

# ВАВИЛОВСКИЙ

## ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

2026 • 30 • 1



Биоинформатика и системная биология / Генетика животных / Генетика и селекция растений /  
Молекулярная и клеточная биология / Генетика человека / Генетика микроорганизмов / Популяционная генетика



Сетевое издание

# ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

## VAVILOV JOURNAL OF GENETICS AND BREEDING

Основан в 1997 г.

Периодичность 8 выпусков в год

doi 10.18699/vjgb-26-01

**Учредители**

Сибирское отделение Российской академии наук

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

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

**Главный редактор**

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

**Заместители главного редактора**

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

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

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

В.К. Шумный – академик РАН, д-р биол. наук, профессор (Россия)

**Ответственный секретарь**

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

**Редакционная коллегия**

Е.Е. Андронов – канд. биол. наук (Россия)

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

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

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

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

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

А. Бёрнер – д-р наук (Германия)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

С.А. Лашин – д-р биол. наук (Россия)

Ж. Ле Гуи – д-р наук (Франция)

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

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

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

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

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

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

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

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

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

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

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

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

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

Е.Ю. Рыкова – д-р биол. наук (Россия)

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

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

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

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

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

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

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

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

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

Н.С. Юдин – канд. биол. наук (Россия)

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

Online edition

# VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

## VAVILOV JOURNAL OF GENETICS AND BREEDING

*Founded in 1997**Publication frequency: 8 issues a year*

doi 10.18699/vjgb-26-01

**Founders**

Siberian Branch of the Russian Academy of Sciences

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

The Vavilov Society of Geneticists and Breeders

**Editor-in-Chief***A.V. Kochetov*, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Professor of the RAS, Russia**Deputy Editor-in-Chief***N.A. Kolchanov*, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia*I.N. Leonova*, Dr. Sci. (Biology), Russia*N.B. Rubtsov*, Professor, Dr. Sci. (Biology), Russia*V.K. Shumny*, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia**Executive Secretary***G.V. Orlova*, Cand. Sci. (Biology), Russia**Editorial board***O.S. Afanasenko*, Full Member of the RAS, Dr. Sci. (Biology), Russia*D.A. Afonnikov*, Associate Professor, Dr. Sci. (Biology), Russia*L.I. Aftanas*, Full Member of the RAS, Dr. Sci. (Medicine), Russia*E.E. Andronov*, Cand. Sci. (Biology), Russia*Yu.S. Aulchenko*, Dr. Sci. (Biology), The Netherlands*L.A. Beshpalova*, Full Member of the RAS, Dr. Sci. (Agricul.), Russia*N.P. Bondar*, Cand. Sci. (Biology), Russia*S.A. Borinskaya*, Dr. Sci. (Biology), Russia*P.M. Borodin*, Professor, Dr. Sci. (Biology), Russia*A. Börner*, Dr. Sci., Germany*M. Chen*, Dr. Sci. (Biology), People's Republic of China*S.A. Demakov*, Dr. Sci. (Biology), Russia*T.A. Gavrilenko*, Dr. Sci. (Biology), Russia*N.E. Gruntenko*, Dr. Sci. (Biology), Russia*S.G. Inge-Vechtomov*, Full Member of the RAS, Dr. Sci. (Biology), Russia*E.K. Khlestkina*, Corr. Member of the RAS, Professor of the RAS,

Dr. Sci. (Biology), Russia

*E.K. Khusnutdinova*, Professor, Dr. Sci. (Biology), Russia*A.V. Kilchevsky*, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology),

Belarus

*A.M. Kudryavtsev*, Corr. Member of the RAS, Dr. Sci. (Biology), Russia*D.M. Larkin*, Cand. Sci. (Biology), Great Britain*S.A. Lashin*, Dr. Sci. (Biology), Russia*J. Le Gouis*, Dr. Sci., France*I.N. Lebedev*, Corr. Member of the RAS, Professor, Dr. Sci. (Biology), Russia*B. Lugtenberg*, Professor, Dr. Sci., Netherlands*L.A. Lutova*, Professor, Dr. Sci. (Biology), Russia*V.Yu. Makeev*, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.),

Russia

*I.V. Maksimov*, Dr. Sci. (Biology), Russia*B.A. Malyarchuk*, Dr. Sci. (Biology), Russia*Yu.G. Matushkin*, Cand. Sci. (Biology), Russia*V.I. Molodin*, Full Member of the RAS, Dr. Sci. (History), Russia*M.P. Moshkin*, Professor, Dr. Sci. (Biology), Russia*L.Yu. Novikova*, Dr. Sci. (Agricul.), Russia*E.K. Potokina*, Dr. Sci. (Biology), Russia*V.P. Puzryev*, Full Member of the RAS, Dr. Sci. (Medicine),

Russia

*D.V. Pyshnyi*, Corr. Member of the RAS, Dr. Sci. (Chemistry),

Russia

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

Dr. Sci. (Biology), Russia

*E.V. Shakhshneider*, Dr. Sci. (Medicine), Russia*S.V. Shekhovtsov*, Dr. Sci. (Biology), Russia*S.V. Shestakov*, Full Member of the RAS, Dr. Sci. (Biology),

Russia

*V.A. Stepanov*, Full Member of the RAS, Dr. Sci. (Biology),

Russia

*I.A. Tikhonovich*, Full Member of the RAS, Dr. Sci. (Biology),

Russia

*A.V. Vasiliev*, Corr. Member of the RAS, Dr. Sci. (Biology), Russia*M.I. Voevoda*, Full Member of the RAS, Dr. Sci. (Medicine),

Russia

*N.K. Yankovsky*, Full Member of the RAS, Dr. Sci. (Biology),

Russia

*N.S. Yudin*, Cand. Sci. (Biology), Russia*I.K. Zakharov*, Professor, Dr. Sci. (Biology), Russia*I.A. Zakharov-Gezekhus*, Corr. Member of the RAS,

Dr. Sci. (Biology), Russia

## Биоинформатика и системная биология

- 5 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Рациональный биоинформатический подход к анализу функциональных свойств метаболитов пробиотических микроорганизмов на основе реконструкции генных сетей. В.А. Иванисенко, Т.М. Хлебодарова, М.А. Клещев, А.Р. Волянская, И.В. Язык, А.В. Адамовская, Т.В. Иванисенко, П.С. Деменков, Н.А. Колчанов
- 15 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Развитие метода филогенетического футпринтинга для распознавания сайтов связывания транскрипционных факторов на основе использования bootstrap-испытаний для анализа больших бактериальных геномных данных. А.М. Мухин, Т.М. Хлебодарова, Д.Ю. Ощепков
- 27 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Детекция колосков в колосе пшеницы на RGB-изображениях с использованием глубокого машинного обучения. М.А. Генаев, И.Д. Бусов, Ю.В. Кручинина, В.С. Коваль, Н.П. Гончаров

## Генетика животных

- 36 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Паттерн распределения *P*-транспозона и *blood* ретротранспозона в отводках референсных линий Harwich и Canton-S из разных лабораторий не влияет на проявление внутривидового РМ гибридного дисгенеза у *Drosophila melanogaster*. Л.П. Захаренко, Ю.Ю. Илинский (на англ. языке)

## Генетика и селекция растений

- 43 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Локус сои qDTF-7 как пример генетической гетерогенности, ассоциированной с продолжительностью цветения и созревания. Р.Н. Перфильев, М.И. Шматова, А.Б. Щербань, Е.А. Салина
- 53 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Гистохимический тест для определения профиля полифенольных соединений в зерне злаков. С.Р. Мурсалимов, О.Ю. Шоева (на англ. языке)

## Молекулярная и клеточная биология

- 61 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Модификация оптогенетической системы VrhP1-QPAS1 для регуляции экспрессии генов в листьях табака *Nicotiana benthamiana* с помощью ближнего инфракрасного света. Э.С. Суркова, Ю.А. Галимова, Н.В. Баттулина, Д.М. Моторина, Е.С. Омелина (на англ. языке)

- 72 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Редактирование оснований в гене *AUTS2* и высокопроизводительное NGS-гено-типирование клонов: стратегия создания клеточной модели. А.П. Ян, П.А. Сальников, А.А. Буздин, В.А. Ковальская, Е.В. Мусатова, П.С. Орлов, О.П. Рыжкова, А.И. Субботовская, М.В. Сунцова, А.Ю. Христинченко, А.А. Хабарова

- 85 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Биоинформатический анализ механизмов жизнеспособности линий опухолевых клеток при делеции генов-супермишеней. Д.А. Четверина, Н.Я. Козельчук, Д.В. Ломаев, А.А. Штиль, М.М. Ерохин

## Генетика человека

- 94 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Особенности полиморфизма гена рецептора тиреоидных гормонов *THRB* у коренного населения Сибири. Б.А. Малайчук, Н.В. Похилук, Г.А. Денисова, А.Н. Литвинов
- 101 **ОБЗОР**  
Транскриптомика тяжелой формы COVID-19. А.А. Гусарова, Е.А. Трифонова, А.А. Бабовская, М.М. Гавриленко, В.А. Степанов

## Генетика микроорганизмов

- 117 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Однонуклеотидные полиморфизмы в геномах вирусов клещевого энцефалита и Западного Нила при тройной природной инфекции у садовой камышовки (*Acrocephalus dumetorum*). Е.П. Пономарева, В.А. Терновой, Е.В. Протопопова, Н.Л. Тупота, В.Б. Локтев

- 126 **ОБЗОР**  
Новые аргументы в дискуссии о природе пикобирнавирuсов. А.Ю. Кашников, Н.В. Епифанова, Н.А. Новикова

## Популяционная генетика

- 136 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Эволюция порядка генов в мтДНК байкальских эндемичных амфипод и ее возможные механизмы. Е.А. Сиротинина, Д.Ю. Щербаков, Е.В. Романова

- 146 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Морфологический и молекулярно-генетический анализ полиморфизма рода *Iris* L. в Республике Башкортостан и Оренбургской области. Э.А. Аухадиева, А.Г. Блинов, Е.И. Вшивцева, З.Х. Шигапов, А.Р. Ишбирдин, Я.М. Голованов, А.В. Крюкова

## Bioinformatics and systems biology

- 5 ORIGINAL ARTICLE  
Rational bioinformatic approach to the analysis of functional properties of metabolites of probiotic microorganisms based on gene network reconstruction. V.A. Ivanisenko, T.M. Khlebodarova, M.A. Kleshchev, A.R. Volyanskaya, I.V. Yatsyk, A.V. Adamovskaya, T.V. Ivanisenko, P.S. Demenkov, N.A. Kolchanov

- 15 ORIGINAL ARTICLE  
Improvement of a phylogenetic footprinting method for transcription factor binding sites recognition based on the use of bootstrap trials for the analysis of large bacterial genomic data. A.M. Mukhin, T.M. Khlebodarova, D.Yu. Oshchepkov

- 27 ORIGINAL ARTICLE  
Wheat spikelet detection on RGB images using deep machine learning. M.A. Genaev, I.D. Busov, Yu.V. Kruchinina, V.S. Koval, N.P. Goncharov

## Animal genetics

- 36 ORIGINAL ARTICLE  
Different patterns of *P* transposon and *blood* retrotransposon distribution in Harwich and Canton-S sub-strains do not affect the manifestation of *Drosophila melanogaster* intraspecific PM hybrid dysgenesis. L.P. Zakharenko, Y.Y. Ilinsky

## Plant genetics and breeding

- 43 ORIGINAL ARTICLE  
Soybean locus qDTF-7 as an example of genetic heterogeneity associated with flowering and maturity time. R.N. Perfil'ev, M.I. Shmatova, A.B. Shcherban, E.A. Salina

- 53 ORIGINAL ARTICLE  
A histochemical assay for polyphenolic profiling in cereal grains. S.R. Mursalimov, O.Yu. Shoeva

## Molecular and cell biology

- 61 ORIGINAL ARTICLE  
Modification of the BphP1-QPAS1 optogenetic system for gene expression regulation in *Nicotiana benthamiana* tobacco leaves using near-infrared light. E.S. Surkova, Y.A. Galimova, N.V. Battulina, D.M. Motorina, E.S. Omelina

- 72 ORIGINAL ARTICLE  
Base editing in the *AUTS2* gene and high-throughput NGS genotyping of clones: a strategy for generating a cellular model. A.P. Yan, P.A. Salnikov, A.A. Buzdin, V.A. Kovalskaia, E.V. Musatova, P.S. Orlov, O.P. Ryzhkova, A.I. Subbotovskaia, M.V. Suntsova, A.U. Khristichenko, A.A. Khabarova

- 85 ORIGINAL ARTICLE  
The computational analysis of tumor cell sensitivity to supertarget deletion. D.A. Chetverina, N.Y. Kozelchuk, D.V. Lomaev, A.A. Shtil, M.M. Erokhin

## Human genetics

- 94 ORIGINAL ARTICLE  
The specific features of the thyroid hormone receptor gene *THRB* polymorphism in indigenous populations of Siberia. B.A. Malyarchuk, N.V. Pokhilyuk, G.A. Denisova, A.N. Litvinov

- 101 REVIEW  
Transcriptomics of severe COVID-19. A.A. Gusarova, E.A. Trifonova, A.A. Babovskaya, M.M. Gavrilenko, V.A. Stepanov

## Microbial genetics

- 117 ORIGINAL ARTICLE  
Single nucleotide polymorphisms are typical for tick-borne encephalitis and West Nile viruses during triple natural mixed infections in Blyth's reed warbler (*Acrocephalus dumetorum*). E.P. Ponomareva, V.A. Ternovoi, E.V. Protopopova, N.L. Tupota, V.B. Loktev

