1.001	69	181203	3024800	66004	16	66.43
11	BZR 736	6097988	34.94	85.014	1.010	16	1408346	5366477	760581	3	35.03
12	BZR 762	4098772	56.49	95.136	1.000	63	486891	3921546	118193	10	56.60
13	BZR 920	3861135	46.54	95.678	0.999	70	267487	3714855	123313	10	46.84
14	BZR 926	6885835	65.11	1.511	1.001	43	737359	6214265	318995	7	64.92
15	BZR 936	3861135	46.54	95.657	1.000	74	283481	3714091	123326	10	46.85

skipped, since they require the locations of the regions in the assembly relative to the reference. The resulting assemblies were submitted to Gfinisher, the result of which was selected as the final assembly.

Genome annotation was performed using the Prokka tool with a user database of annotated amino acid sequences taken from the reference genome, as well as databases of fungicidal and insecticidal proteins. A total of 10 custom databases were used, as the reference genomes coincided for a number of strains. Thus, for strains BZR 201, BZR 206, and BZR 466, the reference is GCF_024806925.1; for BZR 201P and BZR 588 – GCF_041929845.1; for BZR 278 and BZR 762 – GCF_030515105.1; and for BZR 920 and BZR 936 – GCF_014230015.1. Individual user databases were used for the remaining samples.

Table 4 shows info on the annotation of bacterial genomes compared to the corresponding reference genome. The "Predicted proteins" column shows the percentage of predicted proteins relative to all genes predicted in the genome, with the rest annotated as "hypothetical protein". Although the Genome fraction was low for the BZR 635 and BZR 926 genomes, we used the reference data as annotation.

Analysis of the gene repertoire using GenAPI revealed differences between the analyzed strains and the reference genomes used (Fig. 3).

The genomes of the bacteria were annotated, and the gene repertoire of each genome was analyzed relative to the corresponding reference genomes. Analysis of the intersections of gene sets for different samples showed varying results. For example, strains BZR 920 and BZR 936 showed high similarity to *B. velezensis*, and strains BZR 206, BZR 466, and BZR 201 showed high similarity to *L. aridicollis*. However, the fact that 160 genes were found only in the reference genome of *B. velezensis* but not in the studied strains of this species may be due to both the fragmentation of the BZR 920 and BZR 936 genome assemblies and the accumulated changes in the sequences, based on the homology of which the absence and presence of genes is assessed. It is also possible that these strains do not actually have a number of genes relative to the selected reference genome. These considerations apply to all results presented in Figure 3.

BZR 585 showed a high degree of similarity between the gene repertoire of the studied strains and the corresponding reference genomes in relation to *A. phenolicus*; BZR 762 and BZR 278 in relation to *Alcaligenes* sp.; and BZR 588 and BZR 201P in relation to *P. pullorum*. In addition, BZR 736 showed a good degree of correspondence with *B. cereus*, BZR 1159 with *Brevundimonas* sp., and BZR 162 with *O. quorumnocens*. At the same time, BZR 926 showed only an average level of correspondence with the reference genome of *A. marplatensis*, and BZR 635, when compared with the closest reference genome, *Leucobacter* sp., showed only a small number of intersections, which may be due both to different taxonomic identity and to the fragmentation of the genome assembly of this strain.

Comparison of strains, analysis of their insecticidal and fungicidal potential

Below are the results of the analysis of the insecticidal (Fig. 4) and fungicidal (see figure in Supplementary Material 5) potential of the studied strains based on a comparison of their protein repertoires with the corresponding functional activity.

Table 4. Characteristics of bacterial genome annotations from the analyzed samples compared with the corresponding reference genomes

Strain/reference	Identifier	Total length of assembly, bp	Genes	CDS	Predicted proteins, %
Strain	BZR 1159	2772387	2692	2655	79.16
Reference	GCF_003576505.1	3144717	3026	2953	83.58
Strain	BZR 162	4079562	3831	3786	78.86
Reference	GCF_002278035.1	4593578	5344	5118	95.92
Strain	BZR 201	3424442	3088	3034	80.47
	BZR 206	3387765	3052	2997	81.26
	BZR 466	3459663	3147	3092	79.41
Reference	GCF_024806925.1	3459535	3152	3079	81.63
Strain	BZR 201P	3452129	3277	3227	84.50
	BZR 588	3238581	3031	2991	86.26
Reference	GCF_041929845.1	3664841	3523	3355	89.64
Strain	BZR 278	4159417	3862	3807	85.81
	BZR 762	3922022	3606	3551	89.57
Reference	GCF_030515105.1	4098772	3753	3677	93.15
Strain	BZR 585	4184220	3863	3807	85.63
Reference	GCF_040430305.1	4121371	3752	3679	93.58
Strain	BZR 635	3024800	2707	2652	77.39
Reference	GCF_001482305.1	3554188	3404	3285	85.40
Strain	BZR 736	5366477	5456	5407	79.73
Reference	GCF_001635955.1	6097988	6148	5781	87.64
Strain	BZR 920	3715092	3643	3579	88.28
	BZR 936	3714897	3640	3570	88.41
Reference	GCF_014230015.1	3861135	3828	3691	92.74
Strain	BZR 926	6214265	5705	5638	85.17
Reference	GCF_902859635.1	6885835	6393	6280	90.94

Heat maps reflect the results of the search for fungicidal and insecticidal proteins among our strains in the form of iScore, which is the percentage of identity weighted by the proportion of the aligned region of the entire length of the reference protein (see Materials and methods).

A gene encoding the chaperonin GroEl was found in all analyzed strains (Fig. 4). At the same time, the samples belonging to the genus *Bacillus* (BZR 736, BZR 920, BZR 936) had the highest number of genes encoding insecticidal proteins. While strains BZR 920 and BZR 936, which are representatives of *B. velezensis*, demonstrate the presence of fengycin, mycosubtilin, and bacillomycin synthetases, BZR 736, belonging to *B. cereus*, possesses the genes Spp1Aa1 and Spp1Aa2, as well as Vpb1Ab1 and Vpb4Ca1. At the same time, all analyzed strains of the genus *Bacillus* also possess a gene encoding the enzyme N-acetylmuramic acid 6-phosphate esterase (EC 4.2.1.126). In addition, strains BZR 278, BZR 585, and BZR 762, belonging to the genus *Alcaligenes*, have fragments homologous to the gene for the bifunctional protein HldE, while strains BZR 201, BZR 206, and BZR 466, representing *Leucobacter aridicollis*, contain a number of fragments homologous to bacillomycin synthetase genes.

Proteins with fungicidal activity were identified (see Supplementary Material 5; the heat map was constructed based on the values specified in Supplementary Material 6) in samples belonging to the genus *Bacillus* (BZR 736, BZR 920, BZR 936). It can be seen that the most complete proteins matching the sequences from the database are present in two samples: BZR 920 and BZR 936, belonging to *B. velezensis*. These samples also lead in terms of the number of proteins

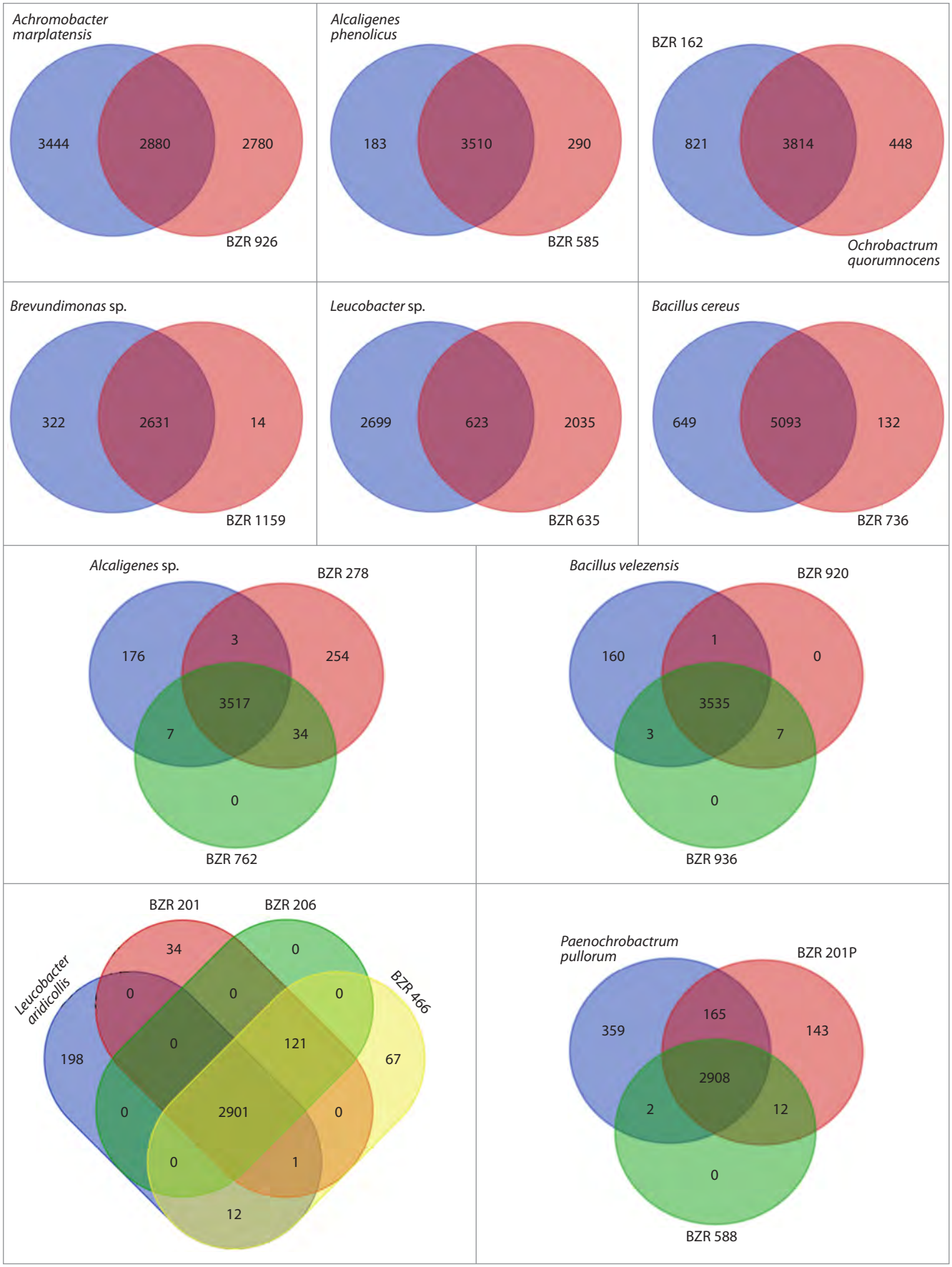


Fig. 3. Venn diagrams for gene sets of the studied strains, grouped by species, analyzed together with gene sets from the corresponding reference genomes.

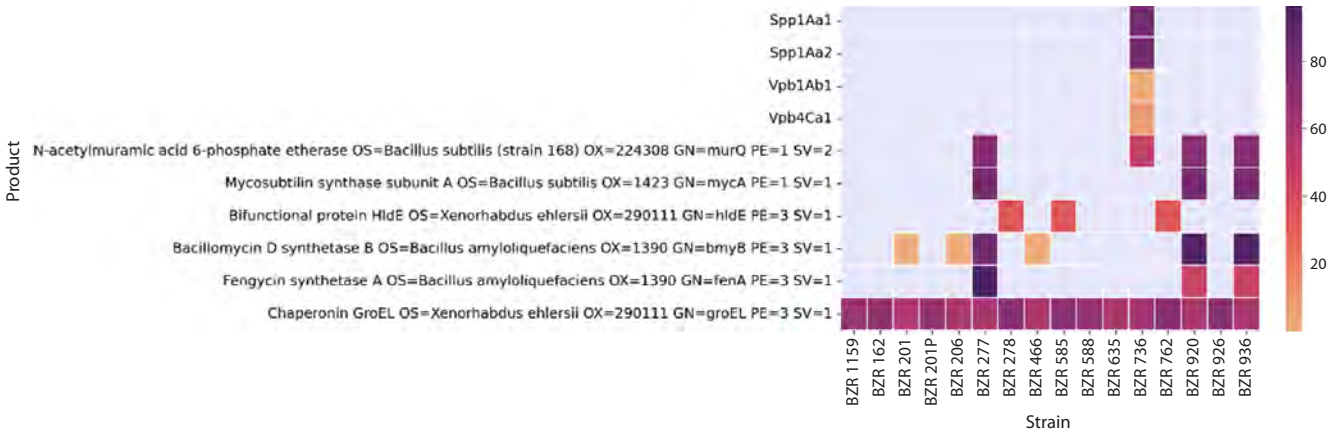


Fig. 4. Repertoire of proteins selected during pairwise alignment of translated genomes of the studied strains to the insecticidal protein database, where the color scale reflects the iScore parameter (see Equation 1). That is, the more intense the color, the more completely the sequence from the studied genome was aligned with the amino acid sequence of the protein from the database. Missing proteins are marked in grey. The heat map is based on the values (see Supplementary Material 4).

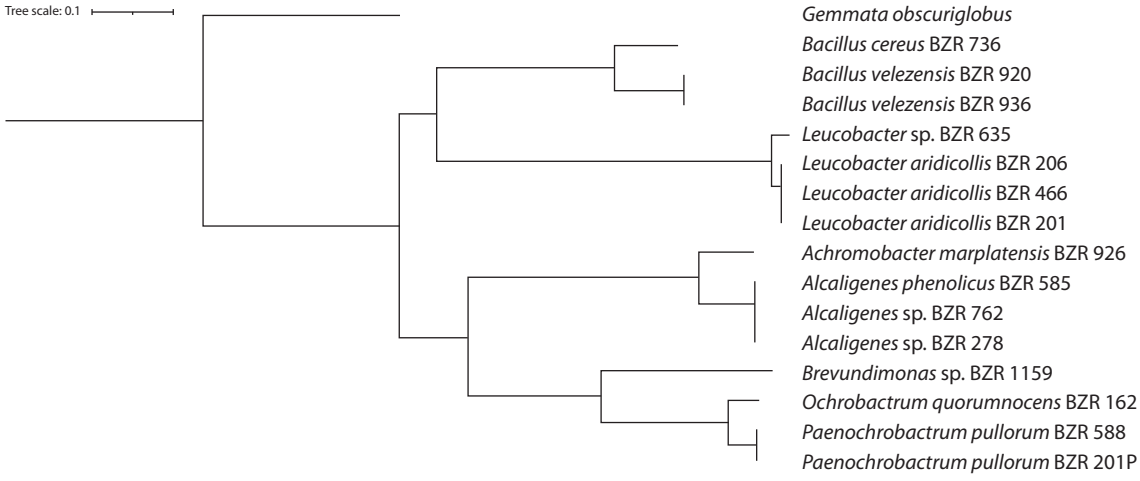


Fig. 5. Phylogenetic tree of the studied strains, constructed based on 120 marker genes found in the bacteria. *Gemmata obscuriglobus* serves as the outgroup.

found. Since the vast majority of sequences in the database of amino acid sequences of fungicidal proteins belonged to the species *B. velezensis* and *B. amyloliquefaciens*, the discovery of proteins in these two samples is a logical consequence. The predominance of data in the collected database of sequences related to *B. velezensis* and *B. amyloliquefaciens* was due to the predominance of data on iturins, fengycins, and surfactins of *B. velezensis* and *B. amyloliquefaciens* relative to other species of the genus *Bacillus* in the UniprotDB and NCBI Protein databases. However, genes encoding synthetases, as well as *YxjF* and *YxjC*, were also found in a strain belonging to *B. cereus* (BZR 736), although unlike the other strains analyzed, BZR 736 does not have the *ScoA* gene, which, along with *ScoB*, is present in all the strains analyzed. In addition, strains BZR 201, BZR 206, and BZR 466, representing *L. aridicollis*, contain a number of fragments homologous to the genes of iturin and fengycin synthetases. It is worth mentioning that the genomes of the entire clade of representatives of the Brucellaceae family (strains BZR 162, BZR 588, and BZR 201P) do not contain genes encoding fungicidal proteins represented in our database.

Phylogenetic analysis

To construct a phylogenetic tree (Fig. 5), a search for a set of 120 bacterial marker genes was performed using GTDB-Tk in all studied samples, and multiple alignment of protein sequences was performed, which was fed into the tree construction. Proteins from *Gemmata obscuriglobus* (GCF_003149495.1) were used as an outgroup.

It should be noted that the constructed phylogenetic tree (see Fig. 5) fully reflects the topology of the previously presented taxonomic tree (see Fig. 2), which relatively confirms the correctness of the established taxonomic identity. At the same time, the gene sequences used to construct the tree are identical in strains BZR 762 and BZR 278, as well as in the trio of strains BZR 466, BZR 206, BZR 201, indicating that the former belong to one species of the genus *Alcaligenes*, and the latter, to *L. aridicolis*.

Discussion

The study of endophytic bacteria in agriculturally important crops is a pressing issue, the resolution of which will enable the regulation of pest populations and effective control of plant diseases arising from interaction with pathogens. Therefore, it is particularly important to identify strains that possess fungicidal and insecticidal potential.

To date, the following antagonistic properties of endophytic bacteria are known for the studied genera. Here, we list the genera of bacteria for which taxonomic identification of genomes was performed in this study.

***Bacillus* spp.** The genomes of bacteria of the genus *B. velezensis* contain nine key NRPS/PKS clusters, synthesized lipopeptides (iturin, fengycin, surfactin) and polyketides (deficidin, macrolactin) capable of suppressing phytopathogenic fungi of the genera *Fusarium*, *Bipolaris*, *Exserohilum*, and others, both *in vitro* and *in vivo* (Wang S. et al., 2024; Yeo et al., 2024). Surfactin, iturin, and fengycin can act as entomicides and nematicides, as evidenced by the high mortality (up to 100 %) of *Aedes aegypti* mosquito larvae and *Agriotes lineatus* click beetles, while pure surfactin causes systemic metabolic disorders in the caterpillars of the Asian cotton bollworm *Spodoptera litura* (Falqueto et al., 2021; Zhang F. et al., 2024; Knežević et al., 2025).

***Alcaligenes* spp.** A binary protein, AfIP-1A/1B, has been discovered in *A. faecalis* bacteria that is capable of forming pores in the intestines of western corn rootworm larvae *Diabrotica virgifera*, including in insect populations resistant to *B. thuringiensis*. Some strains produce an exoprotease that kills the root-knot nematode *Meloidogyne incognita* and the soil nematode *Caenorhabditis elegans*. The *A. faecalis* N1-4 strain is capable of producing dimethyl disulfide and methyl isovalerate, which inhibit the growth of the fungus *Aspergillus flavus* and reduce the amount of mycotoxins in grain during storage (Ju et al., 2016; Gong et al., 2019; Pérez Ortega et al., 2021).

***Achromobacter* spp.** The *A. xylosoxidans* soil strain causes 95 % mortality of larvae and 100 % mortality of adults of the housefly *Musca domestica*. Volatile esters (S-methylthiobutyrate, acetates) reduce the population of the gall nematode *M. javanica* and suppresse gall formation on tomatoes by 60 %. The endophytic strain CTA8689 reduces melon wilt caused by *Fusarium oxysporum*/*F. solani* by 60 % in a greenhouse thanks to siderophores and esterases (Yamaç et al., 2010; Dhaouadi et al., 2019; Deng et al., 2022; Mohamadpoor et al., 2022).

***Brevundimonas* spp.** *B. diminuta* YYC02 produced 42 volatile compounds, of which butyl-2-methylbutanoate and isoamyl butyrate caused 90–100 % mortality of the root-knot nematode *Meloidogyne javanica* within 48 hours, while soil treatment reduced the number of galls by 37 % and increased the mass of cucumber shoots (Sun et al., 2023). *Brevundimonas* is part of the core microbiome of entomopathogenic nematodes (Steinernema, Meloidogyne) and actively adheres to the cuticle of J2 larvae, increasing their mortality and reducing egg hatching, which indirectly confirms the anti-PPN activity of the genus (Topalović et al., 2019). Although no classical Cry/Cyt toxins have been found in *Brevundimona*, a membrane organophosphate hydrolase has been studied in detail in *B. diminuta*, allowing the strain to use organophosphate insecticides as a source of phosphorus; the enzyme is localized in the periplasm by means of a Tat signal (Parthasarathy et al., 2016).

***Leucobacter* spp.** Two strains, Verde1/Verde2, cause nematode death through a rare mechanism in which a sticky exopolymer sticks the nematode tails together into “stars”, leading to the death of the colony. Nematode resistance to these strains is controlled by individual genes, which emphasizes the specialized virulence of the genus. The *L. aridicollis* SASBG215 strain inhibits cucumber anthracnose *Colletotrichum orbiculare* and causes lysis of hyphae, presumably with the help of unknown polyketides (Hodgkin et al., 2013; Abdul Salam et al., 2022).

***Ochrobactrum* spp.** Strain BS-206 synthesizes the glycolipid ochrozin, which kills 90–100 % of storage pests (*Tribolium*, *Sitophilus*, *Callosobruchus*) and has an insecticidal effect against the corn borer *Spodoptera*. The *O. pseudogrignonense* NC1 strain produces dimethyl disulfide and benzaldehyde, which cause up to 100 % mortality of young *M. incognita* individuals and reduce tomato gall formation by more than 60 % (Kumar et al., 2014; Yang T. et al., 2022).

As can be seen from the brief summary above, there is a variety of active substances for different pathogens, which can complicate mass analysis, but offer prospects for future research on little-studied endophytes that may have similar properties.

In this article, we presented 15 genomes of endophytic bacteria isolated from various parts of wheat, soybeans, and rapeseed. We have demonstrated the presence of genes responsible for insecticidal and fungicidal activity in the studied strains, with the largest number of genes encoding insecticidal and fungicidal proteins found in strains BZR 736, BZR 920, and BZR 936 of the genus *Bacillus*. However, for strains BZR 162, BZR 588, and BZR 201P of the Brucellaceae family, no genes encoding fungicidal proteins present in our database were identified.

This may indicate several factors: firstly, bacteria of the genus *Bacillus* are more widely studied and published, and secondly, the database contains a bias in data on insecticidal and fungicidal proteins for this genus. Therefore, when interpreting the results obtained for less represented genera, we make the reservation that the presence of such proteins is possible in these genomes, but their sequences differ significantly from the publicly available sequences that we were able to aggregate into the database. Nevertheless, the bias in the data for the genus *Bacillus* does not affect the search for proteins with insecticidal and fungicidal properties. There is also little aggregated information in the literature on insecticidal and fungicidal proteins, which confirms the need to create databases of proteins associated with insecticidal and fungicidal properties.

Nevertheless, the genomic data obtained in this study are consistent with the results of bioassays previously conducted for some of the strains studied. According to the initial screening of the bioresource collection, strains with pronounced entomopathogenic activity against the wax moth *Galleria mellonella* were identified. The strains BZR 1159, BZR 588, BZR 936, BZR 206, BZR 920, BZR 926, and BZR 277 (65–95 % mortality on the third day and 83–95 % on the fifth day) (Gyrnets (Bondarchuk), Asaturova, 2022).

With regard to the large mealworm beetle *Tenebrio molitor*, strains BZR 201, BZR 278, BZR 1159, BZR 635, BZR 762, BZR 736 showed themselves to be effective (72–98 % on the third day and 81–99 % on the fifth day). In addition, the potential multifunctionality of a number of strains was assessed: BZR 1159, BZR 936, and BZR G3, which showed insecticidal activity against the natural population of the codling moth *Cydia pomonella*, reaching 95.5 %, as well as fungicidal activity against apple disease pathogens of the genera *Fusarium*, *Alternaria*, *Trichothecium*, with mycelium inhibition of up to 84.8 % (Gyrnets (Bondarchuk), Asaturova, 2022; Gyrnets, Asaturova, 2023). The BZR 936 strain is worth noting: it has both insecticidal and fungicidal properties, which directly correlates with its identified lipopeptide synthase genes (iturin, fengycin, surfactin) and other biocontrol markers. Thus, comparison of the genomic composition with experimental biotests confirms that the presence of specific insecticidal protein genes and lipopeptide synthetase clusters is a reliable indicator of the biocontrol potential of strains. This opens up prospects for their further inclusion in programs for the development of biological products for the protection of agricultural crops.

Conclusion

In this article, we present the taxonomic identification, assemblies, and annotations of 15 endophytic bacterial genomes, the samples of which were obtained from the roots and/or leaves of wheat, soybean, and rapeseed.

Taxonomically, strains BZR 736, BZR 920, and BZR 936 belong to the genus *Bacillus*, strains BZR 635, BZR 466, BZR 206, and BZR 201 belong to the genus *Leucobacter*, and the remaining strains belong to the phylum Pseudomonadota. The phylogenetic tree constructed from a set of 120 bacterial marker genes fully reflects the topology of our taxonomic tree, which confirms the correctness of the established taxonomic identity at a relative level.

Genome assembly was performed in two stages: a preliminary stage of primary genome assembly and a hybrid (*de novo* + reference-guided) cascade approach based on a set of closely related references. Genome annotation was performed taking into account the databases of protein sequences with insecticidal properties prepared within the framework of this study, including the BPPRC database, and fungicidal properties, as well as the most complete known genome annotations of the corresponding species/genus.

Analysis of the gene repertoire revealed differences between the analyzed strains and the reference genomes used. A high degree of correspondence between the gene repertoire of the studied strains and the corresponding reference genomes was shown by BZR 585 in relation to *A. phenolicus*, by BZR 762 and BZR 278 in relation to *Alcaligenes* sp., by BZR 588 and BZR 201P with respect to *P. pullorum*, by BZR 920 and BZR 936 in relation to *B. velezensis*, and by BZR 206, BZR 466 and BZR 201 with respect to *L. aridicollis*. This indicator may also confirm the correctness of the established taxonomic identity. At the same time, when compared with the closest reference genome, *Leucobacter* sp., BZR 635 shows only a small number of intersections, which may be due to both its distinctive taxonomic identity and the fragmentation of the genome assembly of this strain.

We demonstrated the presence of genes encoding fungicidal and insecticidal proteins in all strains except BZR 162, BZR 588, and BZR 201P. No genes encoding fungicidal proteins present in our database were identified for these strains. However, the results obtained in this study indicate that the strains under study, which possess a complex of lipopeptide synthetase and insecticidal toxin genes, demonstrate an experimentally confirmed broad spectrum of biological activity against insects and phytopathogens.

Further targeted study of endophytic bacteria with fungicidal and insecticidal genes opens up prospects for identifying candidates for biocontrol agents of various pathogens and using bacteria to protect agricultural plants.

References

Abdul Salam S., Parthiban V.K., Paranidharan V., Johnson I., Karthikeyan M., Kavitha C. *Leucobacter aridicollis* strain SASBG215: a novel biocontrol agent against *Colletotrichum orbiculare*. *Biol Forum Int J.* 2022;14(2):905-911

Ali M.A., Ahmed T., Ibrahim E., Rizwan M., Chong K.P., Yong J.W.H. A review on mechanisms and prospects of endophytic bacteria in biocontrol of plant pathogenic fungi and their plant growth-promoting activities. *Heliyon*. 2024;10(11):e31573. doi 10.1016/j.heliyon.2024.e31573

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., … Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-477. doi 10.1089/cmb.2012.0021

Bateman A., Martin M.J., Orchard S., Magrane M., Agivetova R., Ahmad S., Alpi E., … Vinayaka C.R., Wang Q., Wang Y., Yeh L.S., Zhang J. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49(D1):D480-D489. doi 10.1093/nar/gkaa1100

Berry C., O'Neil S., Ben-Dov E., Jones A.F., Murphy L., Quail M.A., Holden M.T.G., Harris D., Zaritsky A., Parkhill J. Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. *Appl Environ Microbiol.* 2002;68(10):5082-5095. doi 10.1128/AEM.68.10.5082-5095.2002

Blanco-Míguez A., Beghini F., Cumbo F., McIver L.J., Thompson K.N., Zolfo M., Manghi P., … Franzosa E.A., Pasolli E., Asnicar F., Huttenhower C., Segata N. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol.* 2023;41(11):1633-1644. doi 10.1038/s41587-023-01688-w

Bolotin A., Gillis A., Sanchis V., Nielsen-LeRoux C., Mahillon J., Lereclus D., Sorokin A. Comparative genomics of extrachromosomal elements in *Bacillus thuringiensis* subsp. *israelensis*. *Res Microbiol.* 2017;168(4):331-344. doi 10.1016/j.resmic.2016.10.008

Chaumeil P.-A., Mussig A.J., Hugenholtz P., Parks D.H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*. 2022;38(23):5315-5316. doi 10.1093/bioinformatics/btac672

Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*. 2023;2(2):e107. doi 10.1002/imt2.107

Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890. doi 10.1093/bioinformatics/bty560

Dai R., Zhang J., Liu F., Xu H., Qian J.-M., Cheskis S., Liu W., … de Jonge R., Pieterse C.M.J., Levy A., Schlaeppi K., Bai Y. Crop root bacterial and viral genomes reveal unexplored species and microbiome patterns. *Cell*. 2025;188(9):2521-2539.e22. doi 10.1016/j.cell.2025.02.013

Deng X., Wang X., Li G. Nematicidal effects of volatile organic compounds from microorganisms and plants on plant-parasitic nematodes. *Microorganisms*. 2022;10(6):1201. doi 10.3390/microorganisms10061201

Dhaouadi S., Rouissi W., Mougou-Hamdane A., Nasraoui B. Evaluation of biocontrol potential of *Achromobacter xylosoxidans* against *Fusarium* wilt of melon. *Eur J Plant Pathol.* 2019;154(2):179-188. doi 10.1007/s10658-018-01646-2

Falqueto S.A., Pitaluga B.F., de Sousa J.R., Targanski S.K., Campos M.G., de Oliveira Mendes T.A., da Silva G.F., Silva D.H.S., Soares M.A. *Bacillus* spp. metabolites are effective in eradicating *Aedes aegypti* (Diptera: Culicidae) larvae with low toxicity to non-target species. *J Invertebr Pathol.* 2021;179:107525. doi 10.1016/j.jip.2020.107525

Gabrielaite M., Marvig R.L. GenAPI: a tool for gene absence-presence identification in fragmented bacterial genome sequences. *BMC Bioinformatics*. 2020;21(1):320. doi 10.1186/s12859-020-03657-5

Gong A.-D., Wu N.-N., Kong X.-W., Zhang Y.-M., Hu M.-J., Gong S.-J., Dong F.-Y., Wang J.-H., Zhao Z.-Y., Liao Y.-C. Inhibitory effect of volatiles emitted from *Alcaligenes faecalis* N1-4 on *Aspergillus flavus* and aflatoxins in storage. *Front Microbiol.* 2019;10:1419. doi 10.3389/fmicb.2019.01419

Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307-321. doi 10.1093/sysbio/syq010

Guizelini D., Raittz R.T., Cruz L.M., Souza E.M., Steffens M.B.R., Pedrosa F.O. GFinisher: a new strategy to refine and finish bacterial genome assemblies. *Sci Rep.* 2016;6(1):34963. doi 10.1038/srep34963

Gurevich A., Saveliev V., Vyahhi N., Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075. doi 10.1093/bioinformatics/btt086

Guyomar C., Delage W., Legeai F., Mougel C., Simon J.-C., Lemaitre C. MinYS: mine your symbiont by targeted genome assembly in symbiotic communities. *NAR Genom Bioinform.* 2020;2(3):lqaa047. doi 10.1093/nargab/lqaa047

Gyrnets (Bondarchuk) E.Yu., Asaturova A.M. Screening of bacterial strains on the criterion of entomopathogenic activity against *Galleria mellonella* L. and *Tenebrio molitor* L. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology in Agro-Industrial Complex*. 2022;36(3):53-60. doi 10.53859/02352451_2022_36_3_53 (in Russian)

Gyrnets E.Yu., Asaturova A.M. Study of the polyfunctional properties of promising bacterial agents in relation to phytophages and pathogens of fruit cenosis. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology in Agro-Industrial Complex*. 2023;37(5):39-45. doi 10.53859/02352451_2023_37_5_39 (in Russian)

Hamane S., El Yemlahi A., Hassani Zerrouk M., El Galiou O., Laglaoui A., Bakkali M., Arakrak A. Plant growth promotion and biocontrol potentiality of endophytes isolated from root nodules of *Sulla flexuosa* L. plants. *Int J Agron.* 2023;2023(1):1-9. doi 10.1155/2023/2451806

Hodgkin J., Félix M.-A., Clark L.C., Stroud D., Gravato-Nobre M.J. Two *Leucobacter* strains exert complementary virulence on *Caenorhabditis* including death by worm-star formation. *Curr Biol.* 2013;23(21):2157-2161. doi 10.1016/j.cub.2013.08.060

Horwich A.L., Fenton W.A., Chapman E., Farr G.W. Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol.* 2007;23(1):115-145. doi 10.1146/annurev.cellbio.23.090506.123555

Hussain H.S., Tabbasum S., Mahreen N., Yahya M., Ejaz K., Aslam Z., Imtiaz M., ul Islam E., Sajid Z.I., Yasmin S. Integrated application of drought-tolerant phosphate solubilizing bacteria and native P source improves P availability with sustainable wheat production. *J Soil Sci Plant Nutr.* 2025;25(2):4723-4746. doi 10.1007/s42729-025-02425-9

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

Ju S., Lin J., Zheng J., Wang S., Zhou H., Sun M. *Alcaligenes faecalis* ZD02, a novel nematicidal bacterium with an extracellular serine protease virulence factor. *Appl Environ Microbiol.* 2016;82(7):2112-2120. doi 10.1128/AEM.03444-15

Knežević M., Dervišević M., Jovković M., Jevđenović G., Maksimović J., Buntić A. Versatile role of *Bacillus velezensis*: biocontrol of *Fusarium poae* and wireworms and barley plant growth promotion. *Biol Control*. 2025;206:105789. doi 10.1016/j.biocontrol.2025.105789

Kolmogorov M., Raney B., Paten B., Pham S. Ragout – a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*. 2014;30(12):i302-i309. doi 10.1093/bioinformatics/btu280

Krzyżanowska D.M., Maciąg T., Ossowicki A., Rajewska M., Kaczyński Z., Czerwicka M., Rąbalski Ł., Czaplewska P., Jafra S. *Ochrobactrum quorumnocens* sp. nov., a quorum quenching bacterium from the potato rhizosphere, and comparative genome analysis with related type strains. *PLoS One*. 2019;14(1):e0210874. doi 10.1371/journal.pone.0210874

Kumar C.G., Sujitha P., Mamidyala S.K., Usharani P., Das B., Reddy C.R. Ochrosin, a new biosurfactant produced by halophilic *Ochrobactrum* sp. strain BS-206 (MTCC 5720): purification, characterization and its biological evaluation. *Process Biochem.* 2014;49(10):1708-1717. doi 10.1016/j.procbio.2014.07.004

Kupper M., Gupta S.K., Feldhaar H., Gross R. Versatile roles of the chaperonin GroEL in microorganism-insect interactions. *FEMS Microbiol Lett.* 2014;353(1):1-10. doi 10.1111/1574-6968.12390

Letunic I., Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 2024;52(W1):W78-W82. doi 10.1093/nar/gkae268

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013. doi 10.48550/arXiv.1303.3997

Liu Y., Yin C., Zhu M., Zhan Y., Lin M., Yan Y. Comparative genomic analysis of *Bacillus velezensis* BRI3 reveals genes potentially associated with efficient antagonism of *Sclerotinia sclerotiorum* (Lib.) de Bary. *Genes* (*Basel*). 2024;15(12):1588. doi 10.3390/genes15121588

Mohamadpoor M., Amini J., Ashengroph M., Azizi A. Evaluation of biocontrol potential of *Achromobacter xylosoxidans* strain CTA8689 against common bean root rot. *Physiol Mol Plant Pathol.* 2022;117:101769. doi 10.1016/j.pmpp.2021.101769

Panneerselvam S., Mishra R., Berry C., Crickmore N., Bonning B.C. BPPRC database: a web-based tool to access and analyse bacterial pesticidal proteins. *Database*. 2022;2022:baac022. doi 10.1093/database/baac022

Parthasarathy S., Parapatla H., Nandavaram A., Palmer T., Siddavattam D. Organophosphate hydrolase is a lipoprotein and interacts with P_i-specific transport system to facilitate growth of *Brevundimonas diminuta* using OP insecticide as source of phosphate. *J Biol Chem.* 2016;291(14):7774-7785. doi 10.1074/jbc.M116.715110

Pellegrinetti T.A., Monteiro G.G.T.N., Lemos L.N., dos Santos R.A.C., Barros A.G., Mendes L.W. *PGPg_finder*: a comprehensive and user-friendly pipeline for identifying plant growth-promoting genes in genomic and metagenomic data. *Rhizosphere*. 2024;30:100905. doi 10.1016/j.rhisph.2024.100905

Pérez Ortega C., Leininger C., Barry J., Poland B., Yalpani N., Altier D., Nelson M.E., Lu A.L. Coordinated binding of a two-component insecticidal protein from *Alcaligenes faecalis* to western corn rootworm midgut tissue. *J Invertebr Pathol.* 2021;183:107597. doi 10.1016/j.jip.2021.107597

Prjibelski A., Antipov D., Meleshko D., Lapidus A., Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinform.* 2020;70(1):e102. doi 10.1002/cpbi.102

Rabbee M.F., Ali M.S., Choi J., Hwang B.S., Jeong S.C., Baek K. *Bacillus velezensis*: a valuable member of bioactive molecules within plant microbiomes. *Molecules*. 2019;24(6):1046. doi 10.3390/molecules24061046

Richter M., Rosselló-Móra R., Oliver Glöckner F., Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*. 2016;32(6):929-931. doi 10.1093/bioinformatics/btv681

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. doi 10.1093/bioinformatics/btu153

Shi H., Zeng H., Yang X., Zhao J., Chen M., Qiu D. An insecticidal protein from *Xenorhabdus ehlersii* triggers prophenoloxidase activation and hemocyte decrease in *Galleria mellonella*. *Curr Microbiol*. 2012;64(6):604-610. doi 10.1007/s00284-012-0114-7

Sun Y., Ran Y., Yang H., Mo M., Li G. Volatile metabolites from *Brevundimonas diminuta* and nematicidal esters inhibit *Meloidogyne javanica*. *Microorganisms*. 2023;11(4):966. doi 10.3390/microorganisms11040966

Topalović O., Elhady A., Hallmann J., Richert-Pöggeler K.R., Heuer H. Bacteria isolated from the cuticle of plant-parasitic nematodes attached to and antagonized the root-knot nematode *Meloidogyne hapla*. *Sci Rep*. 2019;9(1):11477. doi 10.1038/s41598-019-47942-7

Walker B.J., Abeel T., Shea T., Priest M., Abouelliel A., Sakthikumar S., Cuomo C.A., Zeng Q., Wortman J., Young S.K., Earl A.M. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963. doi 10.1371/journal.pone.0112963

Wang S., Jin P., Zheng Y., Kangkang W., Chen J., Liu J., Li Y. *Bacillus velezensis* B105-8, a potential and efficient biocontrol agent in control of maize stalk rot caused by *Fusarium graminearum*. *Front Microbiol*. 2024;15:1462992. doi 10.3389/fmicb.2024.1462992

Wang Z., Lu K., Liu X., Zhu Y., Liu C. Comparative functional genome analysis reveals the habitat adaptation and biocontrol characteristics of plant growth-promoting bacteria in NCBI databases. *Microbiol Spectr.* 2023;11(3):e0500722. doi 10.1128/spectrum.05007-22

Wu Y.-W., Simmons B.A., Singer S.W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2016;32(4):605607. doi 10.1093/bioinformatics/btv638

Yamaç M., Şahin E., Ceyhan E., Amoroso M.J., Cuozzo S.A., Pilatin S. The screening of bacterial insecticides against *Musca domestica* L. (diptera: Muscidae). *Fresenius Environ Bull.* 2010;19(5):862-870

Yang K., Dai X., Maitikadir Z., Zhang H., Hao H., Yan C. Comparative genome analysis of endophytic *Bacillus amyloliquefaciens* MR4: a potential biocontrol agent isolated from wild medicinal plant root tissue. *J Appl Genet.* 2024;65(4):907-923. doi 10.1007/s13353-024-00905-9

Yang T., Xin Y., Liu T., Li Z., Liu X., Wu Y., Wang M., Xiang M. Bacterial volatile-mediated suppression of root-knot nematode (*Meloidogyne incognita*). *Plant Dis.* 2022;106(5):1358-1365. doi 10.1094/PDIS-06-21-1139-RE

Yeo Y.J., Park A.R., Vuong B.S., Kim J.-C. Biocontrol of *Fusarium* head blight in rice using *Bacillus velezensis* JCK-7158. *Front Microbiol.* 2024;15:1358689. doi 10.3389/fmicb.2024.1358689

Zhang F., Liu Q., Wang Y., Yin J., Meng X., Wang J., Zhao W., Liu H., Zhang L. Effects of surfactin stress on gene expression and pathological changes in *Spodoptera litura*. *Sci Rep.* 2024;14(1):30357. doi 10.1038/s41598-024-81946-2

Zhang M., Kong Z., Fu H., Shu X., Xue Q., Lai H., Guo Q. Rhizosphere microbial ecological characteristics of strawberry root rot. *Front Microbiol.* 2023;14:1286740. doi 10.3389/fmicb.2023.1286740

Conflict of interest. The authors declare no conflict of interest.
Received July 2, 2025. Revised October 17, 2025. Accepted October 18, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1318-1324

doi 10.18699/vjgb-25-138

A housekeeping gene search to analyze expression changes of individual genes in *Macaca mulatta*

M.V. Shulskaya, A.Kh. Alieva, I.R. Kumakov, M.I. Shadrina, P.A. Slominsky

National Research Center "Kurchatov Institute", Moscow, Russia
shulskaya.m@yandex.ru

Abstract. Rhesus macaques (*Macaca mulatta*) are the most common non-human primates living in captivity. The use of rhesus macaques as model objects is determined, first of all, by their phylogenetic and physiological closeness to humans, and, as a consequence, the possibility of extrapolating the obtained results to humans. Currently, it is known that a number of biochemical changes occur under various physiological conditions, including at the transcriptomic level. The real-time polymerase chain reaction is a widely used universal method for gene expression analysis. Carrying out such studies always requires a preliminary selection of "housekeeping genes" (HKGs) – genes necessary for the implementation of basic functions in the cell and stably expressed in different cell types and under different conditions. At present, there are only two systematic studies on the search for HKGs in the rhesus macaque brain, and therefore in this work a search and systematization of HKGs for this species were carried out. As a result, two panels of promising HKGs for *M. mulatta* were formed: an extended panel, consisting of 56 genes, and a small panel, consisting of 8 genes: *ARHGDIA*, *CYB5R1*, *NDUFA7*, *RRAGA*, *TTC1*, *UBA6*, *VPS72*, and *YWHAH*. Both panels of potential HKGs do not have pseudogenes in macaques or humans, are characterized by stable and sufficient expression in the brain of rhesus macaques and can be used to analyze expression not only in the brain but also in peripheral blood. However, it should be noted that the data have not been experimentally verified and require verification in laboratory conditions.
Key words: *Macaca mulatta*; expression analysis; "housekeeping gene"; real-time PCR; expression

For citation: Shulskaya M.V., Alieva A.Kh., Kumakov I.R., Shadrina M.I., Slominsky P.A. A housekeeping gene search to analyze expression changes of individual genes in *Macaca mulatta*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1318-1324. doi 10.18699/vjgb-25-138

Funding. This work was supported by the Ministry of Science and Higher Education of the Russian Federation (the Federal Scientific-technical programme for genetic technologies development for 2019–2030, agreement No. 075-15-2025-491 dated May 30, 2025).