- 126 REVIEW  
New arguments in the discussion about the nature of picobirnaviruses. A.Yu. Kashnikov, N.V. Epifanova, N.A. Novikova

## Population genetics

- 136 ORIGINAL ARTICLE  
Evolution of gene order in mtDNA of Baikal endemic amphipods and its possible mechanisms. E.A. Sirotnina, D.Yu. Sherbakov, E.V. Romanova

- 146 ORIGINAL ARTICLE  
Morphological and molecular genetic analysis of the genus *Iris* L. polymorphism in the Republic of Bashkortostan and the Orenburg Region. E.A. Aukhadieva, A.G. Blinov, E.I. Vshivtseva, Z.Kh. Shigapov, A.R. Ishbirdin, Ya.M. Golovanov, A.V. Kryukova

doi 10.18699/vjgb-26-11

## Рациональный биоинформатический подход к анализу функциональных свойств метаболитов пробиотических микроорганизмов на основе реконструкции генных сетей

В.А. Иванисенко  , Т.М. Хлебодарова, М.А. Клещев , А.Р. Волянская , И.В. Яцык ,  
А.В. Адамовская , Т.В. Иванисенко , П.С. Деменков , Н.А. Колчанов 

Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

 salix@bionet.nsc.ru

**Аннотация.** Важное направление промышленной микробиологии – создание штаммов пробиотиков, обладающих ценными потребительскими свойствами. Индустрия пробиотиков является одним из наиболее динамично развивающихся сегментов пищевой и фармацевтической промышленности. Стеариновая (октадекановая) кислота C18:0 – один из основных метаболитов в клеточно-свободном супернатанте бактерии *Streptococcus thermophilus*, широко используемой в производстве различных ферментированных молочнокислых продуктов, включая йогурт и сыр. *S. thermophilus* влияет не только на текстуру и вкусовые свойства продуктов, но и обладает различными пробиотическими эффектами, в том числе антиоксидантной активностью, модуляцией кишечной микробиоты, ингибированием определенных патогенов и др. Предполагается, что ряд пробиотических эффектов, которыми обладает *S. thermophilus*, может быть опосредован именно через октадекановую кислоту, как основной метаболит. Октадекановая кислота C18:0, как и другие длинноцепочечные жирные кислоты, поступает в организм человека с использованием различных механизмов белок-опосредованного транспорта и пассивной диффузии через мембрану клеток. В клетке стеариновая кислота не только служит субстратом для синтеза триглицеридов и других сложных липидов, но и, как показано на клеточных и *in vivo* моделях, является модулятором сигнальных и стресс-ответов, связанных в том числе с апоптозом. Это один из важных аспектов влияния стеариновой кислоты на функционирование организма, определяющий противовоспалительный и потенциально противоопухолевый эффекты. Однако молекулярно-генетические механизмы влияния октадекановой кислоты как пробиотика на организм человека в этом отношении остаются недостаточно изученными. В настоящем исследовании с помощью разработанной нами ранее информационно-программной системы ANDSystem, использующей методы машинного обучения и искусственного интеллекта и предназначенной для автоматического извлечения знаний из научных текстов и баз данных, были реконструированы генные сети регуляции внутреннего (митохондриального) и внешнего (индуцируемого рецепторами смерти) пути апоптоза клеток человека под влиянием стеариновой кислоты. Для поиска метаболитов, продуцируемых пробиотическими микроорганизмами, обладающих полезными терапевтическими свойствами, разработан новый подход, включающий реконструкцию генных сетей и анализ дифференциально экспрессирующихся генов. На его основе было показано, что стеариновая кислота, продуцируемая *S. thermophilus*, контролирует как внешний, так и внутренний пути апоптоза через регуляцию экспрессии гена *PTGS2*, кодирующего фермент циклооксигеназу-2. Полученные данные позволяют рассматривать циклооксигеназу-2 как один из центральных регуляторов, опосредующих влияние стеариновой кислоты на экспрессию генов апоптоза. В работе предложен рациональный биоинформатический подход к поиску новых штаммов, обладающих пробиотическим потенциалом, на основе оценки действия продуцируемых ими метаболитов на целевые биологические процессы в клетках человека через реконструкцию генных сетей.

**Ключевые слова:** индустриальная микробиология; функциональные микроорганизмы; пробиотики; штаммы-продуценты; метаболиты; генные сети; ANDSystem

**Для цитирования:** Иванисенко В.А., Хлебодарова Т.М., Клещев М.А., Волянская А.Р., Яцык И.В., Адамовская А.В., Иванисенко Т.В., Деменков П.С., Колчанов Н.А. Рациональный биоинформатический подход к анализу функциональных свойств метаболитов пробиотических микроорганизмов на основе реконструкции генных сетей. *Вавиловский журнал генетики и селекции*. 2026;30(1):5-14. doi 10.18699/vjgb-26-11

**Финансирование.** Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (Федеральная научно-техническая программа развития генетических технологий на 2019–2030 годы), соглашение № 075-15-2025-516.

# Rational bioinformatic approach to the analysis of functional properties of metabolites of probiotic microorganisms based on gene network reconstruction

V.A. Ivanisenko  , T.M. Khlebodarova, M.A. Kleshchev , A.R. Volyanskaya , I.V. Yatsyk ,  
A.V. Adamovskaya , T.V. Ivanisenko , P.S. Demenkov , N.A. Kolchanov 

Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

 salix@bionet.nsc.ru

**Abstract.** An important direction in industrial microbiology is the development of probiotic strains with valuable consumer properties. The probiotic industry is currently one of the most rapidly developing segments of the food and pharmaceutical sectors. Stearic (octadecanoic) acid C18:0 is one of the major metabolites present in the cell-free supernatant of the bacterium *Streptococcus thermophilus*, which is widely used in the production of fermented dairy products, including yogurt and cheese. *S. thermophilus* affects not only the texture and sensory properties of products, but also exhibits various probiotic effects, including antioxidant activity, modulation of the gut microbiota, inhibition of certain pathogens, and others. It is assumed that a number of probiotic effects exerted by *S. thermophilus* may be mediated through octadecanoic acid as one of its main metabolites. Octadecanoic acid C18:0, like other long-chain fatty acids, enters the human body via several mechanisms, including protein-mediated transport and passive diffusion across cell membranes. Inside the cell, octadecanoic acid serves not only as a substrate for the synthesis of triglycerides and other complex lipids, but, as shown in cell-based and *in vivo* models, also acts as a modulator of signaling and stress responses, including those associated with apoptosis. This is an important aspect of the influence of stearic acid on organism functioning, underpinning its anti-inflammatory and potentially anti-tumor effects. However, the molecular genetic mechanisms by which octadecanoic acid acts as a probiotic on the human organism remain insufficiently understood. In the present study, using our previously developed information – software system ANDSystem (employing machine learning and artificial intelligence for automatic extraction of knowledge from scientific texts and databases), we reconstructed gene networks regulating the intrinsic (mitochondrial) and extrinsic (death receptor-mediated) apoptotic pathways in human cells under the influence of stearic (octadecanoic) acid. To search for metabolites produced by probiotic microorganisms that may have beneficial therapeutic properties, we propose an approach that combines gene network reconstruction with differential gene expression analysis. Using this approach, we show that octadecanoic acid produced by *S. thermophilus* can control the intrinsic and extrinsic apoptotic pathways primarily via regulation of PTGS2 expression; the results indicate that cyclooxygenase-2 is a key regulator mediating the effect of octadecanoic acid on apoptosis-related genes.

**Key words:** industrial microbiology; functional microorganisms; probiotics; producer strains; metabolites; gene networks; ANDSystem

**For citation:** Ivanisenko V.A., Khlebodarova T.M., Kleshchev M.A., Volyanskaya A.R., Yatsyk I.V., Adamovskaya A.V., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. Rational bioinformatic approach to the analysis of functional properties of metabolites of probiotic microorganisms based on gene network reconstruction. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):5-14. doi 10.18699/vjgb-26-11

## Введение

Важное направление промышленной микробиологии – создание штаммов пробиотиков, обладающих ценными потребительскими свойствами. Индустрия пробиотиков является одним из наиболее динамично развивающихся сегментов пищевой промышленности с огромным объемом мирового рынка. Пробиотические микроорганизмы широко используются при производстве ферментированных продуктов, биологически активных добавок и специализированных пищевых продуктов; при этом к промышленным штаммам предъявляются требования не только по безопасности и технологичности, но и по выраженным функциональным эффектам (иммуномодулирующим, противовоспалительным, метаболически опосредованным и др.) (Terrou et al., 2019; Lau, Quek, 2024; Grujović et al., 2025).

*Streptococcus thermophilus* – особый вид стрептококка, применяемый в промышленной биотехнологии для ферментации пищевых продуктов, особенно при производстве йогурта и сыров. Он влияет на скорость подкисления молочных продуктов, текстуру и их вкусовые свойства, а также обладает различными пробиотическими эффектами, включая антиоксидантную активность, модуляцию кишечной микробиоты, ингибирование определенных патогенов и т. д., что делает его привлекательным для промышленного применения (Cui et al., 2016). В недавней работе (Sudheer et al., 2025) было проведено исследование штаммов *S. thermophilus*, выделенных из домашнего и коммерческого кисломолочного продукта дахи, с использованием геномного анализа и профилирования клеточно-свободного супернатанта (КСС) молочнокислых бактерий методом газовой хроматографии – масс-спектрометрии (ГХ–МС).

Показано, что среди основных компонентов КСС присутствует длинноцепочечная жирная октадекановая (стеариновая) кислота (C18:0), формирующая в метаболомном анализе один из крупнейших пиков (RT  $\approx$  29.12 мин) (Sudheer et al., 2025).

Эти результаты позволяют предположить, что ряд пробиотических эффектов *S. thermophilus* может быть опосредован именно через стеариновую кислоту, как один из основных метаболитов, обнаруживаемых в КСС данной бактерии.

Стеариновая (октадекановая) кислота C18:0, как и другие длинноцепочечные жирные кислоты, поступает в организм человека путем белок-опосредованного транспорта и пассивной диффузии через мембрану клеток. В этих процессах участвуют такие белки, как транслоказа CD36 (SR-B2), играющая ключевую роль в гомеостазе жирных кислот (Chen et al., 2022; Glatz et al., 2022), и транспортеры семейства SLC27/FATP, обеспечивающие импорт кислот в сопряженном процессе их ацилирования при участии длинноцепочечных ацил-КоА-синтетаз (Mashek et al., 2007; Anderson, Stahl, 2013). Доставка кислот *in vivo* к мембране обеспечивается их обратимым связыванием с альбумином (Kamp, Hamilton, 1992; Richieri et al., 1993; Richieri, Kleinfeld, 1995).

В клетках человека стеариновая кислота не только служит субстратом для синтеза триглицеридов и других сложных липидов (Paton, Ntambi, 2009; Minville-Walz et al., 2010; Houten et al., 2016), но и, как показано на клеточных и *in vivo* моделях, является модулятором сигнальных путей и стресс-ответов, связанных с апоптозом, пролиферацией опухолевых клеток, лейкоцитоксичностью, а также с провоспалительными ответами макрофагов и микроглии (Evans et al., 2009; Yang et al., 2020; Hung et al., 2023). В обзоре (Shen X. et al., 2025) достаточно подробно обсуждались функции этой жизненно важной молекулы, в том числе ее роль в таких патологических процессах, как сердечно-сосудистые заболевания, развитие диабета и поражения печени. Согласно существующим данным, стеариновая кислота может влиять на функционирование клетки путем взаимодействия с рецептором CD36 на плазматической мембране с последующей модуляцией внутриклеточных сигнальных путей, связанных с этим рецептором (Chen et al., 2022; Glatz et al., 2022). Кроме того, стеариновая кислота, проникая в клетку, может оказывать регуляторное воздействие на экспрессию ряда генов, в частности за счет модуляции активности микроРНК (Shen X. et al., 2025), что можно охарактеризовать как внутриклеточный путь действия октадекановой кислоты. Однако молекулярно-генетические механизмы влияния стеариновой кислоты как пробиотика на организм человека остаются недостаточно изученными. Один из важных аспектов ее влияния – модуляция уровня программируемой клеточной гибели (Yang et al., 2020) как фактора, определяющего противоспалительный и потенциально противоопухолевый эффекты.

Среди возможных подходов для исследования механизмов потенциального влияния метаболитов, продуцируемых микроорганизмами, на функционирование клеток

человека могут быть реконструкция и анализ генных сетей. Генная сеть – это группа координированно функционирующих генов, контролирующих фенотипические признаки организма (Колчанов и др., 2013). Взаимодействия между генами в генной сети осуществляются через их первичные и вторичные продукты – РНК, белки и метаболиты. Реконструкция генных сетей позволяет выявлять конкретные молекулярные пути в клетках человека, функционирование которых изменяется под влиянием различных факторов, включая метаболиты, предсказывать молекулярно-генетические мишени их действия и влияние на профилактику/развитие заболеваний (Saik et al., 2019; Bragina et al., 2023; Ivanisenco V. et al., 2024).