Поиск генов домашнего хозяйства для анализа изменения экспрессии отдельных генов у *Macaca mulatta*

М.В. Шульская, А.Х. Алиева, И.Р. Кумаков, М.И. Шадрина, П.А. Сломинский

Национальный исследовательский центр «Курчатовский институт», Москва, Россия
shulskaya.m@yandex.ru

Аннотация. Макаки резус (*Macaca mulatta*) являются наиболее распространенными нечеловекообразными приматами, их используют в качестве модельных объектов, в первую очередь, из-за филогенетической и физиологической близости к человеку. В настоящее время модельные организмы широко используются для целого ряда исследований, в том числе на уровне транскриптома. При этом для анализа экспрессии отдельных генов применяется универсальный метод – полимеразная цепная реакция в реальном времени. Проведение такого рода исследований всегда требует предварительного подбора «генов домашнего хозяйства» (ГДХ) – генов, необходимых для реализации основных функций в клетке и стабильно экспрессирующихся в различных типах клеток и при разных условиях. На сегодняшний день для макак резус существуют лишь две систематизированные работы по поиску ГДХ, однако эти исследования проводились лишь для тканей мозга и не учитывают такой важный критерий, как связь ГДХ с заболеваниями. В связи с этим в нашей работе были сформулированы ключевые критерии, учитываемые при подборе ГДХ. Проведены поиск и систематизация кандидатных ГДХ. В результате сформированы две панели перспективных ГДХ для *M. mulatta*: расширенная панель на 56 генов и малая панель, состоящая из восьми генов: *ARHGDIA*, *CYB5R1*, *NDUFA7*, *RRAGA*, *TTC1*, *UBA6*, *VPS72* и *YWHAH*. Обе панели соответствуют всем критериям подбора ГДХ (не имеют псевдогенов ни у макаки, ни у человека, характеризуются стабильной и достаточной экспрессией в мозге макак резус и могут быть использованы для анализа экспрессии не только в мозге, но и в периферической крови). Однако необходимо отметить, что данные экспериментально не верифицированы и требуют проверки в лабораторных условиях.
Ключевые слова: *Macaca mulatta*; экспрессионный анализ; «ген домашнего хозяйства»; ПЦР в реальном времени; экспрессия

Introduction

Rhesus macaques (*Macaca mulatta*) have served as a model for studying various human diseases for decades. Their use as a model is primarily explained by the phylogenetic and physiological similarity to humans, and, consequently, the potential for transferring the results obtained. To date, genetic models of cancer (Brammer et al., 2018; Deycmar et al., 2023), cardiovascular diseases (Patterson et al., 2002; Ueda et al., 2019), ophthalmologic diseases (Singh et al., 2009; Liu et al., 2015; Moshiri et al., 2019; Peterson et al., 2019, 2023), skeletal diseases (Colman, 2018; Paschalis et al., 2019), diseases of the reproductive system (Lomniczi et al., 2012; Nair et al., 2016; Abbott et al., 2019), as well as a wide range of neurological diseases (McBride et al., 2018; Sherman et al., 2021) are known in rhesus macaques. In addition, rhesus macaques are used for research as model objects of toxicity (Kaya et al., 2023), radiation (Li et al., 2021; Majewski et al., 2021), hormones (Noriega et al., 2010; Eghlidi, Urbanski, 2015), etc. In addition to studying diseases, this model can be used to test various pharmacological drugs, which is especially important for applied research.

It is now known that a wide range of biochemical changes occur under various physiological conditions, including at the transcriptome level. Relative transcript levels of individual genes can be accurately and reproducibly measured using real-time polymerase chain reaction (RT-PCR). This method is a widely used and versatile tool for analyzing the expression of a small number of genes. RT-PCR is also frequently used to confirm results obtained using whole-transcriptome expression analysis (Ramsköld et al., 2009). However, this type of study is always complicated by variations in the copy number of the target mRNA due to differences in the amount of total RNA between samples, therefore requiring the preliminary selection of control (reference) genes, or "housekeeping genes" (HKGs).

The term HKG most often refers to genes stably expressed in various cell types and under various conditions and required for basic cellular functions. They are often used as reference genes in gene expression studies to normalize mRNA levels between different samples.

In rhesus macaques, there is currently very little systematic data on the use of HKGs (Ahn et al., 2008; Noriega et al., 2010). Noriega et al. (2010) conducted a study only on the brain, while Ahn et al. (2008) worked with both brain tissue and some other tissues (intestine, liver, kidney, lung, and stomach). However, neither of these studies examined the animals' peripheral blood, which is widely used for various expression studies. In this regard, this review conducted a search and systematization of data on HKGs in rhesus macaques for their further use in studying gene expression changes under various conditions.

Modern principles of selection of HKGs

Currently, the selection of HKGs is based on the following main principles. First, the absence of pseudogenes, copies of genes that contain certain defects in the coding region (loss of introns and exons, frameshifts, or premature stop codons, as well as pseudogenes formed as a result of retrotransposition),

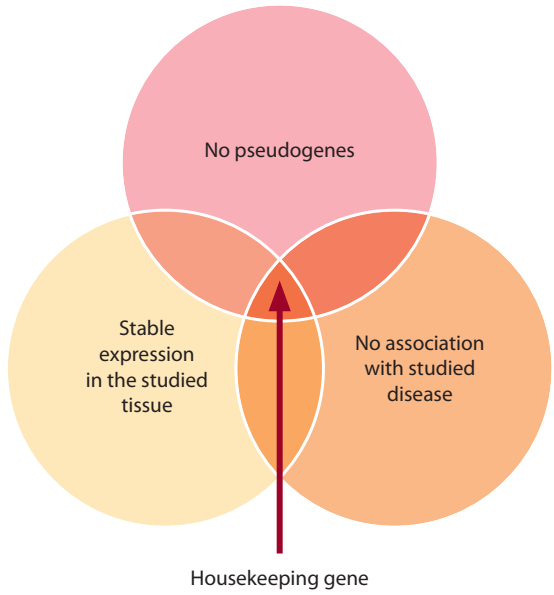


Fig. 1. Main HKG criteria.

is an important criterion for selecting HKGs (Tutar, 2012). Pseudogenes are not involved in protein processing but can be expressed at the RNA level. Furthermore, the number of pseudogenes is known to be unstable in the genomes of different individuals. From a practical standpoint, the presence of pseudogenes may require additional treatment of the analyzed RNA samples with DNases, which is critical for samples with low RNA amounts. Therefore, the presence of pseudogenes is highly undesirable when selecting HKGs.

Second, expression stability is considered to be another important criterion for selecting HKGs, i. e., they should have relatively constant expression levels across different cell types, tissues, and experimental conditions (Tu et al., 2006). However, it is known that HKGs can be expressed differentially in different tissues. For example, well-known HKGs such as beta-actin and *GAPDH* have been shown to vary significantly in expression levels across tissues (Cai J. et al., 2014). Therefore, a high level of HKGs' expression in the specific tissue under study is an important criterion.

Third, there is increasing support for the idea that HKGs should be tailored to specific experimental conditions (Silver et al., 2008). For example, the human *HSPA8* gene is a HKG, but it cannot be used as such in the study of age-related or neurodegenerative diseases, as there is evidence of a decrease in *HSPA8* gene expression with age, as well as an association between this gene and the development of neurodegenerative diseases (Loeffler et al., 2016; Tanaka et al., 2024). Expression profile variability has also been demonstrated for HKGs used in the study of cancer (de Kok et al., 2005; Dheda et al., 2005). To date, no studies have identified all-purpose HKGs, meaning that HKGs' selection for the specific pathology being studied is necessary.

Thus, an ideal HKG should have no pseudogenes, no association with the disease or condition being studied, and it should be stably expressed under specific experimental conditions and tissues (Fig. 1). The optimal HKG should be

Table 1. Names of search queries in the PubMed database (https://pubmed.ncbi.nlm.nih.gov/)

Search query	Result (publications, pcs.)
(gene expression) AND (rhesus macaque)	3,017
(gene expression) AND (*Macaca mulatta*)	2,743
(housekeeping genes) AND (rhesus macaque)	126
(housekeeping genes) AND (*Macaca mulatta*)	112
(reference genes) AND (rhesus macaque)	97
(reference genes) AND (*Macaca mulatta*)	86
((housekeeping genes) AND (rt-pcr)) AND (rhesus macaque)	16
((housekeeping genes) AND (rt-pcr)) AND (*Macaca mulatta*)	16
((reference genes) AND (rt-pcr)) AND (rhesus macaque)	7
((reference genes) AND (rt-pcr)) AND (*Macaca mulatta*)	7

Note. Accessed on April 28, 2025.

carefully selected for each specific experiment. Using multiple HKGs also improves the reliability of the expression data obtained (Vandesompele et al., 2002; Dheda et al., 2005).

Analysis of the published data on HKGs in rhesus macaques

We screened scientific publications in the PubMed database to find papers focused on the analysis of HKGs in rhesus macaques. An initial search using the keywords (gene expression) AND (rhesus macaque) identified 3,017 publications. Since "rhesus macaque" and "*Macaca mulatta*" are synonymous, both terms were used in the analysis of search queries. Due to the relatively large number of publications returned, the search query was specified using the synonymous terms "housekeeping genes" and "reference genes", which yielded 126 and 97 search results, respectively. Further narrowing the search by refining it using the keyword "rt-pcr" revealed 16 and 7 publications (Table 1).

A detailed analysis of these seven studies identified two most relevant systematic studies to date on the selection of HKGs in rhesus macaques (Ahn et al., 2008; Noriega et al., 2010). Five of the seven remaining publications analyzed did not mention HKGs and were therefore not included in the analysis.

Next, a block of 126 open-access publications found in PubMed using the keywords (housekeeping genes) AND (rhesus macaque) was manually analyzed. It was found that 107 publications, for one reason or another, did not mention any HKGs, while 16 publications used genes recommended by the authors of the two main studies on the selection of HKGs in rhesus macaques (Ahn et al., 2008; Noriega et al., 2010). These two types of publications were excluded from further analysis. Our search yielded only one additional publication (Robinson et al., 2018). Supplementary Table S1[1] summarizes the data from these three key studies and describes 115 genes expressed in the rhesus macaque brain that could be considered as HKGs. These genes were selected for further analysis.

Due to periodic database updates, some gene names were updated and given with names different from those used in (Ahn et al., 2008; Noriega et al., 2010) when compiling this list. Four sequences that were homologous to human sequences but were absent in the Ensembl database for rhesus macaques (Genome assembly: Mmul_10 (GCA_003339765.3)) (Table S1) and five *M. mulatta* genes currently identified as having pseudogenes (*LDHB*, *RPL37*, *RPS27A*, *SNRPA*, and *SUI1*) were also excluded.

Due to the underannotation of modern rhesus macaque genome assemblies (for example, we found that the nucleotide sequence of the *M. mulatta YWHAH* gene in the Ensembl database corresponds to the sequence of the unannotated *DEPCD5* gene in the NCBI database), we assessed the presence of pseudogenes not only in rhesus macaques but also in humans using the Ensembl database (www.ensembl.org). As a result, we excluded 58 genes with human orthologs having pseudogenes.

This procedure allows us to identify all-purpose HKGs for both humans and macaques, while also avoiding problems associated with the low level of annotation of the rhesus macaque genome assembly. For example, the *RPL19* gene, currently the most widely used HKG in rhesus macaques, is not recommended for use as an all-purpose HKG because it has pseudogenes in human genome.

The genes selected after the previous screening steps can be used for studies on brain tissue. However, peripheral blood, widely used in human studies, is of particular interest. Peripheral blood is promising for expression studies due to its availability and low invasiveness. Therefore, we considered it necessary to select candidate HKGs for peripheral blood, for the purpose of which the selected genes were further analyzed for acceptable expression levels in peripheral blood (Table S2).

[1] Supplementary Tables S1 and S2 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Shulskaya_Engl_29_8.pdf

Since peripheral blood expression data are currently completely lacking for *M. mulatta*, and due to the similarity between the macaque and human transcriptomes, publicly available mRNA expression data were analyzed in human whole blood and lymphoblasts. We also included expression data in mice, as these animals are a well-studied model object (due to the lack of peripheral blood data, tissues with similar expression patterns, such as bone marrow, lymph nodes, and spleen, were used). Expression data in the brain and spleen of rhesus macaques were added from the Ensembl database (Table S2).

This analysis was conducted using the BioGPS database (http://biogps.org/), where we selected genes with expression above the median in the tissues of interest. "Median expression" represents the 50th percentile of the expression data, meaning that half of the tissues have expression levels below the median, and the other half have expression levels above the median. BioGPS uses this metric to provide a summary of how a gene is expressed in different tissues, conditions, or data sets.

As a result of the analysis, the list of genes was divided into three groups: genes with expression levels above the median in both humans and mice, genes with expression levels above the median in only one of the two species, and genes with expression levels below the median in both humans and mice (Fig. 2, Table S2). Genes from all three groups can be considered as candidate HKGs. However, their use will limit the number of model objects compared based on their expression profiles. Genes from the first group are the most promising. It should also be noted that the expression data presented in BioGPS require experimental verification in the laboratory.

However, it is important to note that the median value is not always a good indicator for selecting candidate genes, since the mRNA abundance in the tissue under study may be higher than the median, but the absolute expression levels are quite low. Therefore, all analyzed genes were ranked according to their relative expression levels in the analyzed tissues. The results of this analysis are presented as a heat map (Fig. 3). Ultimately, we formed a group of 25 most promising candidate HKGs (genes with high or moderate expression levels in humans, mice, and rhesus macaques).

Since HKGs can be used to study changes in the expression of various genes in various diseases, potential HKGs should not be implicated in the development of the disease under study. A selected group of 25 genes was analyzed using the MalaCards database (www.malacards.org). MalaCards is a searchable, integrated knowledge base containing comprehensive information on human diseases, medical conditions, and disorders. We searched for associations between the gene and currently known disease models in rhesus macaques (Table 2). Six genes associated with oncological diseases (*AHSA1*, *B4GALT3*, *HPCAL1*, *TBP*, *TMED9*, and *SSR2*), six genes associated with neurological diseases (*CSNK2B*, *DIAPH1*, *MAPKAPK2*, *NDUFA1*, *RAD23A*, and *UBB*), as well as genes associated with eye diseases (*ARL2* and *PRPF8*) and some other diseases (*GPX4* and *LDHA*) were excluded.

As a result, at this final stage of the selection of candidate HKGs, we selected eight genes (*ARHGDIA*, *CYB5R1*, *NDUFA7*, *RRAGA*, *TTC1*, *UBA6*, *VPS72*, and *YWHAH* – highlighted bold in Table 2), characterized by the absence of pseudogenes, the absence of data on the involvement of these genes in the development of diseases modeled in rhesus macaques, as well as stable and high expression in the analyzed tissues (brain, peripheral blood, spleen, lymph nodes, bone marrow).

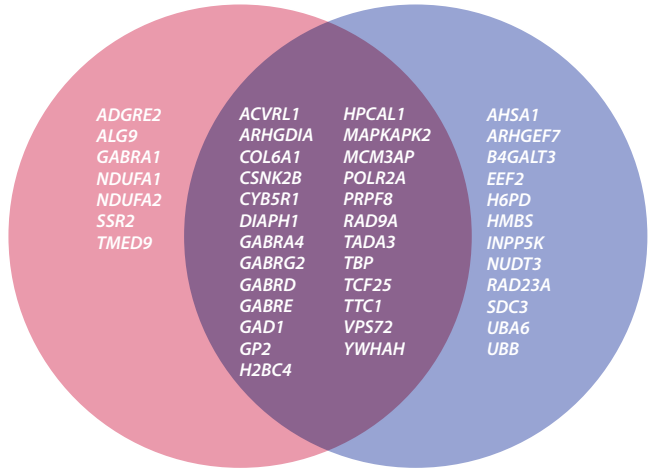


Fig. 2. Expression of candidate HKGs in selected human and mouse tissues.

Genes expressed predominantly in humans are shown in pink, and genes expressed predominantly in mice are shown in purple. The overlapping area indicates genes expressed in specific tissues of both species.

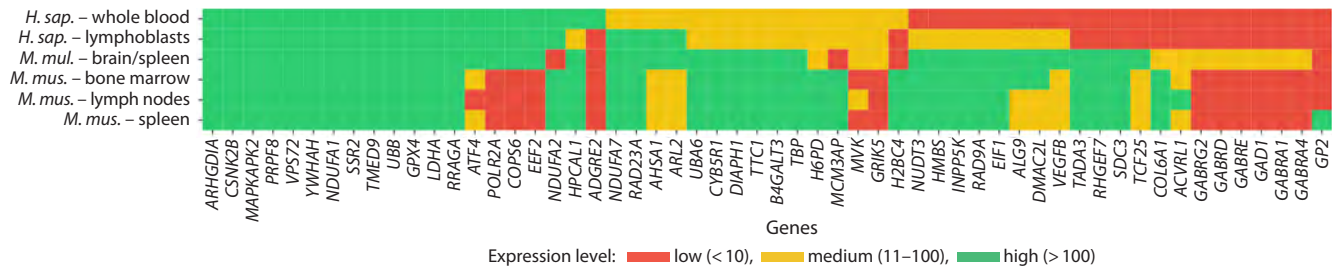


Fig. 3. Heatmap of relative expression levels of candidate HKGs.

Median-normalized values for each gene in the BioGPS resource (http://biogps.org) were used as the basis.

Table 2. Association of the selected highly expressed potential HKGs with disease groups modeled in rhesus macaques

Gene	Disease*	Reference
AHSA1	Osteosarcoma and hepatocellular carcinoma	Wei et al., 2024
ARL2	Rod-cone dystrophy, cataracts, and posterior staphyloma	Cai X.B. et al., 2019
ARHGDIA	–	–
B4GALT3	Cancers	Sun et al., 2016
CSNK2B	Myoclonic epilepsy	Poirier et al., 2017
CYB5R1	–	–
DIAPH1	Microcephaly	Esmaeilzadeh et al., 2024
GPX4	Spondylometaphyseal dysplasia of the Sedaghatian type	Smith et al., 2014
H6PD	Glioblastoma	Zhang Y.B. et al., 2022
HPCAL1	Glioblastoma	Zhang D. et al., 2019
LDHA	Fanconi–Bickel syndrome	Serrano-Lorenzo et al., 2022
MAPKAPK2	Pheochromocytoma, ataxia, telangiectasia	Liang et al., 2015
NDUFA1	Mitochondrial encephalomyopathy	Fernandez-Moreira et al., 2007
NDUFA7	–	–
PRPF8	Retinitis pigmentosa, retinal dystrophy	Tanackovic et al., 2011; Georgiou et al., 2021
RAD23A	Machado–Joseph disease	Doss-Pepe et al., 2003
RRAGA	–	–
SSR2	Hepatocellular carcinoma	Chen et al., 2022
TBP	Ataxia, phenotype associated with Huntington's disease	Zühlke et al., 2001; Stevanin et al., 2003
TMED9	Cancers	Mishra et al., 2019; Wang et al., 2024
TTC1	–	–
UBA6	–	–
UBB	Alzheimer's disease	Maniv et al., 2023
VPS72	–	–
YWHAH	–	–

*Pubmed (https://pubmed.ncbi.nlm.nih.gov/) has no published data for 2000–2025.

Conclusion

Thus, two panels of promising HKGs for *M. mulatta* were formed: an extended panel consisting of 56 genes (Table S2) and a small panel consisting of 8 genes (Table 2). Both panels of potential HKGs have no pseudogenes either in macaques or in humans, and they are characterized by stable and sufficient expression in the rhesus macaque brain. However, the specialized panel is more all-purpose, as it is suitable for selecting HKGs for parallel studies on several model organisms (mouse, macaque, and human) or for studying several different diseases simultaneously by a single research group. The small panel is of interest for further development of a working HKGs panel to study changes in the expression of various genes in various diseases in *M. mulatta*. At the same time, the extended panel of potential HKGs is also quite promising.

References

Abbott D.H., Rogers J., Dumesic D.A., Levine J.E. Naturally occurring and experimentally induced rhesus macaque models for polycystic ovary syndrome: translational gateways to clinical application. *Med Sci* (*Basel*). 2019;7(12):107. doi 10.3390/medsci7120107

Ahn K., Huh J.W., Park S.J., Kim D.S., Ha H.S., Kim Y.J., Lee J.R., Chang K.T., Kim H.S. Selection of internal reference genes for SYBR green qRT-PCR studies of rhesus monkey (*Macaca mulatta*) tissues. *BMC Mol Biol*. 2008;9:78. doi 10.1186/1471-2199-9-78

Brammer D.W., Gillespie P.J., Tian M., Young D., Raveendran M., Williams L.E., Gagea M., … Pasqualini R., Arap W., Rogers J., Abee C.R., Gelovani J.G. MLH1-rheMac hereditary nonpolyposis colorectal cancer syndrome in rhesus macaques. *Proc Natl Acad Sci USA*. 2018;115(11):2806-2811. doi 10.1073/pnas.1722106115

Cai J., Li T., Huang B., Cheng H., Ding H., Dong W., Xiao M., Liu L., Wang Z. The use of laser microdissection in the identification of suitable reference genes for normalization of quantitative real-time

PCR in human FFPE epithelial ovarian tissue samples. *PLoS One.* 2014;9(4):e95974. doi 10.1371/journal.pone.0095974

Cai X.B., Wu K.C., Zhang X., Lv J.N., Jin G.H., Xiang L., Chen J., Huang X.F., Pan D., Lu B., Lu F., Qu J., Jin Z.B. Whole-exome sequencing identified ARL2 as a novel candidate gene for MRCS (microcornea, rod-cone dystrophy, cataract, and posterior staphyloma) syndrome. *Clin Genet.* 2019;96(1):61-71. doi 10.1111/cge.13541

Chen F., Wang J., Zhang S., Chen M., Zhang X., Wu Z. Overexpression of SSR2 promotes proliferation of liver cancer cells and predicts prognosis of patients with hepatocellular carcinoma. *J Cell Mol Med.* 2022;26(11):3169-3182. doi 10.1111/jcmm.17314

Colman R.J. Non-human primates as a model for aging. *Biochim Biophys Acta Mol Basis Dis.* 2018;1864(9):2733-2741. doi 10.1016/j.bbadis.2017.07.008

de Kok J.B., Roelofs R.W., Giesendorf B.A., Pennings J.L., Waas E.T., Feuth T., Swinkels D.W., Span P.N. Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest.* 2005;85(1):154-159. doi 10.1038/labinvest.3700208

Deycmar S., Gomes B., Charo J., Ceppi M., Cline J.M. Spontaneous, naturally occurring cancers in non-human primates as a translational model for cancer immunotherapy. *J Immunother Cancer.* 2023;11(1):e005514. doi 10.1136/jitc-2022-005514

Dheda K., Huggett J.F., Chang J.S., Kim L.U., Bustin S.A., Johnson M.A., Rook G.A., Zumla A. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal Biochem.* 2005;344(1):141-143. doi 10.1016/j.ab.2005.05.022

Doss-Pepe E.W., Stenroos E.S., Johnson W.G., Madura K. Ataxin-3 interactions with rad23 and valosin-containing protein and its associations with ubiquitin chains and the proteasome are consistent with a role in ubiquitin-mediated proteolysis. *Mol Cell Biol.* 2003;23(18):6469-6483. doi 10.1128/MCB.23.18.6469-6483.2003

Eghlidi D.H., Urbanski H.F. Effects of age and estradiol on gene expression in the rhesus macaque hypothalamus. *Neuroendocrinology.* 2015;101(3):236-245. doi 10.1159/000381063

Esmaeilzadeh E., Biglari S., Mosallaei M., Khorshid H.R.K., Vahidnezhad H., Tabatabaiefar M.A. A novel homozygote pathogenic variant in the DIAPH1 gene associated with seizures, cortical blindness, and microcephaly syndrome (SCBMS): report of a family and literature review. *Mol Genet Genomic Med.* 2024;12(11):e70031. doi 10.1002/mgg3.70031

Fernandez-Moreira D., Ugalde C., Smeets R., Rodenburg R.J., Lopez-Laso E., Ruiz-Falco M.L., Briones P., Martin M.A., Smeitink J.A., Arenas J. X-linked *NDUFA1* gene mutations associated with mitochondrial encephalomyopathy. *Ann Neurol.* 2007;61(1):73-83. doi 10.1002/ana.21036

Georgiou M., Ali N., Yang E., Grewal P.S., Rotsos T., Pontikos N., Robson A.G., Michaelides M. Extending the phenotypic spectrum of *PRPF8*, *PRPH2*, *RP1* and *RPGR*, and the genotypic spectrum of early-onset severe retinal dystrophy. *Orphanet J Rare Dis.* 2021;16(1):128. doi 10.1186/s13023-021-01759-8

Kaya I., Nilsson A., Luptáková D., He Y., Vallianatou T., Bjärterot P., Svenningsson P., Bezard E., Andrén P.E. Spatial lipidomics reveals brain region-specific changes of sulfatides in an experimental MPTP Parkinson's disease primate model. *NPJ Parkinsons Dis.* 2023;9(1):118. doi 10.1038/s41531-023-00558-1

Li Y., Singh J., Varghese R., Zhang Y., Fatanmi O.O., Cheema A.K., Singh V.K. Transcriptome of rhesus macaque (*Macaca mulatta*) exposed to total-body irradiation. *Sci Rep.* 2021;11(1):6295. doi 10.1038/s41598-021-85669-6

Liang N., Zhong R., Hou X., Zhao G., Ma S., Cheng G., Liu X. Ataxia-telangiectasia mutated (ATM) participates in the regulation of ionizing radiation-induced cell death *via* MAPK14 in lung cancer H1299 cells. *Cell Prolif.* 2015;48(5):561-572. doi 10.1111/cpr.12203

Liu D.X., Gilbert M.H., Wang X., Didier P.J., Shields C.L., Lackner A.A. Coats-like retinopathy in a Young Indian Rhesus Macaque (*Macaca mulatta*). *J Med Primatol.* 2015;44(2):108-112. doi 10.1111/jmp.12166

Loeffler D.A., Klaver A.C., Coffey M.P., Aasly J.O., LeWitt P.A. Age-related decrease in heat shock 70-kDa protein 8 in cerebrospinal fluid is associated with increased oxidative stress. *Front Aging Neurosci.* 2016;8:178. doi 10.3389/fnagi.2016.00178

Lomniczi A., Garcia-Rudaz C., Ramakrishnan R., Wilmot B., Khouangsathiene S., Ferguson B., Dissen G.A., Ojeda S.R. A single-nucleotide polymorphism in the EAP1 gene is associated with amenorrhea/oligomenorrhea in nonhuman primates. *Endocrinology.* 2012;153(1):339-349. doi 10.1210/en.2011-1540

Majewski M., Ostheim P., Gluzman-Poltorak Z., Vainstein V., Basile L., Schüle S., Haimerl M., Stroszczynski C., Port M., Abend M. Gene expression changes in male and female rhesus macaque 60 days after irradiation. *PLoS One.* 2021;16(7):e0254344. doi 10.1371/journal.pone.0254344

Maniv I., Sarji M., Bdarneh A., Feldman A., Ankawa R., Koren E., Magid-Gold I., … Michaelson D., Van Leeuwen F.W., Verheijen B.M., Fuchs Y., Glickman M.H. Altered ubiquitin signaling induces Alzheimer's disease-like hallmarks in a three-dimensional human neural cell culture model. *Nat Commun.* 2023;14(1):5922. doi 10.1038/s41467-023-41545-7

McBride J.L., Neuringer M., Ferguson B., Kohama S.G., Tagge I.J., Zweig R.C., Renner L.M., … Sherman L.S., Domire J.S., Ducore R.M., Colgin L.M., Lewis A.D. Discovery of a CLN7 model of Batten disease in non-human primates. *Neurobiol Dis.* 2018;119:65-78. doi 10.1016/j.nbd.2018.07.013

Mishra S., Bernal C., Silvano M., Anand S., Ruiz i Altaba A. The protein secretion modulator TMED9 drives CNIH4/TGFα/GLI signaling opposing TMED3-WNT-TCF to promote colon cancer metastases. *Oncogene.* 2019;38(29):5817-5837. doi 10.1038/s41388-019-0845-z

Moshiri A., Chen R., Kim S., Harris R.A., Li Y., Raveendran M., Davis S., … Gopalakrishna K.N., Boyd K., Artemyev N.O., Rogers J., Thomasy S.M. A nonhuman primate model of inherited retinal disease. *J Clin Invest.* 2019;129(2):863-874. doi 10.1172/JCI123980

Nair H.B., Baker R., Owston M.A., Escalona R., Dick E.J., Vandeberg J.L., Nickisch K.J. An efficient model of human endometriosis by induced unopposed estrogenicity in baboons. *Oncotarget.* 2016;7(10):10857-10869. doi 10.18632/oncotarget.7516

Noriega N.C., Kohama S.G., Urbanski H.F. Microarray analysis of relative gene expression stability for selection of internal reference genes in the rhesus macaque brain. *BMC Mol Biol.* 2010;11:47. doi 10.1186/1471-2199-11-47

Paschalis E.P., Gamsjaeger S., Condon K., Klaushofer K., Burr D. Estrogen depletion alters mineralization regulation mechanisms in an ovariectomized monkey animal model. *Bone.* 2019;120:279-284. doi 10.1016/j.bone.2018.11.004

Patterson M.M., Jackson L.R., Brooks M.B., Catalfamo J.L. Type-3 von willebrand's disease in a rhesus monkey (*Macaca mulatta*). *Comp Med.* 2002;52(4):368-371

Peterson S.M., Mcgill T.J., Puthussery T., Stoddard J., Renner L., Lewis A.D., Colgin L.M.A., Gayet J., Wang X., Prongay K., Cullin C., Dozier B.L., Ferguson B., Neuringer M. Bardet-Biedl Syndrome in rhesus macaques: a nonhuman primate model of retinitis pigmentosa. *Exp Eye Res.* 2019;189:107825. doi 10.1016/j.exer.2019.107825

Peterson S.M., Watowich M.M., Renner L.M., Martin S., Offenberg E., Lea A., Montague M.J., Higham J.P., Snyder-Mackler N., Neuringer M., Ferguson B. Genetic variants in melanogenesis proteins TYRP1 and TYR are associated with the golden rhesus macaque phenotype. *G3* (*Bethesda*). 2023;13(10):jkad168. doi 10.1093/g3journal/jkad168

Poirier K., Hubert L., Viot G., Rio M., Billuart P., Besmond C., Bienvenu T. *CSNK2B* splice site mutations in patients cause intellectual

disability with or without myoclonic epilepsy. *Hum Mutat.* 2017; 38(8):932-941. doi 10.1002/humu.23270

Ramsköld D., Wang E.T., Burge C.B., Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol.* 2009;5(12):e1000598. doi 10.1371/journal.pcbi.1000598

Robinson A.A., Abraham C.R., Rosene D.L. Candidate molecular pathways of white matter vulnerability in the brain of normal aging rhesus monkeys. *GeroScience.* 2018;40(1):31-47. doi 10.1007/s11357-018-0006-2

Serrano-Lorenzo P., Rabasa M., Esteban J., Hidalgo Mayoral I., Domínguez-González C., Blanco-Echevarría A., Garrido-Moraga R., Lucia A., Blázquez A., Rubio J.C., Palma-Milla C., Arenas J., Martín M.A. Clinical, biochemical, and molecular characterization of two families with novel mutations in the *LDHA* gene (GSD XI). *Genes* (*Basel*). 2022;13(10):1835. doi 10.3390/genes13101835

Sherman L.S., Su W., Johnson A.L., Peterson S.M., Cullin C., Lavinder T., Ferguson B., Lewis A.D. A novel non-human primate model of Pelizaeus-Merzbacher disease. *Neurobiol Dis.* 2021;158:105465. doi 10.1016/j.nbd.2021.105465

Silver N., Cotroneo E., Proctor G., Osailan S., Paterson K.L., Carpenter G.H. Selection of housekeeping genes for gene expression studies in the adult rat submandibular gland under normal, inflamed, atrophic and regenerative states. *BMC Mol Biol.* 2008;9:64. doi 10.1186/1471-2199-9-64

Singh K.K., Krawczak M., Dawson W.W., Schmidtke J. Association of *HTRA1* and *ARMS2* gene variation with drusen formation in rhesus macaques. *Exp Eye Res.* 2009;88(3):479-482. doi 10.1016/j.exer.2008.10.019

Smith A.C., Mears A.J., Bunker R., Ahmed A., Mackenzie M., Schwartzentruber J.A., Beaulieu C.L., Ferretti E., Majewski J., Bulman D.E., Celik F.C., Boycott K.M., Graham G.E. Mutations in the enzyme glutathione peroxidase 4 cause Sedaghatian-type spondylometaphyseal dysplasia. *J Med Genet.* 2014;51(7):470-474. doi 10.1136/jmedgenet-2013-102218

Stevanin G., Fujigasaki H., Lebre A.S., Camuzat A., Jeannequin C., Dode C., Takahashi J., San C., Bellance R., Brice A., Durr A. Huntington's disease-like phenotype due to trinucleotide repeat expansions in the *TBP* and *JPH3* genes. *Brain.* 2003;126:1599-1603. doi 10.1093/brain/awg155

Sun Y., Yang X., Liu M., Tang H. B4GALT3 up-regulation by miR-27a contributes to the oncogenic activity in human cervical cancer cells. *Cancer Lett.* 2016;375(2):284-292. doi 10.1016/j.canlet.2016.03.016

Tanackovic G., Ransijn A., Thibault P., Abou Elela S., Klinck R., Berson E.L., Chabot B., Rivolta C. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum Mol Genet.* 2011;20: 2116-2130. doi 10.1093/hmg/ddr094

Tanaka M., Fujikawa R., Sekiguchi T., Hernandez J., Johnson O.T., Tanaka D., Kumafuji K., … Hattori K., Mashimo T., Kuwamura M., Gestwicki J.E., Kuramoto T. A missense mutation in the *Hspa8* gene encoding heat shock cognate protein 70 causes neuroaxonal dystrophy in rats. *Front Neurosci.* 2024;18:1263724. doi 10.3389/fnins.2024.1263724

Tu Z., Wang L., Xu M., Zhou X., Chen T., Sun F. Further understanding human disease genes by comparing with housekeeping genes and other genes. *BMC Genomics.* 2006;7:31. doi 10.1186/1471-2164-7-31

Tutar Y. Pseudogenes. *Comp Funct Genomics.* 2012;2012:424526. doi 10.1155/2012/424526

Ueda Y., Slabaugh T.L., Walker A.L., Ontiveros E.S., Sosa P.M., Reader R., Roberts J.A., Stern J.A. Heart rate and heart rate variability of rhesus macaques (*Macaca mulatta*) affected by left ventricular hypertrophy. *Front Vet Sci.* 2019;6:1. doi 10.3389/fvets.2019.00001

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3(7):research0034. doi 10.1186/gb-2002-3-7-research0034

Wang H., Wang Y., Tan P., Liu Y., Zhou S., Ma W. Prognostic value and anti-tumor immunity role of *TMED9* in pan-cancer: a bioinformatics study. *Transl Cancer Res.* 2024;13(10):5429-5445. doi 10.21037/tcr-24-258

Wei H., Zhang Y., Jia Y., Chen X., Niu T., Chatterjee A., He P., Hou G. Heat shock protein 90: biological functions, diseases, and therapeutic targets. *MedComm.* 2024;5(2):e470. doi 10.1002/mco2.470

Zhang D., Liu X., Xu X., Xu J., Yi Z., Shan B., Liu B. HPCAL1 promotes glioblastoma proliferation via activation of Wnt/β-catenin signalling pathway. *J Cell Mol Med.* 2019;23:3108-3117. doi 10.1111/jcmm.14083

Zhang Y.B., Zheng S.F., Ma L.J., Lin P., Shang-Guan H.C., Lin Y.X., Kang D.Z., Yao P.S. Elevated hexose-6-phosphate dehydrogenase regulated by OSMR-AS1/hsa-miR-516b-5p axis correlates with poor prognosis and dendritic cells infiltration of glioblastoma. *Brain Sci.* 2022;12(8):1012. doi 10.3390/brainsci12081012

Zühlke C., Hellenbroich Y., Dalski A., Kononowa N., Hagenah J., Vieregge P., Riess O., Klein C., Schwinger E. Different types of repeat expansion in the TATA-binding protein gene are associated with a new form of inherited ataxia. *Eur J Hum Genet.* 2001;9:160-164. doi 10.1038/sj.ejhg.5200617

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

Received July 21, 2025. Revised October 19, 2025. Accepted October 20, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1325-1337

doi 10.18699/vjgb-25-139

Genes representing the stress-dependent component in arterial hypertension development

D.Yu. Oshchepkov [1] ✉, Yu.V. Makovka [1], I.V. Chadaeva [1, 2], A.G. Bogomolov [1], L.A. Fedoseeva [1], A.A. Seryapina [1], M.P. Ponomarenko [1], A.L. Markel [1, 2], O.E. Redina [1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
✉ diman@bionet.nsc.ru

Abstract. Hypertension is among the major risk factors of many cardiovascular diseases. Chronic psychoemotional stress is one of its key causes. Studies of molecular mechanisms of human hypertension development are conducted in animals, including artificial rat strains that model various forms of the disease. The RatDEGdb database, used in our work, includes 144 hypothalamic genes that represent the common response to single short-term restraint stress in hypertensive ISIAH and normotensive WAG rats. These rat genes were annotated with changes in the expression of the human orthologs using data on 17,458 differentially expressed genes (DEGs) from patients with hypertension compared to normotensive subjects. We applied principal component analysis to orthologous pairs of DEGs identified in hypertensive patients and rat hypothalamic DEGs upon single short-term restraint stress. Two principal components, corresponding to a linear combination of log2 expression changes associated with the similarity (PC1) and difference (PC2) in the response to psychoemotional stress in two rat strains, on the one hand, and different forms of human hypertension, on the other, explained 64 % and 33 % of the variance in differential gene expression, respectively. The significant correlation revealed between PC1 and PC2 values for the group of DEGs with stress-induced downregulation indicates that psychoemotional stress and hypertension share a common molecular mechanism. Functional annotation suggests that stress-induced downregulation of genes involved in the plasma membrane function and, simultaneously, interactions with the extracellular matrix is the most likely contribution of psychoemotional stress to the development of the hypertensive status in patients, and the SMARCA4 transcription factor is the most likely mediator in the epigenetic modification affecting gene expression under chronic stress. Peripheral blood markers for the diagnosis of psychoemotional stress are proposed.

Key words: rat; human; differentially expressed gene (DEG); arterial hypertension; stress; biomedical model; principal component analysis

For citation: Oshchepkov D.Yu., Makovka Yu.V., Chadaeva I.V., Bogomolov A.G., Fedoseeva L.A., Seryapina A.A., Ponomarenko M.P., Markel A.L., Redina O.E. Genes representing the stress-dependent component in arterial hypertension development. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(7):1325-1337. doi 10.18699/vjgb-25-139

Funding. The work was supported by State Budgeted Project FWNR-2022-0020.

Acknowledgements. We are grateful to the Bioinformatics Shared Access Center for access to computational resources and to the Shared Access Center for Gene Pools of Laboratory Animals of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, for the propagation and maintenance of the rats studied.

Гены, представляющие стресс-зависимую компоненту при развитии артериальной гипертонии

Д.Ю. Ощепков [1] ✉, Ю.В. Маковка [1], И.В. Чадаева [1, 2], А.Г. Богомолов [1], Л.А. Федосеева [1], А.А. Серяпина [1], М.П. Пономаренко [1], А.Л. Маркель [1, 2], О.Е. Редина [1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ diman@bionet.nsc.ru

Аннотация. Гипертония считается ведущим фактором риска развития многих сердечно-сосудистых заболеваний. Одним из ключевых факторов, способствующих развитию гипертонии, является хронический психоэмоциональный стресс. Изучение молекулярно-генетических механизмов развития гипертонии человека проводят на животных, в том числе на специально созданных инбредных линиях крыс, моделирующих различные формы гипертонии человека. В настоящей работе использованы данные из базы данных RatDEGdb о 144 генах гипоталамуса, которые пред-

ставляют общий ответ на рестрикционный стресс у гипертензивных крыс НИСАГ и нормотензивных крыс WAG. Эти гены крыс были аннотированы изменениями экспрессии ортологичных им генов человека с использованием данных о 17 458 дифференциально экспрессирующихся генах (ДЭГ) пациентов с артериальной гипертензией по сравнению с нормотензивными пациентами. Для выявленных пар ортологов между ДЭГ гипоталамуса крысы после рестрикционного стресса и пациентов с артериальной гипертензией применили анализ главных компонент. Две главные компоненты, соответствующие линейной комбинации значений log2 изменений экспрессии, связанные со сходством (PC1) и различием (PC2) ответа на психоэмоциональный стресс двух линий крыс, с одной стороны, и разными формами гипертонии человека, с другой, объясняли соответственно 64 и 33 % дисперсии дифференциальной экспрессии генов. Выявленная значимая корреляция между значениями PC1 и PC2 для группы ДЭГ со стресс-индуцированным снижением экспрессии указывает на существование общего молекулярного механизма между психоэмоциональным стрессом и гипертонией. Их функциональная аннотация позволила предположить, что стресс-индуцированное снижение экспрессии генов, участвующих в функционировании плазматической мембраны и одновременно во взаимодействии с межклеточным пространством, является наиболее вероятным вкладом психоэмоционального стресса в формирование гипертензивного статуса пациентов, а транскрипционный фактор SMARCA4 – наиболее вероятным участником эпигенетической модификации экспрессии генов в результате хронического стресса. Также предложены маркеры периферической крови для диагностики психоэмоционального стресса.

Ключевые слова: крыса; человек; дифференциально экспрессирующийся ген (ДЭГ); артериальная гипертония; стресс; биомедицинская модель; метод главных компонент

Introduction

Arterial hypertension is a multifactorial disease. Its development is caused by both hereditary and environmental factors. It is contributed by chronic psychoemotional stress, which can be induced by occupational factors, social isolation, low socioeconomic status, anxiety, distress, and other lifestyle factors (Liu M.Y. et al., 2017).

Chronic stress factors hyperactivate the sympathetic system. This process is accompanied by neuroinflammation and mitochondrial dysfunction. In the brain, it may result in the accumulation of reactive oxygen species (ROS), which exceed the neuronal antioxidant capacity (Lambert A.J., Brand, 2009; Hovatta et al., 2010; Picard, McEwen, 2018). Oxidative stress triggers the cell membrane damage cascade, enhances lipid peroxidation, and impairs neural conduction (Che et al., 2015; Montezano et al., 2015). Neuron damage affects the activity of neuronal circuits; in particular, it reduces the GABAergic inhibitory activity and results in the predominance of glutamatergic exciting signals in sympathetic nuclei. This enhances impulse transmission in the cardiovascular system and causes persisting blood pressure (BP) increase (Lambert E.A., Lambert G.W., 2011; Hering et al., 2015). The emerging closed cycle of neurogenic and oxidative stresses promotes the formation of a pathological vegetative pattern, in which even minor stressful events cause pronounced BP increase (Fontes et al., 2023).

Thus, not only hereditary predisposition can cause hypertension. Some patients develop hypertension due to a combination of other causes. Their multitude hampers the understanding of underlying molecular processes and determination of causes that are associated with hereditary predisposition. Hereditary causes may include high stress sensitivity; other processes may be triggered by, e. g., elevated salt sensitivity or related to endothelial or endocrine dysfunction. Knowledge of these processes may be helpful in seeking approaches to the pathogenetic therapy of hypertensive disease. Our study is focused on this issue, in particular, the identification of the stress-associated component of hypertension development.

Strains of hypertensive rats characterized by a broad range of pathophysiological changes in the cardiovascular system are often used in studies of the molecular mechanisms of hypertension. Each of such strains models a certain form of arterial hypertension. One of them is ISIAH (Inherited Stress-Induced Arterial Hypertension), prone to stress-induced arterial hypertension. It models spontaneously developing hypertension, marked by severe response (BP increase) to psychoemotional stress (Markel, 1992; Markel et al., 1999, 2007).

Stress adaptation in cells involves a significant remodeling of gene expression programs (de Nadal et al., 2011). However, the molecular mechanisms underlying stress adaptation are still poorly understood. Formerly, we showed that the transcription levels of many genes in the hypothalamus of hypertensive ISIAH and normotensive WAG/GSto-Icgn (Wistar Albino Glaxo, hereinafter WAG) rats changed when the animals were exposed to short-term (2 h) restraint stress in tight wire-mesh cages. These changes may affect a great number of biologic processes and metabolic pathways (Oshchepkov et al., 2024). The focus on hypothalamic genes is due to the fact that this brain region is among the key ones regulating the neuroendocrine response to stress. It integrates the central and peripheral components involved in blood pressure regulation and arterial hypertension development by controlling glucocorticoid secretion (Carmichael, Wainford, 2015; Burford et al., 2017; Kinsman et al., 2017; Fontes et al., 2023). Actually, it is a key link between stress and hypertension development. The studied stress model induced a significant BP increase in ISIAH but not in WAG rats, although blood corticosterone levels significantly increased in both rat strains; thus, both hypertensive and normotensive animals responded to stress (Oshchepkov et al., 2024). Therefore, we presume that the differentially expressed genes (DEGs) associated with the

common (hypertension genotype-independent) response of the rat hypothalamus on restraint stress (considered psychoemotional) are involved in the stress-dependent component.

In human daily routine, stress can be induced by many factors to activate the sympathetic nervous system and cause hypertension, as well as other cardiovascular diseases. As we mentioned, the hypothalamus takes a significant part in these processes. The study of molecular mechanisms of hypothalamic response to stress in humans is difficult, and there is no relevant information on gene expression in brain regions in hypertensive patients. Nevertheless, PubMed presents commonly available independent sets of experimental data on gene expression in other organs and tissues, including peripheral blood, in hypertensive patients and in biomedical cellular models of hypertension (Oshchepkov et al., 2022; Shikhevich et al., 2023). Here we employ these data to reveal human genes orthologous to rat genes associated with the response to restraint (psychoemotional) stress and to identify genes forming the stress-dependent component in hypertension development in the human.

Materials and methods

Experimental animals. Experiments were conducted with 3-month old male rats of the hypertensive ISIAH and normotensive WAG strains in the conventional vivarium of the Center for Experimental Animal Genetic Resources, Institute of Cytology and Genetics, Novosibirsk, Russia. The animals were kept under standard conditions at the light schedule 12:12. Water and standard diet were given ad libitum.

Transcriptome analysis in the hypothalamus by the RNA-Seq method was done in four groups of seven animals each: (1) ISIAH_control, (2) WAG_control, (3) ISIAH_stress, and (4) WAG_stress. Basal systolic arterial BP was measured by the tail-cuff method (Markel et al., 2007). Rats were semi-narcotized with ether to avoid emotional stress during the measurement. The experimental rats in ISIAH_stress and WAG_stress groups were exposed to restraint (emotional) stress seven days after BP measurement. In this procedure, an animal was placed into a tight wire-mesh cage for 2 h; for details, see (Oshchepkov et al., 2024).

All international guidelines for the care and use of laboratory animals were followed. Animal protocols were approved by the ICG Bioethics Committee, Novosibirsk, Russia, protocol No. 115 of December 20, 2021.

RNA-Seq. RNA was isolated at the Institute of Genomic Analysis, Moscow, Russia. Hypothalamus sample preparation and transcriptome sequencing were conducted at BGI Hongkong Tech Solution NGS Lab. following manufacturer's recommendations (MGI Tech Co., Ltd., China). Paired-end sequencing of cDNA libraries was performed by DNBSEQ Technology with read length 150 bp and sequencing depth over 30,000,000 uniquely mapped reads. All samples were analyzed as biological replicates.

The sequencing results were preprocessed with FastQC software version 0.11.5 (Andrews, 2010) to check the quality. The overall number of nucleotide reads in the libraries after the preprocessing was 1,287,393,367; of them, 1,267,436,623 (98.45 %) were mapped on the reference rat genome mRatBN7.2/rn7 (rn7 assembly, Wellcome Sanger Institute Nov, 2020) with STAR 2.7.10a software (Dobin et al., 2013).

The mapping data were statistically processed to calculate differential expression of genes in the R environment for statistical computing. We applied surrogate variable analysis SVA (Leek et al., 2012) to take into account undesirable variations in data caused by inadvertent systematic deviations during sample preparation. Prior to SVA, the expression data were normalized and transformed with the vst function in DESeq2 v1.30.1 (Love et al., 2014) according to online documentation. Relevant surrogate variables were then included as factors in the differential expression analysis with DESeq2. The differential expression analysis was conducted separately for each pair of groups.

Differential expression was calculated for all genes exhibiting significant expression levels above the threshold: sum of gene coverages in all libraries over 10 reads. The significance level for DEG detection was chosen with the consideration of the correction for multiple comparisons. It corresponded to adjusted p-value < 5 % and Log2 fold change $\geq |0.585|$ (1.5 fold). The information on the revealed DEGs had been described in (Makovka et al., 2024; Oshchepkov et al., 2024) and uploaded to the RatDEGdb database (Chadaeva et al., 2023). The data are presented as transcription upon stress normalized to transcription at rest.

Choice of human genes whose expression changes in hypertension. We used generally accessible sets of experimental data on patients with hypertensive disease and data on cellular hypertension models available from PubMed (Lu, 2011). The sample included only data reported as statistically significant according to Fisher's Z test with correction for multiple comparisons ($P_{ADJ} < 0.05$). We selected 16 publications with data on 17,458 genes differentially expressed in tissues and cells of hypertensive and normotensive subjects. This list of DEGs presents 16 tissues and 15 hypertension forms (Table 1). The threshold for a significant change in transcription level was set to be 1.5 times. The data are presented as the transcription level in hypertensive patients normalized to normotensive subjects.

Bioinformatical analysis of orthologous genes in the rat and human. Orthologs were sought with the Gene and Ortholog Location Finder (GOLF) module in the Rat Genome database (https://rgd.mcw.edu/rgdweb/ortholog/start.html) (Vedi et al., 2023). Pairwise combinations of log2 changes in the expression of orthologous genes were analyzed by the principal component method with the conventional Past v.4.04 software (Hammer et al., 2001). Two values corresponded to log2 expression changes for ISIAH and WAG rats and one for the human orthologous gene. The same software was applied to the calculation of factor loadings and the statistical significance of the explained portion of component variance from 1,000 bootstrap samples (Efron et al., 1996). Pearson correlation analysis was conducted with the two-tailed test of significance. Groups of

Table 1. DEGs in hypertensive patients and biomedical cell models of hypertension found in PubMed

#	Hypertension form	Datum	Tissue	N_{DEG}	References
1	Essential hypertension	Norm	Renal cortex (♂,♀)	46	Marques et al., 2011
2			Renal cortex (♂)	6	
3			Renal cortex (♀)	10	
4	Pulmonary artery hypertension (PAH)	Norm	Lung biopsy	5,685	Yao et al., 2021
5			Lung biopsy	49	Qiu et al., 2021
6	Pulmonary fibrosis associated with hypertension	Norm	Lung biopsy	6,936	Yao et al., 2021
7	Lung transplantation for PAH	Norm	Lung	5	Mura et al., 2019
8	Hypertensive nephrosclerosis	Norm	Renal biopsy	16	Neusser et al., 2010
9	Cerebrorenal hypertension	Norm	Renal medulla	13	Wu et al., 2013
10	Preeclampsia	Norm	Venous blood	64	Textoris et al., 2013
11			Decidua	372	Yong et al., 2015
12			Amniotic fluid	10	Jung et al., 2019
13			Placenta	1,228	Saei et al., 2021
14	Hypertensive heart failure	Norm	Peripheral blood	248	Maciejak et al., 2015
15	Hypertensive myocardial infarction	Norm	Peripheral blood	75	Maciejak et al., 2015
16	Hypertensive coronary artery disease	Norm	Peripheral blood	1,524	Zheng, He, 2021
17	Hypertensive atrial fibrillation	Norm	Aortic biopsy	300	Zheng, He, 2021
18	Hypertensive squamous cell carcinoma of the lung	Norm	Squamous cell carcinoma cells	119	Koper et al., 2017
19	Sedentary teenagers with obesity and hypertension	Athletes	General blood analysis	250	Plaza-Florido et al., 2021
20	mir-201 excess as a preeclampsia model	Normal Swan71 cells	Swan71 human trophoblast cell line	19	Ahn et al., 2017
21	*BMPR2* knockdown as a PAH model	Normal endothelial cells	Pulmonary artery endothelial cells	483	Awad et al., 2016
Σ	15 hypertension forms	4 norms	16 tissues	17,458	16 publications

Note. N_{DEG}, number of differentially expressed genes (DEGs); Σ, total.

gene pairs for which the first principal component PC1 was above zero and groups for which it was below zero were analyzed separately.

Functional annotation of DEGs. Functional enrichment networks were constructed with the STRING database (Szklarczyk et al., 2023). The functional annotation of DEGs was done with Enrichr-KG (Evangelista et al., 2023). Analysis of the enrichment of DEG promoter regions with transcription factor binding sites was done with Enrichr (Xie et al., 2021).

Results

We studied genes that change their transcription levels by more than 1.5 fold in the hypothalamus of hypertensive ISIAH and normotensive WAG rats upon short-term (2 h) restraint stress. There were 257 and 229 such genes, respectively. Of them, 144 DEGs produced common responses, 113 DEGs showed significant expression changes only in the ISIAH hypothalamus, and 85 genes significantly change their expression only in the WAG hypothalamus (Oshchepkov et al., 2024). To reveal genes that might form the stress-sensitive component in human hypertension development, we sought orthologous human genes differently expressed in subjects with hypertension and normal arterial blood pressure.

Analysis of the common response

The search for orthologs of the 144 rat DEGs in the GOLF Rat Genome Database and their comparison with the list of

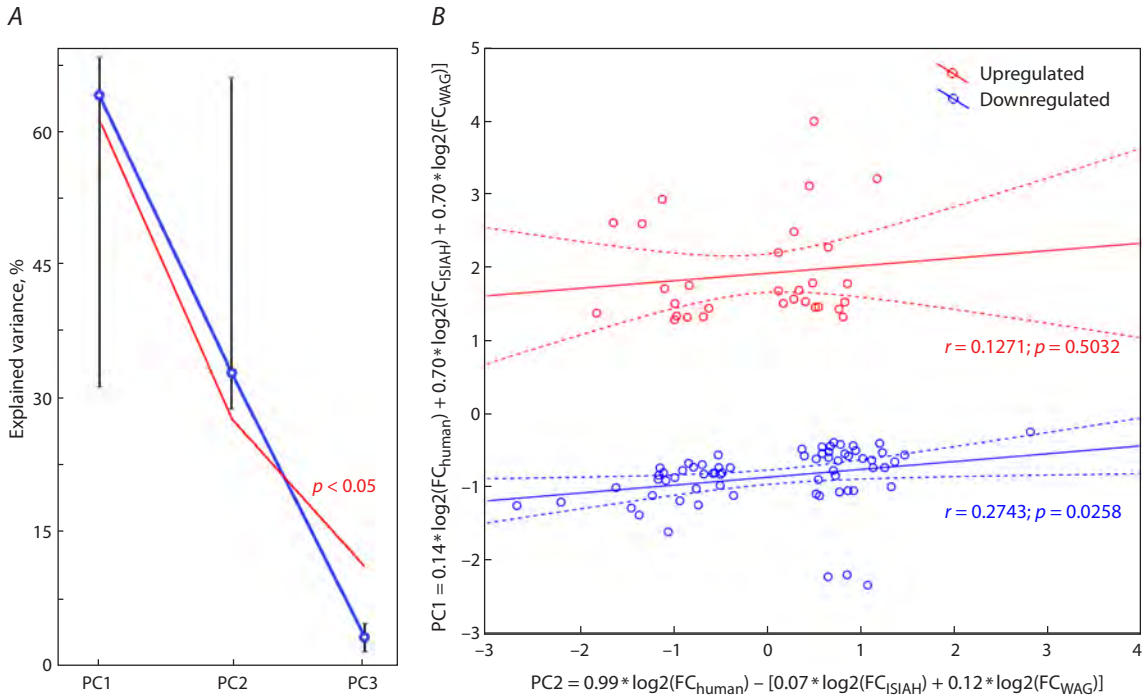


Fig. 1. Factor analysis of differentially expressed genes (DEGs) in hypertensive subjects (Table 1) and stressed rats.

A, choice of statistically significant ($p < 0.05$) principal components PC1 and PC2 in the factor analysis of DEGs in hypertensive subjects and stressed rats. Designations: ○, mean; whiskers, standard error of the mean; red line, lower boundary of the 95 % confidence range of the statistical significance of the explained variance. *B*, principal component and regression analysis. Designations: PC1 and PC2, the first (Y-axis) and second (X-axis) principal components as the corresponding linear combinations of original variables with regard to the calculated factor loadings. Red marks (Upregulated): group of DEGs with stress-induced upregulation; blue marks (Downregulated): group of DEGs with stress-induced downregulation according to the PC1 estimate. Dashed lines: boundaries of the 95 % confidence range for the regression line (solid). Letters *r* and *p*: linear correlation coefficients and their statistical significance, respectively, assessed with conventional Statistica (Statsoft™, USA).

human DEGs found in PubMed (see Materials and methods) revealed 96 orthologous pairs. This set included data on six tissues of nine hypertension forms from seven publications (Supplementary Table S1, 96 common DEGs)[1].

In succession to our earlier papers on the factor analysis of DEGs (Chadaeva et al., 2021; Shikhevich et al., 2023), we processed data on the changes in the expression of rat genes upon stress regardless of the hypertensive genotype and data on orthologous genes of hypertensive humans by the principal component method. As seen in Figure 1*A*, the 96 pairs of human genes and their rat orthologs are characterized by two principal components: PC1 and PC2. They explain 64 and 33 % of the variance, and both are statistically significant according to 1,000 bootstrap samples (Efron et al., 1996).

The results of the principal component analysis are shown in Figure 1*B*. We see that PC1, plotted along the Y-axis, is proportional to a linear combination of the original variables with regard to the calculated factor loadings. The combination summarizes the log2 values of changes measured in independent experiments with hypertensive subjects and log2 values of changes in rats of the two strains exposed to a single restraint stress. Thus, this component may characterize the similarity between the responses of two rat strains on psychoemotional stress, on the one side, and different forms of human hypertension, on the other side.

The second principal component PC2, plotted along the X-axis, is proportional to a linear combination of the original variables with regard to the calculated factor loadings. The combination corresponds to the positive contribution of values obtained in independent experiments on humans and negative contribution of stress response values in the two rat strains. It may characterize differences between the species and features of particular impairments in hypertensive human patients. Figure 1*B* clearly shows that all DEGs studied are divided into two disjoint groups with respect to PC1 = 0. Red and blue colors mark genes that respond to stress by expression increase (Upregulated) and decrease (Downregulated), respectively. Regression analysis demonstrates a significant ($p < 0.05$) correlation between PC1 and PC2 in a group of 66 pairs of 52 DEGs in the Downregulated cluster (Fig. 1*B*, blue marks). This correlation may be indicative of a common molecular mechanism of gene suppression in the formation of stress response in the rat and a similar process in the human, which leads to hypertension.

The corresponding analysis of the group of DEGs with stress-induced upregulation (Upregulated) shows no significant correlation. Thus, the mechanisms activating this group of genes may be different in different tissues and disorder forms.

[1] Supplementary Table S1 is available at:
https://vavilovj-icg.ru/download/pict-2025-29/appx54.xlsx

Functional annotation of DEGs associated with stress-induced expression downregulation

To understand the molecular mechanism downregulating genes in stress response formation common to the human and rat, we performed functional annotation of 52 DEGs associated with stress-induced downregulation (the Downregulated cluster in Fig. 1*B*). Eight genes associated with hypertension were found among the 52 genes of the Downregulated cluster: *ALOX12*, *ATP2B4*, *CX3CR1*, *GRK3*, *KDR*, *NOS1*, *RASGRP3*, and *SMAD9*. Five of them (*ALOX12*, *GRK3*, *KDR*, *NOS1*, and *RASGRP3*) showed unidirectional changes in the rat and human in all cases (Table S1, 96 common DEGs).

Analysis of DEG enrichment in the STRING database. Analysis of the group of 52 DEGs in STRING revealed the most enriched terms: Postsynapse, Cell periphery, and Plasma membrane (Fig. 2). Thus, the process bringing these DEGs together is associated with extracellular matrix and the membrane, directly involved in its interaction with the cell.

Analysis with the Enrichr-KG resource. We analyzed biologic processes and metabolic pathways with the Enrichr-KG resource (Table 2). Signal Transduction was the most enriched term. It included 17 DEGs. This group included several hypertension-associated genes. The most enriched metabolic pathways were Adrenergic signaling in cardiomyocytes, Salivary secretion, Aldosterone synthesis and secretion, Glucagon signaling pathway, Insulin resistance, Differentiation of white and brown adipocyte, Constitutive Androstane Receptor Pathway, and Neovascularisation processes. They involve several DEGs under consideration associated with hypertension (Table 2). Some DEGs associated with the biological processes and metabolic pathways listed in Table 2 were found in peripheral blood.

DEGs in peripheral blood in the Downregulated cluster

We found 14 DEGs belonging to the Downregulated cluster in the peripheral blood of patients (Table 3). These DEGs are promising as candidate peripheral markers associated with psychoemotional stress.

Analysis of transcription factors involved in the regulation of DEGs of the Downregulated cluster

Analysis of DEGs of the Downregulated cluster with the Enrichr resource revealed transcription factors that may be involved in the regulation of these DEGs in the rat hypothalamus and in tissues of hypertensive patients (Table 4). The most enriched terms in the ChEA_2022 library are associated with transcription factors SMARCA4 (SWI/SNF related BAF chromatin remodeling complex subunit ATPase 4) and glucocorticoid receptor NR3C1 (nuclear receptor subfamily 3 group C member 1). Receptor NR3C1 and six transcription factors (PPARG, ESR1, AR, NFE2L2, BRD4, and STAT3) presented in Table 4 are associated with hypertension.

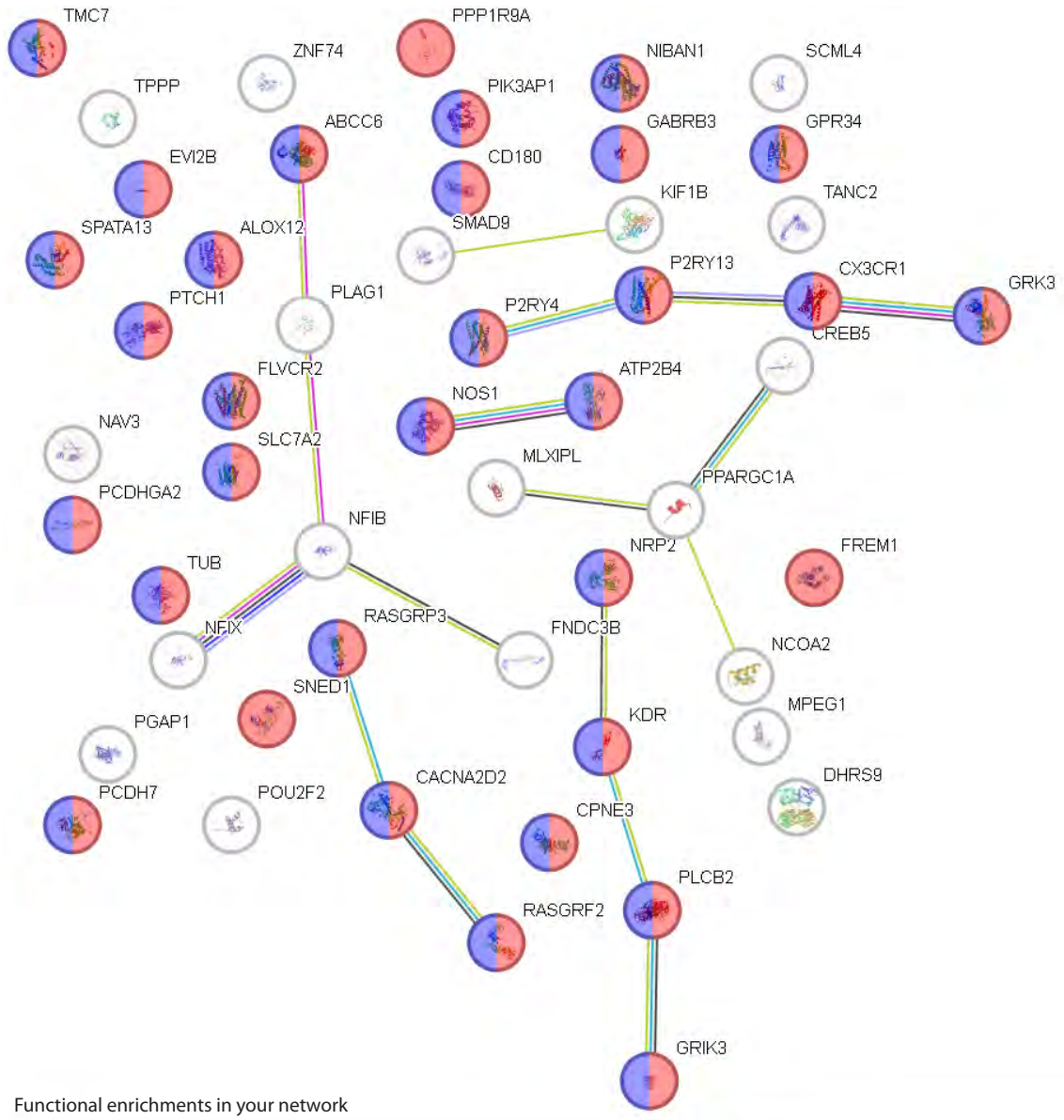
Discussion

The hypothesis of an association between chronic psychosocial stress and hypertension was put forward long ago (Spruill, 2010). It has been confirmed in several studies (Liu M.Y. et al., 2017; Bautista et al., 2019), but the molecular mechanisms responsible for this association remain obscure.

Here we considered 144 genes for common stress response in the hypothalamus of hypertensive and normotensive rats and compared this response with changes in gene transcription levels in different organs and tissues in patients with different hypertension forms. It follows from our results that the transcription changes observed may point to some common features in the restructuring of the gene expression pattern, which reflect the contribution of psychoemotional stress to the manifestation of the hypertensive condition.

We found human DEGs orthologous to 144 genes that change their transcription in response to single short-term restraint stress in the hypothalamus of hypertensive ISIAH and normotensive WAG rats. Analysis by the principal component method revealed two DEG clusters and demonstrated a significant correlation between PC1 and PC2 values for the DEG cluster with stress-induced downregulation. This correlation suggests the existence of a molecular mechanism suppressing gene expression in the formation of stress response shared by the human and rat. It is in good agreement with the notion that adaptation to stress is accompanied by remodeling of the functional transcriptome and arrest of translational processes (Advani, Ivanov, 2019; Baymiller, Moon, 2023). It should be mentioned that transcription up- or downregulation under stress can be both compensatory and pathogenetic, as can be elucidated by functional analysis of a particular gene. Gene downregulation can stimulate many processes, contributing to the activation of the sympathetic nervous system and enhancing oxidative stress in tissues. It may also participate in neuroprotection, as we discussed earlier in connection with some DEGs (Makovka et al., 2024; Oshchepkov et al., 2024).

The enrichment in the linked terms Plasma membrane and Cell periphery found in the group with stress-induced downregulation encompasses 30 and 33 genes of 52, respectively. It is consistent with the idea that the change in the composition of extracellular matrix associated with the release of corticosterone and catecholamines resulting from the activation of the hypothalamic–pituitary–adrenal (HPA) axis and the sympathetic nervous system is the most common feature in the model of human chronic stress. This change triggers various signaling pathways, including free radical generation, apoptosis, elevated glutamate release in synapses, and production/release of cytotoxic cytokines (Bali et al., 2013). These intercellular signals, in turn, may affect gene expression, leading to structural and functional changes in plasma membranes (Bali et al., 2013), such as changes in the lipid profile of platelet membranes (Bikulciene et al., 2024) and to neuron excitability due to modifications in membrane structure and composition (Rosenkranz et al., 2010; Matovic et al., 2020), including the postsynaptic membrane, as indicated by the enrichment of the term Postsynapse in our study. Analysis of biological processes and metabolic pathways in the DEG cluster with stress-induced downregulation also emphasizes the role of signal transduction and postsynaptic signal transmission (Table 2).



Functional enrichments in your network

Cellular Component (Gene Ontology)					
GO-term	*description*	*count in network*	*strenght*	*signal*	*false discovery rate*
GO:0098794	Postsynapse	9 of 621	0.74	0.42	0.0238
GO:0071944	Cell periphery	33 of 6015	0.32	0.37	0.0021
GO:0005886	Plasma membrane	30 of 5544	0.31	0.33	0.0083

Fig. 2. Screenshot of the annotation of the Downregulated cluster in Fig. 1*B* with the STRING database (Szklarczyk et al., 2023). Designations follow the STRING annotation.

Stress can also impair the expression of genes for autophagy marker proteins and disturb the concentrations of lysosomal proteins and enzymes (Ulecia-Moron et al., 2025). Generally speaking, the considerable number of genes downregulated in response to stress that are common in different organs and tissues characterizes the scale of changes affecting the plasma membrane in response to stress-induced release of active signaling molecules into the extracellular matrix.

The DEG cluster with stress-induced downregulation was found to include eight hypertension-associated genes. Their expression changes may influence biological processes and metabolic pathways involved in blood pressure regulation: Adrenergic signaling in cardiomyocytes (Maltsev et al., 2019), Aldosterone synthesis and secretion (Ferreira et al., 2021), and Lipid oxidation (Leong, 2021).

Our study revealed genes encoding transcription factors that may be involved in the regulation of genes of the DEG cluster with stress-induced downregulation. We suppose that SMARCA4 (SWI/SNF-related BAF chromatin remodeling complex subunit ATPase 4) and NR3C1 (nuclear receptor

Table 2. Analysis of biological processes and metabolic pathways of the Downregulated cluster

Library	Term	q-value	Genes
Reactome 2022	Signal Transduction R-HSA-162582	0.0206	*GABRB3*, *CX3CR1*#, *NCOA2*, *P2RY13*, *NRP2*, *PTCH1**, *RASGRF2**, *PCDH7*, *SMAD9*#, *RASGRP3*#, *GRK3*#, *P2RY4*, *DHRS9*, *SPATA13*, *KDR*#, *PIK3AP1**, *PLCB2*
	P2Y Receptors R-HSA-417957	0.0295	*P2RY13*, *P2RY4*
	Transmission Across Chemical Synapses R-HSA-112315	0.0295	*GABRB3*, *RASGRF2**, *GRIK3*, *CACNA2D2*, *PLCB2*
	Neurotransmitter Receptors And Postsynaptic Signal Transmission R-HSA-112314	0.0428	*GABRB3*, *RASGRF2**, *GRIK3*, *PLCB2*
KEGG 2021 Human	Adrenergic signaling in cardiomyocytes	0.0510	*ATP2B4*#, *CACNA2D2*, *PLCB2*, *CREB5*
	Salivary secretion	0.0510	*ATP2B4*#, *NOS1*#, *PLCB2*
	Aldosterone synthesis and secretion	0.0510	*ATP2B4*#, *PLCB2*, *CREB5*
	Glucagon signaling pathway	0.0510	*PPARGC1A*, *PLCB2*, *CREB5*
	Insulin resistance	0.0510	*MLXIPL**, *PPARGC1A*, *CREB5*
WikiPathway 2021 Human	Differentiation of white and brown adipocyte WP2895	0.0877	*SMAD9*#, *PPARGC1A*
	Constitutive Androstane Receptor Pathway WP2875	0.0877	*NCOA2*, *PPARGC1A*
	Neovascularisation processes WP4331	0.0877	*KDR*#, *SMAD9*#
GO Biological Process 2021	Positive regulation of transcription by RNA polymerase II (GO:0045944)	0.1140	*NCOA2*, *MLXIPL**, *NFIX**, *NFIB*, *PLAG1**, *SMAD9*#, *NOS1*#, *POU2F2**, *PPARGC1A*
	Lipid oxidation (GO:0034440)	0.1140	*ALOX12*#, *PPARGC1A*
	Positive regulation of purine nucleotide biosynthetic process (GO:1900373)	0.1140	*NOS1*#, *PPARGC1A*
	Cellular response to growth factor stimulus (GO:0071363)	0.1140	*KDR*#, *SMAD9*#, *CPNE3*, *NOS1*#
	Positive regulation of fatty acid metabolic process (GO:0045923)	0.1140	*MLXIPL**, *PPARGC1A*

* DEGs found in peripheral blood, # genes associated with hypertension.

Table 3. Peripheral blood markers in the Downregulated cluster

Hypertension form	Human gene	Human log2FC HT/norm	Rat ortholog	log2FC Stress/rest		Gene name	References
				ISIAH	WAG		
CAD	*DHRS9*	1.35	*Dhrs9*	–1.60	–1.40	Dehydrogenase/reductase 9	Zheng, He, 2021
	MLXIPL	1.39	*Mlxipl*	–0.46	–0.61	MLX interacting protein like	
	NFIX	1.20	*Nfix*	–0.60	–0.25	Nuclear factor I X	
	NRP2	1.08	*Nrp2*	–0.59	–0.40	Neuropilin 2	
	PGAP1	–1.41	*Pgap1*	–0.62	–0.31	Post-GPI attachment to proteins 1	
	PIK3AP1	1.27	*Pik3ap1*	–0.63	–0.95	Phosphoinositide-3-kinase adaptor protein 1	
	PLAG1	–1.70	*Plag1*	–0.62	–0.49	PLAG1 zinc finger	
	POU2F2	–1.36	*Pou2f2*	–1.00	–0.56	POU class 2 homeobox 2	
	PTCH1	–1.70	*Ptch1*	–0.64	–0.82	Patched 1	
	RASGRF2	–1.33	*Rasgrf2*	–0.72	–0.28	Ras protein specific guanine nucleotide releasing factor 2	
	SCML4	–1.20	*Scml4*	–0.36	–0.61	Sex comb on midleg-like 4 (Drosophila)	
MI	*GPR34*	1.355	*Gpr34*	–0.56	–1.03	G protein-coupled receptor 34	Maciejak et al., 2015
HF	*FLVCR2*	1.797	*Flvcr2*	–0.70	–0.59	Feline leukemia virus subgroup C cellular receptor family, member	
	GPR34	2.026	*Gpr34*	–0.56	–1.03	G protein-coupled receptor 34	

Note. CAD, hypertension-induced coronary artery disease; MI, hypertension-induced myocardial infarction; HF, hypertension-induced heart failure; HT/norm, transcription level in hypertensive subjects normalized to the transcription level in subjects with normal arterial blood pressure.

Table 4. Analysis of transcription factors involved in the regulation of DEGs of the Downregulated cluster

ChEA_2022 term	TF	Number of DEGs	Adjusted *p*-value	DEGs of the Downregulated cluster
SMARCA4 23332759 ChIP-Seq OLIGODENDROCYTES Mouse	SMARCA4	20	0.000068	*CX3CR1, NCOA2, NRP2, PGAP1, PCDH7, CD180, PPP1R9A, TANC2, SLC7A2, RASGRP3, NAV3, NFIB, SPATA13, FLVCR2, KIF1B, TPPP, PIK3AP1, FREM1, PPARGC1A, CREB5*
NR3C1 27076634 ChIP-Seq BEAS2B Human Lung Inflammation	NR3C1#	21	0.00165	*NRP2, NFIX, PCDH7, CD180, FNDC3B, ATP2B4, CACNA2D2, PPP1R9A, POU2F2, TANC2, SLC7A2, NAV3, DHRS9, NFIB, KDR, FLVCR2, KIF1B, TPPP, PIK3AP1, PPARGC1A, CREB5*
OLIG2 23332759 ChIP-Seq OLIGODENDROCYTES Mouse	OLIG2	15	0.00172	*NCOA2, PGAP1, PCDH7, PTCH1, CD180, FNDC3B, PPP1R9A, TANC2, RASGRP3, NAV3, SPATA13, KIF1B, PIK3AP1, FREM1, CREB5*
SOX2 21211035 ChIP-Seq LN229 Gbm	SOX2	19	0.00205	*CX3CR1, TUB, NCOA2, NRP2, PGAP1, RASGRF2, PCDH7, CD180, ABCC6, GRIK3, FNDC3B, ATP2B4, PPP1R9A, SLC7A2, RASGRP3, NAV3, NFIB, PPARGC1A, CREB5*
MTF2 20144788 ChIP-Seq MESCs Mouse	MTF2	18	0.00213	*TUB, SNED1, NRP2, NFIX, RASGRF2, PCDH7, PTCH1, GRIK3, FNDC3B, CACNA2D2, SMAD9, TMC7, SLC7A2, MLXIPL, NFIB, KDR, TPPP, PIK3AP1*
PPARG 20176806 ChIP-Seq 3T3-L1 Mouse	PPARG#	13	0.00213	*NRP2, PGAP1, CD180, ATP2B4, MPEG1, NAV3, NFIB, KIF1B, NOS1, TPPP, PIK3AP1, PPARGC1A, CREB5*
TCF4 22108803 ChIP-Seq LS180 Human	TCF4	14	0.00213	*SNED1, NCOA2, P2RY13, NFIX, PTCH1, ATP2B4, ALOX12, TANC2, MLXIPL, NAV3, NFIB, NOS1, TPPP, CREB5*
YAP1 20516196 ChIP-Seq MESCs Mouse	YAP1	15	0.00242	*NCOA2, NFIX, RASGRF2, PTCH1, CD180, ABCC6, GRIK3, PPP1R9A, TANC2, SLC7A2, NAV3, NFIB, FLVCR2, NOS1, PPARGC1A*
IRF8 27001747 Chip-Seq BMDM Mouse	IRF8	13	0.00377	*CX3CR1, P2RY13, NRP2, PGAP1, GPR34, CD180, MPEG1, SCML4, DHRS9, PIK3AP1, EVI2B, PLCB2, CREB5*
BACH1 22875853 ChIP-PCR HELA AND SCP4 Human	BACH1	11	0.00488	*TUB, SNED1, NCOA2, PGAP1, PCDH7, ABCC6, KDR, ATP2B4, KIF1B, PLCB2, CREB5*
TCF4 23295773 ChIP-Seq U87 Human	TCF4	19	0.00872	*NCOA2, NRP2, PGAP1, NFIX, RASGRF2, PCDH7, ABCC6, GRIK3, ATP2B4, CACNA2D2, ALOX12, PPP1R9A, TANC2, RASGRP3, FLVCR2, KIF1B, FREM1, PPARGC1A, CREB5*
CEBPD 21427703 ChIP-Seq 3T3-L1 Mouse	CEBPD	11	0.0177	*CX3CR1, SNED1, NRP2, PGAP1, NFIX, NFIB, PCDH7, CD180, FNDC3B, TANC2, CREB5*
MYCN 21190229 ChIP-Seq SHEP-21N Human	MYCN	5	0.0190	*PCDH7, PIK3AP1, MPEG1, SLC7A2, CREB5*
ESR1 22446102 ChIP-Seq UTERUS Mouse	ESR1#	10	0.0220	*SNED1, NCOA2, SCML4, NAV3, CD180, KDR, FNDC3B, ATP2B4, FLVCR2, CACNA2D2*
RUNX2 22187159 ChIP-Seq PCA Human	RUNX2	16	0.0277	*SNED1, PGAP1, PCDH7, PTCH1, GRIK3, FNDC3B, CACNA2D2, TMC7, PPP1R9A, RASGRP3, NFIB, SPATA13, KDR, TPPP, PPARGC1A, CREB5*
AR 27270436 ChIP-Seq LNCaP Human Prostate Carcinoma	AR#	4	0.0317	*GABRB3, NFIB, FNDC3B, PIK3AP1*
FOXA2 19822575 ChIP-Seq HepG2 Human	FOXA2	15	0.0347	*SNED1, NCOA2, NRP2, NFIX, PTCH1, ABCC6, FNDC3B, CACNA2D2, TMC7, PPP1R9A, MLXIPL, SPATA13, KIF1B, PIK3AP1, PPARGC1A*
ISL1 27105846 Chip-Seq CPCs Mouse	ISL1	10	0.0347	*NRP2, NAV3, NFIB, PCDH7, PTCH1, CD180, KDR, FNDC3B, FREM1, RASGRP3*
NFE2L2 20460467 ChIP-Seq MEFs Mouse	NFE2L2#	8	0.0347	*PGAP1, SCML4, PCDH7, KIF1B, TMC7, MPEG1, TANC2, RASGRP3*
LEF1 29337183 ChIP-Seq mESC Mouse Stem	LEF1	14	0.0396	*NCOA2, NRP2, PLAG1, PCDH7, PTCH1, GRIK3, CACNA2D2, ALOX12, SLC7A2, SCML4, NFIB, SPATA13, FREM1, CREB5*
BRD4 27068464 Chip-Seq AML-cells Mouse	BRD4#	11	0.0411	*NCOA2, P2RY13, NRP2, SCML4, DHRS9, PTCH1, PIK3AP1, EVI2B, MPEG1, TANC2, RASGRP3*
STAT1 17558387 ChIP-Seq HELA Human	STAT1	10	0.0411	*CX3CR1, SCML4, GPR34, DHRS9, SPATA13, PPP1R9A, POU2F2, MPEG1, SLC7A2, RASGRP3*

Table 4 (end)

ChEA_2022 term	TF	Number of DEGs	Adjusted *p*-value	DEGs of the Downregulated cluster
STAT3 23295773 ChIP-Seq U87 Human	STAT3[#]	15	0.0411	*GABRB3*, *SNED1*, *NRP2*, *CD180*, *GRIK3*, *FNDC3B*, *CACNA2D2*, *TMC7*, *ALOX12*, *TANC2*, *RASGRP3*, *MLXIPL*, *KIF1B*, *FREM1*, *CREB5*
KDM2B 26808549 Chip-Seq REH Human	KDM2B	10	0.0449	*TUB*, *PGAP1*, *NFIB*, *PCDH7*, *CD180*, *SPATA13*, *KIF1B*, *PPP1R9A*, *CREB5*, *RASGRP3*
TCF4 18268006 ChIP-ChIP LS174T Human	TCF4	5	0.0449	*NRP2*, *PGAP1*, *DHRS9*, *ATP2B4*, *RASGRP3*
EZH2 27294783 Chip-Seq ESCs Mouse	EZH2	11	0.0474	*GABRB3*, *SNED1*, *NRP2*, *NFIX*, *SCML4*, *PTCH1*, *CACNA2D2*, *ALOX12*, *TPPP*, *PPP1R9A*, *PPARGC1A*

[#]Transcription factors encoded by genes associated with hypertension; TF, transcription factor.

subfamily 3 group C member 1, the glucocorticoid receptor) are essential for the processes.