В настоящем исследовании с помощью разработанной нами ранее информационно-программной системы ANDSystem, использующей методы машинного обучения и искусственного интеллекта (Ivanisenco V. et al., 2015, 2019), были реконструированы генные сети регуляции процесса апоптоза клеток человека под влиянием стеариновой (октадекановой) кислоты, присутствующей в клеточно-свободном супернатанте *S. thermophilus*. Анализ этих генных сетей показал, что стеариновая кислота контролирует апоптоз главным образом через регуляцию экспрессии гена *PTGS2*, кодирующего фермент циклооксигеназу-2. Проведенный дополнительный анализ дифференциальной экспрессии генов в клетках культуры HepG2 (Vendel Nielsen et al., 2013) при воздействии на эти клетки стеариновой кислоты показал, что дифференциально экспрессирующиеся гены вошли в состав реконструированных генных сетей.

Полученные результаты формируют рациональный биоинформатический подход к оценке функциональных эффектов метаболитов микроорганизмов на целевые биологические процессы в клетках человека через реконструкцию генных сетей, который может быть использован при поиске новых штаммов, обладающих пробиотическим потенциалом.

## Материалы и методы

**Реконструкция и анализ генных сетей** регуляции экспрессии генов человека стеариновой кислотой осуществлялись с помощью информационно-программной системы ANDSystem (Ivanisenco V. et al., 2019, 2024; Ivanisenco T. et al., 2020). Система ANDSystem, предназначенная для реконструкции ассоциативных генных сетей, осуществляет в автоматическом режиме извлечение знаний и фактов из научных публикаций и биологических баз данных (Ivanisenco V. et al., 2015, 2019). Одним из ключевых модулей ANDSystem является постоянно пополняемая база знаний. В настоящее время она содержит информацию о более чем 150 млн взаимодействиях между 12 различными типами молекулярно-генетических объектов (гены, РНК, белки, метаболиты, лекарственные вещества и др.), автоматически извлеченную из текстов более 30 млн научных статей и патентов, а также фактографических баз данных. В базе знаний представлены 49 типов взаимодействий между объектами, включая регуляцию экспрессии генов, физические взаимодействия

(белок–белок, белок–лиганд), химические взаимодействия (каталитические реакции, посттрансляционные модификации) и другие (Ivanisenko T. et al., 2022). Эффективность ANDSystem для изучения молекулярных механизмов заболеваний, действия лекарств и метаболитов, в том числе на основе анализа омиксных данных, таких как метаболом и транскриптом, была показана в целом ряде исследований: например, приоритизации генов апоптоза при лимфедеме (Saik et al., 2019), поиске генов апоптоза как основы для коморбидности болезни Хантингтона и рака (Bragina et al., 2023), идентификации метаболических маркеров постоперативного делирия (Ivanisenko V. et al., 2024).

**Анализ дифференциальной экспрессии генов и их функциональная аннотация.** Для изучения влияния стеариновой кислоты на экспрессию генов был проведен анализ доступных транскриптомных данных, полученных с использованием технологии микрочипов (DNA microarrays), в экспериментальном исследовании (Vendel Nielsen et al., 2013), посвященном влиянию жирных кислот (эллаидиновой, олеиновой и стеариновой) на метаболизм клеток культуры HepG2 (идентификатор исследования в GEO Onimbus Database – GSE34045). К настоящему времени это единственное экспериментальное исследование, включающее анализ транскриптомного профиля клеток при воздействии стеариновой кислоты. Выявление дифференциально экспрессирующихся генов (ДЭГ) проводили с использованием пакета Limma (Ritchie et al., 2015). Статистическую значимость различия в уровнях экспрессии генов человека в контроле и при добавлении стеариновой кислоты оценивали с помощью меры FDR (false discovery rate). Различия с FDR менее 0.05 считались статистически значимыми.

Анализ перепредставленности биологических процессов для сформированного списка ДЭГ, выявленных описанным выше методом, выполнялся с использованием веб-сервера DAVID, версия 2021 (<https://david.ncifcrf.gov/>; Sherman et al., 2022), с настройками по умолчанию. DAVID оценивал статистическую значимость пересечения списка изучаемых генов (в нашем случае ДЭГ) со списками генов, соответствующих биологическим процессам, описанным в онтологии генов (GO) (Sherman et al., 2022).

## Результаты и обсуждение

### Реконструкция генных сетей регуляции апоптоза октадекановой кислотой

Известно, что стеариновая кислота может влиять на функционирование клетки как через рецептор CD36 на плазматической мембране (Chen et al., 2022; Glatz et al., 2022), так и проникая в клетку и действуя на экспрессию генов внутри клетки (Shen X. et al., 2025). В связи с этим на основе информации, представленной в базе знаний программно-информационной системы ANDSystem, были реконструированы две генные сети регуляции апоптоза октадекановой кислотой: 1) внутриклеточный путь действия стеариновой кислоты, не включающий белок-рецептор CD36 по информации базы знаний ANDSystem (генная сеть ГС-ВПДО); 2) CD36-опосредованный путь (генная сеть ГС-CD36).

Прежде всего, был сформирован список белок-кодирующих генов человека, участвующих в процессе апоптоза (748 генов), полученный из базы данных онтологии генов (<https://geneontology.org/>) для термина GO:0006915 apoptotic process (табл. S1 Приложения)<sup>1</sup>. Этот список был использован в качестве входных данных в системе ANDSystem для реконструкции генных сетей.

Генная сеть ГС-ВПДО регуляции экспрессии генов апоптоза стеариновой (октадекановой) кислотой, поступающей в клетку, приведена на рис. 1.

Анализ генной сети ГС-ВПДО показал, что эта сеть включала 33 гена апоптоза и дополнительно 11 генов, участвующих в регуляции их экспрессии, а также 48 регуляторных взаимодействий между генами апоптоза и белками-регуляторами и 2 регуляторных взаимодействия между стеариновой (октадекановой) кислотой и генами апоптоза (отмечены на рис. 1 бирюзовыми рамками).

Белок циклооксигеназа-2 имел наибольшее число (25) регуляторных связей с генами апоптоза (табл. 1) в генной сети ГС-ВПДО. Кроме того, к числу ключевых регуляторов генов апоптоза в этой сети можно отнести белки PROX1 и CD276, каждый из которых регулирует по 5 генов (см. табл. 1).

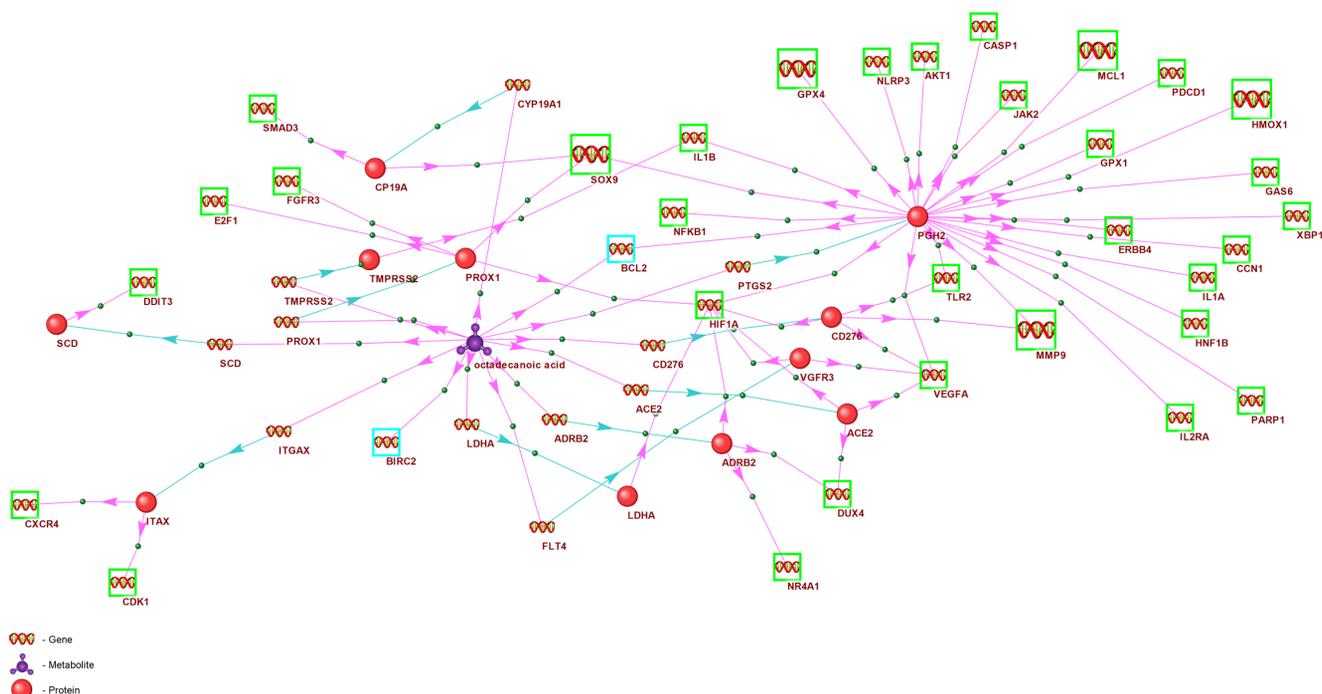
Несколько примеров путей регуляторного влияния октадекановой кислоты на экспрессию генов апоптоза из генной сети ГС-ВПДО приведено на рис. 2.

**Ген *BIRC2*.** На рис. 2 показано, что экспрессия гена *BIRC2* находится под регуляторным влиянием стеариновой кислоты. Подавление экспрессии этого гена, обладающего антиапоптотической активностью, стеариновой кислотой в преадипоцитах является одним из механизмов редукции жировой ткани при действии этого метаболита (Shen M.C. et al., 2014).

**Ген *SCD*.** Согласно данным ANDSystem, стеариновая кислота регулирует экспрессию гена *SCD*, экспрессирующего белок SCD, который, в свою очередь, регулирует ген *DDIT3*, непосредственно вовлеченный в апоптоз (см. рис. 2). По данным литературы (Aardema et al., 2017), стеариновая кислота активирует экспрессию гена *SCD*, продукт которого – белок SCD – подавляет (Minville-Walz et al., 2010) ген *DDIT3*, один из важнейших индукторов апоптоза (Bento et al., 2009).

**Ген *BCL2*.** Экспрессия гена *PTSG2*, кодирующего белок циклооксигеназу-2 (PHG2), находится под регуляторным влиянием стеариновой (октадекановой) кислоты (см. рис. 2). По данным литературы (Liu J. et al., 2014), стеариновая кислота индуцирует экспрессию гена *PTSG2* циклооксигеназы-2. Циклооксигеназа-2, согласно ANDSystem, в свою очередь, усиливает экспрессию гена *BCL2*, что подтверждается результатами экспериментальных исследований (Lin et al., 2019). Отметим, что, по информации из базы знаний ANDSystem и литературы (Shen M.C. et al., 2014), экспрессия гена *BCL2* может тоже подавляться стеариновой кислотой. Таким образом, по данным реконструкции генных сетей экспрессия гена *BCL2*, кодирующего ключевой ингибитор апоптоза (Newton et al., 2024), может как активироваться стеариновой кислотой при участии

<sup>1</sup> Табл. S1–S4 и рис. S1 Приложения доступны по адресу: <https://vavilovj-icg.ru/download/pict-2026-30/appx1.zip>



**Рис. 1.** Генная сеть ГС-ВПДО регуляции экспрессии генов апоптоза стеариновой (октадекановой) кислотой.

Зелеными рамками показаны гены, вовлеченные в процесс апоптоза по данным онтологии генов (GO:0006915 apoptotic process), бирюзовыми рамками отмечены гены апоптоза, находящиеся под регуляторным влиянием октадекановой кислоты без участия посредников. Гены, не имеющие рамок, – это гены-посредники, экспрессия которых регулируется октадекановой кислотой, их белковые продукты (красные шары на рисунке) регулируют экспрессию генов апоптоза. Крупными рамками маркированы гены, дифференциально экспрессирующиеся в культуре клеток НерG2 в ответ на октадекановую кислоту (Vendel Nielsen et al., 2013). Стеариновая (октадекановая) кислота обозначена как "octadecanoic acid". Синие стрелки – процессы экспрессии генов, ведущие к наработке кодируемых ими белков; красные – регуляторное влияние белков на экспрессию генов апоптоза.

**Таблица 1.** Список пяти белков-регуляторов с наибольшим числом регуляторных связей с генами апоптоза в генной сети ГС-ВПДО

№ п/п	Символ белка	Название белка	Число связанных генов апоптоза
1	PTSG2	Циклооксигеназа-2	25
2	PROX1	Проперго гомеобокс 1	5
3	CD276	Белок CD276	5
4	ADRB2	Адренорецептор бета 2	4
5	ACE2	Ангиотензин-превращающий фермент 2	4

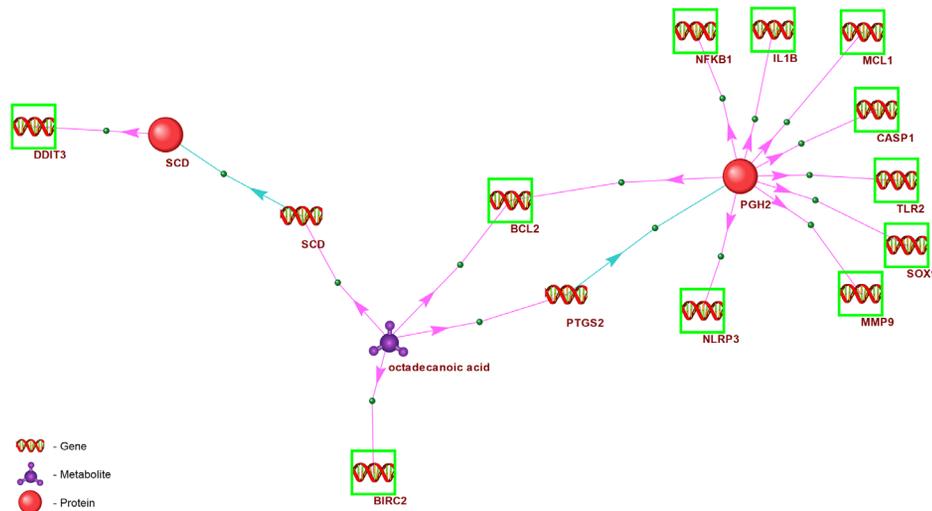
циклооксигеназы-2, так и подавляться этим метаболитом, вероятно, без участия этого фермента. Следовательно, модуляция экспрессии или активности циклооксигеназы-2 может изменять баланс проапоптотических и антиапоптотических эффектов стеариновой кислоты. Имеющиеся экспериментальные данные свидетельствуют о том, что циклооксигеназа-2 ингибирует апоптоз; при этом повышенная экспрессия гена циклооксигеназы-2 является одним из механизмов уклонения опухолевых клеток от апоптоза (Liu C.H. et al., 2001).

**Гены *NFKB1*, *NLRP3*, *CASP1*, *MCL1*.** По данным ANDSystem (см. рис. 2), экспрессия генов *NFKB1*, *MCL1*, кодирующих антиапоптотические белки *NFKB1*, *MCL1* (Newton et al., 2024), а также генов *NLRP3*, *CASP1*, кодирующих проапоптотические белки (Yu et al., 2023), регули-

руется стеариновой (октадекановой) кислотой при участии упомянутого выше белка циклооксигеназа-2. Показано, что этот белок активирует экспрессию антиапоптотических генов *NFKB1*, *MCL1* и *BCL2* (Lin et al., 2019), что может обуславливать подавление апоптоза циклооксигеназой-2.

Реконструированная нами генная сеть ГС-CD36 регуляции апоптоза содержала 187 молекулярно-генетических объектов, включая 98 генов, вовлеченных в процесс апоптоза, собственно белок-рецептор CD36 и 44 гена, кодирующих белки, оказывающие регуляторное влияние на экспрессию генов апоптоза. В связи с тем, что полная генная сеть ГС-CD36 содержит большое число регуляторных связей, она приведена в Приложении (рис. S1).

При анализе полной генной сети ГС-CD36 (см. рис. S1) был выявлен ряд белков – ключевых регуляторов апоптоза:



**Рис. 2.** Фрагменты генной сети регуляторных связей стеариновой (октадекановой) кислоты с генами *BIRC2*, *SCD*, *BCL2*, *NFKB1*, *NLRP3*, *CASP1*, *MCL1*, вовлеченными в процесс апоптоза.

Красными шарами показаны белки, кодируемые генами-посредниками, зелеными рамками – гены апоптоза. Гены, не имеющие рамок, – гены-посредники, экспрессия которых регулируется октадекановой кислотой и белковые продукты которых (красные шары на рисунке) регулируют экспрессию генов апоптоза. Стеариновая (октадекановая) кислота обозначена как "octadecanoic acid". Синие стрелки – процессы экспрессии генов, ведущие к наработке кодируемых ими белков; красные – регуляторное влияние белков на экспрессию генов апоптоза.

PPARG, регулирующий 32 гена апоптоза; циклооксигеназа-2, регулирующая 25 генов; SIR1, регулирующий 26 генов (табл. 2).

На рис. 3 показано несколько примеров путей регуляторного влияния рецептора CD36 на экспрессию генов апоптоза, которые описаны ниже.

**Ген *NFKB1*.** Выявлено два пути регуляции гена *NFKB1*, вовлеченного в процесс апоптоза. В первом случае CD36 регулирует экспрессию гена-посредника *SIRT1*, экспрессирующего белок SIR1, который оказывает регуляторное воздействие на ген *NFKB1*. Согласно (Yin et al., 2020), белок SIR1 обладает активирующим влиянием на регуляторный путь NF-κB. Второй путь включает регуляторное воздействие CD36 на ген-посредник *TRAF6* и его белковый продукт – транскрипционный фактор TRAF6, который, в свою очередь, регулирует экспрессию гена *NFKB1*. Белок TRAF6, экспрессия гена которого стимулируется при участии CD36 (Cao et al., 2019), активирует также экспрессию гена *NFKB1* (Saba et al., 2014).

**Ген *MYD88*.** CD36 регулирует экспрессию гена *MYD88*, вовлеченного в апоптоз (короткий путь регуляции). Кроме

того, экспрессия гена *MYD88* регулируется рецептором CD36 через более длинный путь, проходящий через ген-посредник *TRAF6* и кодируемый им белок, оказывающий регуляторное воздействие на экспрессию гена *MYD88*.

**Ген *MMP9*.** Как отмечено выше, белок-рецептор CD36 регулирует экспрессию гена *TRAF6*, кодирующего белок TRAF6. В свою очередь, TRAF6 регулирует экспрессию гена металлопротеиназы *MMP9* (Luo et al., 2016) и, помимо него, еще пяти генов апоптоза: *IFIT2*, *HIF1A*, *IRF3*, *MYD88*, *NFKB1*. Таким образом, транскрипционный фактор TRAF6 играет роль касетного активатора шести генов апоптоза.

Сравнение генных сетей ГС-ВПДО и ГС-CD36 показало, что 30 генов, вовлеченных в процесс апоптоза, были общими для этих сетей; среди них экспрессия 25 генов регулировалась циклооксигеназой-2 (табл. S2). Следовательно, этот белок может рассматриваться как важнейший регулятор генов апоптоза в каждой из двух реконструированных нами генных сетей. В список генов апоптоза, общих для генных сетей ГС-ВПДО и ГС-CD36 и регулируемых циклооксигеназой-2, вошли, в частности, такие гены, как *BCL2*, *MCL1* и *NFKB1*, белковые продукты

**Таблица 2.** Список пяти белков-регуляторов с наибольшим числом регуляторных связей с генами апоптоза в генной сети ГС-CD36

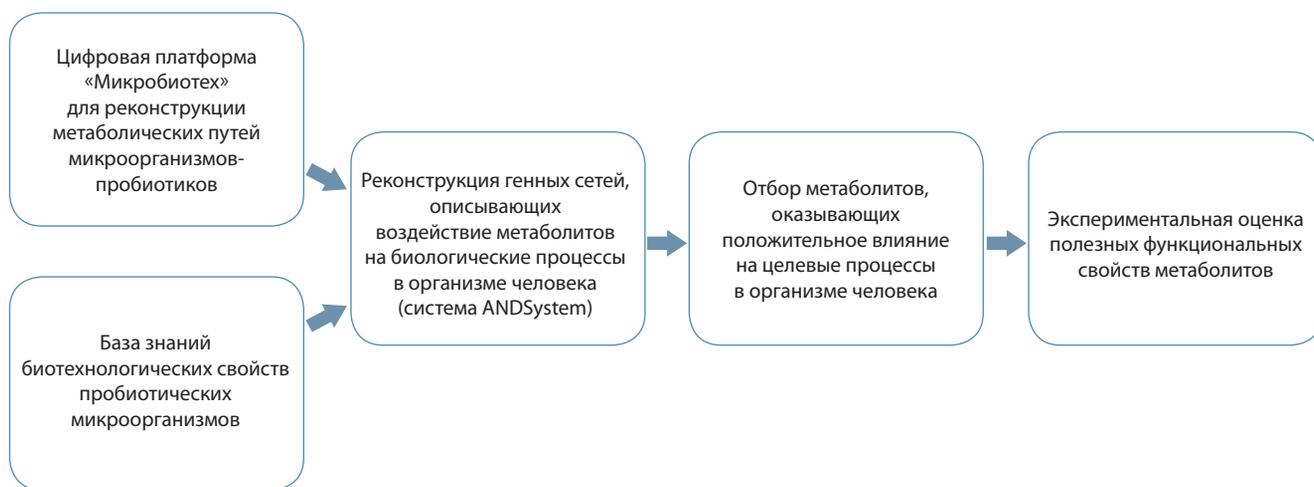
№ п/п	Символ белка	Название белка	Количество регулируемых генов апоптоза
1	PPARG	Гамма-рецептор, активируемый пролифератором пероксисом	32
2	PTGS2	Циклооксигеназа-2	25
3	SIR1	Сиртуин 1	26
4	PPARA	Альфа-рецептор, активируемый пролифератором пероксисом	19
5	GLI2	Белок GLI2 с доменом «цинковые пальцы»	19



**Таблица 3.** Результаты анализа перепредставленности биологических процессов GO для генов, дифференциально экспрессирующихся в клетках культуры HepG2, при воздействии октадекановой кислоты

№ п/п	Термин геной онтологии	Число генов, вовлеченных в биологический процесс	FDR
1	Протон-зависимый митохондриальный синтез АТФ (proton motive force-driven mitochondrial ATP synthesis)	28	9.14E-07
2	Аэробное дыхание (aerobic respiration)	28	0.00000334
3	Сборка комплекса I митохондриальной дыхательной цепи (mitochondrial respiratory chain complex I assembly)	28	0.00000334
4	Митохондриальный перенос электронов, превращение NADH в убихинон (mitochondrial electron transport, NADH to ubiquinone)	20	0.000194
5	Переход G2/M в митотическом клеточном цикле (G2/M transition of mitotic cell cycle)	19	0.00373
7	Удаление активных форм кислорода (cellular oxidant detoxification)	23	0.00798
6	Высвобождение цитохрома С из митохондрий (release of cytochrome c from mitochondria)	12	0.0162
8	Бета-окисление жирных кислот (fatty acid beta-oxidation)	17	0.0171
9	Аутофагия (autophagy)	36	0.0392
10	Апоптоз (apoptotic process)	108	0.0396

Примечание. FDR (false discovery rate) – мера статистической значимости перепредставленности биологического процесса с учетом множественных сравнений.



**Рис. 4.** Общая схема платформы для поиска продуцируемых микроорганизмами-пробиотиками функциональных метаболитов, оказывающих положительное влияние на целевые процессы в организме человека.

поиска функциональных метаболитов, продуцируемых микроорганизмами-пробиотиками, оказывающих положительное влияние на целевые процессы в организме человека. Принципиальные блоки этой платформы представлены на рис. 4.

Исходная информация для платформы поступает из двух источников. Первый – цифровая платформа «Микробиотех», разработанная в рамках Курчатовского геномного центра ИЦиГ СО РАН (Деменков и др., 2024). Она обеспечивает реконструкцию метаболических путей микроорганизмов на основе компьютерного анализа их геномов. Второй источник – База знаний биотехнологических

свойств микроорганизмов, разрабатываемая в рамках Курчатовского геномного центра ИЦиГ СО РАН с применением методов искусственного интеллекта. Эти методы позволяют автоматически извлекать знания и факты из научных публикаций и патентов.

Сформированные данные используются для реконструкции генных сетей, описывающих воздействие метаболитов на биологические процессы в организме человека, с применением системы ANDSystem (Ivanisenko V. et al., 2015, 2019). Анализ реконструированных генных сетей позволяет определить критерии отбора метаболитов, оказывающих положительное влияние на целевые процессы в

организме человека. На заключительном этапе проводится экспериментальная оценка полезных функциональных свойств отобранных метаболитов.