SMARCA4 belongs to the family of proteins with helicase and ATPase activities. These proteins play the key role in the arrangement of chromatin conformation, which is crucial for gene regulation. The SWI/SNF (BRG1) complex has been shown to participate in the epigenetic and transcription regulation in vascular smooth muscle cells, thus being associated with the development of cardiovascular disorders (Liu H. et al., 2024). Genetic polymorphisms in the *SMARCA4* gene are associated with hypertension risk (Ma et al., 2019). The SWI/SNF complex can be directly recruited by glucocorticoid receptor NR3C1 for remodeling chromatin in glucocorticoid-dependent gene activation to potentiate the glucocorticoid receptor action afterwards (Fryer, Archer, 1998; Wallberg et al., 2000). This consideration is in line with the inference by Patel et al. (2025) that SMARCA4 may be the key regulator in glucocorticoid-induced increase in intraocular tension (regarded as ocular hypertension by the authors), which can result in secondary glaucoma.

Glucocorticoid receptor NR3C1, upregulated by steroid hormones, is associated with hypertension. It acts as a transcription factor (Timmermans et al., 2019). Our previous study demonstrated *Nr3c1* downregulation in the hypothalamus of both rat strains under stress: log2FC = −0.133 in ISIAH and log2FC = −0.113 in WAG. The decrease in glucocorticoid receptor activity after the increase in glucocorticoid level upon stress can be interpreted as regulation of the homeostasis of glucocorticoid receptors (Burnstein et al., 1991). The results of this work suggest the involvement of NR3C1 in the regulation of 21 genes associated with the DEG cluster with stress-induced downregulation. The fact that the lists of target genes for NR3C1 and SMARCA4 intersect significantly (14 common targets in Table 4, p-value = 0.0007) agrees well with the aforementioned data that the regulation networks of these TFs are interrelated and, correspondingly, many genes we found can undergo SMARCA4-mediated epigenetic modification under stress, whose key signal is NR3C1 activation.

Our approach to data analysis revealed 14 DEGs in peripheral blood (Table 3). The genes listed in Table 3 have not been found to be associated with hypertension thus far, but our study suggests that they may be candidate peripheral markers associated with psychoemotional stress in hypertensive patients. Our results are insufficient for a reliable choice of peripheral blood markers suitable for clinical practice. This issue demands a special study aimed at the corroboration of the role of the markers in the blood of patients with chronic hypertension. Such a study would also help in determining their reference values in clinical use.

Conclusions

Our work revealed stress-induced downregulation of genes involved in plasma membrane function and, simultaneously, in interaction with the extracellular matrix. This downregulation reflects the significant contribution of psychoemotional stress to the formation of hypertension in humans and appears to be a fundamental process linking chronic stress and hypertension, primarily in the hypothalamus, which is a key component of this link and the focus of our study. Numerous publications confirm the effect of stress on the plasma membrane. This process is universal, even when comparing stress responses in the rat and fruit fly (Podkolodnaya et al., 2025). Furthermore, according to the membrane concept of arterial hypertension pathogenesis (Orlov, 2019), changes in membrane structure and permeability are among the chief processes underlying the impairments that promote the development of hypertensive disease. The identified target genes of the SMARCA4 transcription factor, which is likely to be involved in their epigenetic regulation, are the most probable common factor mediating long-term alteration of gene expression patterns caused by chronic stress. Its further investigation looks promising. On the grounds of our data, we propose candidate peripheral blood markers that may be helpful in clinical practice to diagnose psychoemotional stress.

References

Advani V.M., Ivanov P. Translational control under stress: reshaping the translatome. *BioEssays*. 2019;41(5):e1900009. doi 10.1002/bies.201900009

Ahn S., Jeong E., Min J.W., Kim E., Choi S.S., Kim C.J., Lee D.C. Identification of genes dysregulated by elevation of microRNA-210

levels in human trophoblasts cell line, Swan 71. *Am J Reprod Immunol*. 2017;78(5):e12722. doi 10.1111/aji.12722

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Awad K.S., Elinoff J.M., Wang S., Gairhe S., Ferreyra G.A., Cai R., Sun J., Solomon M.A., Danner R.L. Raf/ERK drives the proliferative and invasive phenotype of BMPR2-silenced pulmonary artery endothelial cells. *Am J Physiol Lung Cell Mol Physiol*. 2016;310(2):L187-L201. doi 10.1152/ajplung.00303.2015

Bali A., Gupta S., Singh N., Jaggi A.S. Implicating the role of plasma membrane localized calcium channels and exchangers in stress-induced deleterious effects. *Eur J Pharmacol*. 2013;714(1-3):229-238. doi 10.1016/j.ejphar.2013.06.010

Bautista L.E., Bajwa P.K., Shafer M.M., Malecki K.M.C., McWilliams C.A., Palloni A. The relationship between chronic stress, hair cortisol and hypertension. *Int J Cardiol Hypertens*. 2019;2:100012. doi 10.1016/j.ijchy.2019.100012

Baymiller M., Moon S.L. Stress granules as causes and consequences of translation suppression. *Antioxid Redox Signal*. 2023;39(4-6):390-409. doi 10.1089/ars.2022.0164

Bikulciene I., Baleisis J., Mazgelyte E., Rudys R., Vosyliute R., Simkunaite-Rizgeliene R., Kaminskas A., Karciauskaite D. Impact of chronic psychological stress on platelet membrane fatty acid composition in a rat model of type 1 diabetes Mellitus. *Lipids Health Dis*. 2024;23(1):69. doi 10.1186/s12944-024-02067-3

Burford N.G., Webster N.A., Cruz-Topete D. Hypothalamic-pituitary-adrenal axis modulation of glucocorticoids in the cardiovascular system. *Int J Mol Sci*. 2017;18(10):2150. doi 10.3390/ijms18102150

Burnstein K.L., Bellingham D.L., Jewell C.M., Powell-Oliver F.E., Cidlowski J.A. Autoregulation of glucocorticoid receptor gene expression. *Steroids*. 1991;56(2):52-58. doi 10.1016/0039-128x(91)90124-e

Carmichael C.Y., Wainford R.D. Hypothalamic signaling mechanisms in hypertension. *Curr Hypertens Rep*. 2015;17(5):39. doi 10.1007/s11906-015-0550-4

Chadaeva I., Ponomarenko P., Kozhemyakina R., Suslov V., Bogomolov A., Klimova N., Shikhevich S., Savinkova L., Oshchepkov D., Kolchanov N.A., Markel A., Ponomarenko M. Domestication explains two-thirds of differential-gene-expression variance between domestic and wild animals; the remaining one-third reflects intraspecific and interspecific variation. *Animals* (*Basel*). 2021;11(9):2667. doi 10.3390/ani11092667

Chadaeva I.V., Filonov S.V., Zolotareva K.A., Khandaev B.M., Ershov N.I., Podkolodnyy N.L., Kozhemyakina R.V., … Stefanova N.A., Kolosova N.G., Markel A.L., Ponomarenko M.P., Oshchepkov D.Y. RatDEGdb: a knowledge base of differentially expressed genes in the rat as a model object in biomedical research. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):794-806. doi 10.18699/VJGB-23-92

Che Y., Zhou Z., Shu Y., Zhai C., Zhu Y., Gong S., Cui Y., Wang J.F. Chronic unpredictable stress impairs endogenous antioxidant defense in rat brain. *Neurosci Lett*. 2015;584:208-213. doi 10.1016/j.neulet.2014.10.031

de Nadal E., Ammerer G., Posas F. Controlling gene expression in response to stress. *Nat Rev Genet*. 2011;12(12):833-845. doi 10.1038/nrg3055

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi 10.1093/bioinformatics/bts635

Efron B., Halloran E., Holmes S. Bootstrap confidence levels for phylogenetic trees. *Proc Natl Acad Sci USA*. 1996;93(23):13429-13434. doi 10.1073/pnas.93.23.13429

Evangelista J.E., Xie Z., Marino G.B., Nguyen N., Clarke D.J.B., Ma'ayan A. Enrichr-KG: bridging enrichment analysis across multiple libraries. *Nucleic Acids Res*. 2023;51(W1):W168-W179. doi 10.1093/nar/gkad393

Ferreira N.S., Tostes R.C., Paradis P., Schiffrin E.L. Aldosterone, inflammation, immune system, and hypertension. *Am J Hypertens*. 2021;34(1):15-27. doi 10.1093/ajh/hpaa137

Fontes M.A.P., Marins F.R., Patel T.A., de Paula C.A., Dos Santos Machado L.R., de Sousa Lima E.B., Ventris-Godoy A.C., Viana A.C.R., Linhares I.C.S., Xavier C.H., Filosa J.A., Patel K.P. Neurogenic background for emotional stress-associated hypertension. *Curr Hypertens Rep*. 2023;25(7):107-116. doi 10.1007/s11906-023-01235-7

Fryer C.J., Archer T.K. Chromatin remodelling by the glucocorticoid receptor requires the BRG1 complex. *Nature*. 1998;393(6680):88-91. doi 10.1038/30032

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

Hering D., Lachowska K., Schlaich M. Role of the sympathetic nervous system in stress-mediated cardiovascular disease. *Curr Hypertens Rep*. 2015;17(10):80. doi 10.1007/s11906-015-0594-5

Hovatta I., Juhila J., Donner J. Oxidative stress in anxiety and comorbid disorders. *Neurosci Res*. 2010;68(4):261-275. doi 10.1016/j.neures.2010.08.007

Jung Y.W., Shim J.I., Shim S.H., Shin Y.J., Shim S.H., Chang S.W., Cha D.H. Global gene expression analysis of cell-free RNA in amniotic fluid from women destined to develop preeclampsia. *Medicine* (*Baltimore*). 2019;98(3):e13971. doi 10.1097/MD.0000000000013971

Kinsman B.J., Nation H.N., Stocker S.D. Hypothalamic signaling in body fluid homeostasis and hypertension. *Curr Hypertens Rep*. 2017;19(6):50. doi 10.1007/s11906-017-0749-7

Koper A., Zeef L.A., Joseph L., Kerr K., Gosney J., Lindsay M.A., Booton R. Whole transcriptome analysis of pre-invasive and invasive early squamous lung carcinoma in archival laser microdissected samples. *Respir Res*. 2017;18(1):12. doi 10.1186/s12931-016-0496-3

Lambert A.J., Brand M.D. Reactive oxygen species production by mitochondria. In: Stuart J.A. (Ed.) Mitochondrial DNA. Methods in Molecular Biology™. Vol. 554. Humana Press, 2009;165-181. doi 10.1007/978-1-59745-521-3_11

Lambert E.A., Lambert G.W. Stress and its role in sympathetic nervous system activation in hypertension and the metabolic syndrome. *Curr Hypertens Rep*. 2011;13(3):244-248. doi 10.1007/s11906-011-0186-y

Leek J.T., Johnson W.E., Parker H.S., Jaffe A.E., Storey J.D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883. doi 10.1093/bioinformatics/bts034

Leong X.F. Lipid oxidation products on inflammation-mediated hypertension and atherosclerosis: a mini review. *Front Nutr*. 2021;8:717740. doi 10.3389/fnut.2021.717740

Liu H., Zhao Y., Zhao G., Deng Y., Chen Y.E., Zhang J. SWI/SNF complex in vascular smooth muscle cells and its implications in cardiovascular pathologies. *Cells*. 2024;13(2):168. doi 10.3390/cells13020168

Liu M.Y., Li N., Li W.A., Khan H. Association between psychosocial stress and hypertension: a systematic review and meta-analysis. *Neurol Res*. 2017;39(6):573-580. doi 10.1080/01616412.2017.1317904

Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi 10.1186/s13059-014-0550-8

Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database* (*Oxford*). 2011;baq036. doi 10.1093/database/baq036

Ma H., He Y., Bai M., Zhu L., He X., Wang L., Jin T. The genetic polymorphisms of *ZC3HC1* and *SMARCA4* are associated with hypertension risk. *Mol Genet Genomic Med*. 2019;7(11):e942. doi 10.1002/mgg3.942

Maciejak A., Kiliszek M., Michalak M., Tulacz D., Opolski G., Matlak K., Dobrzycki S., Segiet A., Gora M., Burzynska B. Gene ex-

pression profiling reveals potential prognostic biomarkers associated with the progression of heart failure. *Genome Med.* 2015;7(1):26. doi 10.1186/s13073-015-0149-z

Makovka Y.V., Oshchepkov D.Y., Fedoseeva L.A., Markel A.L., Redina O.E. Effect of short-term restraint stress on the expression of genes associated with the response to oxidative stress in the hypothalamus of hypertensive ISIAH and normotensive WAG rats. *Antioxidants* (*Basel*). 2024;13(11):1302. doi 10.3390/antiox13111302

Maltsev A.V., Evdokimovskii E.V., Kokoz Y.M. α2-Adrenoceptor signaling in cardiomyocytes of spontaneously hypertensive rats starts to impair already at early age. *Biochem Biophys Res Commun.* 2019; 512(4):908-913. doi 10.1016/j.bbrc.2019.03.117

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

Markel A.L., Maslova L.N., Shishkina G.T., Mahanova N.A., Jacobson G.S. Developmental influences on blood pressure regulation in ISIAH rats. In: McCarty R., Blizard D.A., Chevalier R.L. (Eds) Development of the Hypertensive Phenotype: Basic and Clinical Studies. In the series Handbook of Hypertension. Amsterdam: Elsevier, 1999;493-526

Markel A.L., Redina O.E., Gilinsky M.A., Dymshits G.M., Kalashnikova E.V., Khvorostova Y.V., Fedoseeva L.A., Jacobson G.S. Neuroendocrine profiling in inherited stress-induced arterial hypertension rat strain with stress-sensitive arterial hypertension. *J Endocrinol*. 2007;195(3):439-450. doi 10.1677/JOE-07-0254

Marques F.Z., Campain A.E., Tomaszewski M., Zukowska-Szczechowska E., Yang Y.H., Charchar F.J., Morris B.J. Gene expression profiling reveals renin mRNA overexpression in human hypertensive kidneys and a role for microRNAs. *Hypertension.* 2011;58(6): 1093-1098. doi 10.1161/HYPERTENSIONAHA.111.180729

Matovic S., Ichiyama A., Igarashi H., Salter E.W., Sunstrum J.K., Wang X.F., Henry M., Kuebler E.S., Vernoux N., Martinez-Trujillo J., Tremblay M.E., Inoue W. Neuronal hypertrophy dampens neuronal intrinsic excitability and stress responsiveness during chronic stress. *J Physiol.* 2020;598(13):2757-2773. doi 10.1113/JP279666

Montezano A.C., Dulak-Lis M., Tsiropoulou S., Harvey A., Briones A.M., Touyz R.M. Oxidative stress and human hypertension: vascular mechanisms, biomarkers, and novel therapies. *Can J Cardiol*. 2015;31(5):631-641. doi 10.1016/j.cjca.2015.02.008

Mura M., Cecchini M.J., Joseph M., Granton J.T. Osteopontin lung gene expression is a marker of disease severity in pulmonary arterial hypertension. *Respirology*. 2019;24(11):1104-1110. doi 10.1111/resp.13557

Neusser M.A., Lindenmeyer M.T., Moll A.G., Segerer S., Edenhofer I., Sen K., Stiehl D.P., Kretzler M., Grone H.J., Schlondorff D., Cohen C.D. Human nephrosclerosis triggers a hypoxia-related glomerulopathy. *Am J Pathol*. 2010;176(2):594-607. doi 10.2353/ajpath.2010.090268

Orlov S.N. Membrane theory of the pathogenesis of arterial hypertension: what do we know about this, half a century later? *Bulletin of Siberian Medicine.* 2019;18(2):234-247. doi 10.20538/1682-0363-2019-2-234-247 (in Russian)

Oshchepkov D., Chadaeva I., Kozhemyakina R., Zolotareva K., Khandaev B., Sharypova E., Ponomarenko P., … Kolosova N.G., Nazarenko M., Kolchanov N.A., Markel A., Ponomarenko M. Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients. *Int J Mol Sci*. 2022;23(5):2835. doi 10.3390/ijms23052835

Oshchepkov D.Y., Makovka Y.V., Fedoseeva L.A., Seryapina A.A., Markel A.L., Redina O.E. Effect of short-term restraint stress on the hypothalamic transcriptome profiles of rats with Inherited Stress-Induced Arterial Hypertension (ISIAH) and normotensive Wistar Albino Glaxo (WAG) rats. *Int J Mol Sci*. 2024;25(12):6680. doi 10.3390/ijms25126680

Patel P.D., Patel G.C., Millar J.C., Feris S., Curry S., Geisert E.E., Clark A.F. Mechanistic insights into glucocorticoid-induced ocular hypertension using differences in mouse strain responsiveness. *bioRxiv.* 2025. doi 10.1101/2025.07.02.662542

Picard M., McEwen B.S. Psychological stress and mitochondria: a systematic review. *Psychosom Med.* 2018;80(2):141-153. doi 10.1097/PSY.0000000000000545

Plaza-Florido A., Altmae S., Esteban F.J., Lof M., Radom-Aizik S., Ortega F.B. Cardiorespiratory fitness in children with overweight/obesity: insights into the molecular mechanisms. *Scand J Med Sci Sports.* 2021;31(11):2083-2091. doi 10.1111/sms.14028

Podkolodnaya O.A., Deryuzhenko M.A., Tverdokhleb N.N., Zolotareva K.A., Makovka Yu.V., Podkolodny N.L., Suslov V.V., … Kondratyuk E.Yu., Redina O.E., Markel A.L., Gruntenko N.E., Ponomarenko M.P. FlyDEGdb knowledge base on differentially expressed genes of *Drosophila melanogaster*, a model object in biomedicine. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2025;29(7):952-962. doi 10.18699/vjgb-25-101

Qiu X., Lin J., Liang B., Chen Y., Liu G., Zheng J. Identification of hub genes and microRNAs associated with idiopathic pulmonary arterial hypertension by integrated bioinformatics analyses. *Front Genet*. 2021;12:667406. doi 10.3389/fgene.2021.636934

Rosenkranz J.A., Venheim E.R., Padival M. Chronic stress causes amygdala hyperexcitability in rodents. *Biol Psychiatry*. 2010;67(12): 1128-1136. doi 10.1016/j.biopsych.2010.02.008

Saei H., Govahi A., Abiri A., Eghbali M., Abiri M. Comprehensive transcriptome mining identified the gene expression signature and differentially regulated pathways of the late-onset preeclampsia. *Pregnancy Hypertens.* 2021;25:91-102. doi 10.1016/j.preghy.2021.05.007

Shikhevich S., Chadaeva I., Khandaev B., Kozhemyakina R., Zolotareva K., Kazachek A., Oshchepkov D., … Markel A., Savinkova L., Kolchanov N.A., Kozlov V., Ponomarenko M. Differentially expressed genes and molecular susceptibility to human age-related diseases. *Int J Mol Sci*. 2023;24(4):3996. doi 10.3390/ijms24043996

Spruill T.M. Chronic psychosocial stress and hypertension. *Curr Hypertens Rep.* 2010;12(1):10-16. doi 10.1007/s11906-009-0084-8

Szklarczyk D., Kirsch R., Koutrouli M., Nastou K., Mehryary F., Hachilif R., Gable A.L., Fang T., Doncheva N.T., Pyysalo S., Bork P., Jensen L.J., von Mering C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 2023;51(D1):D638-D646. doi 10.1093/nar/gkac1000

Textoris J., Ivorra D., Ben Amara A., Sabatier F., Menard J.P., Heckenroth H., Bretelle F., Mege J.L. Evaluation of current and new biomarkers in severe preeclampsia: a microarray approach reveals the *VSIG4* gene as a potential blood biomarker. *PLoS One*. 2013; 8(12):e82638. doi 10.1371/journal.pone.0082638

Timmermans S., Souffriau J., Libert C. A general introduction to glucocorticoid biology. *Front Immunol.* 2019;10:1545. doi 10.3389/fimmu.2019.01545

Ulecia-Moron C., Bris A.G., MacDowell K.S., Cervero-Garcia P., Madrigal J.L.M., Garcia-Bueno B., Pereira M.P., Leza J.C., Caso J.R. Chronic mild stress dysregulates autophagy, membrane dynamics, and lysosomal status in frontal cortex and hippocampus of rats. *Eur Neuropsychopharmacol*. 2025;94:24-35. doi 10.1016/j.euroneuro.2025.02.005

Vedi M., Smith J.R., Thomas Hayman G., Tutaj M., Brodie K.C., De Pons J.L., Demos W.M., … Tutaj M.A., Wang S.J., Zacher S., Dwinel M.R., Kwitek A.E. 2022 updates to the Rat Genome Database: a Findable, Accessible, Interoperable, and Reusable (FAIR) resource. *Genetics*. 2023;224(1):iyad042. doi 10.1093/genetics/iyad042

Wallberg A.E., Neely K.E., Hassan A.H., Gustafsson J.A., Workman J.L., Wright A.P. Recruitment of the SWI-SNF chromatin remodeling complex as a mechanism of gene activation by the glucocorticoid receptor τ1 activation domain. *Mol Cell Biol*. 2000;20(6): 2004-2013. doi 10.1128/MCB.20.6.2004-2013.2000

Wu Y.B., Zang W.D., Yao W.Z., Luo Y., Hu B., Wang L., Liang Y.L. Analysis of FOS, BTG2, and NR4A in the function of renal medul-

lary hypertension. *Genet Mol Res*. 2013;12(3):3735-3741. doi 10.4238/2013.September.19.4

Xie Z., Bailey A., Kuleshov M.V., Clarke D.J.B., Evangelista J.E., Jenkins S.L., Lachmann A., Wojciechowicz M.L., Kropiwnicki E., Jagodnik K.M., Jeon M., Ma'ayan A. Gene set knowledge discovery with Enrichr. *Curr Protoc*. 2021;1(3):e90. doi 10.1002/cpz1.90

Yao X., Jing T., Wang T., Gu C., Chen X., Chen F., Feng H., Zhao H., Chen D., Ma W. Molecular characterization and elucidation of pathways to identify novel therapeutic targets in pulmonary arterial hypertension. *Front Physiol*. 2021;12:694702. doi 10.3389/fphys.2021.694702

Yong H.E., Melton P.E., Johnson M.P., Freed K.A., Kalionis B., Murthi P., Brennecke S.P., Keogh R.J., Moses E.K. Genome-wide transcriptome directed pathway analysis of maternal pre-eclampsia susceptibility genes. *PLoS One*. 2015;10(5):e0128230. doi 10.1371/journal.pone.0128230

Zheng Y., He J.Q. Common differentially expressed genes and pathways correlating both coronary artery disease and atrial fibrillation. *EXCLI J*. 2021;20:126-141. doi 10.17179/excli2020-3262

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

Received October 7, 2025. Revised October 24, 2025. Accepted October 28, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1338-1347

doi 10.18699/vjgb-25-140

Overcoming the problem of heterologous proteins folding to improve the efficiency of yeast bioproducers

N.V. Dorogova, S.A. Fedorova

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

Abstract. In the last few decades, yeasts have been successfully engineered to be an excellent microbial cell factory for producing recombinant proteins with desired properties. This was due to their cost-effective characteristics and the successful application of genomic modification technologies. In addition, yeasts have a conserved post-translational modification pathway among eukaryotic organisms, which ensures the correct folding of recombinant proteins. However, the folding machinery cannot always cope with the load caused by the overexpression of recombinant genes, leading to the accumulation of misfolded proteins, the formation of aggregates and low production. Therefore, the protein-folding capacity of the endoplasmic reticulum (ER) remains one of the main limitations for heterologous protein production in yeast host organisms. However, thanks to many years of effective research of the fundamental mechanisms of protein folding, these limitations have been largely overcome. The study of folding in both model organisms and bioproducers has allowed to identify the molecular factors and cellular mechanisms that determine how a nascent polypeptide chain acquires its three-dimensional functional structure. This knowledge has become the basis for developing new effective techniques for engineering highly productive yeast strains. In this review, we examined the main cellular mechanisms associated with protein folding, such as ER transition, chaperone binding, oxidative folding, glycosylation, protein quality control. We discuss the effectiveness of applying this knowledge to the development of various engineering techniques aimed at overcoming bottlenecks in the protein folding system. In particular, selection of optimal signal peptides, co-expression with chaperones and foldases, modification of protein quality control, inhibition of proteolysis, and other techniques have allowed to enhance the ability of yeast bioproducers to effectively secrete heterologous proteins.

Key words: yeast bioproducers; protein folding; endoplasmic reticulum; molecular chaperones; recombinant proteins

For citation: Dorogova N.V., Fedorova S.A. Overcoming the problem of heterologous proteins folding to improve the efficiency of yeast bioproducers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1338-1347. doi 10.18699/vjgb-25-140

Funding. This study was supported by the Ministry of Science and Higher Education of the Russian Federation, project No. FWNR-2025-0031.

Решения проблем фолдинга белков для повышения эффективности дрожжевых биопродуцентов

Н.В. Дорогова, С.А. Федорова

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

Аннотация. За последние несколько десятилетий дрожжи стали наиболее эффективными биопродуцентами рекомбинантных белков с различными потребительскими свойствами. Это стало возможным благодаря их экономически выгодным характеристикам и успешному применению генно-инженерных технологий. Кроме того, дрожжи обладают консервативным для эукариотических организмов механизмом посттрансляционной модификации белков, который обеспечивает их корректный фолдинг, необходимый для дальнейшей секреции и функциональной активности. Однако аппарат фолдинга не всегда справляется с нагрузкой, вызванной сверхэкспрессией рекомбинантных генов, что приводит к накоплению неправильно свернутых белков, образованию агрегатов и низкой продуктивности дрожжевых штаммов. Таким образом, способность к фолдингу белков в эндоплазматическом ретикулуме по-прежнему остается одним из основных ограничений при синтезе рекомбинатных белков в дрожжевых клетках. Эти ограничения были в значительной степени преодолены благодаря многолетним эффективным исследованиям фундаментальных механизмов белкового фолдинга. Изучение фолдинга как у модельных организмов, так и у биопродуцентов позволило выявить молекулярные факторы и клеточные механизмы, определяющие формирование трехмерной функциональной структуры растущей пептидной цепи. Полученные знания легли в основу разработки новых эффективных методов конструирования высокопродуктивных штаммов дрожжей. В данном обзоре мы рассмотрели основные клеточные механизмы, связанные с фолдингом белков, такие как транспорт через эндоплазматический

ретикулум, взаимодействие с шаперонами, окислительный фолдинг, гликозилирование и контроль качества белков. Мы обсудили эффективность применения этих знаний при разработке различных инженерных методов, направленных на преодоление узких мест в системе белкового фолдинга. В частности, подбор оптимальных сигнальных пептидов, коэкспрессия с шаперонами и фолдазами, модификация клеточных механизмов контроля качества белков, ингибирование протеолиза и некоторые другие приемы позволили повысить возможности использования дрожжей-продуцентов в качестве эффективной производительной платформы для экспрессии и секреции рекомбинантных белков.

Ключевые слова: дрожжи-продуценты; фолдинг белков; эндоплазматический ретикулум; молекулярные шапероны; рекомбинантные белки

Introduction

Yeast expression systems are excellent for the production of valuable recombinant proteins and peptides widely used as biopharmaceuticals and industrial enzymes. They are commercially viable bioproducers due to their high growth rate, resistance to harmful microbiota, ability to assimilate many food sources, and fairly easy to cultivate in industrial conditions (Thak et al., 2020; Madhavan et al., 2021; De Brabander et al., 2023). The development of genetic and metabolic engineering has increased the efficiency of yeast strains, mainly due to the use of genomic technologies: strong promoters, new vector elements with improved inducers and enhancers, targeted mutagenesis, signaling molecules, high-performance devices for cloning, screening and fermentation (De Brabander et al., 2023; Tsuda, Nonaka, 2024). However, expression at the transcriptional and translational levels often does not correlate with the level of secretion of heterologous proteins, which is due to the insufficient efficiency of the folding mechanism (Ishiwata-Kimata, Kimata, 2023; Zahrl et al., 2023). Therefore, the search for new technological methods for optimizing the synthesis and increasing the yield of heterologous proteins in yeast cells remains an urgent task. A significant direction for its solution is overcoming the problem of folding target proteins into the correct three-dimensional structure.

Proper folding is necessary for the functional activity of synthesized proteins, their intracellular transport and further secretion. Recombinant proteins are secretory and go through a secretory pathway, beginning with folding in the ER and ending with release into the extracellular environment (culture media) (Raschmanová et al., 2021). Proteins are translocated to the ER in an unfolded state and then undergo modification and folding involving ER-resident chaperones, folding enzymes, and glycosylation (Hartl et al., 2011; Saibil, 2013). Disruptions in this machinery result in the accumulation of misfolded proteins in the cytoplasm, where they are recognized by the Protein Quality Control (PQC) system that regulates cellular homeostasis (Korennykh, Walter, 2012). This system includes the Unfolded Protein Response (UPR) signaling pathway, which can trigger refolding of misfolded proteins or initiate their proteolysis.

In some cases, misfolded proteins clump together to form aggregates or Inclusion Bodies (IBs), which can cause cell damage (Yamaguchi, Miyazaki, 2014). Such IBs often contain potentially active proteins with a normal secondary structure, which can be recovered from these aggregates under appropriate conditions (Burgess, 2009; Yamaguchi, Miyazaki, 2014; Singhvi et al., 2021). However, their isolation and refolding procedures are complex, expensive, and inefficient (Singhvi et al., 2021). It is therefore clear that the solution to the folding problem of heterologous proteins must be directly linked to their production process: protein folding engineering and quality control in yeast host strains. In this article, we review various cellular mechanisms and signaling pathways that influence heterologous protein folding and discuss the latest updates to biotechnological strategy allowing to address this issue in order to maximize the yield of recombinant proteins.

Post-translational modifications and folding

Preparation for folding

In yeast, as in other eukaryotes, a newly synthesized peptide must undergo post-translational modification and folding in the endoplasmic reticulum to form the correct spatial conformation. A stable 3D structure determines the functional activity of the protein and its subsequent traffic through the secretory pathway (Schwarz, Blower, 2016).

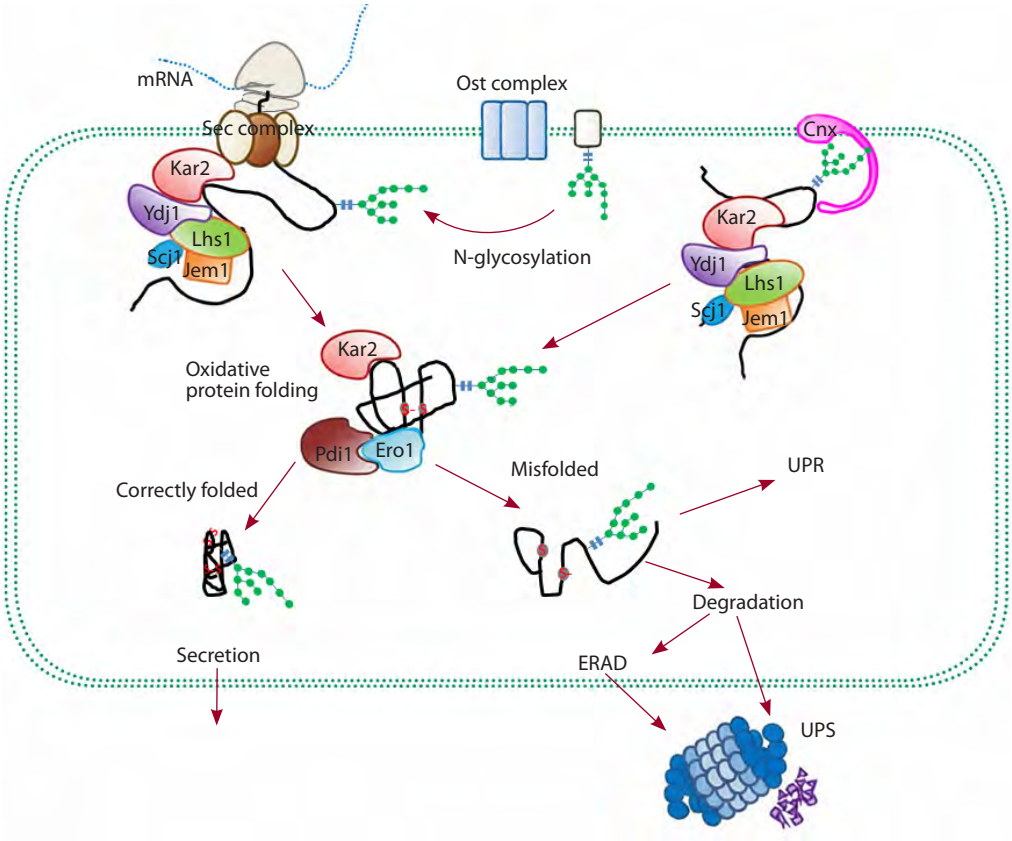
Translation of secretory proteins occurs on cytosolic ribosomes and they are then co-translationally or post-translationally directed to the ER by a specific protein-RNA complex – Signal Recognition Particle (SRP). SRP binds to the N-terminal sequence of a precursor protein, which is called the signal peptide (SP). Transfer across the ER membrane occurs in an unfolded state of the nascent protein and is dependent on the Sec water channel, a multiprotein complex that spans the membrane (Berner et al., 2018; O'Keefe et al., 2022) (see the Figure).

When the unfolded nascent chain appears in the ER lumen, its hydrophobic sequence elements are recognized by ER-resident HSP chaperones: Kar2 (yeast Hsp70) and Ydj1 (yeast Hsp40) (Braakman, Hebert, 2013; Hendershot et al., 2024). They bind to the hydrophobic amino acid side chains exposed by the unfolded proteins. There is evidence that Kar2 assists in the folding of nascent proteins as they enter the ER and remains bound until folding is complete. The Ydj1 chaperone forms transient complexes with Kar2, facilitating its binding to non-native polypeptides (see the Figure). Thus, chaperones prevent premature misfolding of the immature polypeptide chain and protect it from aggregation (Omkar et al., 2024; Ruger-Herreros et al., 2024).

According to some data, the chaperone function of Kar2 in these events also depends on members of the DnaJ-like protein family, such as Jem1 and Scj1, and the nucleotide exchange factor Lhs1 (see the Figure). These co-chaperones promote the ATPase cycle and thus maintain Kar2 activity (Schlenstedt et al., 1995; Steel et al., 2004; de Keyzer et al., 2009).

Oxidative protein folding

3D structure formation begins with the process called Oxidative Protein Folding (OPF), which results in the formation of disulfide bonds due to the oxidation of thiol groups of cysteine



Nascent secretory protein is transferred co-translationally into the ER lumen through the Sec multiprotein complex that spans the ER membrane. At the ER membrane, the protein undergoes N-glycosylation involving the OST complex. The N-glycan-modified peptide chain binds to calnexin (Cnx). When the unfolded nascent chain appears in the ER lumen, it interacts with Kar2 and Ydj1. Jem1, Scj1 and Lhs1 assist in this process. In the ER lumen, the nascent peptide undergoes an OPF process involving Pdi and Ero, which results in disulfide bonds. Accumulation of misfolded proteins in the cytoplasm induces the UPR response which can trigger refolding of misfolded proteins or initiate ERAD.

(Hatahet, Ruddock, 2009; Palma et al., 2023). OPF is carried out by Protein disulfide isomerase (Pdi) (see the Figure). Pdi not only is responsible for the formation of the disulfide bonds in unfolded eukaryotic proteins, but also catalyzes the rearrangement of incorrect disulfide bonds (isomerase activity) (Gross et al., 2006). The OPF pathway, like Pdi, is conserved across eukaryotic organisms. Yeast Pdi1 protein is encoded by the *pdi1* gene, and *pdi1*-deleted yeast strains have low viability and accumulate secretory proteins within the ER lamellae (Mizunaga et al., 1990; Frand, Kaiser, 1998).

The transition of Pdi to the active form is catalyzed by ER sulfhydryl oxidase 1 (Ero1). Ero1 is an oxidoreductase that oxidizes Pdi1 via direct thiol-disulfide exchange (conversion of cystine to cysteine) (Gross et al., 2006; Sevier, Kaiser, 2006). Yeast contains a single *ero1* gene that encodes the Ero1 protein. The loss of *ero1* is lethal for yeast (Niu et al., 2016). Thus, Pdi and Ero act synergistically; for the correct catalysis of disulfide bonds in proteins, a balance of these two factors is necessary (Niu et al., 2016; Wang L., Wang C.C., 2023). The need to maintain such a balance is also due to the fact that the activation of Pdi and Ero occurs via the oxidation–reduction type and the high rate of disulfide bond formation in cells and tissues should create dangerous levels of oxidative stress (Gasser et al., 2008).

Role of glycosylation in promoting protein folding

Proper folding of most proteins requires post-translational modification known as glycosylation. In yeast, secretory proteins are glycoproteins and they contain covalently linked oligosaccharides, which are mainly mannose residues (see the Figure). Predominantly, yeast polypeptides are N-glycosylated, i. e. mannose is N-glycosidically linked to the β-amido group of asparagine. This reaction is catalyzed by the Oligosaccharyltransferase Enzyme Complex (OST) (Kelleher, Gilmore, 2006) and occurs on ER membranes. OST transfers Glc3Man9GlcNAc2-oligosaccharide (where Glc is glucose, Man is mannose, and GlcNAc is N-acetylglucosamine) from the lipid-pyrophosphate donor, dolichol diphosphate, to asparagine residues of nascent polypeptide chains. The protein-linked oligosaccharide is called N-glycan. This glycan has maintained a well-conserved structure throughout evolution and is characteristic of all eukaryotes (Qi et al., 2020). Glycans undergo further processing in the ER by glycosidases. In *Saccharomyces cerevisiae*, the following glycosidases have been described: alpha-glucosidase I encoded by CWH41, alpha-glucosidase II encoded by ROT2, alpha1,2-mannosidase encoded by MNS1. Glycosidases partially deglycosylate and shorten the N-glycan (Herscovics, 1999; Lehle et al., 2006).

Glycosylation promotes folding by enhancing the solubility and stability of the proteins in the ER and the Golgi. Oligosaccharide residues are a marker for interactions with certain chaperones, also assisting in the folding of glycoproteins (Parodi, 2000; Xu C., Ng, 2015).

The N-glycan-modified polypeptide chain is recognized by lectin chaperones (Ware et al., 1995; Caramelo, Parodi, 2015). In mammalian cells, two related ER lectin chaperones, calnexin (Cnx) and calreticulin (Crt), are important for the proper folding of newly synthesized glycoproteins. Calnexin is an integral membrane protein, and calreticulin is a soluble protein found in the ER lumen (see the Figure). They retain glycoproteins in the ER during translation by inhibiting their aggregation and formation of non-canonical disulfide bridges, and also promote their association with other chaperones. In yeast (*S. cerevisiae*), only the calnexin homologue Cne1 has been identified. Cne1 is structurally similar to mammalian calnexin except that it lacks a cytoplasmic tail and does not bind calcium (Parlati et al., 1995). Yeast calnexin has been shown to function as a molecular chaperone similar to mammalian calnexin (Xu X. et al., 2004). A calnexin-like transmembrane protein has also been identified in *S. pombe* (Núñez et al., 2015). However, genes encoding a calreticulin homologue have not yet been identified in the yeast genome.

Protein Quality Control

All eukaryotes have conserved mechanisms that control cellular proteostasis and protect the cell from stress. These are the three main pathways that comprise the quality control system: UPR, ERAD (ER-associated degradation), and autophagy. There are essentially two alternative cellular responses to the accumulation of abnormal proteins. UPR promotes their refolding repair through additional activation of ER chaperones and folding enzymes, while ERAD and autophagy target them for degradation (see the Figure).