## Список литературы / References

- Деменков П.С., Мухин А.М., Иванисенко В.А., Лашин С.А., Колчанов Н.А. Цифровая платформа «Микробиотек»: архитектура и назначение. *Проблемы информатики*. 2024;4:27-36. doi 10.24412/2073-0667-2024-4-27-36
- [Demenkov P.S., Mukhin A.M., Ivanisenko V.A., Lashin S.A., Kolchanov N.A. The "Microbiotech" digital platform: architecture and purpose. *Problems of Informatics*. 2024;4:27-36. doi 10.24412/2073-0667-2024-4-27-36 (in Russian)]
- Колчанов Н.А., Игнатъева И.В., Подколотная О.А., Лихошвай В.А., Матушкин В.Г. Генные сети. *Вавиловский журнал генетики и селекции*. 2013;17(4/2):833-850
- [Kolchanov N.A., Ignatyeva I.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin V.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2013;17(4/2):833-850 (in Russian)]
- Aardema H., van Tol H.T.A., Wubbolts R.W., Brouwers J.F.H.M., Gadella B.M., Roelen B.A.J. Stearoyl-CoA desaturase activity in bovine cumulus cells protects the oocyte against saturated fatty acid stress. *Biol Reprod*. 2017;96(5):982-992. doi 10.1095/biolreprod.116.146159
- Anderson C.M., Stahl A. SLC27 fatty acid transport proteins. *Mol Aspects Med*. 2013;34(2-3):516-528. doi 10.1016/j.mam.2012.07.010
- Bento C., Andersson M.K., Aman P. DDIT3/CHOP and the sarcoma fusion oncoprotein FUS-DDIT3/TLS-CHOP bind cyclin-dependent kinase 2. *BMC Cell Biol*. 2009;10:89. doi 10.1186/1471-2121-10-89
- Bragina E.Y., Gomboeva D.E., Saik O.V., Ivanisenko V.A., Freidin M.B., Nazarenko M.S., Puzyrev V.P. Apoptosis genes as a key to identification of inverse comorbidity of Huntington's disease and cancer. *Int J Mol Sci*. 2023;24(11):9385. doi 10.3390/ijms24119385
- Cao D., Luo J., Chen D., Xu H., Shi H., Jing X., Zang W. CD36 regulates lipopolysaccharide-induced signaling pathways and mediates the internalization of *Escherichia coli* in cooperation with TLR4 in goat mammary gland epithelial cells. *Sci Rep*. 2016;6:23132. doi 10.1038/srep23132. Erratum in: *Sci Rep*. 2019;9(1):6457. doi 10.1038/s41598-019-42156-3
- Chen Y., Zhang J., Cui W., Silverstein R.L. CD36, a signaling receptor and fatty acid transporter that regulates immune cell metabolism and fate. *J Exp Med*. 2022;219(6):e20211314. doi 10.1084/jem.20211314
- Cui Y., Xu T., Qu X., Hu T., Jiang X., Zhao C. New insights into various production characteristics of *Streptococcus thermophilus* strains. *Int J Mol Sci*. 2016;17(10):1701. doi 10.3390/ijms17101701
- Evans L.M., Cowey S.L., Siegal G.P., Hardy R.W. Stearate preferentially induces apoptosis in human breast cancer cells. *Nutr Cancer*. 2009;61(5):746-753. doi 10.1080/01635580902825597
- Glatz J.F.C., Nabben M., Luiken J.J.F.P. CD36 (SR-B2) as master regulator of cellular fatty acid homeostasis. *Curr Opin Lipidol*. 2022;33(2):103-111. doi 10.1097/MOL.0000000000000819
- Grujović M.Ž., Semedo-Lemsaddek T., Marković K.G. Application of probiotics in foods: a comprehensive review of benefits, challenges, and future perspectives. *Foods*. 2025;14(17):3088. doi 10.3390/foods14173088
- Gupta R., Kadhim M.M., Turki J.A., Obayes A.M., Aminov Z., Alsaikhan F., Ramirez-Coronel A.A., Ramaiah P., Tayyib N.A., Luo X. Multifaceted role of NF-κB in hepatocellular carcinoma therapy: molecular landscape, therapeutic compounds and nanomaterial approaches. *Environ Res*. 2023;228:115767. doi 10.1016/j.envres.2023.115767
- Houten S.M., Violante S., Ventura F.V., Wanders R.J.A. The biochemistry and physiology of mitochondrial fatty acid β-oxidation and its genetic disorders. *Annu Rev Physiol*. 2016;78:23-44. doi 10.1146/annurev-physiol-021115-105045
- Hung T.-Y., Chen H.-H., Wu C.-Y., Hsu H.-T., Hsueh Y.-P., Kuan Y.-H. Stearic acid activates microglia and impairs memory via NF-κB signaling. *Neurobiol Dis*. 2023;182:106121. doi 10.1016/j.nbd.2023.106121
- 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(Suppl.11):228. doi 10.1186/s12859-020-03557-8
- Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. New ANDDigest with improved AI-based short-name recognition. *Int J Mol Sci*. 2022;23(23):14934. doi 10.3390/ijms232314934
- Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Likhoshvai V.A., Kolchanov N.A. ANDSystem: an Associative Network Discovery System. *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., Saik O.V., Kolchanov N.A. A new version of ANDSystem with tissue-specific functionality. *BMC Bioinformatics*. 2019;20(Suppl.1):34. doi 10.1186/s12859-018-2567-6
- Ivanisenko V.A., Rogachev A.D., Makarova A.A., Basov N.V., Gaisler E.V., Kuzmicheva I.N., Demenkov P.S., ... Kolchanov N.A., Plesko V.V., Moroz G.B., Lomivorotov V.V., Pokrovsky A.G. AI-assisted identification of primary and secondary metabolomic markers for postoperative delirium. *Int J Mol Sci*. 2024;25(21):11847. doi 10.3390/ijms252111847
- Kamp F., Hamilton J.A. pH gradients across phospholipid membranes caused by fast flip-flop of un-ionized fatty acids. *Proc Natl Acad Sci USA*. 1992;89(23):11367-11370. doi 10.1073/pnas.89.23.11367
- Lau L.Y.J., Quek S.Y. Probiotics: health benefits, food application, and colonization in the human gastrointestinal tract. *Food Bioeng*. 2024;3(1):41-64. doi 10.1002/fbe2.12078
- Lin X.M., Luo W., Wang H., Li R.Z., Huang Y.S., Chen L.K., Wu X.P. The role of prostaglandin-endoperoxide synthase-2 in chemoresistance of non-small cell lung cancer. *Front Pharmacol*. 2019;10:836. doi 10.3389/fphar.2019.00836
- Liu C.H., Chang S.H., Narko K., Trifan O.C., Wu M.T., Smith E., Haudenschild C., Lane T.F., Hla T. Overexpression of cyclooxygenase-2 is sufficient to induce tumorigenesis in transgenic mice. *J Biol Chem*. 2001;276(21):18563-18569. doi 10.1074/jbc.M010787200
- Liu J., Hu S., Cui Y., Sun M.K., Xie F., Zhang Q., Jin J. Saturated fatty acids up-regulate COX-2 expression in prostate epithelial cells via toll-like receptor 4/NF-κB signaling. *Inflammation*. 2014;37(2):467-477. doi 10.1007/s10753-013-9760-6
- Luo Z., Zhang X., Zeng W., Su J., Yang K., Lu L., Lim C.B., Tang W., Wu L., Zhao S., Jia X., Peng C., Chen X. TRAF6 regulates melanoma invasion and metastasis through ubiquitination of Basigin. *Oncotarget*. 2016;7(6):7179-7192. doi 10.18632/oncotarget.6886
- Mashek D.G., Li L.O., Coleman R.A. Long-chain acyl-CoA synthetases and fatty acid channeling. *Future Lipidol*. 2007;2(4):465-476. doi 10.2217/17460875.2.4.465
- Minville-Walz M., Pierre A.S., Pichon L., Bellenger S., Fèvre C., Bellenger J., Tessier C., Narce M., Rialland M. Inhibition of stearoyl-CoA desaturase 1 expression induces CHOP-dependent cell death in human cancer cells. *PLoS One*. 2010;5(12):e14363. doi 10.1371/journal.pone.0014363
- Newton K., Strasser A., Kayagaki N., Dixit V.M. Cell death. *Cell*. 2024;187(2):235-256. doi 10.1016/j.cell.2023.11.044
- Paton C.M., Ntambi J.M. Biochemical and physiological function of stearoyl-CoA desaturase. *Am J Physiol Endocrinol Metab*. 2009;297(1):E28-E37. doi 10.1152/ajpendo.90897.2008
- Richieri G.V., Kleinfeld A.M. Unbound free fatty acid levels in human serum. *J Lipid Res*. 1995;36(2):229-240. doi 10.1016/S0022-2275(20)39899-0
- Richieri G.V., Anel A., Kleinfeld A.M. Interactions of long-chain fatty acids and albumin: determination of free fatty acid levels using the fluorescent probe ADIFAB. *Biochemistry*. 1993;32(29):7574-7580. doi 10.1021/bi00080a032

- Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47. doi 10.1093/nar/gkv007
- Saba R., Sorensen D.L., Booth S.A. MicroRNA-146a: a dominant, negative regulator of the innate immune response. *Front Immunol.* 2014;5:578. doi 10.3389/fimmu.2014.00578
- Saik O.V., Nimaev V.V., Usmonov D.B., Demenkov P.S., Ivanisenko T.V., Lavrik I.N., Ivanisenko V.A. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSystem. *BMC Med Genomics.* 2019;12(Suppl.2):47. doi 10.1186/s12920-019-0492-9
- Shen M.C., Zhao X., Siegal G.P., Desmond R., Hardy R.W. Dietary stearic acid leads to a reduction of visceral adipose tissue in athymic nude mice. *PLoS One.* 2014;9(9):e104083. doi 10.1371/journal.pone.0104083
- Shen X., Miao S., Zhang Y., Guo X., Li W., Mao X., Zhang Q. Stearic acid metabolism in human health and disease. *Clin Nutr.* 2025; 44:222-238. doi 10.1016/j.clnu.2024.12.012
- Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C., Imamichi T., Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 2022;50(W1):W216-W221. doi 10.1093/nar/gkac194
- Sudheer A., Dastidar D.G., Ghosh G., Taj Z., Nidhin I.K., Chattopadhyay I. Comprehensive genomics, probiotic, and antibiofilm potential analysis of *Streptococcus thermophilus* strains isolated from homemade and commercial dahi. *Sci Rep.* 2025;15(1):7089. doi 10.1038/s41598-025-90999-w
- Terpou A., Papadaki A., Lappa I.K., Kachrimanidou V., Bosnea L.A., Kopsahelis N. Probiotics in food systems: significance and emerging strategies towards improved viability and delivery of enhanced beneficial value. *Nutrients.* 2019;11(7):1591. doi 10.3390/nu11071591
- Vendel Nielsen L., Krogager T.P., Young C., Ferreri C., Chatgialioglou C., Jensen O.N., Enghild J.J. Effects of elaidic acid on lipid metabolism in HepG2; dataset GSE34045. *PLoS One.* 2013;8(9):e74283. doi 10.1371/journal.pone.0074283
- Yang Y., Huang J., Li J., Yang H., Yin Y. Effects of stearic acid on proliferation, differentiation, apoptosis, and autophagy in porcine intestinal epithelial cells. *Curr Mol Med.* 2020;20(2):157-166. doi 10.2174/1566524019666190917144127
- Yin Q., Shen L., Qi Y., Song D., Ye L., Peng Y., Wang Y., Jin Z., Ning G., Wang W., Lin D., Wang S. Decreased SIRT1 expression in the peripheral blood of patients with Graves' disease. *J Endocrinol.* 2020;246(2):161-173. doi 10.1530/JOE-19-0501
- Yu C., Chen P., Miao L., Di G. The role of the NLRP3 inflammasome and programmed cell death in acute liver injury. *Int J Mol Sci.* 2023;24(4):3067. doi 10.3390/ijms24043067

---

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 15.11.2025. После доработки 24.11.2025. Принята к публикации 08.12.2025.

doi 10.18699/vjgb-26-10

# Развитие метода филогенетического футпринтинга для распознавания сайтов связывания транскрипционных факторов на основе использования bootstrap-испытаний для анализа больших бактериальных геномных данных

А.М. Мухин <sup>1,2,3</sup> , Т.М. Хлебодарова<sup>1,2</sup>, Д.Ю. Ощепков <sup>1,2</sup><sup>1</sup> Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия<sup>2</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия mukhin@bionet.nsc.ru

**Аннотация.** Активное развитие технологий высокопроизводительного секвенирования привело к взрывообразному накоплению высококачественных данных по последовательностям бактериальных геномов – их число приближается к трем миллионам, и дальнейший рост продолжается. Это, в свою очередь, дает дополнительный импульс развитию технологий для более эффективной их аннотации аналитическими методами с прицелом на применение таких больших геномных данных, а также получение нового качества аннотаций. Одним из таких аналитических подходов стал метод филогенетического футпринтинга, направленный на выявление мотивов, соответствующих сайтам связывания транскрипционных факторов в промоторных областях бактериальных геномов путем сравнения соответствующих выборок регуляторных последовательностей генов-ортологов для родственных организмов. Дальнейшее накопление геномных данных стало стимулом для развития подхода. Так, было обнаружено, что избыточное число последовательностей в выборке, анализируемой с использованием филогенетического футпринтинга, лишь ухудшает точность метода, тогда как включение этапа отбора последовательностей в анализируемую выборку с учетом данных о взаимных эволюционных расстояниях повышает качество работы метода. В настоящей статье нами предложен и реализован следующий шаг развития метода филогенетического футпринтинга, основанный на множественном запуске описанного выше этапа отбора для формирования различающихся подвыборок, последующего запуска конвейера для каждой из подвыборок и на статистическом анализе получаемых результатов множественных запусков конвейера. Предложенный подход, реализованный в методе MotifsOnFly, позволяет повысить устойчивость получаемых результатов распознавания мотивов, выявляемых в многократных запусках конвейера. Эффективность метода MotifsOnFly продемонстрирована на примере анализа хорошо аннотированного промотора гена *OmpW Escherichia coli*.

**Ключевые слова:** филогенетический футпринтинг; бактериальный геном; сайты связывания транскрипционных факторов; мотивы; бутстреп; Python

**Для цитирования:** Мухин А.М., Хлебодарова Т.М., Ощепков Д.Ю. Развитие метода филогенетического футпринтинга для распознавания сайтов связывания транскрипционных факторов на основе использования bootstrap-испытаний для анализа больших бактериальных геномных данных. *Вавиловский журнал генетики и селекции*. 2026;30(1):15-26. doi 10.18699/vjgb-26-10

**Финансирование.** Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (Федеральная научно-техническая программа развития генетических технологий на 2019–2030 годы, соглашение № 075-15-2025-516).

**Благодарности.** Авторы выражают благодарность Центру коллективного пользования (ЦКП) «Биоинформатика» за предоставление вычислительных ресурсов.