Unfolded protein response

In eukaryotes, UPR includes ER-localized molecular chaperones that participate in sensing misfolded proteins, activating downstream signaling cascades, and mitigating proteotoxic ER stress: Inositol-requiring enzyme 1 (Ire1), Activating transcription factor 6 (Atf6), and Protein kinase R-like ER kinase (PERK). However, only Ire1 is highly conserved and has been found in unicellular eukaryotes, including yeast (Schroder et al., 2003; Mori K., 2022). Presumably, through its luminal domain, Ire1 enables direct interaction with exposed hydrophobic groups of misfolded proteins in the ER. This reaction results in the activation of Ire1, and as a result, its C-terminal RNase domain mediates splicing of the Hac1 gene mRNA removing a 252-nucleotide intron near the 3′ end. Then Rlg1, a tRNA ligase, ligates the *hac1* transcript cleaved by Ire1p. Its spliced and unspliced forms are termed Hac1i and Hac1u, respectively ("i" and "u" for induced and uninduced) (Schroder et al., 2003; Xia, 2019). Hac1i mRNA is translated into the transcription factor Hac1. It is then transported into the nucleus where it induces transcription of a large number of genes involved in the UPR mechanism (Hernández-Elvira et al., 2018) including those encoding ER-located molecular chaperones and protein modification enzymes such as *kar2*, *pdi1*, *ero1*, *ecj1*, *lhs1*, *jem1*. Additionally, the Ire1/Hac1 pathway is essential for activating genes that carry out ERAD functions promoting selective removal of terminally damaged proteins (Friedlander et al., 2000; Travers et al., 2000).

Hac1 orthologs have been identified in various yeast species, such as *Pichia pastoris*, *Hansenula polymorpha*, *Kluyveromyces lactis*, *Yarrowia lipolytica*, *Candida albicans*, and *Candida parapsilosis*. All of these species share the Ire1-dependent mechanism of splicing the Hac1 transcripts in response to ER stress (Hernández-Elvira et al., 2018; Fauzee et al., 2020; Ishiwata-Kimata, Kimata, 2023).

However, it has been reported that *S. pombe* probably does not contain a Hac1 ortholog. Its Ire1 triggers a process called Regulated Ire-dependent decay (Kimmig et al., 2012). In *S. pombe*, stress-activated IRE1 cleaves mRNAs located in the ER, leading to their exonuclease-mediated degradation (Hernández-Elvira et al., 2018).

ER-associated degradation

Proteins that have not achieved their native conformation after repeated folding remain in the ER, bound to ER chaperones that prevent their aggregation. If a protein remains unfolded in the ER for too long, it is identified as potentially harmful and eliminated via ER-associated degradation. The process involves several steps: the recognition of substrates in the lumen and membrane of the ER, their translocation into the cytosol, ubiquitination and degradation in the 26S proteasome. Thus, ERAD is associated with the highly conserved ubiquitin proteasome system (UPS) (Ruggiano et al., 2014; Krshnan et al., 2022). In all eukaryotes, it includes the following main components: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), ubiquitin ligase (E3), and 26S proteasome (Pickart, 2001). If a misfolded glycoprotein stays in the ER for a critically long time, its N-linked glycans are trimmed by yeast 1,2-mannosidase1 – Htm1. This trimmed N-glycan is recognized by a lectin chaperone known as Yos9. At the same time, the Hrd3 protein (HMG-CoA reductase degradation protein 3) associates with hydrophobic amino acid residues exposed on the surface of misfolded glycoproteins (Thibault, Ng, 2012; Berner et al., 2018). After recognition and binding to Hrd3 and Yos9, the substrate protein is transferred to the E3 ubiquitin ligases complex responsible for ubiquitination – Hrd1 (HMG-CoA reductase degradation protein 1). Hrd1 is embedded in the ER membrane and is in conjunction with Der1 (Degradation in the endoplasmic reticulum protein 1) and Usa 1 (U1 SNP1-associating protein 1), which mediates retrotranslocation of misfolded proteins through the ER membrane from the lumen side of the cytosolic membrane. The Hrd1 E3 ubiquitin ligase contains a RING domain that accepts ubiquitin from the membrane-associated protein Ubc7 (E2 ubiquitin-conjugating ligase). Finally, Hrd1 transfers ubiquitin to the substrate protein. Proteins covalently linked to one or more ubiquitin molecules are recognized by proteasome (Preston, Brodsky, 2017; Berner et al., 2018; Krshnan et al., 2022).

Ways to solve the protein folding problem in yeast synthetic biology

Selection of ER-targeting signal peptides

ER-targeting signal peptides (SPs) are critical components for the secretion of heterologous proteins because they are required for their correct transport and localization to the ER,

where modification and folding occur (Zha et al., 2023). SPs are short sequences, 15 to 30 amino acids in length, mostly located at the N-terminus of the secreted proteins. SPs typically consist of a series of hydrophobic amino acids (core region) that embed into membranes, positively charged residues at the N-terminus (necessary for proper topology of the polypeptide), and a cleavage site at the C-terminus. When an SP is cleaved by a signal peptidase, it releases the secreted protein into the ER lumen (Zha et al., 2023).

Recent research has revealed that the secretion efficiency of heterologous proteins is strongly dependent on their successful combination with SPs. Therefore, various bioinformatic and experimental studies have been carried out to elucidate the optimal SP sequence, allowing to maximize the efficiency of protein secretion.

A wide range of signal sequences, both from native genes (including those from different organisms) and synthetic molecules, are used to improve heterologous expression in yeast. For example, in *Komagataella phaffii* (*P. pastoris*), a widespread host microorganism, the most commonly used secretion signal is the mating pheromone of the α-factor (MFα1) from *S. cerevisiae* (Eskandari et al., 2023). Although this SP has proven its effectiveness in increasing the yield of heterologous proteins, work continues on its modification and the search for more productive variants. In particular, it was shown that some single amino acid substitutions of the MFα signal sequences provided a significant increase in the production of secreted proteins (Ito et al., 2022). An improved secretion signal was also obtained by creating a chimeric construct combining the MFα1 leader region and the Ost1 signal sequence. This hybrid variant turned out to be more effective compared to the original MFα1 SP (Barrero et al., 2018). The emergence of new improved modifications of MFα1 SPs was facilitated by the analysis of mutations accumulated in MFα signal sequences during yeast evolution. This allowed to identify specific motifs as well as their combinations (mutation synergism) that are important for enhancing yeast enzyme secretion (Aza et al., 2021). Promising results were obtained by combining bioinformatic prediction of the efficiency of certain signal sequences and subsequent experimental validation. For example, Duan and colleagues (2019) discovered four new endogenous signal peptides, including Dan4, Gas1, Msb2, and Fre2, according to the reported secretome and genome of *P. pastoris*. Their properties were investigated experimentally using three reporter proteins, and these SPs were shown to be superior to α-MFs in the production of heterologous proteins (Duan et al., 2019).

The SignalP databases (www.signalpeptide.de) can be used as a resource for selecting suitable SPs. Mori and colleagues (2015) created a library of 60 *S. cerevisiae* SPs that were identified in SignalP 3.0 using SOSUI software. The authors experimentally showed that six those SPs can maximize secretion of heterologous proteins (Mori A. et al., 2015).

While some progress has been made in bioinformatically predicting the efficiency of particular signal peptides for recombinant proteins, their potential has only been explored in laboratory settings. There is no reliable information yet on their successful application in biotech platforms, and it is unclear how they will function in specific yeast strains and in combination with specific target proteins.

Increased activity of the ER folding network

Positive effects of chaperone addition

Traditionally (since the 90s), yeast strain engineering has used the method of simultaneous expression (co-expression) of genes encoding heterologous proteins and genes encoding chaperones and folding enzymes. It has been shown for a variety of heterologous proteins that the introduction of extra copies of these genes into yeast host cells increases the secretion and decreases the aggregation of recombinant proteins. For example, Robinson et al. (1994) demonstrated that overexpression of Pdi in *S. cerevisiae* led to a four- to tenfold increase in secretion yields of human protein. Shusta and coauthors (1998) reported 2–6-fold increased secretion titers for single-chain antibody fragments upon co-overexpression of the Kar2 chaperone or Pdi. Simultaneous overexpression of Pdi and Kar2 resulted in a synergistic up to eightfold increase (Shusta et al., 1998).

The addition of Pdi and Kar2, both together and separately, increased the yield of recombinant proteins used for medical purposes: antithrombotic factor, hirudin (Kim et al., 2003); mammalian peptide recognition proteins (Yang et., 2016); *Necator americanus* secretory protein (Inan et al., 2007); fragment of single-chain antibody A33 (Damasceno et al., 2007); hydrophobins (Sallada et al., 2019), virus glycoprotein, RABV-G (Ben Azoun et al., 2016).

In some cases, chaperone activity is enhanced by their co-expression with molecular partners and cofactors. For example, Kar2 functionality depends on the ATPase cycle, which is promoted by the Jem1 co-chaperone and the Lhs1p nucleotide exchange factor (Steel et al., 2004). Joint expression of genes encoding these factors has been shown to stimulate Kar2 activity and increase production of recombinant human proteins (Payne et al., 2008).

As we wrote above, yeast strains overexpressing Pdi are the most frequently used in the production of correctly folded and functionally secreted recombinant proteins. However, in the oxidative protein folding pathway, Pdi acts in partnership with Ero1-oxidase. Therefore, co-overexpression of these genes can improve the efficiency of protein folding and secretion. Beal and colleagues (2019) developed a new methodology enabling the quantitative assessment of the interaction of Pdi1 and Ero1, and based on it provided a platform for the design of more efficient heterologous protein expression systems in yeast. The high efficiency of co-expression of Pdi and its molecular partner Ero1 was also shown by other authors (Ben Azoun et al., 2016; Sallada et al., 2019).

Thus, addition of known ER network folding factors to enhance heterologous expression has become a basic approach. However, this is not a universal solution; there is no guarantee that chaperones are always able to improve the secretion of recombinant proteins. In particular, a combination of different chaperones does not always lead to a synergistic effect.

Limitations of the approach

Recent investigations suggest that the efficiency of chaperones depends on the dose of the gene encoding the recombinant protein. For example, overexpression of Pdi1 or Ero1 caused a statistically significant increase in recombinant hydrophobin in *P. pastoris*, but only in the 3-copy gene strain, and

not in the 1- and 2-copy strains (Sallada et al., 2019). In the same experiment, it was shown that hydrophobin production increased with co-expression of Kar2 14-fold for the 1-copy strain, 9.8-fold for the 2-copy and 22-fold for the 3-copy ones (Sallada et al., 2019).

A correlation between chaperone activity and the copy number of the recombinant gene has been found in other experiments as well. In lipase-producing *P. pastoris*, overexpression of Pdi1 led to enhanced productivity in the strain carrying four copies of the lipase gene, whereas in the two-copy strain, it remained unchanged (Huang J. et al., 2020). Efficient expression of mammalian peptidoglycan recognition proteins in *P. pastoris* also depended on the combination of recombinant gene copy number and folding enzymes (Yang et al., 2016). Thus, these experiments support the need for empirical selection of optimal combinations of recombinant protein-encoding genes and folding factor-encoding genes.

Recently, an increasing number of published data have demonstrated that the activity of the folding factors described above is selective for certain substrates and does not improve the expression of some important recombinant proteins. In particular, it was shown that Pdi1, Ero1, Kar2 did not have a beneficial effect on the secretion of antibodies in *S. cerevisiae* (de Ruijter et al., 2016). Similar conclusions were also drawn from other experiments (Smith et al., 2004; Payne et al., 2008). These findings demonstrate that there is no single suitable strategy that provides optimal conditions for the secretion of all recombinant proteins, and very often a separate productivity enhancement program must be set up for each protein.

Forced activation of the UPR system

An effective way to solve the folding problem is the activation of the ER chaperone network by overexpression of the Hac1 transcription factor, which is the main regulating UPR pathway. Theoretically, an artificial increase in the level of Hac1 can be achieved in two ways. The first is the intensification of Hac1 mRNA splicing by overexpression of Ire1. The second way is overexpression of genetic sequences encoding Hac1.

The first approach is associated with additional expression of Ire1, which should lead to an increase in the number of events associated with Hac1 mRNA splicing and, as a consequence, promote transcription of genes encoding ER chaperones and folding catalysts (see above "Unfolded protein response"). However, at present, the effectiveness of this approach is confirmed by only one experiment. Only one study convincingly showed that overexpression of Ire1 improved the capabilities of the yeast expression system, which in particular led to an increase in hepatitis B small antigen (HBsAg) production in *S. cerevisiae* (Sheng et al., 2017). The lack of positive results may be due to the complex mechanism of Hac1 activation, which involves many molecular factors.

The second approach aimed at using Hac1 overexpression is the most preferable in yeast biotechnology. As we wrote above, adding Hac1 to yeast host strains leads to up-regulation of a large number of its target genes that mainly function for the protein secretory machinery, and in particular, almost all these genes are required for protein folding (see above "Unfolded protein response"). For heterologous protein secretion, both the active (spliced) form and the inactive (unspliced) form of Hac1 can be used. However, activation of the unspliced form of Hac1 mRNA requires additional cellular resources, in particular, molecular factors involved in its processing. The genes encoding these factors must be added to the yeast strains. The engineered strains with such a modification increased protein secretion significantly and showed better performance than strains with only overexpression of Hac1 (Lin et al., 2023).

Most strains used in biotechnological practice contain an insertion of the Hac1 gene with a previously removed 3′ end intron, i.e. encoding the active form of the transcript. For such strains, a high level of transcriptional activity of Hac1 target genes involved in the UPR pathway and responsible for correct protein folding was shown. In *P. pastoris* strains overexpressing Hac1, electron microscopy revealed an expansion of the intracellular membranes (Guerfal et al., 2010), which probably indicates an increase in the folding capacity of the ER and may contribute to the alleviation of ER stress (Schuck et al., 2009).

To date, a large number of research groups have reported an increase in the productivity of secretory proteins due to artificial and high-level expression of the Hac1 protein (Raschmanová et al., 2021; Lin et al., 2023; Khlebodarova et al., 2024). Hac1 overexpression successfully increased the yield of recombinant proteins: phytase (a product of the *Phy* gene from *C. amalonaticus*) in *P. pastoris* (Li C. et al., 2015); α-amylase (Valkonen et al., 2003; Lin et al., 2023); chitosanase (from *Bacillus subtilis*) (Han et al., 2021); kringle fragment of human apolipoprotein (which inhibits endothelial cell migration) in *S. cerevisiae* (Lee et al., 2012); xylanase (a microbial hydrolase used to hydrolyze xylan) in *S. cerevisiae* (Li C. et al., 2015; Bao et al., 2020).

However, to achieve efficient heterologous expression in Hac-modified cells, it is important to select the optimal combination of the copy number of Hac-encoding genes and genes encoding recombinant proteins (Valkonen et al., 2003; Guerfal et al., 2010; Huang M. et al., 2017; Huang J. et al., 2020). Some studies have found that increasing Hac1 doses also leads to increased ER stress (Gasser et al., 2006; Guerfal et al., 2010; Li C. et al., 2015). It is also necessary to take into account that Hac1 is a positive regulator of the ERAD pathway (see above "ER-associated degradation") and its overactivation can promote protein degradation.

Thus, the addition of Hac allows the yeast strain to be adapted to large-scale protein expression caused by an excessive dose of transgenic constructs. However, many researchers note that the effectiveness of this approach must be assessed in each specific case when designing a specific producer strain.

An additional advantage in strain engineering may be gained by using Hac orthologs from other yeasts, and even phylogenetically more distant eukaryotes. In particular, Valkonen et al. (2003) reported that the secretion of α-amylase was increased by overexpressing *Trichoderma reesei*-derived Hac1 in *S. cerevisiae*. Bankefa and colleagues (2018) showed that for *P. pastoris*, Hac1 orthologs of other species and even mammalian ones may be more effective than the native one. They investigate the effects of overexpressing Hac1 orthologs from *S. cerevisiae* (ScHac1p), *Trichoderma reesei* (TrHac1p) and *Homo sapiens* (HsXbp1) on the secretory expression levels of three reporter proteins, b-galactosidase, b-mannanase and glucose oxidase. The authors reported diverse effects of

these orthologs on heterologous expression levels, but HsXbp1 remarkably improved the enzyme production levels, both in shake flask and fermenter culture, both in single- and four-copy strains, which demonstrated its great application potential (Bankefa et al., 2018).

Thus, artificial activation of the UPR pathway by overexpression of the Hac1 transcription factor has demonstrated an obvious positive effect on improving the secretory protein productivity. It allowed to remove bottlenecks in the engineered yeast strain, arising due to abnormal accumulation of unfolded/misfolded proteins in the endoplasmic reticulum. However, this positive experience cannot be extrapolated to all recombinant proteins. Therefore, the real effect for each product must be assessed in experiments, often based on trial and error.

Prevention of ERAD pathway activation

As we wrote above, the UPR system is in crosstalk with the ERAD pathway. ERAD is activated when ER chaperones and folding enzymes are unable to form tertiary or quaternary structures of proteins. Sometimes ERAD is excessively activated during heterologous expression, so depletion of some components of this system can contribute to an increase in the yield of recombinant proteins.

De Ruijter and Frey (2015) analyzed the effect of deletions of genes involved in ERAD on the production of human IgG in *S. cerevisiae*. It was shown that deletion of only one gene, HTM1, contributed to a slight improvement in heterologous secretion, whereas deletions of the *yos9*, *hrd1*, *hrd3*, and *ubc7* genes either did not affect or negatively affected the recombinant protein yield (de Ruijter, Frey , 2015).

In *P. pastoris*, excessive activation of ERAD enhanced intracellular degradation of recombinant antibody fragment Fab. In the work of Pfeffer et al., it was shown that most of the newly synthesized Fab is not secreted but undergoes intracellular degradation via the ubiquitin-proteasome system (Pfeffer et al., 2012; Zahrl et al., 2019). The yield of the recombinant protein was increased by inhibiting proteasome components (Pfeffer et al., 2012). However, subsequent work aimed at reducing proteolysis through ERAD gene disruptions did not yield significant increases in Fab secretion (Zahrl et al., 2019).

Thus, the reduction of ERAD activity can be considered as a potential strategy for improving the secretion of recombinant proteins. However, the current level of research in this area does not yet allow for the transition to engineering producer strains protected from the proteolysis system.

Search for new solutions

The potential of the ER folding network can be enhanced by factors that are not directly involved in it, but create favorable conditions for its active functioning.

Zahrl et al. (2023) proposed to combine the transcriptional programs induced by Hac1 and Msn4 in one strain. Msn4 is a transcriptional factor involved in the response to various forms of stress (heat, oxidative, osmotic, etc.). Co-expression of Hac1 and Msn4 (both native and synthetic) revealed synergistic effects resulting in increased titers of recombinant proteins. This strategy was tested for scFv and VHH antibody fragments expressed in *P. pastoris* (Zahrl et al., 2023).

In another work, the same research group identified the most relevant chaperones of the Hsp70 network, both cytosolic and ER-localized, and investigated the impact of their combined overexpression on recombinant protein secretion (Zahrl et al., 2022). In their work , they implemented a principle they called the push-and-pull strategy. The addition of cytosolic chaperones allowed to increase the translocation competency of the recombinant protein and its targeting to the ER membrane (= push). At the same time resident ER chaperones improved the folding process (= pull). This allowed to successfully engineer strains and improve protein secretion up to 5-fold for the antibody Fab fragment and scFv (Zahrl et al., 2022).

The screening of new molecules involved in the folding system may improve yeast expression systems and expand the range of heterologous proteins. In particular, analysis of the reported *P. pastoris* secretome and genome predicted novel folding factors: Mpd1 and Pdi2 (members of the Pdi family), as well as Sil1 (nucleotide exchange factor for Kar2) (Duan et al., 2019). Subsequent experimental studies showed that all of the novel folding factors enhanced total production of reporter proteins, with Sil1 showing the highest efficiency (Duan et al., 2019). This work is an example of a successful combination of the achievements of yeast omics technologies and metabolic engineering, but this experience has not yet been widely applied.

Conclusion

One of the main limitations of heterologous protein production in yeast hosts is the ability of proteins to fold in the endoplasmic reticulum. The folding system is subject to unbalanced stress due to overexpression of recombinant genes, leading to the accumulation of misfolded proteins, aggregate formation, and low productivity. However, thanks to years of effective research into the fundamental mechanisms of protein folding, these limitations have been largely overcome. Studying folding in both model organisms and bioproducers has enabled the identification of molecular factors and cellular mechanisms that determine how a nascent polypeptide chain acquires its three-dimensional functional structure. This knowledge has formed the basis for the development of new, efficient methods for constructing highly productive yeast strains. Many problems arising from insufficient folding systems have been overcome by selecting optimal signal peptides, coexpressing with chaperones and foldases, modifying the ubiquitin-proteasome system (UPS), and preventing the ERAD pathway. Modern engineering solutions utilize combinations of these factors, but for each protein of interest, the expression strain is typically developed individually. In practice, optimized folding conditions for one protein often do not work for another. Therefore, no general strategy for overcoming protein folding bottlenecks that would be applicable to a wide range of proteins has yet been proposed.

In the future, some problems can be minimized by analyzing data obtained using omics technologies and modeling the secretion pathway *in silico*. An example of such a development is the pcSecYeast model designed for *S. cerevisiae* (Li F. et al., 2022). Such models allow choosing a combination of factors, both known and unknown, to generate new engineering strategies in designing strains with high protein yields.

References

Aza P., Molpeceres G., de Salas F., Camarero S. Design of an improved universal signal peptide based on the α-factor mating secretion signal for enzyme production in yeast. *Cell Mol Life Sci.* 2021;78(7): 3691-3707. doi 10.1007/s00018-021-03793-y

Bankefa O.E., Wang M., Zhu T., Li Y. Hac1p homologues from higher eukaryotes can improve the secretion of heterologous proteins in the yeast *Pichia pastoris*. *Biotechnol Lett*. 2018;40(7):1149-1156. doi 10.1007/s10529-018-2571-y

Bao C., Li J., Chen H., Sun Y., Wang G., Chen G., Zhang S. Expression and function of an *Hac1*-regulated multi-copy xylanase gene in *Saccharomyces cerevisiae*. *Sci Rep.* 2020;10(1):11686. doi 10.1038/s41598-020-68570-6

Barrero J.J., Casler J.C., Valero F., Ferrer P., Glick B.S. An improved secretion signal enhances the secretion of model proteins from *Pichia pastoris*. *Microb Cell Fact.* 2018;17(1):161. doi 10.1186/s12934-018-1009-5

Beal D.M., Bastow E.L., Staniforth G.L., von der Haar T., Freedman R.B., Tuite M.F. Quantitative analyses of the yeast oxidative protein folding pathway *in vitro* and *in vivo*. *Antioxid Redox Signal*. 2019;31(4):261-274. doi 10.1089/ars.2018.7615

Ben Azoun S., Belhaj A.E., Göngrich R., Gasser B., Kallel H. Molecular optimization of rabies virus glycoprotein expression in *Pichia pastoris. Microb Biotechnol*. 2016;9(3):355-368. doi 10.1111/1751-7915.12350

Berner N., Reutter K.R., Wolf D.H. Protein quality control of the endoplasmic reticulum and ubiquitin-proteasome-triggered degradation of aberrant proteins: yeast pioneers the path. *Annu Rev Biochem.* 2018;20(87):751-782. doi 10.1146/annurev-biochem-062917-012749

Braakman I., Hebert D.N. Protein folding in the endoplasmic reticulum. *Cold Spring Harb Perspect Biol*. 2013;5(5):a013201. doi 10.1101/cshperspect.a013201

Burgess R.R. Refolding solubilized inclusion body proteins. *Methods Enzymol*. 2009;463:259-282. doi 10.1016/S0076-6879(09)63017-2

Caramelo J.J., Parodi A.J. A sweet code for glycoprotein folding. *FEBS Lett*. 2015;589(22):3379-3387. doi 10.1016/j.febslet.2015.07.021

Damasceno L.M., Anderson K.A., Ritter G., Cregg J.M., Old L.J., Batt C.A. Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris*. *Appl Microbiol Biotechnol*. 2007;74(2):381-389. doi 10.1007/s00253-006-0652-7

De Brabander P., Uitterhaegen E., Delmulle T., De Winter K., Soetaert W. Challenges and progress towards industrial recombinant protein production in yeasts: a review. *Biotechnol Adv*. 2023;64: 108121. doi 10.1016/j.biotechadv.2023.108121

de Keyzer J., Steel G.J., Hale S.J., Humphries D., Stirling C.J. Nucleotide binding by Lhs1p is essential for its nucleotide exchange activity and for function *in vivo*. *J Biol Chem.* 2009;284(46):31564-31571. doi 10.1074/jbc.M109.055160

de Ruijter J.C., Frey A.D. Analysis of antibody production in *Saccharomyces cerevisiae*: effects of ER protein quality control disruption. *Appl Microbiol Biotechnol.* 2015;99:9061-9071. doi 10.1007/s00253-015-6807-7

de Ruijter J.C., Koskela E.V., Frey A.D. Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum. *Microb Cell Fact.* 2016;23(15):87. doi 10.1186/s12934-016-0488-5

Duan G., Ding L., Wei D., Zhou H., Chu J., Zhang S., Qian J. Screening endogenous signal peptides and protein folding factors to promote the secretory expression of heterologous proteins in *Pichia pastoris*. *J Biotechnol*. 2019;306:193-202. doi 10.1016/j.jbiotec.2019.06.297

Eskandari A., Nezhad N.G., Leow T.C., Rahman M.B.A., Oslan S.N. Current achievements, strategies, obstacles, and overcoming the challenges of the protein engineering in *Pichia pastoris* expression system. *World J Microbiol Biotechnol*. 2023;40(1):39. doi 10.1007/s11274-023-03851-6

Fauzee Y.N.B.M., Taniguchi N., Ishiwata-Kimata Y., Takagi H., Kimata Y. The unfolded protein response in *Pichia pastoris* without external stressing stimuli. *FEMS Yeast Res*. 2020;20(7):foaa053. doi 10.1093/femsyr/foaa053

Frand A.R., Kaiser C.A. The *Ero1* gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum. *Mol Cell.* 1998;1(2):161-170. doi 10.1016/s1097-2765(00)80017-9

Friedlander R., Jarosch E., Urban J., Volkwein C., Sommer T. A regulatory link between ER-associated protein degradation and the unfolded-protein response. *Nat Cell Biol.* 2000;2(7):379-384. doi 10.1038/35017001

Gasser B., Maurer M., Gach J., Kunert R., Mattanovich D. Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol Bioeng.* 2006;94(2):353-361. doi 10.1002/bit.20851

Gasser B., Saloheimo M., Rinas U., Dragosits M., Rodríguez-Carmona E., Baumann K., Giuliani M., … Porro D., Ferrer P., Tutino M.L., Mattanovich D., Villaverde A. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact*. 2008;4(7):11. doi 10.1186/1475-2859-7-11

Gross E., Sevier C.S., Heldman N., Vitu E., Bentzur M., Kaiser C.A., Thorpe C., Fass D. Generating disulfides enzymatically: reaction products and electron acceptors of the endoplasmic reticulum thiol oxidase Ero1p. *Proc Natl Acad Sci USA.* 2006;103(2):299-304. doi 10.1073/pnas.0506448103

Guerfal M., Ryckaert S., Jacobs P.P., Ameloot P., Van Craenenbroeck K., Derycke R., Callewaert N. The *HAC1* gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. *Microb Cell Fact.* 2010;9:49-60. doi 10.1186/1475-2859-9-49

Han M., Wang W., Gong X., Zhou J., Xu C., Li Y. Increased expression of recombinant chitosanase by co-expression of Hac1p in the yeast *Pichia pastoris*. *Protein Pept Lett.* 2021;28(12):1434-1441. doi 10.2174/0929866528666211105111155

Hartl F.U., Bracher A., Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. *Nature*. 2011;475(7356):324-332. doi 10.1038/nature10317

Hatahet F., Ruddock L.W. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. *Antioxid Redox Signal.* 2009;11(11):2807-2850. doi 10.1089/ars.2009.2466

Hendershot L.M., Buck T.M., Brodsky J.L. The essential functions of molecular chaperones and folding enzymes in maintaining endoplasmic reticulum homeostasis. *J Mol Biol.* 2024;436(14):168418. doi 10.1016/j.jmb.2023.168418

Hernández-Elvira M., Torres-Quiroz F., Escamilla-Ayala A., Domínguez-Martin E., Escalante R., Kawasaki L., Ongay-Larios L., Coria R. The unfolded protein response pathway in the yeast *Kluyveromyces lactis*. A comparative view among yeast species. *Cells.* 2018;7(8):106. doi 10.3390/cells7080106

Herscovics A. Processing glycosidases of *Saccharomyces cerevisiae*. *Biochim Biophys Acta.* 1999;1426(2):275-285. doi 10.1016/s0304-4165(98)00129-9

Huang J., Zhao Q., Chen L., Zhang C., Bu W., Zhang X., Zhang K., Yang Z. Improved production of recombinant *Rhizomucor miehei* lipase by coexpressing protein folding chaperones in *Pichia pastoris*, which triggered ER stress. *Bioengineered*. 2020;11(1):375-385. doi 10.1080/21655979.2020.1738127

Huang M., Gao Y., Zhou X., Zhang Y., Cai M. Regulating unfolded protein response activator HAC1p for production of thermostable raw-starch hydrolyzing α-amylase in *Pichia pastoris*. *Bioprocess Biosyst Eng.* 2017;40(3):341-350. doi 10.1007/s00449-016-1701-y

Inan M., Fanders S.A., Zhang W., Hotez P.J., Zhan B., Meagher M.M. Saturation of the secretory pathway by overexpression of a hookworm (*Necator americanus*) protein (Na-ASP1). In: Cregg J.M. (Ed.) Pichia Protocols. Methods in Molecular Biology. Vol. 389. Humana Press, 2007;65-76. doi 10.1007/978-1-59745-456-8_5

Ishiwata-Kimata Y., Kimata Y. Fundamental and applicative aspects of the unfolded protein response in yeasts. *J Fungi* (*Basel*). 2023; 9(10):989. doi 10.3390/jof9100989

Ito Y., Ishigami M., Hashiba N., Nakamura Y., Terai G., Hasunuma T., Ishii J., Kondo A. Avoiding entry into intracellular protein degradation pathways by signal mutations increases protein secretion in *Pichia pastoris*. *Microb Biotechnol*. 2022;15(9):2364-2378. doi 10.1111/1751-7915.14061

Kelleher D.J., Gilmore R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology*. 2006;16(4):47R-62R. doi 10.1093/glycob/cwj066

Khlebodarova T.M., Bogacheva N.V., Zadorozhny A.V., Bryanskaya A.V., Vasilieva A.R., Chesnokov D.O., Pavlova E.I., Peltek S.E. *Komagataella phaffii* as a platform for heterologous expression of enzymes used for industry. *Microorganisms*. 2024;12(2):346. doi 10.3390/microorganisms12020346

Kim M.D., Han K.C., Kang H.A., Rhee S.K., Seo J.H. Coexpression of BiP increased antithrombotic hirudin production in recombinant *Saccharomyces cerevisiae*. *J Biotechnol*. 2003;101(1):81-87. doi 10.1016/s0168-1656(02)00288-2

Kimmig P., Diaz M., Zheng J., Williams C.C., Lang A., Aragón T., Li H., Walter P. The unfolded protein response in fission yeast modulates stability of select mRNAs to maintain protein homeostasis. *eLife*. 2012;1:e00048. doi 10.7554/eLife.00048

Korennykh A., Walter P. Structural basis of the unfolded protein response. *Annu Rev Cell Dev Biol*. 2012;28:251-277. doi 10.1146/annurev-cellbio-101011-155826

Krshnan L., van de Weijer M.L., Carvalho P. Endoplasmic reticulum-associated protein degradation. *Cold Spring Harb Perspect Biol*. 2022;14(12):a041247. doi 10.1101/cshperspect.a041247

Lee T.H., Bae Y.H., Kim M.D., Seo J.H. Overexpression of Hac1 gene increased levels of both intracellular and secreted human kringle fragment in *Saccharomyces cerevisiae*. *Process Biochem*. 2012;47(12):2300-2305. doi 10.1016/j.procbio.2012.09.006

Lehle L., Strahl S., Tanner W. Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. *Angew Chem Int Ed Engl*. 2006;45(41):6802-6818. doi 10.1002/anie.200601645

Li C., Lin Y., Zheng X., Pang N., Liao X., Liu X., Huang Y., Liang S. Combined strategies for improving expression of *Citrobacter amalonaticus* phytase in *Pichia pastoris*. *BMC Biotechnol*. 2015;15:88. doi 10.1186/s12896-015-0204-2

Li F., Chen Y., Qi Q., Wang Y., Yuan L., Huang M., Elsemman I.E., Feizi A., Kerkhoven E.J., Nielsen J. Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints. *Nat Commun*. 2022;13(1):2969. doi 10.1038/s41467-022-30689-7

Lin Y., Feng Y., Zheng L., Zhao M., Huang M. Improved protein production in yeast using cell engineering with genes related to a key factor in the unfolded protein response. *Metab Eng*. 2023;77:152-161. doi 10.1016/j.ymben.2023.04.004

Madhavan A., Arun K.B., Sindhu R., Krishnamoorthy J., Reshmy R., Sirohi R., Pugazhendi A., Awasthi M.K., Szakacs G., Binod P. Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up. *Microb Cell Fact*. 2021;20(1):124. doi 10.1186/s12934-021-01617-z

Mizunaga T., Katakura Y., Miura T., Maruyama Y. Purification and characterization of yeast protein disulfide isomerase. *J Biochem*. 1990;108(5):846-851. doi 10.1093/oxfordjournals.jbchem.a123291

Mori A., Hara S., Sugahara T., Kojima T., Iwasaki Y., Kawarasaki Y., Sahara T., Ohgiya S., Nakano H. Signal peptide optimization tool for the secretion of recombinant protein from *Saccharomyces cerevisiae*. *J Biosci Bioeng*. 2015;120(5):518-525. doi 10.1016/j.jbiosc.2015.03.003

Mori K. Evolutionary aspects of the unfolded protein response. *Cold Spring Harb Perspect Biol*. 2022;14(12):a041262. doi 10.1101/cshperspect.a041262

Niu Y., Zhang L., Yu J., Wang C.C., Wang L. Novel roles of the non-catalytic elements of yeast protein-disulfide isomerase in its interplay with endoplasmic reticulum oxidoreductin 1. *J Biol Chem*. 2016;291(15):8283-8294. doi 10.1074/jbc.M115.694257

Núñez A., Dulude D., Jbel M., Rokeach L.A. Calnexin is essential for survival under nitrogen starvation and stationary phase in *Schizosaccharomyces pombe*. *PLoS One*. 2015;10(3):e0121059. doi 10.1371/journal.pone.0121059

O'Keefe S., Pool M.R., High S. Membrane protein biogenesis at the ER: the highways and byways. *FEBS J*. 2022;289(22):6835-6862. doi 10.1111/febs.15905

Omkar S., Mitchem M.M., Hoskins J.R., Shrader C., Kline J.T., Nitika, Fornelli L., Wickner S., Truman A.W. Acetylation of the yeast Hsp40 chaperone protein Ydj1 fine-tunes proteostasis and translational fidelity. *PLoS Genet*. 2024;20(12):e1011338. doi 10.1371/journal.pgen.1011338

Palma A., Rettenbacher L.A., Moilanen A., Saaranen M., Pacheco-Martinez C., Gasser B., Ruddock L. Biochemical analysis of *Komagataella phaffii* oxidative folding proposes novel regulatory mechanisms of disulfide bond formation in yeast. *Sci Rep*. 2023;13(1):14298. doi 10.1038/s41598-023-41375-z

Parlati F., Dominguez M., Bergeron J.J., Thomas D.Y. Saccharomyces cerevisiae CNE1 encodes an endoplasmic reticulum (ER) membrane protein with sequence similarity to calnexin and calreticulin and functions as a constituent of the ER quality control apparatus. *J Biol Chem*. 1995;270(1):244-253. doi 10.1074/jbc.270.1.244

Parodi A.J. Protein glucosylation and its role in protein folding. *Annu Rev Biochem*. 2000;69:69-93. doi 10.1146/annurev.biochem.69.1.69

Payne T., Finnis C., Evans L.R., Mead D.J., Avery S.V., Archer D.B., Sleep D. Modulation of chaperone gene expression in mutagenized *Saccharomyces cerevisiae* strains developed for recombinant human albumin production results in increased production of multiple heterologous proteins. *Appl Environ Microbiol*. 2008;74(24):7759-7766. doi 10.1128/AEM.01178-08

Pfeffer M., Maurer M., Stadlmann J., Grass J., Delic M., Altmann F., Mattanovich D. Intracellular interactome of secreted antibody Fab fragment in *Pichia pastoris* reveals its routes of secretion and degradation. *Appl Microbiol Biotechnol*. 2012;93(6):2503-2512. doi 10.1007/s00253-012-3933-3

Pickart C.M. Mechanisms underlying ubiquitination. *Annu Rev Biochem*. 2001;70:503-533. doi 10.1146/annurev.biochem.70.1.503

Preston G.M., Brodsky J.L. The evolving role of ubiquitin modification in endoplasmic reticulum-associated degradation. *Biochem J*. 2017;474(4):445-469. doi 10.1042/BCJ20160582

Qi Q., Li F., Yu R., Engqvist M.K.M., Siewers V., Fuchs J., Nielsen J. Different routes of protein folding contribute to improved protein production in *Saccharomyces cerevisiae*. *mBio*. 2020;11(6):e02743-20. doi 10.1128/mBio.02743-20

Raschmanová H., Weninger A., Knejzlík Z., Melzoch K., Kovar K. Engineering of the unfolded protein response pathway in *Pichia pastoris*: enhancing production of secreted recombinant proteins. *Appl Microbiol Biotechnol*. 2021;105(11):4397-4414. doi 10.1007/s00253-021-11336-5

Robinson A.S., Hines V., Wittrup K.D. Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Nat Biotechnol*. 1994;12(4):381-384. doi 10.1038/nbt0494-381

Ruger-Herreros C., Svoboda L., Mogk A., Bukau B. Role of J-domain proteins in yeast physiology and protein quality control. *J Mol Biol*. 2024;436(14):168484. doi 10.1016/j.jmb.2024.168484

Ruggiano A., Foresti O., Carvalho P. Quality control: ER-associated degradation: protein quality control and beyond. *J Cell Biol*. 2014;204(6):869-879. doi 10.1083/jcb.201312042

Saibil H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol*. 2013;14(10):630-642. doi 10.1038/nrm3658

Sallada N.D., Harkins L.E., Berger B.W. Effect of gene copy number and chaperone coexpression on recombinant hydrophobin HFBI biosurfactant production in *Pichia pastoris*. *Biotechnol Bioeng*. 2019;116(8):2029-2040. doi 10.1002/bit.26982

Schlenstedt G., Harris S., Risse B., Lill R., Silver P.A. A yeast DnaJ homologue, Scj1p, can function in the endoplasmic reticulum with

BiP/Kar2p via a conserved domain that specifies interactions with Hsp70s. *J Cell Biol.* 1995;129(4):979-988. doi 10.1083/jcb.129.4.979

Schroder M., Clark R., Kaufman R.J. *IRE1*- and *HAC1*-independent transcriptional regulation in the unfolded protein response of yeast. *Mol Microbiol.* 2003;49(3):591-606. doi 10.1046/j.1365-2958.2003.03585.x

Schuck S., Prinz W.A., Thorn K.S., Voss C., Walter P. Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. *J Cell Biol.* 2009;187(4):525-536. doi 10.1083/jcb.200907074

Schwarz D.S., Blower M.D. The endoplasmic reticulum: structure, function and response to cellular signaling. *Cell Mol Life Sci.* 2016;73(1):79-94. doi 10.1007/s00018-015-2052-6

Sevier C.S., Kaiser C.A. Disulfide transfer between two conserved cysteine pairs imparts selectivity to protein oxidation by Ero1. *Mol Biol Cell.* 2006;17(5):2256-2266. doi 10.1091/mbc.e05-05-0417

Sheng J., Flick H., Feng X. Systematic optimization of protein secretory pathways in *Saccharomyces cerevisiae* to increase expression of hepatitis B small antigen. *Front Microbiol.* 2017;8:875. doi 10.3389/fmicb.2017.00875

Shusta E.V., Raines R.T., Plückthun A., Wittrup K.D. Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat Biotechnol.* 1998;16(8):773-777. doi 10.1038/nbt0898-773

Singhvi P., Saneja A., Ahuja R., Panda A.K. Solubilization and refolding of variety of inclusion body proteins using a novel formulation. *Int J Biol Macromol*. 2021;193(Pt.B):2352-2364. doi 10.1016/j.ijbiomac.2021.11.068

Smith J.D., Tang B.C., Robinson A.S. Protein disulfide isomerase, but not binding protein, overexpression enhances secretion of a non-disulfide-bonded protein in yeast. *Biotechnol Bioeng.* 2004;85(3):340-350. doi 10.1002/bit.10853

Steel G.J., Fullerton D.M., Tyson J.R., Stirling C.J. Coordinated activation of Hsp70 chaperones. *Science*. 2004;303(5654):98-101. doi 10.1126/science.1092287

Thak E.J., Yoo S.J., Moon H.Y., Kang H.A. Yeast synthetic biology for designed cell factories producing secretory recombinant proteins. *FEMS Yeast Res.* 2020;20(2):foaa009. doi 10.1093/femsyr/foaa009

Thibault G., Ng D.T.W. The endoplasmic reticulum-associated degradation pathways of budding yeast. *Cold Spring Harb Perspect Biol*. 2012;4(12):a013193. doi 10.1101/cshperspect.a013193

Travers K.J., Patil C.K., Wodicka L., Lockhart D.J., Weissman J.S., Walter P. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell*. 2000;101(3):249-258. doi 10.1016/s0092-8674(00)80835-1

Tsuda M., Nonaka K. Recent progress on heterologous protein production in methylotrophic yeast systems. *World J Microbiol Biotechnol.* 2024;40(7):200. doi 10.1007/s11274-024-04008-9

Valkonen M., Penttila M., Saloheimo M. Effects of inactivation and constitutive expression of the unfolded-protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 2003;69(4):2065-2072. doi 10.1128/aem.69.4.2065-2072.2003

Wang L., Wang C.C. Oxidative protein folding fidelity and redoxtasis in the endoplasmic reticulum. *Trends Biochem Sci.* 2023;48(1):40-52. doi 10.1016/j.tibs.2022.06.011

Ware F.E., Vassilakos A., Peterson P.A., Jackson M.R., Lehrman M.A., Williams D.B. The molecular chaperone calnexin binds Glc1Man9GlcNAc2 oligosaccharide as an initial step in recognizing unfolded glycoproteins. *J Biol Chem.* 1995;270(9):4697-4704. doi 10.1074/jbc.270.9.4697

Xia X. Translation control of *HAC1* by regulation of splicing in *Saccharomyces cerevisiae*. *Int J Mol Sci.* 2019;20(12):2860. doi 10.3390/ijms20122860

Xu C., Ng D.T.W. Glycosylation-directed quality control of protein folding. *Nat Rev Mol Cell Biol.* 2015;16:742-752. doi 10.1038/nrm4073

Xu X., Kanbara K., Azakam H., Kato A. Expression and characterization of *Saccharomyces cerevisiae* Cne1p, a calnexin homologue. *J Biochem.* 2004;135(5):615-618. doi 10.1093/jb/mvh074

Yamaguchi H., Miyazaki M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules.* 2014;4(1):235-251. doi 10.3390/biom4010235

Yang J., Lu Z., Chen J., Chu P., Cheng Q., Liu J., Ming F., Huang C., Xiao A., Cai H., Zhang L. Effect of cooperation of chaperones and gene dosage on the expression of porcine PGLYRP-1 in *Pichia pastoris*. *Appl Microbiol Biotechnol*. 2016;100(12):5453-5465. doi 10.1007/s00253-016-7372-4

Zahrl R.J., Gasser B., Mattanovich D., Ferrer P. Detection and elimination of cellular bottlenecks in protein-producing yeasts. In: Gasser B., Mattanovich D. (Eds) Recombinant Protein Production in Yeast. Methods in Molecular Biology. Vol. 1923. Humana Press, 2019;75-95. doi 10.1007/978-1-4939-9024-5_2

Zahrl R.J., Prielhofer R., Ata Ö., Baumann K., Mattanovich D., Gasser B. Pushing and pulling proteins into the yeast secretory pathway enhances recombinant protein secretion. *Metab Eng.* 2022;74:36-48. doi 10.1016/j.ymben.2022.08.010

Zahrl R.J., Prielhofer R., Burgard J., Mattanovich D., Gasser B. Synthetic activation of yeast stress response improves secretion of recombinant proteins. *N Biotechnol.* 2023;73:19-28. doi 10.1016/j.nbt.2023.01.001

Zha J., Liu D., Ren J., Liu Z., Wu X. Advances in metabolic engineering of *Pichia pastoris* strains as powerful cell factories. *J Fungi* (*Basel*). 2023;9(10):1027. doi 10.3390/jof9101027

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

Received September 2, 2025. Revised November 25, 2025. Accepted November 25, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1348-1360

doi 10.18699/vjgb-25-141

Cellulases: key properties, natural sources, and industrial applications

A.V. Zadorozhny, N.M. Slynko, S.V. Bannikova, N.V. Bogacheva, V.N. Shlyakhtun, A.R. Vasilieva, E.Yu. Bukatich, V.S. Ushakov, Yu.E. Uvarova, A.V. Korzhuk, A.A. Shipova, D.V. Bochkov, E.Y. Pavlova, D.O. Chesnokov, S.E. Peltek

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

Abstract. This review focuses on cellulases, a subclass of hydrolases that catalyse the breakdown of the polysaccharide cellulose. Cellulases are of immense practical significance, given that cellulose-containing materials are utilised across a multitude of industrial sectors. An overview of the fundamental properties and structure of cellulases is provided. However, primary attention is paid to the industrial application of these enzymes, with other aspects discussed within this context. The most practically significant bacterial and fungal cellulases are analysed, with their key benefits and differences being emphasised. Particular attention is paid to extremophilic (specifically thermo-, psychro-, and halophilic) cellulases, as they possess properties essential for modern technological processes. Given that practical application necessitates mass production and an optimal combination of enzymatic characteristics, the creation of effective producers and the modification of cellulase properties are also assessed. Finally, key trends in cellulase production approaches and their future application potential are summarised.