## Improvement of a phylogenetic footprinting method for transcription factor binding sites recognition based on the use of bootstrap trials for the analysis of large bacterial genomic data

А.М. Mukhin <sup>1,2,3</sup> , Т.М. Khlebodarova<sup>1,2</sup>, D.Yu. Oshchepkov <sup>1,2</sup><sup>1</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia<sup>2</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia mukhin@bionet.nsc.ru

**Abstract.** The rapid development of high-throughput sequencing technologies has led to an explosive accumulation of high-quality bacterial genome sequence data – their number is approaching three million, and this growth continues. This, in turn, provides additional impetus for the development of technologies for more efficient annotation using analytical methods designed to utilize such large-scale genomic data, as well as for achieving new levels of annotation quality. One such analytical approach is phylogenetic footprinting, which aims to identify motifs corresponding to transcription factor binding sites in the promoter regions of bacterial genomes by comparing corresponding sets of regulatory sequences of orthologous genes in related organisms. The continued accumulation of genomic data has served as the basis for further development of this approach. It has been found that an excessive number of sequences in a set analyzed using phylogenetic footprinting only reduces the accuracy of the method, whereas the inclusion of a sequence selection step in the analyzed set based on data on mutual evolutionary distances improves the method's performance. In this paper, we propose and implement a further step in the development of the phylogenetic footprinting method. This step involves multiple runs of the selection step described above to generate distinct subsamples, subsequent pipeline runs for each subsample, and statistical analysis of the results obtained from multiple pipeline runs. The proposed approach, implemented in the MotifsOnFly method, improves the robustness of motif recognition results obtained from multiple pipeline runs. The effectiveness of the MotifsOnFly method is demonstrated using the analysis of the well-annotated promoter of the *Escherichia coli OmpW* gene.

**Key words:** phylogenetic footprinting; bacterial genome; transcription factor binding sites; motifs; bootstrap; Python

**For citation:** Mukhin A.M., Khlebodarova T.M., Oshchepkov D.Yu. Improvement of a phylogenetic footprinting method for transcription factor binding sites recognition based on the use of bootstrap trials for the analysis of large bacterial genomic data. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed.* 2026;30(1):15-26. doi 10.18699/vjgb-26-10

## Введение

Стремительное развитие и массовое применение технологий высокопроизводительного секвенирования в молекулярной генетике стимулировало развитие не только биотехнологий, обеспечив возможность массовой сборки бактериальных геномов для их анализа, модификации и дальнейшего использования штаммов бактерий при решении биотехнологических задач, но и методов биоинформатики для все более качественной их аннотации. Аннотация бактериальных геномов сайтами связывания транскрипционных факторов (ССТФ) является одним из важнейших этапов в задачах биотехнологии и микробиологии, поскольку через связывание транскрипционных факторов со своими сайтами в промоторах генов обеспечивается один из основных механизмов регуляции их экспрессии в бактериях (Browning, Busby, 2004).

Стратегия филогенетического футпринтинга, направленная на выявление ССТФ и впервые предложенная в 1988 г. (Tagle et al., 1988; Katara et al., 2012), оказалась плодотворной для поиска мотивов *de novo* с учетом выросшего на тот момент числа секвенированных геномов. Эта стратегия основана на общем принципе, что регуляторные элементы в промоторах, такие как ССТФ, в подавляющем большинстве случаев эволюционируют с меньшей скоростью и более консервативны на уровне последовательности ДНК, чем окружающие их нефункциональные последовательности (Levy et al., 2001). Дальнейший рост числа секвенированных бактериальных геномов делал применение филогенетического футпринтинга все более эффективным (Blanchette, Tompa, 2002), дав стимул к развитию множества алгоритмов, наиболее результативных для имеющихся к тому времени секвенированных геномов, таких как MotifSuite, FootPrinter, AlignACE, BioProspector, CONSENSUS, MDscan, MEME, CUBIC и BoBro (Hertz, Stormo, 1999; Liu X. et al., 2001, 2002; Blanchette, Tompa, 2003; Olman et al., 2003; Chen et al., 2008; Bailey et al., 2009; Li et al., 2011a; Claeys et al., 2012).

Впоследствии было обнаружено, что ограниченное число правильным образом отобранных референтных про-

моторов может быть достаточным для идентификации ССТФ в гене (McCue et al., 2002), поскольку сходные последовательности, имеющие слишком малые эволюционные расстояния, оказываются малоинформативными для филогенетического футпринтинга в силу малого числа накопленных мутаций в окружающих ССТФ нефункциональных последовательностях. Поэтому дальнейшее накопление геномных данных для бактерий позволило улучшить метод филогенетического футпринтинга путем применения этапа отбора последовательностей в анализируемую выборку с учетом данных о взаимных эволюционных расстояниях между ними. Это обеспечивает получение более информативных наборов ортологичных промоторов для распознавания функциональных мотивов. Описанный подход, реализованный в работе (Liu B. et al., 2016), превосходил по качеству распознавания ССТФ популярные методы поиска мотивов, разработанные ранее и перечисленные выше.

Используя в основе подход, предложенный В. Liu с коллегами (2016), мы ранее разработали вычислительный конвейер для поиска ССТФ в бактериальных геномах (Мухин и др., 2024). Он интегрирует исчерпывающий набор необходимых баз данных и алгоритмов для работы с ними, обеспечивая быстрое применение этих ресурсов для аннотации выбранных бактериальных геномов сайтами связывания транскрипционных факторов. Однако в ходе его использования для поиска ССТФ в геноме *Geobacillus icigianus* (Peltek et al., 2024) мы выявили зависимость результата распознавания от порядка, состава и способа отбора подаваемой на вход программы выборки промоторных районов, не позволяющую сделать однозначный вывод о наиболее вероятном расположении сайта связывания и наиболее вероятном транскрипционном факторе, взаимодействующем с этим сайтом.

В настоящей работе на базе ранее разработанного конвейера (Мухин и др., 2024) нами предложена и реализована модификация метода, названного нами MotifsOnFly. В основе этой модификации лежат: а) множественный запуск этапа конвейера для формирования подвыборок

промоторов из полной выборки с учетом данных о взаимных эволюционных расстояниях для формирования различающихся подвыборок; б) запуск конвейера для каждой подвыборки для поиска перепредставленных мотивов *de novo* в каждой из них; в) статистический анализ получаемых результатов. Такой бутстреп-подход (с возвращением) обеспечивает гораздо более полное использование исходных данных и повышает устойчивость и достоверность получаемых результатов за счет статистического анализа результатов распознавания мотивов, выявляемых в многократных запусках. На примере анализа хорошо аннотированного промотора гена *ompW Escherichia coli* показаны преимущества разработанного метода MotifsOnFly, ставшие результатом предложенного развития подхода филогенетического футпринтинга.

## Материалы и методы

**Вычислительный конвейер.** Для распознавания потенциальных сайтов связывания транскрипционных факторов на основе выявления *de novo* мотивов в бактериальных геномах с помощью реализованного в работе вычислительно-компьютерного конвейера использовались аннотации оперонов для 3850 бактериальных геномов, доступные в базе данных (БД) DOOR2 (Mao et al., 2014), последовательности которых были извлечены из БД NCBI (Sayers et al., 2021).

В конвейере использован интегративный подход для реализации метода филогенетического футпринтинга, предложенный ранее В. Liu с коллегами (2016). В ходе выполнения этапов работы конвейера для исследуемого (целевого) гена производится выявление ортологичных генов среди всех генов, имеющихся в базе (3850) аннотированных геномов. Поиск осуществляется путем выявления сходства белковых последовательностей, для этого применяется программный модуль GOST (Li et al., 2011b). Далее для каждого из выявленных генов-ортологов, исходя из известной оперонной структуры соответствующего генома, извлекается его промоторная последовательность, в результате чего формируется полная выборка промоторных последовательностей генов-ортологов исследуемого целевого гена. Оценка расстояний между каждой парой ортологичных промоторных последовательностей, входящих в анализируемую полную выборку, проводится на основании построенного филогенетического дерева с помощью программы ClustalW2 (Larkin et al., 2007). Формирование различающихся подвыборок последовательностей промоторов путем их отбора в зависимости от их взаимных эволюционных расстояний проводили согласно принципам, эффективность которых была показана ранее (Liu B. et al., 2016). Далее, по аналогии с подходом (Liu B. et al., 2016), для выявления мотивов *de novo* в каждой из подвыборок промоторов применяли принцип голосования между результатами, получаемыми набором базовых поисковых систем, в качестве которых были использованы ранее разработанные алгоритмы AlignACE, BioProspector, CONSENSUS, MDscan, MEME, CUBIC и BoBro (Hertz, Stormo, 1999; Liu X. et al., 2001, 2002; Olman et al., 2003; Chen et al., 2008; Bailey et al., 2009; Li et al., 2011a).

Найденные *de novo* мотивы сравнивались с помощью алгоритма Tomtom (Gupta et al., 2007) с известными ССТФ, содержащимися в двух базах данных – SwissRegulon (Pachkov et al., 2013) для *E. coli* и PRODORIC для бактерий (Dudek, Jahn, 2022). На основе рассчитываемой в Tomtom статистической меры сходства определялась выборка ССТФ, наиболее похожая на выявленный *de novo* мотив, таким образом определялся наиболее вероятный ТФ, взаимодействующий с найденным мотивом.

Выбор двух баз данных обусловлен возможностью дополнительного сравнения выявленных *de novo* мотивов в промоторах *E. coli* как с выборками известных ССТФ именно этого организма, представленных в БД SwissRegulon, так и с обобщенными выборками известных ССТФ бактерий, представленных в БД PRODORIC. Результаты множественного запуска конвейера подвергались статистическому анализу и визуализации.

Аналитические скрипты были реализованы на Python 3.12, 3.6 и 2.7 с использованием платформы Anaconda (<https://anaconda.com>) для управления программными окружениями и различными версиями программ в них. Для хранения необходимых данных и структурированной метаинформации о генах, последовательностях и оперонах применялась система управления базами данных PostgreSQL (<https://www.postgresql.org/>), которая была развернута на инфраструктуре Института цитологии и генетики СО РАН силами ЦКП «Биоинформатика».

**Визуализация и анализ результатов.** Данные по известным ССТФ в геноме *E. coli* были извлечены из БД RegulonDB, содержащей наиболее полный массив знаний о регуляции инициации транскрипции *Escherichia coli* K-12 (Salgado et al., 2024). При расчетах использована аннотация генома NC\_000913.3, извлеченная из БД NCBI.

Для визуализации как данных по аннотации генома из NCBI (последовательность генома) и RegulonDB (экспериментально выявленные сайты связывания транскрипционных факторов), так и проекции результатов работы конвейера на целевой промотор применялся модульный геномный браузер JBrowse2 (Diesh et al., 2023), способный визуализировать множество распространенных форматов файлов с геномными данными. Визуализация дерева, отражающего таксономическую принадлежность анализируемых геномов, представленную в NCBI (Sayers et al., 2021), осуществлялась с помощью программного модуля ETE3 (Huerta-Cepas et al., 2016). Для визуализации результатов множественных запусков конвейера и их статистического анализа использован программный пакет Matplotlib (Hunter, 2007).

## Результаты

### Модификация подхода филогенетического футпринтинга, реализованного в методе MotifsOnFly

В основе реализованной в этой работе модификации подхода лежат: множественный запуск этапа конвейера, обеспечивающего формирование различающихся подвыборок последовательностей промоторов с сохранением принципов отбора, эффективность которых была показана



**Рис. 1.** Схема конвейерного поиска *de novo* функциональных мотивов, соответствующих сайтам связывания транскрипционных факторов, основанная на подходе филогенетического футпринтинга и реализованная в методе MotifsOnFly.

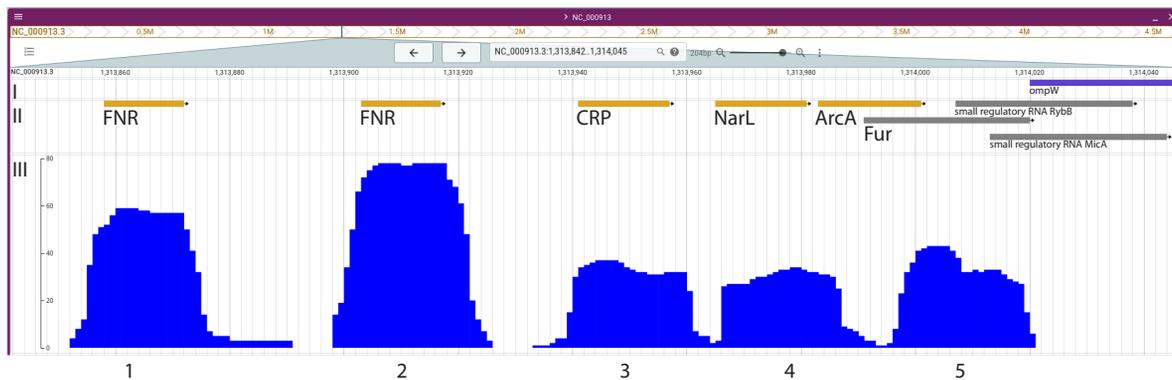
Фиолетовой рамкой (1) обозначены данные от пользователя, синими цветами – программные инструменты (2, 7) и внешние данные (3), использованные в конвейере. Зеленым цветом (4, 6) показаны промежуточные данные, генерируемые в процессе работы конвейера. Красной рамкой отмечены этапы, разработанные в рамках метода MotifsOnFly (5, 8 и 9). Этап, включающий отбор с формированием подвыборок промоторов исходя из данных о взаимных эволюционных расстояниях между ними (5), позволяет проводить множественный запуск конвейера для поиска мотивов *de novo* (7) на этих подвыборках. Получаемые в результате данные о выявляемых *de novo* мотивах (8) и статистике их распределения в промоторах генов-ортологов для каждой позиции выравнивания анализируемого промотора позволяют получить гистограмму частоты найденных *de novo* мотивов, отображаемую в геномном браузере JBrowse2 (9). Сравнение полученных в каждом запуске *de novo* мотивов с помощью инструмента Tomtom с базами данных известных CCTФ SwissRegulon и PRODORIC (8) и дальнейший анализ статистики для этих сравнений позволяют предположить наиболее вероятные транскрипционные факторы-регуляторы исследуемого промотора.