Key words: celulase; fungal and bacterial cellulases; extremophilic cellulases; celulase structure and essential properties; cellulase complex; biotechnology; genetic engineering

For citation: Zadorozhny A.V., Slynko N.M., Bannikova S.V., Bogacheva N.V., Shlyakhtun V.N., Vasilieva A.R., Bukatich E.Yu., Ushakov V.S., Uvarova Yu.E., Korzhuk A.V., Shipova A.A., Bochkov D.V., Pavlova E.Y., Chesnokov D.O., Peltek S.E. Cellulases: key properties, natural sources, and industrial applications. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1348-1360. doi 10.18699/vjgb-25-141

Funding. This work was supported by project No. FWNR-2025-0029, "Development of technologies for scaling up the production of feed enzymes secreted by *Komagataella phaffii* and obtaining enzyme preparations based on them".

Целлюлаза: основные свойства, природные источники и применение в промышленности

А.В. Задорожный, Н.М. Слынько, С.В. Банникова, Н.В. Богачева, В.Н. Шляхтун, А.Р. Васильева, Е.Ю. Букатич, В.С. Ушаков, Ю.Е. Уварова, А.В. Коржук, А.А. Шипова, Д.В. Бочков, Е.Ю. Павлова, Д.О. Чесноков, С.Е. Пельтек

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

Аннотация. В последние годы целлюлаза привлекает огромное внимание как промышленно важный фермент с широким спектром применения. Целлюлазы представляют собой сложную группу ферментов, которые секретируются многообразием микроорганизмов, включая грибы и бактерии. Они относятся к подклассу ферментов гидролаз, субстратом которых является полисахарид целлюлоза. Целлюлазы имеют огромное практическое значение, поскольку содержащие целлюлозу материалы используются во множестве отраслей народного хозяйства. В этом обзоре приведены сведения об основных свойствах и структуре целлюлаз. Но основное внимание уделено применению этих ферментов в промышленности, а прочие аспекты, так или иначе, рассматриваются с учетом этого. Исследованы имеющие наибольшее практическое значение бактериальные и грибные целлюлазы, их основные преимущества и отличия. Отдельно рассмотрены экстремофильные (а именно термо-, психро- и галофильные) целлюлазы как обладающие свойствами, нужными в условиях современных технологических процессов. Поскольку для практического применения необходимы массовая продукция и оптимальное сочетание свойств ферментов, внимание также уделено получению эффективных продуцентов и модификации свойств, производимых целлюлаз. Наконец, обозначены ключевые тенденции в подходах к производству целлюлаз и перспективы практического применения.

Ключевые слова: целлюлаза; грибные и бактериальные целлюлазы; экстремофильные целлюлазы; структура и основные свойства целлюлаз; целлюлазный комплекс; биотехнология; генная инженерия

Introduction

The rise in environmental consciousness and progress in biotechnology have facilitated the replacement of numerous chemical procedures by enzymatic biocatalysts in a variety of industries, including textiles, leather, pulp and paper, fruit and vegetable processing, food processing, and feed production (Wackett, 2019). Enzymes can be isolated from various sources: animal tissues, plants, and microbial cells. Microbial proteins are frequently more stable than enzymes of similar specificity derived from plant or animal sources and can often be stored under less-than-ideal conditions for extended periods without significant loss of biological activity (Singhania et al., 2010). Most commercial enzymes are obtained from microorganisms. These enzymes are strong candidates for efficient biotechnological processes due to properties like thermostability, broad pH range stability, and multifunctionality, which allow them to function under various physicochemical conditions.

Cellulases represent the second largest group of industrial enzymes by market share, and their usage continues to grow alongside increasing demand from a wide range of sectors: food, pulp and paper, textiles, pharmaceuticals, detergents, animal feed, biofuels, and waste management (Ranjan et al., 2023).

Cellulases, similar to other enzymes, are sourced from plants, microorganisms, and animals. Fungi and bacteria are most often used in the production of these enzymes because of traits such as considerable yield and cost reduction. Historically, the principal producers of cellulases have been natural strains of fungi, and their productivity has been enhanced through selection and/or mutagenesis methods. However, the advent and development of recombinant DNA technology, genetic engineering, protein engineering, directed evolution, omics technologies, and high-throughput sequencing have led to the discovery of novel microbes and enzymes for industrial applications (Patel et al., 1994; Kirk et al., 2002; Rubin-Pitel, Zhao, 2006) alongside the development of various recombinant microbial cellulases. These enzymes are of particular significance due to their specific attributes, including cost-effectiveness, energy-efficient catalytic processes, environmental compatibility, non-toxicity, and high efficiency.

In numerous scenarios, it is advantageous to use cellulases that exhibit extremophilic properties, including stability and efficiency across a range of temperatures, pH levels, pressures, or in the presence of organic solvents, detergents, and elevated ionic strengths within the operational environment. Enzymes demonstrating these attributes are typically discovered in research on microorganisms within suitable environments, whereas enzymes with uncommon properties are also observed in mesophilic organisms. The increasing demand has led to a gradual expansion in the selection of cellulases with a wide range of properties. The application of enzymes, which are stable under extreme conditions, will optimise outcomes while minimising enzyme consumption.

Cellulases: key properties and structure

Fungi and bacteria are both employed in the production of various cellulases. Fungi have long been a primary focus due to their ability to secrete significant quantities of enzymes. The recent shift towards bacteria is attributed to their rapid growth rates, multifunctional enzymes, and their ubiquity across diverse ecological niches. Bacteria are not only capable of surviving in harsh conditions but frequently produce stable enzymes that can accelerate catalytic processes more effectively than their fungal counterparts.

Cellulose, a natural polymer, serves as the substrate for cellulase. Cellulases are enzymatic systems that hydrolyse β-1,4-glycoside bonds found within cellulose polymers and its derivatives, yielding soluble oligosaccharides and glucose monomers (Nishida et al., 2007). The biochemical degradation of the cellulose molecule by microorganisms is catalysed by an extracellular enzymatic system comprising three key components (Bhat M.K., Bhat S., 1997). These are: (1) β-1,4-glucan glucanohydrolase (endoglucanase; EC 3.2.1.4), which cleaves the long cellulose chain into shorter fragments; (2) β-1,4-glucan cellobiohydrolase (exoglucanase; EC 3.2.1.91), which acts upon the non-reducing end of the cellulose chain; and (3) β-1,4-glucosidase (EC 3.2.1.21), which breaks the glycosidic bonds of cellobiose and cellodextrins, producing glucose molecules that can readily permeate the cell. In nature, the synergistic action of all three enzymes is required for complete hydrolysis of the cellulose polymer to glucose units (Uhlig, 1998).

The complex spatial structure of the cellulase complex facilitates the positioning of cellulose fibers near the active sites of the enzymes. Cellulase systems are categorised into two distinct types. The first type is exemplified by the extracellular cellulases of filamentous fungi and aerobic bacteria, which act synergistically to decompose cellulose. The second type, found in anaerobic clostridia, attaches to the bacterial cell surface, known as the “cellulosome”.

A typical characteristic of cellulases is a two-domain structure, which includes a catalytic domain and a cellulose-binding domain (CBD), also referred to as a carbohydrate-binding module (CBM), which are usually connected by a peptide linker. The active site is found in the catalytic domain, with the CBD ensuring accurate cellulose fiber placement (Mathew et al., 2008).

Cellulose-degrading aerobic fungi, such as *Hypocrea jecorina* (*Trichoderma reesei*), produce an enzyme complex that acts synergistically and displays all three primary activities: endoglucanase, exoglucanase, and β-glucosidase (Dashtban et al., 2011; Mukherjee et al., 2012). The process begins with endoglucanases cleaving cellulose fibers in the more amorphous regions. Subsequently, exoglucanases can cleave cellobiose molecules from the more crystalline sections of the fibre that were previously inaccessible. Finally, β-glucosidases hydrolyse the cellobiose.

Most anaerobic microorganisms have their cellulosolytic system organised differently. The whole complex of various cellulases and hemicellulases is integrated into a multi-

protein complex, the cellulosome, that is attached to the external part of the cell through the non-catalytic protein scaffoldin. The molecular weight of cellulosomes can reach several thousand kDa (Pinheiro et al., 2009). The surface of a single *Clostridium thermocellum* cell (the most extensively studied cellulolytic bacterium capable of rapid growth on cellulosic substrates) hosts multiple cellulosomes. This ensures the secure attachment of the microorganism to the cellulose fibre, while the spatial arrangement of the various enzymes within the cellulosome facilitates their access to the substrate. Furthermore, clostridia also produce free cellulases not associated with cellulosomes. The presence of a synergistic effect with cellulosomes remains undetermined (Doi, Tamaru, 2001).

This organisational structure of the cellulosolytic system results in simple sugars being produced in the immediate vicinity of the microorganism, significantly facilitating nutrient uptake. In certain scenarios, anaerobic bacteria were observed to regulate cellulosome binding on the cell surface, resulting in free cellulosomes being detected, and to alter the production of various cellulosome components depending on growth conditions (Mohand-Oussaid et al., 1999).

Besides including the complex of cellulolytic enzymes, the cellulosome also comprises carbohydrate fibrils, which can account for up to 90 % of the molecular mass of the cellulosome. They promote the sorption of the enzyme onto the substrate and the sliding of the enzyme along the fibrillar structures of cellulose. Furthermore, the carbohydrate portion protects the protein from the action of denaturing agents and proteases.

The efficacy of different cellulase complexes in cellulose hydrolysis varies significantly when the substrate possesses high crystallinity. A significant reduction in activity occurs in numerous complexes when crystallinity reaches 60–70 %, with hydrolysis limited to the amorphous portion of the substrate. The effectiveness of these complexes in hydrolysing cellulose with high crystallinity is contingent upon the presence of endoglucanases capable of strong adsorption. Consequently, the rate of crystalline cellulose hydrolysis is directly influenced by the quantity of endoglucanase adsorbed.

Industrial applications of cellulases

Cellulases are in high demand across a wide range of industrial sectors. In the food industry, cellulase is traditionally employed for the extraction and clarification of juices (Azman et al., 2021; Ozyilmaz, Gunay, 2023). In addition, it allows the nutritional characteristics of brown rice to be enhanced. Q. Zhang et al. (2019) demonstrated that cyclic treatment with cellulase combined with germination is an effective method for increasing GABA levels in brown rice while improving its culinary and sensory qualities.

Additionally, cellulase is employed to improve dough quality (Hu X. et al., 2022). Adding potato to wheat flour increases the nutritional content of bread. However, the adverse effects caused by the high dietary fibre content of potato flour can impair gluten matrix formation. The addition of cellulase and/or pectinase promotes loaf volume and softness.

The application of cellulase for oil extraction in the food and cosmetic industries also holds promise. D. Yu et al. (2022) introduced a method for extracting oil from rice bran using magnetically immobilised cellulase in combination with magnetically immobilised alkaline protease. Similarly, M.O. Chiwetalu et al. (2022) presented an extraction method for obtaining fat from *Pycnanthus angolensis* seeds via pre-treatment with an enzyme from *Aspergillus niger* strain BC23.

Significant progress has recently been made in lignocellulose biorefinery technologies for the production of biochemicals and biofuels, using cellulase enzymes as a key component (Sharma et al., 2023).

A. Shankar et al. (2024) pre-treated rice straw, rice bran, wheat straw, wood chips, sorghum bagasse, and cotton stalks with the basidiomycete fungus *Ganoderma lucidum*. Due to its laccase activity, the fungus significantly degraded the lignin present in the biomass. The saccharification of pre-treated biomass using cellulase consortia, specifically those isolated from *Aspergillus flavus* MDU-5 and *Trichoderma citrinoviride* MDU-1, resulted in an approximate 70 % increase in saccharide yield and an 89 % increase in ethanol yield. Additionally, H. Sha et al. (2023) developed a novel approach to the anaerobic digestion of corn straw using magnetic cellulase coated with nickel and graphite. This system increased methane production by approximately 74 % compared to an uncoated system. Additionally, an increase in the combined population of electroactive *Bacteroidota* and *Methanomicrobiales* and improved energy conversion efficiency by up to 57 % were reported.

E. Zanuso et al. (2022) demonstrated the efficacy of using cellulase immobilised on magnetic nanoparticles for the hydrolysis of corn cob biomass. Moreover, the magnetic properties of the carrier make this method promising for continuous operation, contributing to reduced overall process costs.

In the pulp and paper industry, cellulase is used in waste-paper recycling (deinking) and the pre-treatment of raw wood materials for pitch removal (Chutani, Sharma, 2016; Lehr et al., 2021; Singh A. et al., 2021), thereby reducing the environmental impact of the industry.

In the pharmaceutical industry, cellulase is employed to extract biologically active compounds from plant raw materials (Puri et al., 2012; Cao et al., 2019; Hu Y. et al., 2021). It is also widely used in the production of enzyme tablet formulations intended to facilitate the digestion of plant-based foods rich in fiber. VeganZyme, for example, is used both to enhance digestion and to manage metabolic disorders (Ranjan et al., 2023). Furthermore, cellulase and papain are used to treat phytobezoars (Iwamuro et al., 2014).

Chemical decomposition and incineration are the most frequently employed techniques for the remediation of envi-

ronmental pollutants. However, the application of microbial enzymes for purification offers a more straightforward and eco-friendly method (Okino-Delgado et al., 2019). Given that *Bacillus cereus* and *B. subtilis* can produce cellulase and lipase, enzymes known for their efficacy in waste degradation, these bacteria were identified by as potentially useful in treating palm oil mill waste (Ranjan et al., 2023).

The study by J. Luo et al. (2021) used cellulase in the simultaneous fermentation of sewage sludge and paper waste, resulting in the production of volatile fatty acids. All these findings collectively suggest the applicability of cellulase for wastewater treatment and its potential for producing value-added products.

Cellulase finds broad application in the manufacture of animal feed. Cellulase and hemicellulase facilitate the hydrolysis of low-protein feeds and β-glucans, consequently improving the nutritional characteristics of the feed. The application of cellulase also allows duodenal viscosity to be reduced and feed consistency to be improved, consequently enhancing digestion and nutrient uptake by animals (Azzaz et al., 2021; Selzer et al., 2021). In poultry farming, the use of cellulase is a viable strategy as it degrades cellulosic bonds, releasing nutrients such as glucose, increasing the energy value of the diet, and thus improving bird performance (Perim et al., 2024).

In the textile industry, cellulase is used to improve the fabric product finish by modifying protruding fibres. Cellulase treatment reduces fabric roughness, increasing smoothness, gloss, and colour brightness (Karmakar, Ray, 2011; Sajith et al., 2016).

The inclusion of cellulase in detergents, alongside proteases and lipases, is intended to enhance washing efficacy. It helps maintain the shape and colour of laundered garments by reducing the formation of fuzz and pills. Additionally, the enzyme contributes to the soil and stain removal process by selectively acting on cellulose fibres from the interior. The cellulase enzyme breaks the bonds between cellulose and dirt particles and alters the fibre surface structure, thereby facilitating soil removal by other detergent ingredients. Currently, both broad-temperature-range (30–60 °C) and mesophilic cellulases are added to detergents (Kasana, Gulati, 2011). Cellulase enzymes are used in detergent compositions to provide cleaning, softening, and colour retention. However, the use of most cellulases has been limited due to the potential negative impact on fabric tensile strength resulting from the hydrolysis of crystalline cellulose. Recently, cellulases with high specificity towards amorphous cellulose have been developed to exploit their cleaning potential without undesirable loss of fabric strength.

Fungal cellulases

Cellulolytic fungi of the genus *Trichoderma* have long been considered the premier source of cellulases (Reese, Mandels, 1963). However, the primary bottleneck associated with *Trichoderma* cellulases is the very low β-glucosidase activity in culture supernatants, coupled with product inhibition of this enzyme (Pérez et al., 2002). Conversely, cellulases produced by the thermophilic fungi *Sporotrichum thermophile* and *Talaromyces emersonii* exhibit activity comparable to that of the mesophilic fungus *H. jecorina* (Coutts, Smith, 1976; Folan, Coughlan, 1978).

Humicola insolens demonstrates exceptional production capabilities of pH-neutral, thermostable cellulases that are industrially relevant (Xu X. et al., 2016).

P. Chellapandi and H.M. Jani investigated the endoglucanase activity of 26 *Streptomyces* strains isolated from garden soil (Chellapandi, Jani, 2008). The two most promising isolates, selected for their potential cellulolytic activity on Bennett's agar medium, were assessed under varying conditions, including carbon and nitrogen sources, and growth conditions. Maximum endoglucanase activity (11.25–11.90 U/ml) was achieved after 72–88 hours in a fermentation medium containing Tween-80, followed by the utilisation of phosphate sources. Both cellulolytic *Streptomyces* isolates produced nearly identical quantities of enzyme in all trials. However, the influence of media ingredients on endoglucanase induction differed somewhat between the strains. Like mesophilic fungi, thermophilic fungi produce all components of the cellulase complex, synergistically degrading cellulose and hemicellulose (Mathew et al., 2008).

Fungi are a source of vital extracellular enzymes used in industrial applications. Genera such as *Trichoderma*, *Penicillium*, and *Aspergillus* are especially recognised in this regard (Zhao C.H. et al., 2018). In addition, fungi are widely used in the production of industrial cellulases because they possess useful characteristics such as the extracellular secretion of the enzyme in large quantities using economical substrates (Niyonzima, 2021). Enzymes can be produced using substrates originating from agricultural byproducts. For example, maize stalks and sugarcane bagasse have been used to produce detergent-compatible cellulases by *Aspergillus* fungi (Imran et al., 2018; El-Baroty et al., 2019). B.R. Dave et al. (2012) used readily available, low-cost de-oiled *Jatropha* seed cake to obtain cellulase from *Thermoascus aurantiacus* RBB-1. Compared to its amorphous or mixed forms, crystalline cellulose has been observed to be a more effective carbon source for cellulase production in thermophilic fungi (Fracheboud, Ganevascini, 1989).

Fungal enzyme genes are easily cloned into bacterial strains for cellulase production, as fungal enzymes are structurally less complex than their bacterial counterparts (Maki et al., 2009; Acharya, Chaudhary, 2012). However, their primary advantage remains that cellulases can be obtained from fungi using relatively simple methods, and the fungi themselves can produce cellulases when cultivated on inexpensive substrates. Moreover, because they produce a full spectrum of cellulases, filamentous fungi remain the most popular choice for industrial cellulase production (Ilić et al., 2023).

Bacterial cellulases

Several studies have demonstrated that bacterial cellulases possess significant advantages over fungal cellulases in specific activity and stability (Ejaz et al., 2021). Bacteria exhibit high growth rates, are adapted to diverse ecological niches, and are amenable to genetic manipulation (Nyathi et al., 2023). Numerous cellulosolytic bacterial strains have been identified, which produce specific enzymes that exhibit resistance to extreme conditions (Bhati et al., 2021). Rapid advancements in bacterial cellulase research indicate a more diverse genetic makeup than fungal cellulases, which are currently more extensively commercialised (Ilić et al., 2023).

The most common cellulolytic bacteria include *Acetivibrio cellulolyticus*, *Bacillus* spp., *Cellulomonas* spp., *Clostridium* spp., *Erwinia chrysanthemi*, *Thermobispora bispara*, *Ruminococcus albus*, *Streptomyces* spp., *Thermonospora* spp., and *Thermobifida* (Sadhu, Maiti, 2013). The search for new cellulolytic bacteria strains is currently attracting increasing attention. Consequently, numerous new species of cellulolytic bacteria have been described. These include *Streptomyces abietis* (Fujii et al., 2013), *Kallotenue papyrolyticum* (Cole et al., 2013), *Ornatilinea apprima* (Podosokorskaya et al., 2013), *Bacteroides luti* (Hatamoto et al., 2014), *Alicyclobacillus cellulosilyticus* (Kusube et al., 2014), *Anaerobacterium chartisolvens* (Horino et al., 2014), *Caldicellulosiruptor changbaiensis* (Bing et al., 2015), *Herbinix hemicellulosilytica* (Koeck et al., 2015), *Pseudomonas coleopterorum* (Menendez et al., 2015), *Siphonobacter aquaeclarae*, *Cellulosimicrobium funkei*, *Paracoccus sulfuroxidans*, *Ochrobactrum cytisi*, *O. haematophilum*, *Kaistia adipata*, *Devosia riboflavina*, *Labrys neptuniae*, and *Citrobacter freundii* (Huang et al., 2012), *Thermotoga naphthophila* (Akram, Haq, 2020), and *Nocardiopsis dassonvillei* (Sivasankar et al., 2022).

Aerobic free-living bacteria secrete extracellular enzymes equipped with binding modules for various cellulose conformations. Enzyme synergism ensures the efficient hydrolysis of cellulose-containing substrates. Anaerobic bacteria, frequently found in the gastrointestinal tracts of herbivorous animals, are characterised by an extracellular multienzyme complex like the cellulosome. Within the cellulosome, diverse cellulolytic enzymes are arranged on a scaffold protein, ensuring the secure attachment of cells to cellulose, promoting elevated local concentrations, and maintaining the proper ratios and sequence of components. The cellulosome of the thermophilic bacterium *C. thermocellum* was the initial subject of study, succeeded by investigations of the cellulosomes of mesophilic clostridia, ruminococci, and additional anaerobes (Schwarz, 2001).

In addition to cellulosome complexes, anaerobes, including clostridia, secrete cellulases and hemicellulases. By contrast, cellobiohydrolases have not been found in the enzyme systems of *Pseudomonas*, *Bacillus*, and *Erwinia*, with their function extending beyond nutrition to include facilitating the penetration of the phytopathogen into the host cell (Rabinovich et al., 2002).

Cellulases of thermophilic microorganisms

Thermostable cellulases exhibit stability at elevated temperatures (Azadian et al., 2016). Thermostable microbial enzymes demonstrate peak functionality within a temperature range of 60 to 80 °C. In the enzymatic hydrolysis of cellulose, thermostable cellulases play a significant role as they can be used immediately following the heating stage without prior cooling, thereby reducing production cycle times and increasing yields (Liu D. et al., 2011).

Thermophilic microorganisms produce specialised proteins, referred to as chaperones, that assist in refolding proteins into their native conformation and restoring their functions (Laksanalamai, Robb, 2004; Singh S.P. et al., 2010). The small DNA-binding protein Sso7d in *Sulfolobus solfataricus* was observed to be involved in maintaining the stability of aggregated proteins (Ciaramella et al., 2002). However, the existence of such mechanisms suggests that enzymes from thermophilic organisms may not inherently exhibit high thermostability, especially outside of their host organism.

The high temperature tolerance of proteins derived from thermophilic bacteria, actinomycetes, and archaea is attributable to elevated electrostatic, disulfide, and hydrophobic interactions within their structural framework (Ladenstein, Ren, 2006; Pedone et al., 2008). Some thermophilic enzymes are stabilised by metal ions and inorganic salts (Vieille, Zeikus, 2001). The cellulolytic activity of certain thermophilic fungi, such as *Chaetomium thermophile*, *Sporotrichum thermophile*, and *T. aurantiacus*, is two to three times greater than that of *Trichoderma viridae* (Tansey, 1971).

Another factor that may affect thermostability is glycosylation (Kahn et al., 2020; Ramakrishnan et al., 2023). Should the bacterial gene be cloned into a protein-glycosylating organism, like fungi, this phenomenon could augment the thermostability of the enzyme.

Various thermophilic bacteria, including representatives of the genera *Bacillus*, *Geobacillus*, *Caldibacillus*, *Acidothermus*, *Caldocellum*, and *Clostridium*, have been reported to produce thermostable cellulolytic enzymes (Ghosh et al., 2020). The hyperthermophilic bacterium *Dictyoglomus turgidum* carries a gene (*Dtur_0671*) encoding a β-glucosidase expressed in *Escherichia coli*. This enzyme exhibits maximum activity at 80 °C and pH 5.4, is extremely stable within the pH 5–8 range, and retains 70 % of its activity after 2 hours at 70 °C. The high tolerance to glucose and ethanol has proven this enzyme to be suitable for industrial bioethanol production (Fusco et al., 2018).

The gene of the cellulosolytic enzyme from *Thermotoga naphthophila* RKU-10T was successfully obtained and expressed in *E. coli*. The purified enzyme, TnCel12B, was demonstrated to exhibit the maximum activity at pH 6.0 and 90 °C. It retained 100 % activity after incubation for 8 hours at 85 °C, as well as across the pH 5.0–9.0 range (Akram, Haq, 2020). A gene from the hyperthermophilic archaeon *Sulfolobus shibatae*, encoding endo-1,4-β-d-glucanase, demonstrated maximum activity at 95–100 °C following

cloning and overexpression in *E. coli*. This enzyme exhibited excellent resistance to high temperatures: it retained full activity after one hour of incubation at temperatures up to 85 °C; 98, 90, and 84 % of initial activity was observed after 2 hours of incubation at 75, 80, and 85 °C, respectively (Boyce, Walsh, 2018).

Cellulases are categorised into two groups depending on their preferred pH ranges. The first group includes thermoacidophilic cellulases, such as those from *Alicyclobacillus*, *Geobacillus*, and *T. aurantiacus* species, that thrive in acidic conditions. The second group includes thermoalkaliphilic cellulases, such as those from *Bacillus*, *Halobacillus*, and archaea, that prefer basic conditions. Both groups of enzymes can operate effectively at high temperatures, but their application varies according to the specific requirements of different industrial sectors (Arya et al., 2024).

Cellulases of psychrophilic microorganisms

The application of cold-active enzymes operating at alkaline pH may, for example, prove to be in demand in detergents, as it preserves the quality of fabrics without the use of hot water. Psychrophilic microorganisms are used to isolate enzymes active in the low-temperature range. Psychrophilic microorganisms represent a substantial segment of saprophytic organisms that inhabit soil, marine environments, freshwater ecosystems, and wastewater.

Metabolically active bacteria capable of surviving at temperatures between –5 and –15 °C have been isolated from Arctic permafrost (Bakermans, Skidmore, 2011; Mykytczuk et al., 2013). These psychrophiles can survive at low temperatures, with intracellular biochemical processes performed by cold-active enzymes. Furthermore, their extracellular enzymes facilitate the degradation of complex materials present in the environment. In mesophilic organisms, the impact of reduced temperatures on enzyme activity is more pronounced because their enzyme molecules possess fewer structural adaptations for functioning under such conditions, resulting in low enzymatic activity (Karan et al., 2012). Conversely, psychrophiles growing in cold environments possess cold-adapted enzymes characterised by high catalytic activity and stability.

Over the past fifteen years, various psychrotrophic microorganisms have been identified. In 1996, the first low-temperature cellulase was isolated from the fungus *Acremonium alcalophilum*. At 40 °C and pH 7.0, this cellulase demonstrated maximum activity, with over 20 % activity retained at 0 °C (Hayashi et al., 1996). Subsequently, cellulases have been isolated and characterised from other psychrophilic microorganisms. The optimal temperature for their activity typically ranges from 20 to 40 °C (see the Table), with higher optimal temperatures occasionally reported.

Microbial strains adapted to cold environments have been primarily isolated from Antarctic and polar regions. Other possible sources of cold-active cellulases are microorganisms found in mud and deep-sea sediments. A *Clostridium* strain isolated from manure biogas was capable of growth at temperatures from 5 to 50 °C, producing a range of xylano- and cellulolytic enzymes, which were most active at 20 °C (Akila, Chandra, 2003). Similarly, the cellulase produced by the fungus *A. alcalophilum* was active even at 0 °C (Hayashi et al., 1996).

CelE1, a novel cold-tolerant cellulase from the GH5 family, was isolated from a soil metagenomic library obtained from a sugarcane field. A functional screening was employed to identify cellulolytic clones within this library (Alvarez et al., 2013). The CelE1 endoglucanase isolated via this method

Cellulases of psychrophilic microorganisms

Microorganism	Enzyme	Opt. temp., °C	Opt. pH	Activator	Repressor	Reference
Acremonium alcalophilum	Cellulase	40	7.0	Glucose	–	Hayashi et al., 1996
Arthrobacter sp.	β-glucosidas	35	–	–	–	Benešová et al., 2005
Clostridium sp.	Endoglucanase, β-glucosidase	20	5–6	–	–	Akila, Chandra, 2003
Fibrobacter succinogenes	Endoglucanase	25	5.5	–	–	Iyo, Forsberg, 1999
Paenibacillus sp. strain C7	β-glucosidase	30–35	7–8	–	–	Shipkowski, Brenchley, 2005
Paenibacillus sp. BME-14	Endoglucanase	35	6.5	Ca^{2+}, Mg^{2+}, Mn^{2+}, dithiothreitol, β-mercaptoethanol	Cu^{2+}, EDTA	Fu et al., 2010
Pseudoalteromonas sp. DY3	Endoglucanase	40	6–7			Zeng et al., 2006
Rhodotorula glutinis	Endoglucanase	50	4.5	Fe^{3+}, Mn^{2+}	Al^{3+}, Ca^{2+}, Fe^{2+}, EDTA, EGTA	Oikawa et al., 1998
Shewanella sp. G5	β-glucosidase	37	8	–	–	Cristóbal et al., 2008
Pseudoalteromonas sp. MB-1	Endoglucanase	35	7.2	–	–	You, Wang, 2005

demonstrated high activity across a broad temperature range and under alkaline conditions.

T.V. Souza et al. (2015) investigated the influence of pH on the secondary and tertiary structure, charge, and activity of CelE1. While the pH change had a minimal impact on the enzyme structure, its activity diminished in acidic conditions. Optimal activity was achieved at pH 8. To assess the suitability of CelE1 as an additive in detergents and cleaning agents, its activity was evaluated in the presence of surfactants. The authors observed no significant inhibitory effect of surfactants on CelE1 endoglucanase activity (Souza et al., 2015). Furthermore, a thermodynamic analysis based on structural stability and the chemical unfolding/refolding process of CelE1 was conducted. The findings indicated that the chemical decomposition occurs through a reversible two-phase process. Thermodynamic analysis data are highly valuable in predicting enzyme stability.

The psychrophilic actinobacterium *Nocardiopsis dassonvillei* PSY13 produces a highly active cold-adapted cellulase with optima at 10 °C and pH 7.5 (Sivasankar et al., 2022).

A cold-active cellulase (Celluzyme®) active at 15 °C has been developed by Novozymes from cold-adapted *H. insolens* fungi and is commercially available as part of a cellulase blend under the trade name Celluclean®.

Cellulases of halophilic microorganisms

The universal application of cellulases necessitates the continuous search for novel enzyme sources. Enzymes of marine origin have recently attracted particular attention, especially for industrial applications. Specialised environments within the marine ecosystem, including estuaries and mangroves, are rich in lignocellulosic biomass and consequently provide a nutrient-rich habitat for cellulose-degrading organisms. Given their ability to survive in environments with limited nutrients and unfavorable circumstances, marine organisms show promise as candidates for various industrial applications (Dalmaso et al., 2015; Barzakar, 2018). The elevated salinity (3 %) of marine ecosystems, as noted by S.T. Jahromi and N. Barzakar, has resulted in a more diverse array of cellulolytic microorganisms than those found in terrestrial environments (Jahromi, Barzakar, 2018).

Investigations into cellulose degradation in the presence of salt have identified new metabolic pathways and enzymes that exhibit cellulolytic activity. Cellulose-degrading organisms in marine ecosystems significantly contribute to the mineralization of organic matter, which enhances the productivity of these ecosystems (Milici et al., 2017). The capacity to degrade cellulose has been observed across various marine organisms, including bacteria (Harshvardhan et al., 2013), yeasts (Rong et al., 2015), filamentous fungi (Liu J. et al., 2012; Batista-García et al., 2017), protists (Bremer, Tabbot, 1995), rotifers (Chun et al., 2997), krill (Tsuji et al., 2012), and echinoderms (Sakamoto et al., 2007).

A halotolerant endoglucanase with a molecular weight of 39 kDa was isolated from the filamentous fungus *Botrytis ricini* URM 5627 (Silva et al., 2018). The optimal operating conditions for the enzyme were 50 °C and pH 5. The enzyme remained stable at 39–60 °C for 60 minutes and at pH 4–6. Enzymatic activity was observed to increase in the presence of Na^+, Mn^{2+}, Mg^{2+}, and Zn^{2+} and decrease in the presence of Ca^{2+}, Cu^{2+}, and Fe^{2+}. The endoglucanase exhibited a halotolerant profile, with its activity increasing proportionally with NaCl concentration. The highest activity was observed at 2 M NaCl, representing a 75 % increase.

Selection of producers most suitable for enzyme production

E. coli and *Bacillus* sp. are widely used as bacterial systems for recombinant protein expression. In addition, other bacteria, including *Zymomonas mobilis* and *Streptomyces lividans*, are also used as platforms. In the industry, *E. coli* is the most commonly employed cellulase expression system, possessing several benefits, including a thoroughly characterised genome, commercial availability, and ease of modification. However, certain drawbacks have to be considered, including limited secretion (due to the thick outer membrane hindering transport across the cell membrane), degradation of linker sequences, reduced cellulolytic activity, and the potential for inclusion body formation. *Zymomonas mobilis* has proven to be an alternative to yeast due to its versatility in fermenting a wide range of sugars. Moreover, it serves as an alternative to *E. coli* due to its capability to express recombinant proteins both intracellularly and extracellularly.

Genetic methodologies for *C. thermocellum* are less advanced than those for model organisms like *E. coli*, and the introduction of single nucleotide polymorphisms (SNPs) has received limited attention. *C. thermocellum* is an obligate thermophilic and anaerobic Gram-positive bacterium that naturally ferments lignocellulose into ethanol and organic acids (Lynd et al., 2005; Olson et al., 2012; Xu Q. et al., 2016; Tian et al., 2019).

There are two main strategies for improving cellulases or components of the cellulase complex through genetic modifications: (1) rational design and (2) directed evolution (Acharya, Chaudhary, 2012). The merits and modifications (classic random mutagenesis or genetic modification) of each approach are tested and employed by various specialists to achieve maximum cellulase yield and efficiency. To obtain the maximum cellulase yield, S. Sadhu et al. performed random mutagenesis on a cellulolytic strain of the genus *Bacillus* using N-methyl-N′-nitro-N-nitrosoguanidine (NTG), resulting in AT and GC transition mutations (Sadhu et al., 2014). This yielded a mutant strain with enhanced carboxymethylcellulase activity. Similar results were obtained using NTG on *Cellulomonas* sp. (Sangkharak et al., 2012), but the characteristics of these mutants have not been reported.

The production of recombinant enzymes involves developing technologies that combine directed evolution and rational design (Zhao H. et al., 2002; Cherry, Fidantsef, 2003). Nonetheless, a significant barrier to rational design

stems from incomplete understanding of cellulase substrates, their enzymatic interactions, interrelationships, and the regulation of cellulase activity, often resulting in the prevalent application of directed evolution (Zhang Y.H.P. et al., 2006).

Nevertheless, there are examples of successful application of the rational design method. In a study by A. Akbarzadeh et al. (2018), directed mutagenesis was employed on endoglucanase-II (Cel5A) derived from *H. jecorina*. This enzyme demonstrates thermal instability due to the presence of four disulfide bonds in its structure. Cysteine amino acid residues at positions 99 and 323 were substituted with valine and histidine, respectively. The loss of two disulfide bonds resulted in increased activity and thermostability of the enzyme. The activity of cellulase from *Gloeophyllum trabeum* (GtCel5) was enhanced using site-directed mutagenesis of loop 6 (Zheng et al., 2018). A.S. Dotsenko et al. (2020) demonstrated that the thermostability of cellobiohydrolase could be improved using rational design by substituting proline. The resulting G415P protein exhibited a 3.5-fold increase in half-life at 60 °C compared to the wild-type protein. A number of authors have reviewed and compared the expression systems of recombinant cellulases (Garvey et al., 2013; Hasunuma et al., 2013; Sadhu, Maiti, 2013; Juturu, Wu, 2014; Lambertz et al., 2014).

Several studies have documented the employment of directed evolution techniques in conjunction with rational design to achieve cellulase overexpression within their native bacterial hosts. Ease of genetic modification and other attributes have allowed species such as *B. subtilis* and *C. thermocellum* to be used as homologous cellulase production systems.

Nevertheless, employing these bacteria presents drawbacks, including limited protein production, elevated manufacturing expenses, and the necessity of culture amplification in an enriched environment (Lambertz et al., 2014). The use of an *E. coli* strain for the expression of β-1,4-endoglucanase and β-1,4-glucosidase from another *E. coli* strain under a constitutive promoter was reported, which allowed for biomass hydrolysates fermentation (Munjal et al., 2015).

D. Chung et al. (2014) engineered *Caldicellulosiruptor bescii*, a bacterium capable of independently degrading lignocellulosic biomass. The study involved the homologous expression and cloning of a multimodular cellulase, CelA, which consists of GH9 and GH48 domains.

A considerable number of studies have also focused on applying the classical method of overproducing the target protein or enzyme by cloning its coding genes into a high copy-number plasmid. For example, this method was used to obtain homologous overexpression of cellulase CelC2 from *Rhizobium leguminosarum* bv. *trifolii* ANU843, which increased its cellulolytic activity 3-fold (Robledo et al., 2011).

An effective *Agrobacterium tumefaciens*-mediated transformation system for *H. insolens* was developed by X. Xu et al. (2016). The authors transformed plasmids carrying the *H. insolens* glyceraldehyde-3-phosphate dehydrogenase gene promoter, which controlled the transcription of genes encoding neomycin phosphotransferase, hygromycin B phosphotransferase, and enhanced green fluorescent protein. T-DNA insertional mutagenesis was used to create a mutant library of *H. insolens*. As a result, a transformant identified as T4, exhibiting elevated cellulase and hemicellulase activity, was isolated. The activities of phospholipase, endoglucanase, cellobiohydrolase, β-glucosidase, and T4 xylanase at the fermentation endpoint exceeded those of the wild-type strain by 60, 440, 320, 41, and 81 %, respectively.

Strategies based on heterologous expression focus on employing non-cellulolytic microorganisms with high production rates for the expression of microbial cellulases (Bhattacharya et al., 2015). In both research and industry, most frequently used are bacteria such as *E. coli*, various species of the genera *Bacillus*, *Pseudomonas fluorescens*, *Ralstonia eutropha*, and *Zymomonas mobilis*, yeasts such as *Saccharomyces cerevisiae* and *Pichia pastoris*, and mycelial fungi from the genera *Aspergillus* and *Trichoderma*. Furthermore, mammalian, plant, or insect cell cultures, as well as transgenic plants and/or animals, are employed for protein expression (Demain, Vaishnav, 2009).

Conclusion

Due to their wide-ranging applications in cellulose-degrading biocatalytic processes, cellulase enzymes have seen an increase in industrial demand over the last few years. The broad applicability and environmental compatibility of cellulase-mediated processes continue to drive research aimed at discovering efficient and cost-effective enzymes.

Filamentous fungi cultures have traditionally been employed in cellulase production. However, most filamentous fungi obtained through natural selection exhibit low secretory capacity for cellulase production, which is insufficient to meet industrial demands.

An effective method for increasing fungal enzyme production is random mutagenesis combined with an adaptive laboratory evolution strategy (Peng et al., 2021). In recent years, research has increasingly focused on bacterial cellulases, due to their diverse properties, enhanced stability, and the potential to integrate multiple activities within a single enzyme.

Extremophilic microorganisms possess a variety of molecular strategies for surviving extreme conditions. Their enzymes exhibit properties such as salt tolerance, thermostability, and cold adaptiveness. Enzymes that are thermophilic, piezophilic, acidophilic and halophilic have been isolated recently. Through the employment of genetic engineering, these genes can be expressed in other organisms, leveraging an extensive array of existing operational methodologies.

Cellulase immobilization technologies, especially the use of the combination of polymeric carriers with nanomaterials, have attracted considerable attention. It is possible for such immobilised cellulases to exhibit enhanced activity, stability, reusability, and processability. The combination of nanomaterials and biocatalysis technologies using immobilised cellulases is currently considered a cutting-edge

field of research and development in enzyme technology (Ranjan et al., 2023).

The complexity of cellulase presents a unique challenge. All three components of the enzyme complex are essential for its proper functioning (Bhat M.K., Bhat S., 1997). This complex nature makes it difficult to clone the enzyme into heterologous systems. Consequently, significant global research efforts are focused on identifying natural producers and devising strategies to enhance the properties of cellulase through genetic modification of isolated organisms. Individual cellulase complex components are cloned and expressed using standard production systems employed in modern biotechnology.

References

Acharya S., Chaudhary A. Bioprospecting thermophiles for cellulase production: a review. *Braz J Microbiol*. 2012;43(3):844-856. doi 10.1590/S1517-83822012000300001

Akbarzadeh A., Pourzardosht N., Dehnavi E., Ranaei Siadat S.O., Zamani M.R., Motallebi M., Nikzad Jamnani F., Aghaeepoor M., Barshan Tashnizi M. Disulfide bonds elimination of endoglucanase II from *Trichoderma reesei* by site-directed mutagenesis to improve enzyme activity and thermal stability: an experimental and theoretical approach. *Int J Biol Macromol*. 2018;120(Pt. B):1572-1580. doi 10.1016/j.ijbiomac.2018.09.164

Akila G., Chandra T.S. A novel cold-tolerant *Clostridium* strain PXYL1 isolated from a psychrophilic cattle manure digester that secretes thermolabile xylanase and cellulase. *FEMS Microbiol Lett.* 2003; 219(1):63-67. doi 10.1016/S0378-1097(02)01196-5

Akram F., Haq I.U. Overexpression and characterization of TnCel12B, a hyperthermophilic GH12 endo-1,4-β-glucanase cloned from *Thermotoga naphthophila* RKU-10^{T}. *Anal Biochem*. 2020;599:113741. doi 10.1016/j.ab.2020.113741

Alvarez T.M., Paiva J.H., Ruiz D.M., Cairo J.P., Pereira I.O., Paixão D.A., de Almeida R.F., Tonoli C.C., Ruller R., Santos C.R., Squina F.M., Murakami M.T. Structure and function of a novel cellulase 5 from sugarcane soil metagenome. *PLoS One*. 2013;8(12): e83635. doi 10.1371/journal.pone.0083635

Arya M., Chauhan G., Fatima T., Verma D., Sharma M. Statistical modelling of thermostable cellulase production conditions of thermophilic *Geobacillus* sp. TP-1 isolated from Tapovan hot springs of the Garhwal Himalayan mountain ranges, India. *Indian J Microbiol*. 2024;64(3):1132-1143. doi 10.1007/s12088-024-01258-x

Azadian F., Badoei-Dalfard A., Namaki-Shoushtari A., Hassanshahian M. Purification and biochemical properties of a thermostable, haloalkaline cellulase from *Bacillus licheniformis* AMF-07 and its application for hydrolysis of different cellulosic substrates to bioethanol production. *Mol Biol Res Commun*. 2016;5(3):143-155. doi 10.22099/mbrc.2016.3743

Azman N.A., Fuzi S.F.Z.M., Talip B.A., Abdullah S. Enzymatic clarification of soursop juice by pectinase/cellulase enzymes ratio. *Enhanced Knowledge Sci Technol*. 2021;1(2):234-243. doi 10.30880/ekst.2021.01.02.028

Azzaz H.H., Abd A.M., Tawab E., Khattab M.S.A., Szumacher-Strabel M., Cieślak A.C., Murad H.A., Kiełbowicz M., El-Sherbiny M. Effect of cellulase enzyme produced from *Penicilliumchrysogenum* on the milk production, composition, amino acid, and fatty acid profiles of egyptian buffaloes fed a high-forage diet. *Animals.* 2021; 11(11):3066. doi 10.3390/ANI11113066

Bakermans C., Skidmore M.L. Microbial metabolism in ice and brine at –5 °C. *Environ Microbiol.* 2011;13(8):2269-2278. doi 10.1111/j.1462-2920.2011.02485.x

Barzkar N., Homaei A., Hemmati R., Patel S. Thermostable marine microbial proteases for industrial applications: scopes and risks. *Extremophiles*. 2018;22(3):335-346. doi 10.1007/s00792-018-1009-8

Batista-García R.A., Sutton T., Jackson S.A., Tovar-Herrera O.E., Balcázar-López E., Sánchez-Carbente M.D., Sánchez-Reyes A., Dobson A.D., Folch-Mallol J.L. Characterization of lignocellulolytic activities from fungi isolated from the deep-sea sponge *Stelletta normani*. *PLoS One*. 2017;12(3):e0173750. doi 10.1371/journal.pone.0173750

Benešová E., Marková M., Králová B. α-Glucosidase and β-glucosidase from psychrotrophic strain *arthrobacter* sp. C2-2. *Czech J Food Sci*. 2005;23(3):116-120. doi 10.17221/3380-CJFS

Bhat M.K., Bhat S. Cellulose degrading enzymes and their potential industrial applications. *Biotechnol Adv.* 1997;15(3-4):583-620. doi 10.1016/s0734-9750(97)00006-2

Bhati N., Shreya, Sharma A.K. Cost-effective cellulase production, improvement strategies, and future challenges. *J Food Process Eng*. 2021;44(2):e13623. doi 10.1111/jfpe.13623

Bhattacharya A.S., Bhattacharya A., Pletschke B.I. Synergism of fungal and bacterial cellulases and hemicellulases: a novel perspective for enhanced bio-ethanol production. *Biotechnol Lett*. 2015;37(6): 1117-1129. doi 10.1007/s10529-015-1779-3

Bing W., Wang H., Zheng B., Zhang F., Zhu G., Feng Y., Zhang Z. *Caldicellulosiruptor changbaiensis* sp. nov., a cellulolytic and hydrogen-producing bacterium from a hot spring. *Int J Syst Evol Microbiol*. 2015;65(Pt. 1):293-297. doi 10.1099/ijs.0.065441-0

Boyce A., Walsh G. Expression and characterisation of a thermophilic endo-1,4-β-glucanase from *Sulfolobus shibatae* of potential industrial application. *Mol Biol Rep*. 2018;45(6):2201-2211. doi 10.1007/s11033-018-4381-7

Bremer G.B., Talbot G. Cellulolytic enzyme activity in the marine protist *Schizochytrium aggregatum*. *Bot Mar.* 1995;38(1-6):37-42. doi 10.1515/botm.1995.38.1-6.37

Cao L., Yu M., Wang C., Bao Y., Zhang M., He P., Zhang Y., Yang T., Li L., Li G., Gong Y. Cellulase-assisted extraction, characterization, and bioactivity against rheumatoid arthritis of *Astragalus* polysaccharides. *Int J Polym Sci.* 2019;2:8514247. doi 10.1155/2019/8514247

Chellapandi P., Jani H.M. Production of endoglucanase by the native strains of *Streptomyces* isolates in submerged fermentation. *Braz J Microbiol.* 2008;39(1):122-127. doi 10.1590/S1517-83822008000 1000026

Cherry J.R., Fidantsef A.L. Directed evolution of industrial enzymes: an update. *Curr Opin Biotechnol*. 2003;14(4):438-443. doi 10.1016/s0958-1669(03)00099-5

Chiwetalu M.O., Ossai E.C., Enechi O.C., Ugwu C.N., Ozogwu V.E., Okpala C.O.R., Njoku O.U. An enhanced fat extraction from *Pycnanthus angolensis* (African nutmeg) seeds using cellulase from *Aspergillus niger* strain BC23. *Qual Assur Saf Crops Foods.* 2022; 14(3):165-175. doi 10.15586/qas.v14i3.997

Chun C.Z., Hur S.B., Kim Y.T. Purification and characterization of an endoglucanase from the marine rotifer, *Brachionus plicatilis*. *Biochem Mol Biol Int*. 1997;43(2):241-249. doi 10.1080/15216549700 204021

Chung D., Cha M., Guss A.M., Westpheling J. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proc Natl Acad Sci USA.* 2014;111(24):8931-8936. doi 10.1073/pnas.1402210111

Chutani P., Sharma K.K. Concomitant production of xylanases and cellulases from *Trichoderma longibrachiatum* MDU-6 selected for the deinking of paper waste. *Bioprocess Biosyst Eng.* 2016;39(5): 747-758. doi 10.1007/s00449-016-1555-3

Ciaramella M., Pisani F.M., Rossi M. Molecular biology of extremophiles: recent progress on the hyperthermophilic *archaeon Sulfolobus*. *Antonie Van Leeuwenhoek*. 2002;81(1-4):85-97. doi 10.1023/a:1020577510469

Cole J.K., Gieler B.A., Heisler D.L., Palisoc M.M., Williams A.J., Dohnalkova A.C., Ming H., Yu T.T., Dodsworth J.A., Li W.-J., Hedlund B.P. *Kallotenue papyrolyticum* gen. nov., sp. nov., a cellulolytic and filamentous thermophile that represents a novel lineage (*Kallotenuales* ord. nov., *Kallotenuaceae* fam. nov.) within the class

Chloroflexia. *Int J Syst Evol Microbiol*. 2013;63(12):4675-4682. doi 10.1099/ijs.0.053348-0

Coutts A.D., Smith R.E. Factors influencing the production of cellulases by *Sporotrichum thermophile*. *Appl Environ Microbiol.* 1976; 31(6):819-825. doi 10.1128/aem.31.6.819-825.1976

Cristóbal H.A., Breccia J.D., Abate C.M. Isolation and molecular characterization of *Shewanella* sp. G5, a producer of cold-active beta-D-glucosidases. *J Basic Microbiol*. 2008;48(1):16-24. doi 10.1002/jobm.200700146

Dalmaso G.Z., Ferreira D., Vermelho A.B. Marine extremophiles: a source of hydrolases for biotechnological applications. *Mar Drugs*. 2015;13(4):1925-1965. doi 10.3390/md13041925

Dashtban M., Buchkowski R., Qin W. Effect of different carbon sources on cellulase production by *Hypocrea jecorina* (*Trichoderma reesei*) strains. *Int J Biochem Mol Biol*. 2011;2(3):274-286

Dave B.R., Sudhir A.P., Pansuriya M., Raykundaliya D.P., Subramanian R.B. Utilization of Jatropha deoiled seed cake for production of cellulases under solid-state fermentation. *Bioprocess Biosyst Eng*. 2012;35(8):1343-1353. doi 10.1007/s00449-012-0723-3

Demain A.L., Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv*. 2009;27(3):297-306. doi 10.1016/j.biotechadv.2009.01.008

Doi R.H., Tamaru Y. The *Clostridium cellulovorans* cellulosome: an enzyme complex with plant cell wall degrading activity. *Chem Rec*. 2001;1(1):24-32. doi 10.1002/1528-0691(2001)1:1<24::AID-TCR5>3.0.CO;2-W

Dotsenko A.S., Dotsenko G.S., Rozhkova A.M., Zorov I.N., Sinitsyn A.P. Rational design and structure insights for thermostability improvement of *Penicillium verruculosum* Cel7A cellobiohydrolase. *Biochimie*. 2020;176:103-109. doi 10.1016/j.biochi.2020.06.007

Ejaz U., Sohail M., Ghanemi A. Cellulases: from bioactivity to a variety of industrial applications. *Biomimetics*. 2021;6(3):44. doi 10.3390/biomimetics6030044

El-Baroty G., Abou Elella F., Moawad H., El Sebai T., Abdulaziz F., Khattab A.A. Optimization and characterization of extracellular cellulase produced by native Egyptian fungal strain. *Notulae Botanicae*. 2019;47(3):743-750. doi 10.15835/nbha47311500

Folan M.A., Coughlan M.P. The cellulase complex in the culture filtrate of the thermophyllic fungus, *Talaromyces emersonii*. *Int J Biochem.* 1978;9(10):717-722. doi 10.1016/0020-711x(78)90038-1

Fracheboud D., Canevascini G. Isolation, purification, and properties of the exocellulase from *Sporotrichum* (*Chrysosporium*) *thermophile*. *Enzyme Microb Technol*. 1989;11(4):220-229. doi 10.1016/0141-0229(89)90096-3

Fu X., Liu P., Lin L., Hong Y., Huang X., Meng X., Liu Z. A novel endoglucanase (Cel9P) from a marine bacterium *Paenibacillus* sp. BME-14. *Appl Biochem Biotechnol*. 2010;160(6):1627-1636. doi 10.1007/s12010-009-8648-2

Fujii K., Satomi M., Fukui Y., Matsunobu S., Morifuku Y., Enokida Y. *Streptomyces abietis* sp. nov., a cellulolytic bacterium isolated from soil of a pine forest. *Int J Syst Evol Microbiol.* 2013;63(Pt. 12): 4754-4759. doi 10.1099/ijs.0.053025-0

Fusco F.A., Ronca R., Fiorentino G., Pedone E., Contursi P., Bartolucci S., Limauro D. Biochemical characterization of a thermostable endomannanase/endoglucanase from *Dictyoglomus turgidum*. *Extremophiles*. 2018;22(1):131-140. doi 10.1007/s00792-017-0983-6

Garvey M., Klose H., Fischer R., Lambertz C., Commandeur U. Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. *Trends Biotechnol*. 2013;31(10):581-593. doi 10.1016/j.tibtech.2013.06.006

Ghosh S., Lepcha K., Basak A., Mahanty A.K. Thermophiles and thermophilic hydrolases. In: Salwan R., Sharma V. (Eds) Physiological and Biotechnological Aspects of Extremophiles. Academic Press, 2020;219-236. doi 10.1016/B978-0-12-818322-9.00016-2

Harshvardhan K., Mishra A., Jha B. Purification and characterization of cellulase from a marine *Bacillus* sp. H1666: a potential agent for single step saccharification of seaweed biomass. *J Mol Catal B Enzym.* 2013;93:51-56. doi 10.1016/j.molcatb.2013.04.009

Hasunuma T., Okazaki F., Okai N., Hara K.Y., Ishii J., Kondo A. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresour Technol*. 2013;135:513-522. doi 10.1016/j.biortech.2012.10.047

Hatamoto M., Kaneshige M., Nakamura A., Yamaguchi T. *Bacteroides luti* sp. nov., an anaerobic, cellulolytic and xylanolytic bacterium isolated from methanogenic sludge. *Int J Syst Evol Microbiol.* 2014;64(Pt. 5):1770-1774. doi 10.1099/ijs.0.056630-0

Hayashi K., Nimura Y., Ohara N., Uchimura T., Suzuki H., Komacata K., Kozaki M. Low-temperature-active cellulase produced by *Acremonium alcalophilum*-JCM 7366. *J Ferment Bioeng*. 1996; 74(1):7-10

Horino H., Fujita T., Tonouchi A. Description of *Anaerobacterium chartisolvens* gen. nov., sp. nov., an obligately anaerobic bacterium from *Clostridium* rRNA cluster III isolated from soil of a Japanese rice field, and reclassification of *Bacteroides cellulosolvens* Murray et al. 1984 as *Pseudobacteroides cellulosolvens* gen. nov., comb. nov. *Int J Syst Evol Microbiol.* 2014;64(Pt. 4):1296-1303. doi 10.1099/ijs.0.059378-0

Hu X., Cheng L., Hong Y., Li Z., Li C., Gu Z. Impact of celluloses and pectins restrictions on gluten development and water distribution in potato-wheat flour dough. *Int J Biol Macromol*. 2022;206:534-542. doi 10.1016/j.ijbiomac.2022.02.150

Hu Y., Kang G., Wang L., Gao M., Wang P., Yang D., Huang H. Current status of mining, modification, and application of cellulases in bioactive substance extraction. *Curr Issues Mol.* 2021;43(2):687-703. doi 10.3390/cimb43020050

Huang S., Sheng P., Zhang H. Isolation and identification of cellulolytic bacteria from the gut of *Holotrichia parallela* larvae (Coleoptera: Scarabaeidae). *Int J Mol Sci*. 2012;13(3):2563-2577. doi 10.3390/ijms13032563

Ilić N., Milić M., Beluhan S., Dimitrijević-Branković S. Cellulases: from lignocellulosic biomass to improved production. *Energies*. 2023;16(8):3598. doi 10.3390/en16083598

Imran M., Anwar Z., Zafar M., Ali A., Arif M. Production and characterization of commercial cellulase produced through *Aspergillus niger* IMMIS1 after screening fungal species. *Pak J Bot.* 2018; 50(4):1563-1570

Iwamuro M., Kawai Y., Shiraha H., Takaki A., Okada H., Yamamoto K. In vitro analysis of gastric phytobezoar dissolubility by Coca-Cola, Coca-Cola Zero, cellulase, and papain. *J Clin Gastroenterol.* 2014;48(2):190-191. doi 10.1097/MCG.0b013e3182a39116

Iyo A.H., Forsberg C.W. A cold-active glucanase from the ruminal bacterium *Fibrobacter succinogenes* S85. *Appl Environ Microbiol.* 1999;65(3):995-998. doi 10.1128/aem.65.3.995-998.1999

Jahromi S.T., Barzkar N. Future direction in marine bacterial agarases for industrial applications. *Appl Microbiol Biotechnol*. 2018; 102(16):6847-6863. doi 10.1007/s00253-018-9156-5

Juturu V., Wu J.C. Microbial cellulases: engineering, production and applications. *Renew Sustain Energy Rev.* 2014;33:188-203. doi 10.1016/j.rser.2014.01.077

Kahn A., Moraïs S., Chung D., Sarai N.S., Hengge N.N., Kahn A., Himmel M.E., Bayer E.A., Bomble Y.J. Glycosylation of hyperthermostable designer cellulosome components yields enhanced stability and cellulose hydrolysis. *FEBS J*. 2020;287(20):4370-4388. doi 10.1111/febs.15251

Karan R., Capes M.D., Dassarma S. Function and biotechnology of extremophilic enzymes in low water activity. *Aquat Biosyst*. 2012; 8(1):4. doi 10.1186/2046-9063-8-4

Karmakar M., Ray R.R. Current trends in research and application of microbial cellulases. *Res J Microbiol*. 2011;6(1):41-53. doi 10.3923/jm.2011.41.53

Kasana R.C., Gulati A. Cellulases from psychrophilic microorganisms: a review. *J Basic Microbiol*. 2011;51(6):572-579. doi 10.1002/jobm.201000385

Kirk O., Borchert T.V., Fuglsang C.C. Industrial enzyme applications. *Curr Opin Biotechnol*. 2002;13(4):345-351. doi 10.1016/s0958-1669(02)00328-2

Koeck D.E., Ludwig W., Wanner G., Zverlov V.V., Liebl W., Schwarz W.H. *Herbinix hemicellulosilytica* gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Int J Syst Evol Microbiol*. 2015;65(8):2365-2371. doi 10.1099/ijs.0.000264

Kusube M., Sugihara A., Moriwaki Y., Ueoka T., Shimane Y., Minegishi H. *Alicyclobacillus cellulosilyticus* sp. nov., a thermophilic, cellulolytic bacterium isolated from steamed Japanese cedar chips from a lumbermill. *Int J Syst Evol Microbiol*. 2014;64(7):2257-2263. doi 10.1099/ijs.0.061440-0

Ladenstein R., Ren B. Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles. *FEBS J*. 2006;273(18):4170-4185. doi 10.1111/j.1742-4658.2006.05421.x

Laksanalamai P., Robb F.T. Small heat shock proteins from extremophiles: a review. *Extremophiles*. 2004;8(1):1-11. doi 10.1007/s00792-003-0362-3

Lambertz C., Garvey M., Klinger J., Heesel D., Klose H., Fischer R., Commandeur U. Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol Biofuels*. 2014;7(1):135. doi 10.1016/j.rser.2014.01.077

Lehr M., Miltner M., Friedl A. Removal of wood extractives as pulp (pre-)treatment: a technological review. *SN Appl Sci*. 2021;3:886. doi 10.1007/s42452-021-04873-1

Liu D., Zhang R., Yang X., Xu Y., Tang Z., Tian W., Shen Q. Expression, purification and characterization of two thermostable endoglucanases cloned from a lignocellulosic decomposing fungi *Aspergillus fumigatus* Z5 isolated from compost. *Protein Expr Purif*. 2011;79(2):176-186. doi 10.1016/j.pep.2011.06.008

Liu J., Xue D., He K., Yao S. Cellulase production in solid-state fermentation by marine *Aspergillus* sp. ZJUBE-1 and its enzymological properties. *Adv Sci Lett.* 2012;16(1):381-386. doi 10.1166/asl.2012.3304

Luo J., Li Y., Li Y., Li H., Fang X., Li Y., Huang W., Cao J., Wu Y. Waste-to-energy: cellulase induced waste activated sludge and paper waste co-fermentation for efficient volatile fatty acids production and underlying mechanisms. *Bioresour Technol*. 2021;341:125771. doi 10.1016/j.biortech.2021.125771

Lynd L.R., van Zyl W.H., McBride J.E., Laser M. Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol*. 2005;16(5):577-583. doi 10.1016/j.copbio.2005.08.009

Maki M., Leung K.T., Qin W. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *Int J Biol Sci*. 2009;5(5):500-516. doi 10.7150/ijbs.5.500

Mathew G.M., Sukumaran R.K., Singhania R.R., Pandey A. Progress in research on fungal cellulases for lignocellulose degradation. *J Sci Ind Res*. 2008;67(11):898-907

Menendez E., Ramírez-Bahena M.H., Fabryová A., Igual J.M., Benada O., Mateos P.F., Peix A., Kolařík M., García-Fraile P. *Pseudomonas coleopterorum* sp. nov., a cellulase-producing bacterium isolated from the bark beetle *Hylesinus fraxini*. *Int J Syst Evol Microbiol.* 2015;65(9):2852-2858. doi 10.1099/ijs.0.000344

Milici M., Vital M., Tomasch J., Badewien T.H., Giebel H.-A., Plumeier I., Wang H., Pieper D.H., Wagner-Döbler I., Simon M. Diversity and community composition of particle-associated and free-living bacteria in mesopelagic and bathypelagic Southern Ocean water masses: evidence of dispersal limitation in the Bransfield Strait. *Limnol Oceanogr*. 2017;62(3):1080-1095. doi 10.1002/lno.10487

Mohand-Oussaid O., Payot S., Guedon E., Gelhaye E., Youyou A., Petitdemange H. The extracellular xylan degradative system in *Clostridium cellulolyticum* cultivated on xylan: evidence for cell-free cellulosome production. *J Bacteriol*. 1999;181(13):4035-4040. doi 10.1128/jb.181.13.4035-4040.1999

Mukherjee P.K., Horwitz B.A., Kenerley C.M. Secondary metabolism in *Trichoderma* – a genomic perspective. *Microbiology*. 2012;158(Pt. 1):35-45. doi 10.1099/mic.0.053629-0

Munjal N., Jawed K., Wajid S., Yazdani S.S. A constitutive expression system for cellulase secretion in *Escherichia coli* and its use in bioethanol production. *PLoS One*. 2015;10(3):e0119917. doi 10.1371/journal.pone.0119917

Mykytczuk N.C.S., Foote S.J., Omelon C.R., Southam G., Greer C.W., Whyteet L.G. Bacterial growth at −15 °C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J*. 2013;7(6):1211-1226. doi 10.1038/ismej.2013.8

Nishida Y., Suzuki K., Kumagai Y., Tanaka H., Inoue A., Ojima T. Isolation and primary structure of a cellulase from the Japanese sea urchin *Strongylocentrotus nudus*. *Biochimie*. 2007;89(8):1002-1011. doi 10.1016/j.biochi.2007.03.015

Niyonzima F.N. Detergent-compatible fungal cellulases. *Folia Microbiol*. 2021;66(1):25-40. doi 10.1007/s12223-020-00838-w

Nyathi M., Dhlamini Z., Ncube T. Bioprospecting for cellulase-producing bacteria from Victoria falls rainforest decaying logs. *Am J Microbiol Res*. 2023;11(3):73-78. doi 10.12691/ajmr-11-3-2

Oikawa T., Tsukagawa Y., Soda K. Endo-β-glucanase secreted by a psychrotrophic yeast: purification and characterization. *Biosci Biotechnol Biochem*. 1998;62(9):1751-1756. doi 10.1271/bbb.62.1751

Okino-Delgado C.H., Zanutto-Elgui M.R., do Prado D.Z., Pereira M.S., Fleuri L.F. Enzymatic bioremediation: current status, challenges of obtaining process, and applications. In: Arora P. (Ed.) Microbial Metabolism of Xenobiotic Compounds. Microorganisms for Sustainability. Vol. 10. Springer, 2019;79-101. doi 10.1007/978-981-13-7462-3_4

Olson D.G., McBride J.E., Shaw A.J., Lynd L.R. Recent progress in consolidated bioprocessing. *Curr Opin Biotechnol*. 2012;23(3):396-405. doi 10.1016/j.copbio.2011.11.026

Ozyilmaz G., Gunay E. Clarification of apple, grape and pear juices by co-immobilized amylase, pectinase and cellulose. *Food Chem*. 2023;398:133900. doi 10.1016/j.foodchem.2022.133900

Patel R.N., Banerjee A., Ko R.Y., Howell J.M., Li W.S., Comezoglu F.T., Partyka R.A., Szarka F.T. Enzymic preparation of (3R-cis)-3-(acetyloxy)-4-phenyl-2-azetidinone: a taxol side-chain synthon. *Biotechnol Appl Biochem*. 1994;20(1):23-33. doi 10.1111/j.1470-8744.1994.tb00304.x

Pedone E., Limauro D., Bartolucci S. The machinery for oxidative protein folding in thermophiles. *Antioxid Redox Signal*. 2008;10(1):157-169. doi 10.1089/ars.2007.1855

Peng Z.Q., Li C., Lin Y., Wu S.S., Gan L.H., Liu J., Yang S.L., Zeng X.H., Lin L. Cellulase production and efficient saccharification of biomass by a new mutant *Trichoderma afroharzianum* MEA-12. *Biotechnol Biofuels*. 2021;14(1):219. doi 10.1186/s13068-021-02072-z

Pérez J., Muñoz-Dorado J., de la Rubia T., Martínez J. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int Microbiol*. 2002;5(2):53-63. doi 10.1007/s10123-002-0062-3

Perim F.D.S., da Silva W.J., de Souza D.O., Ulhoa C.J., Rezende C.F., Dos Santos L.F., Dos Santos F.R., Silva F.G., Minafra C.S. Effects of the addition of *Trichoderma reesei* cellulase to broiler chicken diets for a 21-day period. *Animals* (*Basel*). 2024;14(10):1467. doi 10.3390/ani14101467

Pinheiro B.A., Gilbert H.J., Sakka K., Sakka K., Fernandes V.O., Prates J.A., Alves V.D., Bolam D.N., Ferreira L.M., Fontes C.M. Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *J Biochem*. 2009;424(3):375-384. doi 10.1042/BJ20091152

Podosokorskaya O.A., Bonch-Osmolovskaya E.A., Novikov A.A., Kolganova T.V., Kublanov I.V. *Ornatilinea apprima* gen. nov., sp. nov., a cellulolytic representative of the class *Anaerolineae*. *Int J Syst Evol Microbiol*. 2013;63(Pt. 1):86-92. doi 10.1099/ijs.0.041012-0

Puri M., Sharma D., Barrow C.J., Tiwary A.K. Optimisation of novel method for the extraction of steviosides from *Stevia rebaudiana* leaves. *Food Chem*. 2012;132(3):1113-1120. doi 10.1016/j.foodchem.2011.11.063

Rabinovich M.L., Melnick M.S., Bolobova A.V. The structure and mechanism of action of cellulolytic enzymes. *Biochemistry* (*Mosc*). 2002;67(8):850-871. doi 10.1023/a:1019958419032

Ramakrishnan K., Johnson R.L., Winter S.D., Worthy H.L., Thomas C., Humer D.C., Spadiut O., Hindson S.H., Wells S., Barratt A.H., Menzies G.E., Pudney C.R., Jones D.D. Glycosylation increases active site rigidity leading to improved enzyme stability and turnover. *FEBS J*. 2023;290(15):3812-3827. doi 10.1111/febs.16783

Ranjan R., Rai R., Bhatt S.B., Dhar P. Technological road map of Cellulase: a comprehensive outlook to structural, computational, and industrial applications. *Biochem Eng J*. 2023;198:109020. doi 10.1016/j.bej.2023.109020

Reese E.T., Mandels M. Enzymic hydrolysis of cellulose and its derivatives. *Methods Carbohydr Chem*. 1963;3:139-143

Robledo M., Jiménez-Zurdo J.I., Soto M.J., Velázquez E., Dazzo F., Martínez-Molina E., Mateos P.F. Development of functional symbiotic white clover root hairs and nodules requires tightly regulated production of rhizobial cellulase CelC2. *Mol Plant Microbe Interact*. 2011;24(7):798-807. doi 10.1094/MPMI-10-10-0249

Rong Y., Zhang L., Chi Z., Wang X. A carboxymethyl cellulase from a marine yeast (*Aureobasidium pullulans* 98): its purification, characterization, gene cloning and carboxymethyl cellulose digestion. *J Ocean Univ China.* 2015;14:913-921. doi 10.1007/s11802-015-2574-4

Rubin-Pitel S.B., Zhao H. Recent advances in biocatalysis by directed enzyme evolution. *Comb Chem High Throughput Screen*. 2006; 9(4):247-257. doi 10.2174/138620706776843183

Sadhu S., Maiti T.K. Cellulase production by bacteria: a review. *Br Microbiol Res J*. 2013;3(3):235-258. doi 10.9734/BMRJ/2013/2367

Sadhu S., Ghosh P.K., Aditya G., Maiti T.K. Optimization and strain improvement by mutation for enhanced cellulase production by *Bacillus* sp. (MTCC10046) isolated from cow dung. *J King Saud Univ Sci*. 2014;26(4):323-332. doi 10.1016/j.jksus.2014.06.001

Sajith S., Priji P., Sreedevi S., Benjamin S. An overview on fungal cellulases with an industrial perspective. *J Nutr Food Sci*. 2016;6(1): 461. doi 10.4172/2155-9600.1000461

Sakamoto K., Touhata K., Yamashita M., Kasai A., Toyohara H. Cellulose digestion by common Japanese freshwater clam *Corbicula japonica*. *Fish Sci*. 2007;73(3):675-683. doi 10.1111/j.1444-2906.2007.01381.x

Sangkharak K., Vangsirikul P., Janthachat S. Strain improvement and optimization for enhanced production of cellulase in *Cellulomonas* sp. TSU-03. *Afr J Microbiol Res*. 2012;6(5):1079-1084. doi 10.5897/AJMR11.1550

Schwarz W.H. The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol*. 2001;56:634-649. doi 10.1007/s002530100710

Selzer K., Hassen A., Akanmu A.M., Salem A.Z.M. Digestibility and rumen fermentation of a high forage diet pre-treated with a mixture of cellulase and xylanase enzymes. *South Afr J Anim Sci*. 2021; 51(3):399-406. doi 10.4314/sajas.v51i3.14

Sha H., Zhao B., Yang Y., Zhang Y., Zheng P., Cao S., Wang Q., Wang G. Enhanced anaerobic digestion of corn stover using magnetized cellulase combined with Ni-graphite coating. *Energy*. 2023; 262(B):125532. doi 10.1016/j.energy.2022.125532

Shankar A., Saini S., Sharma K.K. Fungal-integrated second-generation lignocellulosic biorefinery: utilization of agricultural biomass for co-production of lignocellulolytic enzymes, mushroom, fungal polysaccharides, and bioethanol. *Biomass Convers Biorefin.* 2024; 14(1):1117-1131. doi 10.1007/s13399-022-02969-1

Sharma V., Tsai M.L., Nargotra P., Chen C.W., Sun P.P., Singhania R.R., Patel A.K., Dong C.D. Journey of lignin from a roadblock to bridge for lignocellulose biorefineries: a comprehensive review. *Sci Total Environ.* 2023;861:160560. doi 10.1016/j.scitotenv.2022.160560

Shipkowski S., Brenchley J.E. Characterization of an unusual cold-active beta-glucosidase belonging to family 3 of the glycoside hydrolases from the psychrophilic isolate *Paenibacillus* sp. strain C7. *Appl Environ Microbiol*. 2005;71(8):4225-4232. doi 10.1128/AEM.71.8.4225-4232.2005

Silva T.P., de Albuquerque F.S., Dos Santos C.W.V., Franco M., Caetano L.C., Pereira H.J.V. Production, purification, characterization and application of a new halotolerant and thermostable endoglucanase of *Botrytis ricini* URM 5627. *Bioresour Technol*. 2018;270:263-269. doi 10.1016/j.biortech.2018.09.022

Singh A., Bajar S., Devi A., Pant D. An overview on the recent developments in fungal cellulase production and their industrial applications. *Bioresour Technol Rep*. 2021;14:100652. doi 10.1016/j.biteb.2021.100652

Singh S.P., Purohit M.K., Aoyagi C., Kitaoka M., Hayashi K. Effect of growth temperature, induction, and molecular chaperones on the solubilization of over-expressed cellobiose phosphorylase from *Cellvibrio gilvus* under *in vivo* conditions. *Biotechnol Bioproc E*. 2010; 15(2):273-276. doi 10.1007/s12257-009-0023-1

Singhania R.R., Patel A.K., Pandey A. The industrial production of enzymes. In: Soetaert W., Vandamme E.J. (Eds) Industrial Biotechnology. Sustainable Growth and Economic Success. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2010;207-225

Sivasankar P., Poongodi S., Sivakumar K., Al-Qahtani W.H., Arokiyaraj S., Jothiramalingam R. Exogenous production of cold-active cellulase from polar *Nocardiopsis* sp. with increased cellulose hydrolysis efficiency. *Arch Microbiol*. 2022;204(4):218. doi 10.1007/s00203-022-02830-z

Souza T.V., Araujo J.N., da Silva V.M., Liberato M.V., Pimentel A.C., Alvarez T.M., Squina F.M., Garcia W. Chemical stability of a cold-active cellulase with high tolerance toward surfactants and chaotropic agent. *Biotechnol Rep* (*Amst*). 2015;9:1-8. doi 10.1016/j.btre.2015.11.001

Tansey M.R. Agar-diffusion assay of cellulolytic ability of thermophilic fungi. *Arch Mikrobiol*. 1971;77(1):1-11. doi 10.1007/BF00407983

Tian L., Conway P.M., Cervenka N.D., Cui J., Maloney M., Olson D.G., Lynd L.R. Metabolic engineering of *Clostridium thermocellum* for *n*-butanol production from cellulose. *Biotechnol Biofuels*. 2019;12: 186. doi 10.1186/s13068-019-1524-6

Tsuji A., Sato S., Kondo A., Tominaga K., Yuasa K. Purification and characterization of cellulase from North Pacific krill (*Euphausia pacifica*). Analysis of cleavage specificity of the enzyme. *Comp Biochem Physiol B Biochem Mol Biol*. 2012;163(3-4):324-333. doi 10.1016/j.cbpb.2012.08.005

Uhlig H. (Ed.) Industrial Enzymes and Their Applications. John Wiley & Sons, 1998

Vieille C., Zeikus G.J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*. 2001;65(1):1-43. doi 10.1128/MMBR.65.1.1-43.2001

Wackett L.P. Microbial industrial enzymes: an annotated selection of World Wide Web sites relevant to the topics in microbial biotechnology. *Microb Biotechnol*. 2019;12(2):405-406. doi 10.1111/1751-7915.13389

Xu Q., Resch M.G., Podkaminer K., Yang S., Baker J.O., Donohoe B.S., Wilson C., … Brown S.D., Lynd L.R., Bayer E.A., Himmel M.E., Bomble Y.J. Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. *Sci Adv*. 2016;2(2):e1501254. doi 10.1126/sciadv.1501254

Xu X., Li J., Shi P., Ji W., Liu B., Zhang Y., Yao B., Fan Y., Zhang W. The use of T-DNA insertional mutagenesis to improve cellulase production by the thermophilic fungus *Humicola insolens* Y1. *Sci Rep*. 2016;6:31108. doi 10.1038/srep31108

You Y.W., Wang T.H. Cloning and expression of endoglucanase of marine cold-adapted bacteria *Pseudoalteromonas* sp. MB-1. *Wei Sheng Wu Xue Bao*. 2005;45(1):142-144 (in Chinese)

Yu D., Ma X., Huang Y., Jiang L., Wang L., Han C., Yang F. Immobilization of cellulase on magnetic nanoparticles for rice bran oil extraction in a magnetic fluidized bed. *Int J Food Eng*. 2022;18(1):15-26. doi 10.1515/ijfe-2021-0111

Zanuso E., Ruiz H.A., Domingues L., Teixeira J.A. Magnetic nanoparticles as support for cellulase immobilization strategy for enzymatic hydrolysis using hydrothermally pretreated corn cob biomass. *BioEnergy Res*. 2022;15(4):1946-1957. doi 10.1007/s12155-021-10384-z

Zeng R., Xiong P., Wen J. Characterization and gene cloning of a cold-active cellulase from a deep-sea psychrotrophic bacterium *Pseudoalteromonas* sp. DY3. *Extremophiles*. 2006;10(1):79-82. doi 10.1007/s00792-005-0475-y

Zhang Q., Liu N., Wang S., Liu Y., Lan H. Effects of cyclic cellulase conditioning and germination treatment on the γ-aminobutyric acid content and the cooking and taste qualities of germinated brown rice. *Food Chem.* 2019;289:232-239. doi 10.1016/j.foodchem.2019.03.034

Zhang Y.H.P., Himmel M.E., Mielenz J.R. Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv.* 2006;24(5):452-481. doi 10.1016/j.biotechadv.2006.03.003

Zhao C.H., Liu X., Zhan T., He J. Production of cellulase by *Trichoderma reesei* from pretreated straw and furfural residues. *RSC Advances.* 2018;8(63):36233-36238. doi 10.1039/C8RA05936E

Zhao H., Chockalingam K., Chen Z. Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr Opin Biotechnol*. 2002;13(2):104-110. doi 10.1016/s0958-1669(02)00291-4

Zheng F., Tu T., Wang X., Wang Y., Ma R., Su X., Xie X., Yao B., Luo H. Enhancing the catalytic activity of a novel GH5 cellulase *Gt*Cel5 from *Gloeophyllum trabeum* CBS 900.73 by site-directed mutagenesis on loop 6. *Biotechnol Biofuels*. 2018;11:76. doi 10.1186/s13068-018-1080-5

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

Received November 17, 2025. Revised December 2, 2025. Accepted December 5, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1361-1368

doi 10.18699/vjgb-25-142

Breeding for the absence of proanthocyanidins in grain of barley (*Hordeum vulgare* L.): molecular genetic and technological aspects

C.A. Molobekova, I.V. Totsky, N.V. Trubacheeva, O.Yu. Shoeva

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

Abstract. Phenolic compounds constitute a significant group of secondary metabolites in barley grain and influence its technological qualities when used in brewing, feed production, and food manufacturing. Proanthocyanidins – polymeric flavonoids localized in the seed coat – play a particularly important role among them. These compounds are responsible for several production issues, such as colloidal haze in beer and browning of groats after heat treatment. Although proanthocyanidins possess health-beneficial properties based on their antioxidant activity, they can act as antinutritional factors due to their ability to bind proteins. In this regard, the breeding of barley varieties completely lacking proanthocyanidins in the grain was initiated, primarily for use in the brewing industry. Initially, it was assumed that their absence would not be critical for the plant, since wheat, corn, and rice varieties lacking proanthocyanidins in the grain had been identified. However, accumulated evidence indicates that proanthocyanidins perform important physiological functions: they contribute to the maintenance of seed dormancy, provide protection against fungal and bacterial pathogens and pests, and their absence negatively affects agronomic traits. For instance, proanthocyanidin-free barley mutants obtained through induced mutagenesis exhibit reduced productivity and pathogen resistance, an increased risk of pre-harvest sprouting, and deterioration of several technologically important properties. Nevertheless, these mutant lines are actively used in breeding programs to develop varieties for various purposes. This review aims to systematize and analyze global experience in breeding proanthocyanidin-free barley varieties, describing achieved results to identify the most successful approaches and define future research directions. The work examines challenges faced by breeders when using mutant lines, as well as strategies that have helped minimize negative side effects. It is demonstrated that through targeted crossing and optimal selection of mutant alleles, competitive varieties have been developed that combine the required technological qualities with satisfactory agronomic performance, meeting the demands of both the brewing and food industries.