ранее (Liu B. et al., 2016); накопление и статистический анализ результатов множественного запуска алгоритма на этих подвыборках. Принципиальная схема анализа, реализованного в разработанном конвейере метода MotifsOnFly, представлена на рис. 1. В соответствии с подробно описанными в разделе «Материалы и методы» этапами работы конвейера, исходя из белковой последовательности исследуемого гена, подаваемой на вход конвейера (1), происходит выявление ортологичных генов программным модулем GOST (2) для имеющихся в базе (3) аннотированных геномов. Далее, используя эти же геномные последовательности и разметку оперонной структуры (3), извлекаются все промоторные последовательности генов-ортологов (4).

Базовым для разработанной модификации является этап создания множественных подвыборок (5) с сохранением условий отбора промоторов в анализируемую подвыборку исходя из данных об их взаимных эволюционных расстояниях. Эффективность использования таких условий отбора для реализации метода филогенетического футпринтинга была показана ранее (Liu B. et al., 2016). Сформированная полная выборка промоторов (4) по рассчитанным с помощью ClustalW2 эволюционным расстояниям от целево-

го промотора разделялась на три подгруппы, расстояния которых до целевого промотора составляли от 0.05 до 0.31, от 0.31 до 0.55 и от 0.55 до 0.73 соответственно. Далее на основании размеров этих подгрупп определялось максимальное число подвыборок  $M$ , способных вместить все промоторы полной выборки в подвыборки по 12 штук с пропорцией промоторов из соответствующих подгрупп 3:3:6. На последнем этапе формировались  $N = 2M$  подвыборок путем случайного отбора, исходя из пропорции 3:3:6 и дополнительных принципов отбора, которые учитывают расстояние между входящими в итоговую выборку последовательностями, не допуская попадания в нее слишком схожих промоторов. Такой бутстреп-подход (с возвращением) обеспечивал максимальное, но не полное вхождение промоторов исходной выборки в анализ, поскольку сохранение принципов отбора, ограничивающих вхождение промоторов с малым информационным вкладом для каждой подвыборки, является приоритетным для эффективности метода (McCue et al., 2002).

Сформированные описанным выше способом подвыборки (5), дополненные целевым промотором, используются для многократного запуска этапа конвейера для поиска мотивов *de novo* (7) (см. раздел «Материалы и методы»).



**Рис. 2.** Скриншот геномного браузера JBrowse2.

I – аннотация генома в области промотора гена *ompW* согласно данным NCBI; II – экспериментально подтвержденная локализация ССТФ из БД RegulonDB; III – результаты работы конвейера: гистограмма частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора. Гистограмма формирует пики, которые пронумерованы в соответствии с их упоминанием в тексте.

В результате таких множественных запусков алгоритма (7) на подвыборках собирается статистика (8), суммирующая, в каких позициях целевого промотора какие *de novo* мотивы были выявлены, а также в каких позициях каких промоторов генов-ортологов они одновременно были обнаружены. Собранный таким образом гистограмма частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора позволяет оценить надежность получаемого *de novo* мотива на позицию (Tompa et al., 2005). Данные гистограммы сохраняются в файл формата BigWig для дальнейшего отображения в геномный браузер JBrowse2 (9).

### Анализ промотора гена *ompW* *Escherichia coli* K-12 на основе разработанного метода MotifsOnFly

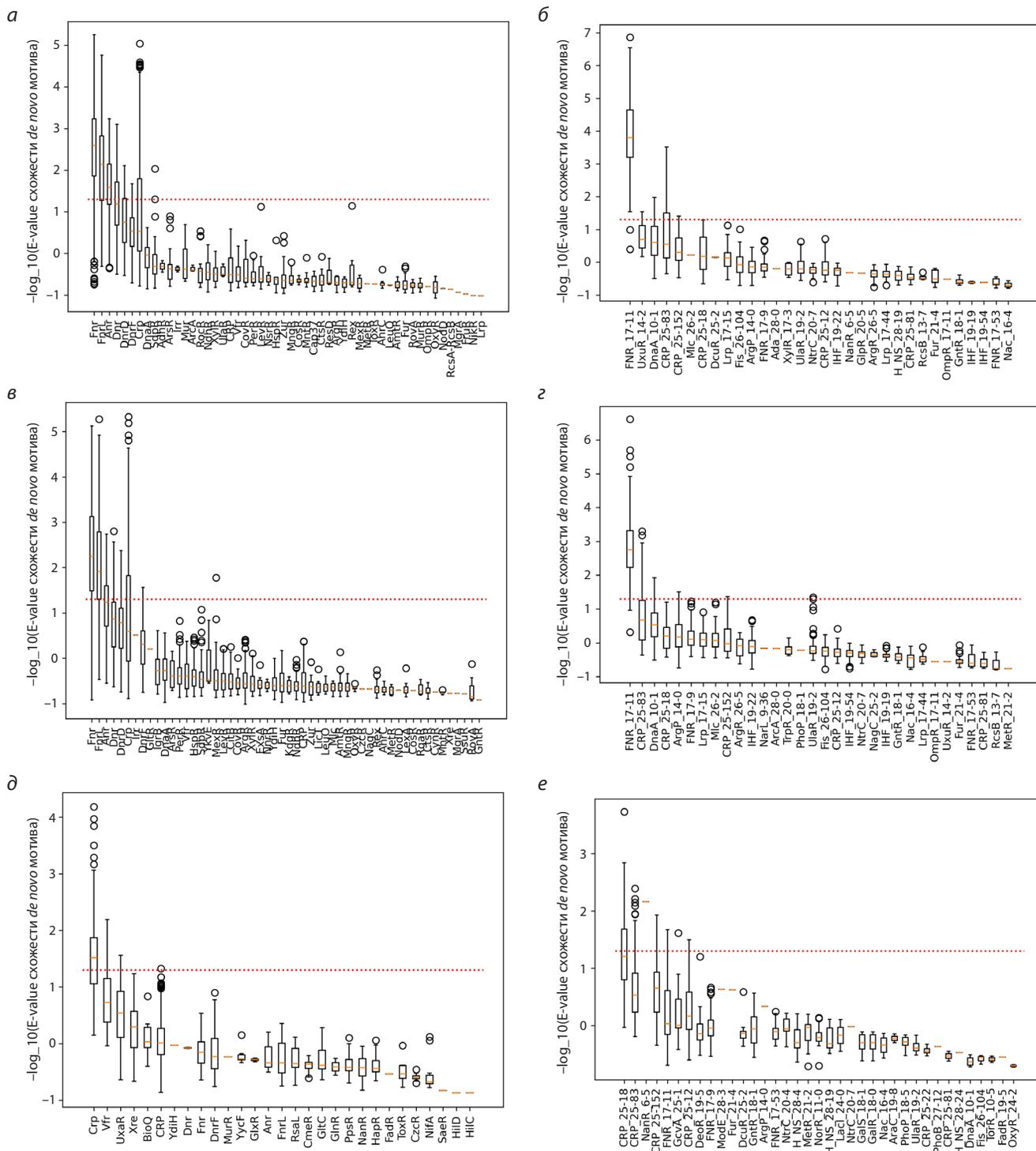
Наиболее хорошо изученным и полно аннотированным бактериальным геномом на текущий момент остается *E. coli* K-12 (Salgado et al., 2024). Для демонстрации возможностей разработанного нами метода MotifsOnFly, расширяющего потенциал подхода, реализованного В. Liu с соавторами (2016), был выбран ген *ompW* *E. coli*, в контроле экспрессии которого участвуют шесть транскрипционных факторов с пятью известными сайтами связывания. Ген *ompW* кодирует белок наружной мембраны и обладает широким спектром физиологических функций, включая устойчивость бактерий к различным антибиотикам, гербицидам, осмотическому стрессу и поддержку роста бактерий в суровых условиях окружающей среды, таких как гипоксия и повышенная температура (Zhang et al., 2020). Характерной особенностью промотора этого оперона, содержащего только ген *ompW*, является наличие как отдельно расположенных сайтов связывания ТФ, так и групп пересекающихся сайтов разных ТФ. В качестве входных данных конвейера использовались белковая и промоторная последовательности гена, как описано в разделе «Материалы и методы».

На рис. 2 представлены результаты работы конвейера, визуализированные с помощью JBrowse2, включая: аннотацию генома в области гена *ompW* согласно данным

NCBI (I), экспериментально подтвержденную локализацию ССТФ из БД RegulonDB (II) и гистограмму частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора (III). Гистограмма формирует отчетливые пики (см. рис. 2), указывающие на то, что в соответствующих позициях промотора гена *ompW* в значительной доле запусков были найдены мотивы *de novo*. При этом пики 1 и 2 с хорошей точностью совпадают с экспериментально выявленными сайтами связывания транскрипционного фактора FNR, а пик 3 – с известной локализацией сайта связывания транскрипционного фактора CRP. Область промотора, в которой локализован пик 4, пересекается с экспериментально выявленным сайтом связывания ТФ NarL. Что же касается пика 5, то он пересекается с локализацией как сайта связывания ТФ ArcA, так и сайта связывания ТФ Fur, что говорит о сложности исследования регуляторных сигналов вблизи старта транскрипции и требует отдельного рассмотрения. Также легко заметить «хвост» справа от пика 1, свидетельствующий о том, что для этого района в некоторых запусках конвейера были идентифицированы мотивы *de novo*. Последнее подтверждает неустойчивость результатов, получаемых в однократных запусках.

Сравнение выявленных *de novo* мотивов с помощью инструмента Tomtom с базами данных известных ССТФ SwissRegulon и PRODORIC позволяет предположить наиболее вероятные транскрипционные факторы-регуляторы исследуемого промотора. Для каждого такого сравнения рассчитывается значимость E-value сходства выявленного мотива с каждым из известных ССТФ. Нами были рассчитаны статистики значений E-value для всех возможных кандидатов всех пиков (рис. 3). Как видно из диаграмм, полученная статистика позволяет однозначно определить ТФ FNR в качестве наиболее вероятного кандидата на связывание для пиков 1 (см. рис. 3, а, б) и 2 (см. рис. 3, в, г), а ТФ CRP – для пика 3 (см. рис. 3, д, е) при использовании любой из баз данных.

Для пика 4 значимые совпадения обнаружены только с сайтами из БД SwissRegulon (рис. 4) для ТФ NanR и NarL.



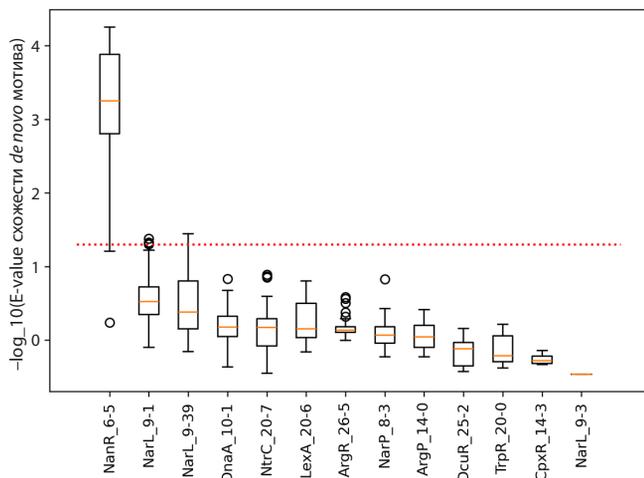
**Рис. 3.** Диаграммы типа «ящик с усами» для статистики сравнения выявленных мотивов для пика 1 (а, б), пика 2 (в, г), пика 3 (д, е) с базами данных известных ССТФ PRODORIC (а, в, д) и SwissRegulon (б, г, е).

Красным пунктиром обозначен порог значимости 5 %. Для визуализации результатов статистического анализа множественных запусков был применен программный пакет Matplotlib.

Для пика 5 такой расчет не дает значимых совпадений с известными ССТФ из обеих баз данных, что предполагает необходимость дополнительного анализа.

Следует отметить, что для пика 1 в случае однократного запуска конвейера лишь в 72 % случаев при использовании

БД SwissRegulon и в 43 % случаев при использовании БД PRODORIC приоритетным результатом был бы именно FNR. Аналогичные оценки для пика 2 дают 90 и 65 % случаев правильного распознавания FNR, а для пика 3 – 49 и 54 % случаев правильного распознавания CRP со-



**Рис. 4.** Диаграмма типа «ящик с усами» для статистики сравнения выявленных мотивов для пика 4 с базой данных известных ССТФ SwissRegulon.

Красным пунктиром обозначен порог значимости 5 %. Для визуализации результатов статистического анализа множественных запусков был применен программный пакет Matplotlib.

ответственно. Предложенный нами модифицированный подход позволяет преодолеть эту неустойчивость работы алгоритма филогенетического футпринтинга при однократном запуске конвейера и однозначно расставить приоритеты для полученных результатов в полном согласии с экспериментальными данными.

Метод также дает возможность проследить наличие *de novo* мотивов, соответствующих каждому пику, в каждом из промоторов генов-ортологов и отобразить эти данные на таксономическом дереве геномов, рассмотренных в процессе анализа. В случае, если в промоторе гена-ортолога определенного вида бактерий при выравнивании с целевым промотором в позиции пика также найден сходный мотив, это отображается красной точкой на узле дерева, соответствующем данному виду. На рис. 5 в качестве примера показано наличие сходных мотивов в исследованных промоторах генов-ортологов для позиций, соответствующих пику 1 выравнивания с целевым промотором.