Key words: barley; grain; proanthocyanidins; beer haze; proanthocyanidin-free; brewing

For citation: Molobekova C.A., Totsky I.V., Trubacheeva N.V., Shoeva O.Yu. Breeding for the absence of proanthocyanidins in grain of barley (*Hordeum vulgare* L.): molecular genetic and technological aspects. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1361-1368. doi 10.18699/vjgb-25-142

Funding. The work was carried out under agreement No. 075-15-2025-580 dated July 25, 2025 between the Ministry of Science and Higher Education of the Russian Federation and the ICG SB RAS.

Селекция на отсутствие проантоцианидинов в зерне ячменя (*Hordeum vulgare* L.): молекулярно-генетический и технологический аспекты

К.А. Молобекова, И.В. Тоцкий, Н.В. Трубачеева, О.Ю. Шоева

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

Аннотация. Фенольные соединения составляют значимую группу вторичных метаболитов зерна ячменя и оказывают влияние на его технологические качества при использовании в пивоварении, производстве кормов и пищевых продуктов. Особую роль среди них играют проантоцианидины – полимерные флавоноиды, локализованные в семенной оболочке. Эти соединения обусловливают ряд производственных проблем, таких как коллоидное помутнение пива, а также потемнение крупы после термообработки. Хотя проантоцианидины обладают полезными для здоровья человека свойствами, основанными на их антиоксидантной активности, они могут выступать как антипитательные факторы из-за своей способности связывать белки. В связи с этим была инициирована селекция сортов ячменя, полностью лишенных проантоцианидинов в зерне, в первую очередь для использования в пивоваренной промышленности. Первоначально предполагалось, что их отсутствие

не критично для растения, поскольку у пшеницы, кукурузы и риса были выявлены образцы, не содержащие проантоцианидинов в зерне. Однако накопленные данные свидетельствуют, что проантоцианидины выполняют важные физиологические функции: участвуют в поддержании покоя семян, обеспечивают защиту от грибных и бактериальных патогенов и вредителей, и их отсутствие негативно сказывается на агрономических характеристиках. Так, у беспроантоцианидиновых мутантов ячменя, полученных методами индуцированного мутагенеза, отмечено снижение продуктивности и устойчивости к патогенам, повышение риска прорастания зерна на корню и ухудшение ряда технологически важных свойств. Тем не менее мутантные линии активно используются в селекционных программах для создания сортов различного назначения. Цель данного обзора – систематизировать и проанализировать мировой опыт селекции беспроантоцианидиновых сортов ячменя, описать достигнутые результаты для выявления наиболее успешных подходов и определения направлений дальнейших исследований. В работе рассматриваются проблемы, с которыми столкнулись селекционеры при использовании мутантных линий, а также стратегии, позволившие минимизировать негативные побочные эффекты. Показано, что за счет целенаправленного скрещивания и оптимального подбора мутантных аллелей удалось создать конкурентоспособные сорта, сочетающие требуемые технологические качества с удовлетворительными агрономическими характеристиками и отвечающие запросам как пивоваренной, так и пищевой промышленности.

Ключевые слова: ячмень; зерно; проантоцианидины; коллоидное помутнение; беспроантоцианидиновый; пивоварение

Introduction

Barley (*Hordeum vulgare* L.) is an important agricultural crop widely used in brewing, forage production, and the food industry. Its grain quality is determined not only by protein and starch content but also by secondary metabolites, particularly phenolic compounds, which accumulate in the grain husk (Van Hung, 2016). Among them, proanthocyanidins (PAs) are of particular interest due to their significant role in plant physiology. The deposition of PAs in the seed coat is associated with their role in maintaining seed dormancy (Debeaujon et al., 2000) and protecting the developing seed from various factors, including fungal and bacterial pathogens, insect pests, and heavy metal exposure (Dixon et al., 2005).

The nutritional and feed value of PAs is controversial. On the one hand, numerous studies describe their potentially beneficial properties for human health (Santos-Buelga, Scalbert, 2000). In animal husbandry, these compounds are being studied as a feed additive alternative to antibiotics (Kumar K. et al., 2022). In ruminants, a moderate amount of PAs in feed (2–4 %) can have a beneficial effect by reducing inefficient protein degradation in the rumen. On the other hand, PAs act as antinutritional factors when feeding monogastric animals, since their digestive system is unable to effectively digest PA complexes with proteins (Dixon et al., 2005). PA causes undesirable darkening when barley porridge is prepared, thereby negatively affecting the product's consumer quality (Kohyama et al., 2009).

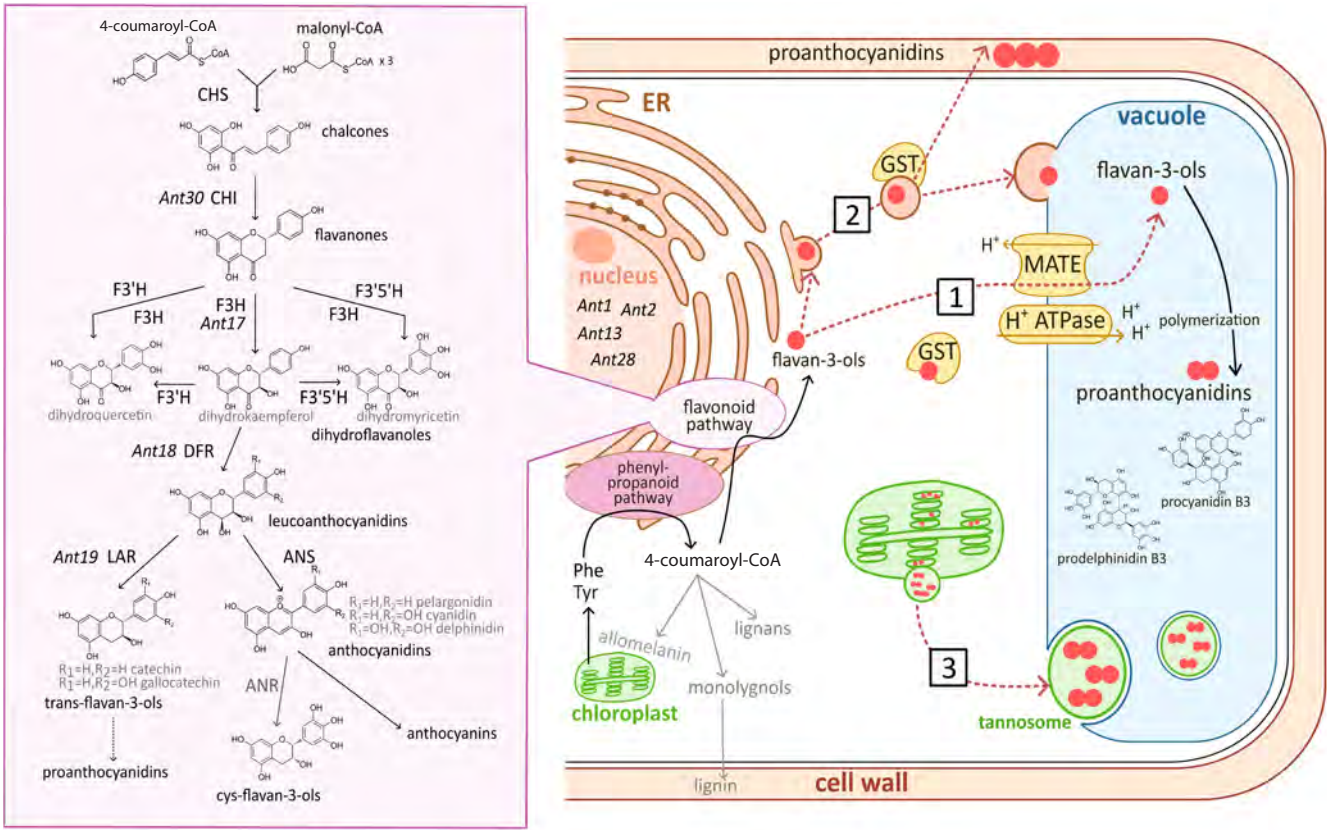
In the brewing industry, PAs are the main cause of colloidal haze, which occurs during cooling of beer and impairs its stability (Delcour et al., 1984). Proteases and selective absorbents are used to solve this problem; however, these methods are not specific enough and can affect the taste and quality of beer (Wang, Ye, 2021). A more effective approach is to develop barley cultivars with reduced levels of protein or PAs in the grain, which helps prevent haze without the use of stabilizers. Since reducing the protein fraction of grain can have a more harmful effect on the plant, breeding of proanthocyanidin-free (PA-free) cultivars is considered a priority (Erdal, 1986; von Wettstein, 2007). This became possible due to the creation of induced barley mutants with impaired flavonoid synthesis, which have served as donors of the target trait in breeding programs (Jende-Strid, 1993).

This review discusses the challenges and prospects associated with breeding PA-free barley cultivars. Since PAs are involved in the regulation of seed dormancy and stress protection, their absence may be accompanied by changes in grain germination rate, plant resistance to pathogens, malt modification, and other important traits. Understanding these relationships is essential for developing commercial barley cultivars that combine the absence of PAs with high agronomic and technical performance.

Molecular and genetic basis of proanthocyanidin synthesis in barley

Flavonoid synthesis begins with the formation of chalcone by the condensation of three malonyl-CoA molecules with one 4-coumaroyl-CoA molecule. 4-Coumaroyl-CoA serves as a precursor not only for flavonoids but also for lignans, allomelanins, and lignin, which are also found in barley grain (Bartłomiej et al., 2012; Shoeva et al., 2020; Yu et al., 2023). The flavonoid synthetic pathway leads to the formation of leucoanthocyanidins, which are common precursors of PAs and anthocyanins (see the Figure). At this stage, the pathway branches: leucoanthocyanidin reductase (LAR) catalyzes the synthesis of flavan-3-ols, the monomers of PA, while the competitive enzyme anthocyanidin synthase (ANS) oxidizes leucoanthocyanidins to anthocyanidins (Saito et al., 1999; Tanner et al., 2003). Subsequent glycosylation of anthocyanidins leads to the formation of anthocyanins. Anthocyanidins can also be reduced to flavan-3-ols by anthocyanidin reductase (ANR), thus forming an alternative pathway for PA synthesis (Xie et al., 2003).

The polymerization of flavan-3-ols to form PAs probably occurs non-enzymatically, although the involvement of an unknown flavanol-condensing enzyme has been previously



Scheme of synthesis and intracellular transport of proanthocyanidins.

Phenylalanine (Phe) and tyrosine (Tyr) serve as substrates for the synthesis of 4-coumaroyl-CoA in the phenylpropanoid pathway. From 4-coumaroyl-CoA, with the participation of enzymes of the flavonoid synthetic pathway, flavan-3-ols are formed. Transport of flavan-3-ols into the vacuole is possible by three potential mechanisms: (1) via transporters of the MATE family, (2) via vesicular transport, and (3) as part of specialized organelles – tannosomes, where their polymerization may occur. Proteins of the GST family can bind flavan-3-ols and facilitate their transport into vesicles or vacuole. Polymerization of flavan-3-ols into PA in the vacuole can occur spontaneously due to the acidic environment or with the participation of an unknown enzyme. Enzymes of the flavonoid synthesis pathway: ANR – anthocyanidin reductase; ANS – anthocyanidin synthase; CHI – chalcone-flavanone isomerase (*Ant30*); CHS – chalcone synthase; DFR – dihydroflavonol-4-reductase (*Ant18*); F3H – flavanone-3-hydroxylase (*Ant17*); F3'H – flavonoid 3'-hydroxylase; F3'5'H – flavonoid 3',5'-hydroxylase; LAR – leucoanthocyanidin reductase (*Ant19*). The activity of the flavonoid pathway enzymes is controlled by the regulatory genes *Ant1*, *Ant2*, *Ant13*, *Ant28*.

suggested (Jende-Strid, 1993; He et al., 2008; Yu et al., 2023). Flavan-3-ols are proposed as the starting units of polymerization, while leucoanthocyanidins and their derivatives are considered to be the extention units. Another important aspect of PA biosynthesis is their intracellular transport (see the Figure). Flavan-3-ols are synthesized in the cytosol, while the final site of PA accumulation is the vacuole (Winkel, 2004). The transport of flavan-3-ols into the vacuole is mediated by MATE family proteins, the functioning of which depends on the proton gradient on the vacuolar membrane generated by H^+-ATPases (Yu et al., 2023). In addition, proteins of the glutathione-S-transferase (GST) family are involved in the transport of flavonoids, including PA, which presumably perform the function of binding and delivering flavonoids to vacuolar transporters (Pérez-Díaz et al., 2016). A tannosome model of PA synthesis was also proposed according to which PA polymerization occurs in specialized organelles – tannosomes, formed from chloroplast thylakoids (Brillouet et al., 2013).

PA biosynthesis is under complex transcriptional control. The key regulatory module is the MBW complex, consisting of transcription factors of the MYB, bHLH, and WD40 families (Bulanov et al., 2025). The complex activates the expression of structural genes of the flavonoid biosynthesis pathway, providing their spatiotemporal regulation. In barley, a MYB factor encoded by the *HvMyb10* (or *Ant28*) gene has been characterized; it specifically controls PA synthesis by regulating the expression of *Dfr* and *Lar* genes (Jende-Strid, 1993; Himi et al., 2011). In addition to the MBW complex, transcription factors of the WRKY, MADS, and WIP families are involved in the regulation of PA synthesis (He et al., 2008).

Many genes involved in barley flavonoid synthesis were identified through analysis of the *Anthocyanin-less* (*Ant*) mutant collection. This collection was created in the 1970s through induced mutagenesis of various cultivars and lines from Europe, USA, and Japan and comprises about 900 mutants with impaired flavonoid synthesis. Based on allelism testing of 566 mutants, 30 *Ant* loci were described (Jende-Strid, 1993). To date, the molecular functions of only seven of them have been established (see the Figure). Of particular interest are the *Ant19* and *Ant25–Ant29* loci, mutations in

which specifically suppress PA synthesis. Although the molecular functions for many loci remain unknown, a potential function has been proposed for some of them based on biochemical analysis of mutants. For example, the *Ant25*, *Ant27*, and *Ant29* loci presumably encode regulatory factors, since the activity of several enzymes of the biosynthetic pathway is impaired in the corresponding mutants. In contrast, *Ant26* is likely a structural gene controlling the final stages of synthesis, since *ant26* mutant grains accumulate monomeric flavan-3-ols in the absence of PA (Jende-Strid, 1993).

Mutant lines not only became an important tool for studying PA synthesis genes but also formed the basis for the development of PA-free brewing cultivars. The first such line was *ant13.13*, obtained from the Foma cultivar. Despite the reduced yield of *ant13.13* compared to the original cultivar, beer produced from the grain of this line demonstrated excellent colloidal stability (von Wettstein et al., 1977). Furthermore, using PA-free mutants, it was shown that the absence of PA in beer does not affect its organoleptic characteristics (von Wettstein et al., 1977; Delcour et al., 1984). These results confirmed the potential of this approach and stimulated further breeding work to develop PA-free barley cultivars, which, in addition to PA content, must meet the general requirements for malting.

Quality parameters of brewing barley cultivars

Malting barley is subject to stringent requirements (GOST 5060–2021) that must be taken into account during selection. A key quality criterion is protein content, which should not exceed 11.5–12.0 % of dry weigh (DW). Excess protein inhibits starch degradation and reduces extractability, while a protein deficiency limits yeast nutrition. Suffitient starch content, which is converted into fermentable sugars, is equally important. For malting barley, starch content should be at least 60–65 % of DW (Golovin et al., 2008).

The grain nature (grain weight per liter) of malting barley, which characterizes its plumpness, must be no less than 660 g/L. The 1,000-kernel weight depends on the grain size and is optimal within the range of 40–50 g, as grains that are too large do not malt quickly enough. The grain size, determined by the proportion of grains passing through a 2.5×20 mm sieve, should be no less than 85 % for first-class malting barley, as larger grains contain more starch and have higher extractability. Husk content, or the percentage of husks in the total grain weight, should not exceed 9 %, as an excess husk reduces the starch content and extractability and impairs beer taste, although a moderate amount of husk is necessary for the formation of a filter layer (Khokonova, 2015). The moisture content of raw grain should not exceed 12 % to prevent mold growth and mycotoxin accumulation during storage (Chi et al., 2003). Germination energy and germination capacity are critical for uniform malting, as starch in ungerminated grains is not fully fermented, which reduces extract and beer yield. Good-quality grain should have a germination capacity of over 95 %.

A crucial stage in beer production is malting, during which a complex of hydrolytic enzymes (proteases and amylases) is formed. The resulting malt serves not only as a substrate for fermentation but also as a source of coloring and aromatic components for beer (Bamforth, 2009). The main quality indicator of malt is malt extract – the proportion of dry matter transferred to the wort – which is at least 79 % for high-quality grain. Diastatic power, measured in Windesch–Kolbach units (WK) and characterizing the activity of amylolytic enzymes, should exceed 220 WK units for spring barley and 350 WK units for winter barley. The Kolbach index is an indicator of the degree of protein degradation and is expressed as the ratio of soluble nitrogen to the total nitrogen content in the malt. Deviation from the optimal values of 35–49 % lead to filtration problems or taste deterioration (Kumar V. et al., 2023). Free amino nitrogen content is essential for normal yeast metabolism during fermentation. The optimal concentration is 140–180 mg/L. The β-glucan content in malt should not exceed 200 mg/100 g, as its excess increases wort viscosity, which impairs filtration and reduces beer quality. These indicators are quantitative traits and are polygenically controlled (Trubacheeva, Pershina, 2021), which significantly complicates barley breeding for brewing purposes.

Advances in breeding proanthocyanidin-free barley cultivars

Developing PA-free barley cultivars that meet stringent brewing standards has been challenging because the original PA-free mutants exhibited a number of disadvantages, including low grain plumpness, reduced yield and 1,000-kernel weight (Figueroa et al., 1989; Bregitzer et al., 1995; Wu, 1995). Since PAs are involved in maintaining seed dormancy, reduced grain dormancy was observed in the mutants (Himi et al., 2011). Although accelerated and uniform germination may be an advantage for malting, reduced dormancy increases the risk of pre-harvest sprouting, especially under high humidity conditions.

Technological limitations were also of particular importance. Low protein content is desirable for brewing cultivars, but PA-free mutants tended to have higher protein content than the original cultivars (von Wettstein et al., 1977; Øverland et al., 1994; Wu, 1995). Malt produced from the mutant grain exhibited reduced malt extract, diastatic power, and degree of attenuation (Bregitzer et al., 1995). Despite these limitations, mutant lines were actively used in breeding of PA-free cultivars, which were registered and cultivated in Europe, USA, Japan, and the Republic of Korea (see the Table).

In Denmark, the Galant cultivar was bred from the *ant17.148* mutant. In trials conducted in 1982–1983 in 13 European regions, Galant showed yield only 1–7 % lower than the standard (Erdal, 1986). However, Galant had reduced synthesis of enzymes that degrade cell wall polysaccharides and starch compared to the original cultivar, which led to a decrease in the proportion of fermentable sugars and an increase in wort viscosity (Palmer, 1988). The Japanese line Mokkei 92-130, derived from the *ant13.347* mutant, was relatively successful, demonstrating high malt extract and

Registered proanthocyanidin-free barley cultivars

Name of the cultivar	Pedigree	Year of creation	Country	Organization	Reference
Galant	*ant17.148* (Triumph)	1985	Denmark	Carlsberg Laboratory (Copenhagen)	Erdal, 1986
NFC8808	*ant27.488* (Zenit) × Sewa × Fergie	1994		Seet Plantbreeding (Horsens)	Wettstein, 1995
Caminant	*ant28.484* (Grit) × Blenheim	1994			
Clarity	–	1993	United Kingdom	CRISP Malting Group (Norfolk)	Clarity. Proantho-cyanidin-free malt *
Radiant	*ant29.667* (Harrington) × Baronesse	2003	USA	Washington State University Agricultural Research Center (Washington)	von Wettstein et al., 2004
Fritz	*ant-499* (Apex) × Alexis × Baronesse × Cellar	2014			Rustgi et al., 2020
Yeongbaekchal	Radiant (*ant29.667*) × Saechalssal	2013	Korea	National Institute of Crop Science, RDA (Wanju)	Lee et al., 2016
Tochinoibuki	*ant28.494* (Catrin) × Tochikei216 × Tochikei253	2007	Japan	Tochigi Prefectural Agricultural Experimental Station (Tochigi)	Takayama et al., 2011
Shiratae Nijo	*ant28.494* × Nishinohoshi	2009		National Agricultural Research Center for Kyushu Okinawa Region (Kumamoto), National Institute of Crop Science (Tsukuba)	Tonooka et al., 2010
Kirari-mochi	Tochinoibuki (*ant28.494*) × Yumesakiboshi × Shikoku-hadaka	2009		National Agriculture Research Center for the Western Region (Fukuyama)	Yanagisawa et al., 2011
Mochikinuka	Daikei RF0831 (*ant28.494*) × Daikei LM1 × Sachiho Golden × Daikei HL9-2-6	2017		Tochigi Prefectural Agricultural Experimental Station (Tochigi)	Yamaguchi et al., 2019
Fukumi Fiber	Yon R kei 3755 / Yon kei 9814 (*ant28.2131* (Alexis))	2018		National Agriculture Research Center for the Western Region (Fukuyama)	Takahashi et al., 2025

* Clarity. Proanthocyanidin-free malt. 1999. CRISP Malting Group. Available at: https://www.yumpu.com/en/document/read/4110511/

diastatic power. However, the beer prepared from it was characterized by an accelerated deterioration of organoleptic properties (Fukuda et al., 1999), which was probably due to the absence of phenolic compounds that usually ensure oxidative stability of beer. In the USA, more than 40 promising PA-free lines were created based on the *ant-517* mutant (Wesenberg et al., 1989). Many of these lines were equal to or superior to the standard Klages cultivar in most malting quality indicators, but were inferior in yield, percentage of malt extract, and also had an excessive amount of protein (Wesenberg et al., 1989). Also in the USA, PA-free lines with improved traits were created based on the *ant13.582* line and the Azure, Glenn, and Hazen cultivars: earlier heading, increased roundness of grains and higher soluble protein content. However, due to the negative impact of the mutant *ant13* allele on malt quality and other traits, the authors considered the *ant13.582* line unsuitable for further breeding (Horsley et al., 1991).

Thus, practice has shown that mutations blocking the initial stages of PA biosynthesis are generally unsuitable for cultivar development due to their pleiotropic effects on yield or brewing quality. The use of mutants with impaired PA synthesis at the final stages is more promising, as it minimizes the side effects of mutations. For example, mutants at the *Ant26* locus, which have impaired PA synthesis at the final stages, demonstrate productivity at the level of the original cultivar Grit (Totsky et al., 2024). In Denmark, the NFC8808 and Caminant cultivars were developed based on the *ant27.488* and *ant28.484* mutants, respectively. The Caminant cultivar exceeded the standard yield by 4 %, including in conditions where there was no fungicidal protection (Wettstein, 1995). It met the European Alexis standard in key parameters such as germination index, malt extract, diastatic power, nitrogen content, and β-glucan content. In 1999, the Clarity cultivar, characterized by high colloidal stability and yield, was registered in the United Kingdom (see the Table). The authors

did not report the cultivar's pedigree, but the presence of anthocyanin pigmentation in the vegetative organs of Clarity indicates that the synthesis was disrupted at the final steps.

It is worth noting that the breeding of feed and food PA-free cultivars is actively developing in addition to malting cultivars. In 2003, the Radiant cultivar, based on the *ant29.667* mutant, was officially approved in the USA. Due to its heat-stable β-amylase, Radiant possessed high diastatic power (von Wettstein et al., 2004). The cultivar also demonstrated high yield, resistance to several pathogens, and improved nutritional characteristics, making it suitable for use not only in the brewing but also in the food industry. Korean breeders, based on Radiant, developed the food cultivar Yeongbaekchal (Lee et al., 2016). Porridge made from Yeongbaekchal grain did not darken after heat treatment, making the product more attractive to consumers. The cultivar was resistant to barley yellow mosaic virus (BaYMV), and its yield was only 5 % lower than the control.

In the USA, the Fritz cultivar, which exhibits excellent germination and high resistance to yellow and leaf rust and powdery mildew, was developed from the *ant-499* mutant (Rustgi et al., 2020). The locus carrying the mutation in the original mutant *ant-499* has not been determined. However, the presence of anthocyanins in the stem of Fritz indicates a defect in PA biosynthesis specifically at late stages. Although Fritz was registered as a forage cultivar due to its high protein and β-glucan content, it also demonstrated good malting qualities, making it a potential dual-purpose cultivar.

Japanese researchers developed the two-row Shiratae Nijo cultivar using the *ant28.494* mutant (Tonooka et al., 2010). Despite the initial late maturity anf susceptibility of the *ant28.494* mutant to BaYMV, they obtained a cultivar not inferior to the original Nishinohoshi in yield and resistance to BaYMV and powdery mildew. Notably, despite reduced levels of flavan-3-ols and PAs, compounds with antifungal activity, Shiratae Nijo's susceptibility to fusarium head blight did not exceed that of the original cultivar. The same *ant28.494* mutant was used to create the Tochinoibuki cultivar (Takayama et al., 2011), which is similar to the standard Sukai Golden in heading and ripening dates and 1,000-kernel weight. Subsequently, the Tochinoibuki cultivar was used to develop the naked, waxy, PA- and amylose-free Kirari-mochi cultivar (Yanagisawa et al., 2011). Pearled barley from Kirari-mochi grain contained 1.5 times more β-glucans than the standard Ichibanboshi cultivar, making it particularly valuable for the production of functional food products. A waxy, PA-free Mochikinuka cultivar was also developed from Tochinoibuki. The absence of the enzyme lipoxygenase-1, which catalyzes lipid oxidation, in this cultivar led to a more pleasant taste of the groats (Yamaguchi et al., 2019). A recent achievement of Japanese breeding was the six-row naked cultivar Fukumi Fiber, which combines the mutant alleles *wax* and *amo1*, which control the content of β-glucans, and *ant28* (Takahashi et al., 2025). Due to this, the β-glucan content in Fukumi Fiber grain reached 13.2 % – three times higher than the standard Ichibanboshi and twice as high as the waxy Kirari-mochi.

Thus, despite the significant shortcomings inherent in PA-free mutants, targeted breeding work has proved the possibility of creating competitive barley cultivars based on them.

Conclusion

Breeding malting barley cultivars that do not accumulate PA in grain has been associated with a number of challenges related to reduced yield and grain quality. However, based on world experience in developing PA-free cultivars, it can be concluded that the choice of a mutant line as a trait donor largely determines the success of breeding. It has been shown that mutations in genes specifically controlling the PA synthesis branch have the least negative impact on agronomic traits, as the synthesis of other physiologically important flavonoids is preserved. Based on breeding experience and comparative morphological studies, mutants for the *Ant26*, *Ant28*, and *Ant29* genes currently represent the most promising donors for breeding PA-free cultivars. Despite the high breeding potential of mutants in these genes, the molecular functions of only one of them, *Ant28*, are known. The lack of data on the functions of the remaining genes makes it impossible to develop DNA markers for selection, which are effectively used in breeding for the *Ant28* gene. This highlights the need for fundamental research into the molecular genetic mechanisms of PA biosynthesis, since a deep understanding of these fundamentals is the key to the creation of competitive commercial cultivars.

References

Bamforth C. Current perspectives on the role of enzymes in brewing. *J Cereal Sci*. 2009;50(3):353-357. doi 10.1016/j.jcs.2009.03.001

Bartłomiej S., Justyna R.-K., Ewa N. Bioactive compounds in cereal grains – occurrence, structure, technological significance and nutritional benefits – a review. *Food Sci Technol Int*. 2012;18(6):559-568. doi 10.1177/1082013211433079

Bregitzer P., Wesenberg D.M., Jones B.L. Effect of the ANT-13 locus (proanthocyanidin-free) on the malting quality and agronomic performance of barley. *J Am Soc Brew Chem*. 1995;53(4):191-193. doi 10.1094/ASBCJ-53-0191

Brillouet J.-M., Romieu C., Schoefs B., Solymosi K., Cheynier V., Fulcrand H., Verdeil J.-L., Conéjéro G. The tannosome is an organelle forming condensed tannins in the chlorophyllous organs of Tracheophyta. *Ann Bot*. 2013;112(6):1003-1014. doi 10.1093/aob/mct168

Bulanov A.N., Andreeva E.A., Tsvetkova N.V., Zykin P.A. Regulation of flavonoid biosynthesis by the MYB-bHLH-WDR (MBW) complex in plants and its specific features in cereals. *Int J Mol Sci*. 2025;26(2):734. doi 10.3390/ijms26020734

Ci X.-F., Wu D.-X., Lou X.-Y., Xia Y., Shu Q. Comparative studies on the starch gelatinization characteristics of five cereal crops. *Acta Agron Sin*. 2003;29(2):300-304

Debeaujon I., Léon-Kloosterziel K.M., Koornneef M. Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiol*. 2000;122(2):403-413. doi 10.1104/pp.122.2.403

Delcour J.A., Schoeters M.M., Meysman E.W., Dondeyne P., Moerman E. The intrinsic influence of catechins and procyanidins on beer haze formation. *J Inst Brew*. 1984;90(6):381-384. doi 10.1002/j.2050-0416.1984.tb04293.x

Dixon R.A., Xie D.-Y., Sharma S.B. Proanthocyanidins – a final frontier in flavonoid research? *New Phytol*. 2005;165(1):9-28. doi 10.1111/j.1469-8137.2004.01217.x

Erdal K. Proanthocyanidin-free barley – malting and brewing. *J Inst Brew*. 1986;92(3):220-224. doi 10.1002/j.2050-0416.1986.tb04404.x

Figueroa J.D.C., Madson M.A., D'Appolonia B.L. The malting and brewing quality of crosses of barley anthocyanogen-free mutants. *J Am Soc Brew Chem*. 1989;47(2):44-48. doi 10.1094/ASBCJ-47-0044

Fukuda K., Saito W., Arai S., Aida Y. Production of a novel proanthocyanidin-free barley line with high quality. *J Inst Brew*. 1999; 105(3):179-184. doi 10.1002/j.2050-0416.1999.tb00017.x

Golovin V.V., Levakova O.V., Artem'eva E.A. Innovative Technology of Spring Barley Cultivation for Brewing Purposes Using Modern and Promising Varieties. Ryazan', 2008 (in Russian)

GOST 5060–2021. Regional standard. Malting Barley: Specification. Promulgated by the RF Federal Agency for Technical Regulation and Measurement, Directive No. 1714-st of December 6, 2021. (in Russian)

He F., Pan Q.-H., Shi Y., Duan C.-Q. Biosynthesis and genetic regulation of proanthocyanidins in plants. *Molecules*. 2008;13(10):2674-2703. doi 10.3390/molecules13102674

Himi E., Yamashita Y., Haruyama N., Yanagisawa T., Maekawa M., Taketa S. *Ant28* gene for proanthocyanidin synthesis encoding the R2R3 MYB domain protein (Hvmyb10) highly affects grain dormancy in barley. *Euphytica*. 2011;188(1):141-151. doi 10.1007/s10681-011-0552-5

Horsley R.D., Schwarz P.B., Foster A.E. Effects of gene *ant13* on agronomic and malt quality traits of barley. *Crop Sci*. 1991;31(3): 593-598. doi 10.2135/cropsci1991.0011183X003100030009x

Jende-Strid B. Genetic control of flavonoid biosynthesis in barley. *Hereditas*. 1993;119(2):187-204. doi 10.1111/j.1601-5223.1993.00187.x

Khokonova M. Influence of barley harvesting method on grain yield and brewing qualities. *Innovation Science*. 2015;(8-2):87-88 (in Russian)

Kohyama N., Fujita M., Ono H., Ohnishi-Kameyama M., Matsunaka H., Takayama T., Murata M. Effects of phenolic compounds on the browning of cooked barley. *J Agric Food Chem*. 2009;57(14): 6402-6407. doi 10.1021/jf901944m

Kumar K., Sinha R.R.K., Kumar S., Nirala R.K., Kumari S., Sahu S.P. Significance of tannins as an alternative to antibiotic growth promoters in poultry production. *Pharma Innovation J*. 2022;11(11S): 1435-1440

Kumar V., Chaturvedi S., Singh G. Brief review of malting quality and frontier areas in barley. *Cereal Res Commun*. 2023;51(1):45-59. doi 10.1007/s42976-022-00292-z

Lee M.J., Kim Y.-K., Kim K.-H., Seo W.-D., Kang H.-J., Park J.-C., Hyun J.-N., Park K.-H. Quality characteristics and development of naked waxy barley (*Hordeum vulgare* L.) cultivar "Yeongbaekchal" without discoloration of cooked barley. *Korean J Breed Sci*. 2016; 48(4):529-534. doi 10.9787/KJBS.2016.48.4.529

Øverland M., Heintzman K.B., Newman C.W., Newman R.K., Ullrich S.E. Chemical composition and physical characteristics of proanthocyanidin-free and normal barley isotypes. *J Cereal Sci*. 1994; 20(1):85-91. doi 10.1006/jcrs.1994.1048

Palmer G. Enzyme development in the aleurone and embryos of Galant and Triumph barleys. *J Inst Brew*. 1988;94(2):61-63. doi 10.1002/j.2050-0416.1988.tb04557.x

Pérez-Díaz R., Madrid-Espinoza J., Salinas-Cornejo J., González-Villanueva E., Ruiz-Lara S. Differential roles for VviGST1, VviGST3, and VviGST4 in proanthocyanidin and anthocyanin transport in *Vitis vinífera*. *Front Plant Sci*. 2016;7:1166. doi 10.3389/fpls.2016.01166

Rustgi S., Brouwer B., von Wettstein D., Reisenauer P.E., Lyon S., Ankrah N., Jones S., Guy S.O., Chen X. Registration of 'Fritz', a two-row spring barley. *J Plant Regist*. 2020;14(3):242-249. doi 10.1002/plr2.20046

Saito K., Kobayashi M., Gong Z., Tanaka Y., Yamazaki M. Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red forma of *Perilla frutescens*. *Plant J Cell Mol Biol*. 1999;17(2):181-189. doi 10.1046/j.1365-313x.1999.00365.x

Santos-Buelga C., Scalbert A. Proanthocyanidins and tannin-like compounds – nature, occurrence, dietary intake and effects on nutrition and health. *J Sci Food Agric*. 2000;80(7):1094-1117. doi 10.1002/(SICI)1097-0010(20000515)80:7%3C1094::AID-JSFA569%3E3.0.CO;2-1

Shoeva O.Yu., Mursalimov S.R., Gracheva N.V., Glagoleva A.Yu., Börner A., Khlestkina E.K. Melanin formation in barley grain occurs within plastids of pericarp and husk cells. *Sci Rep*. 2020; 10(1):179. doi 10.1038/s41598-019-56982-y

Takahashi A., Yoshioka T., Yanagisawa T., Nagamine T., Sugita T. Breeding of Fukumi Fiber, a new six-rowed waxy hull-less barley cultivar containing high levels of β-glucan with a proanthocyanidin-free gene. *Breed Sci*. 2025;75(3):236-243. doi 10.1270/jsbbs.24080

Takayama T., Sotome T., Oozeki M., Haruyama N., Yamaguchi M., Okiyama T., Nagamine T., Kato T., Watanabe H., Oono K. Breeding of a new two-rowed pearling barley cultivar "Tochinoibuki". *Bull Tochigi Agric Exp Station*. 2011;66:53-66

Tanner G.J., Francki K.T., Abrahams S., Watson J.M., Larkin P.J., Ashton A.R. Proanthocyanidin biosynthesis in plants. Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. *J Biol Chem*. 2003;278(34):31647-31656. doi 10.1074/jbc.M302783200

Tonooka T., Kawada N., Yoshida M., Yoshioka T., Oda S., Hatta K., Hatano T., Fujita M., Kubo K. Breeding of a new food barley cultivar "Shiratae Nijo" exhibiting no after-cooking discoloration. *Breed Sci*. 2010;60(2):172-176

Totsky I.V., Li R., Shoeva O.Yu. The effect of the *Ant25*, *Ant26* and *Ant27* loci controlling proanthocyanidin synthesis in barley (*Hordeum vulgare* L.) grain on plant growth and development. *Proceedings on Applied Botany, Genetics and Breeding*. 2024;185(2): 138-146. doi 10.30901/2227-8834-2024-2-138-146 (in Russian)

Trubacheeva N.V., Pershina L.A. Problems and possibilities of studying malting quality in barley using molecular genetic approaches. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2021;25(2):171-177. doi 10.18699/VJ21.021

Van Hung P. Phenolic compounds of cereals and their antioxidant capacity. *Crit Rev Food Sci Nutr*. 2016;56(1):25-35. doi 10.1080/10408398.2012.708909

von Wettstein D. Breeding of value added barley by mutation and protein engineering. In: Induced Mutations and Molecular Techniques for Crop Improvement: Proceedings of an international symposium. Vienna, 1995;67-76

Von Wettstein D. From analysis of mutants to genetic engineering. *Annu Rev Plant Biol*. 2007;58:1-19. doi 10.1146/annurev.arplant.58.032806.104003

von Wettstein D., Jende-Strid B., Ahrenst-Larsen B., Sørensen J.A. Biochemical mutant in barley renders chemical stabilization of beer superfluous. *Carlsberg Res Commun*. 1977;42(5):341-351. doi 10.1007/BF02906119

von Wettstein D., Cochran J., Ullrich S., Kannangara C., Jitkov V., Burns J., Reisenauer P., Chen X., Jones B. Registration of 'Radiant' barley. *Crop Sci*. 2004;44(5):1859-1861. doi 10.2135/cropsci2004.1859

Wang Y., Ye L. Haze in beer: its formation and alleviating strategies, from a protein–polyphenol complex angle. *Foods*. 2021;10(12): 3114. doi 10.3390/foods10123114

Wesenberg D.M., Jones B.L., Robbins G.S., Cochran J. Malting quality and agronomic characteristics of selected proanthocyanidin-free barleys. *J Am Soc Brew Chem*. 1989;47(3):82-86. doi 10.1094/ASBCJ-47-0082

Winkel B.S.J. Metabolic channeling in plants. *Annu Rev Plant Biol*. 2004;55:85-107. doi 10.1146/annurev.arplant.55.031903.141714

Wu G. Associations between three proanthocyanidin-free genes and some important characteristics in barley (*Hordeum vulgare* L.). PhD thesis. Univ. of Saskatchewan, 1995

Xie D.-Y., Sharma S.B., Paiva N.L., Ferreira D., Dixon R.A. Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. *Science*. 2003;299(5605):396-399. doi 10.1126/science.1078540

Yamaguchi M., Oozeki M., Sotome T., Oyama M., Kato T., Sekiwa T., Mochizuki T., Okiyama T., Haruyama N., Takayama T. "Mochikinuka", a new two-rowed waxy hulled barley cultivar with superior characteristics of food quality. *Bull Tochigi Prefect Agric Exp Stn Jpn*. 2019;79:1-22

Yanagisawa T., Nagamine T., Takahashi A., Takayama T., Doi Y., Matsunaka H., Fujita M. Breeding of Kirari-mochi: a new two-rowed waxy hull-less barley cultivar with superior quality characteristics. *Breed Sci*. 2011;61(3):307-310. doi 10.1270/jsbbs.61.307

Yu K., Song Y., Lin J., Dixon R.A. The complexities of proanthocyanidin biosynthesis and its regulation in plants. *Plant Commun*. 2023;4(2):100498. doi 10.1016/j.xplc.2022.100498

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

Received October 28, 2025. Revised December 22, 2025. Accepted December 22, 2025.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» до 2011 г. выходил под названием «Информационный вестник ВОГиС»/ "The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Сетевое издание «Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» – реестровая запись СМИ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, 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 01.12.2025. Формат 60 × 84 1/8. Уч.-изд. л. 32.9