Информация, дающая возможность отследить наличие сходных участков по всем промоторам генов-ортологов (см. рис. 5) в пределах одного пика, также позволяет объединить выравненные участки только тех промоторов генов-ортологов, для которых в любом из запусков конвейера одновременно с целевым промотором был выявлен сходный мотив. Благодаря этому можно получить объединенное выравнивание всех сходных консервативных участков ДНК в пределах одного пика, что обеспечивает максимально полное их представительство в такой выравненной выборке. Границы информативных участков выравнивания, очевидно, определяются шириной каждого из пиков, что и стало критерием для выбора длины объединенного выравнивания и позволило использовать также информацию, содержащуюся на флангах сайтов,

для более качественного сравнения с выборками известных ССТФ. Собранные таким образом объединенные выравнивания для каждого пика были дополнительно проанализированы в программе Tomtom. В результате сравнения полученных объединенных выравниваний для пиков 1–3 с известными ССТФ максимальный уровень достоверности сходства был достигнут при использовании БД SwissRegulon (рис. 6). При этом наиболее вероятными кандидатами для связывания, в полном соответствии с экспериментальными данными, снова оказались ТФ FNR для пиков 1 (см. рис. 6, а) и 2 (см. рис. 6, б) и ТФ CRP для пика 3 (см. рис. 6, в) с уровнем значимости (E-value)  $3.13 \times 10^{-8}$ ,  $9.17 \times 10^{-8}$  и  $1.80 \times 10^{-4}$  соответственно. Это согласуется с результатами, приведенными выше для анализа статистики сравнения выявленных мотивов (см. рис. 3), и показывает при этом значительно более высокую достоверность сходства мотивов с соответствующими выборками ССТФ для полученных объединенных выравниваний.

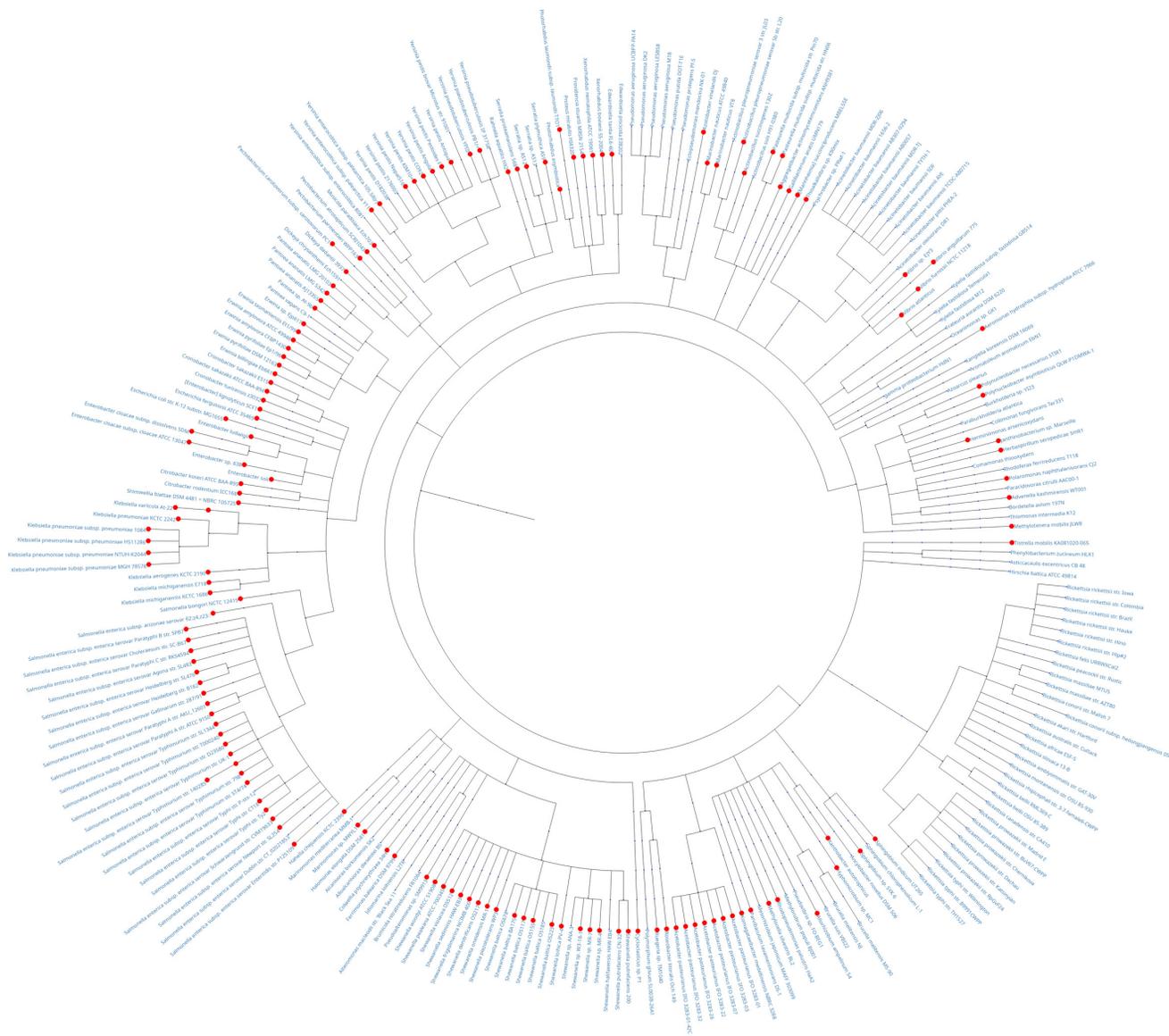
Анализ программой Tomtom объединенного выравнивания для пика 4 дает результаты лишь при использовании БД SwissRegulon – выявлено его значимое сходство с ССТФ NanR (рис. 7, а) и незначимое сходство с выборкой ССТФ NarL (см. рис. 7, б). Других ТФ, имеющих сходные с пиком 4 сайты связывания, выявлено не было.

Объединенное выравнивание для участка пика 5 при анализе программой Tomtom дает результаты лишь при сравнении с БД PRODORIC. При этом ни один ССТФ не показывает значимого сходства с этой выборкой. Тем не менее было выявлено слабое сходство объединенного выравнивания с ССТФ ArcA и Fur (рис. 8), сайты связывания которых имеют значительное пересечение.

## Обсуждение

Нами разработан вычислительно-компьютерный конвейер для поиска функциональных мотивов, соответствующих ССТФ, основанный на модификации метода филогенетического футпринтинга, разработанного ранее В. Liu с коллегами (2016). Ключевая идея модификации состоит в множественном запуске эффективной процедуры отбора промоторов, учитывающей их взаимные эволюционные расстояния, для формирования различающихся подвыборок промоторов, последующих запусков конвейера для каждой из таких подвыборок и в статистическом анализе результатов, получаемых при множественных запусках конвейера (см. рис. 1). В предложенном подходе, реализованном в методе MotifsOnFly, применена схема бутстреп-подхода с возвращением, что позволяет использовать максимальное число имеющихся на текущий момент геномных данных, одновременно учитывая статистику результатов на различающихся подвыборках промоторов.

Эффективность работы реализованного конвейера продемонстрирована на примере анализа наиболее полно аннотированного промотора гена *ompW*, кодирующего белок наружной мембраны *Escherichia coli* K-12. Для него известно шесть транскрипционных факторов-регуляторов, у которых экспериментально определены пять сайтов связывания (Salgado et al., 2024). Подробный анализ регуляторных взаимодействий этого промотора был пред-

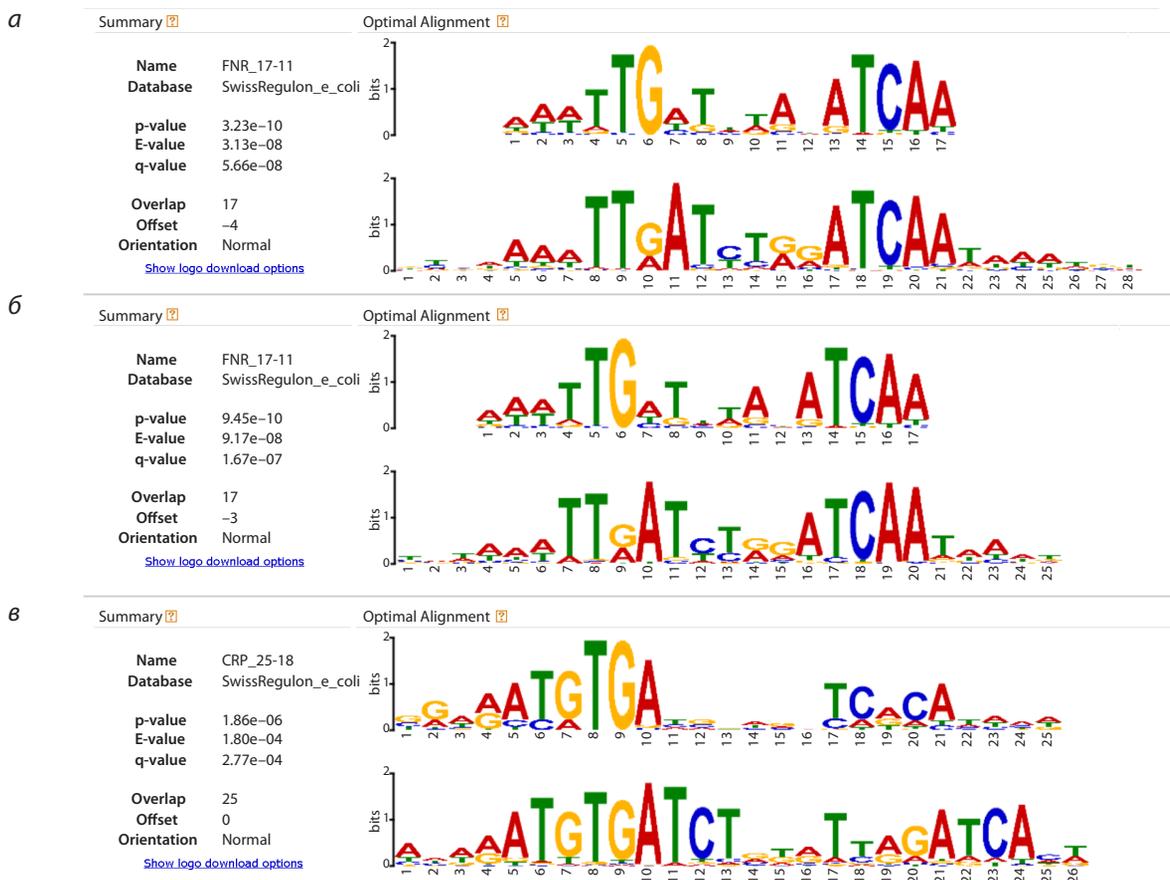


**Рис. 5.** Отображение наличия сходных мотивов (отмечены красными точками) в исследованных промоторах генов-ортологов разных видов бактерий для позиций, соответствующих пику 1 выравнивания с целевым промотором.

Дерево построено в соответствии с данными по таксономической принадлежности анализируемых геномов, представленными в NCBI. Для визуализации дерева использован пакет ETE3 (Huerta-Cepas et al., 2016).

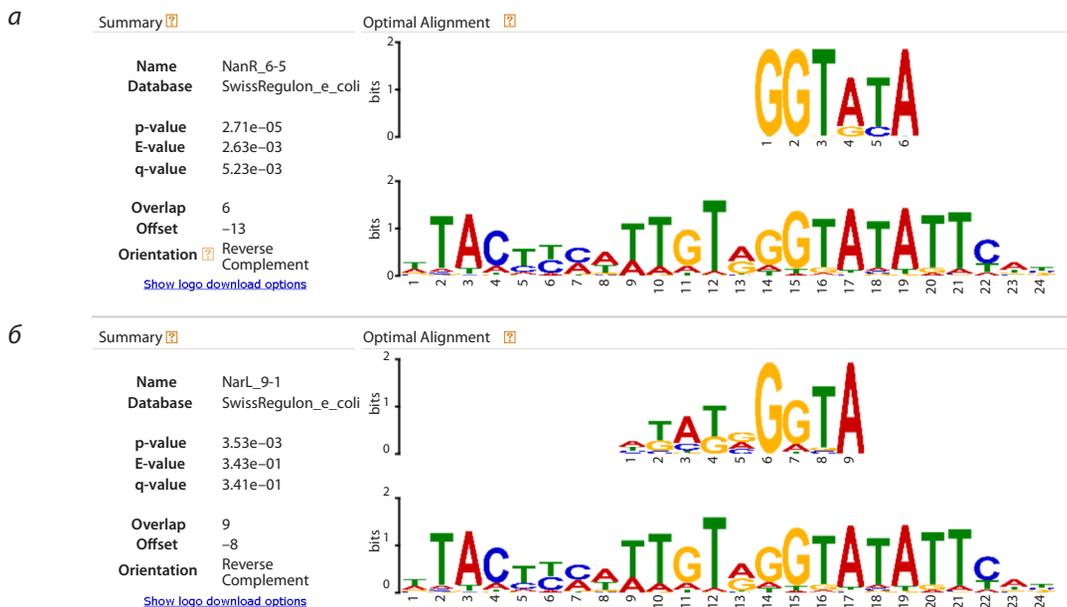
ставлен ранее (Xiao et al., 2016). Статистика, посчитанная в результате работы конвейера, позволила получить гистограмму частоты найденных *de novo* мотивов в промоторах генов-ортологов для каждой позиции анализируемого промотора, что, в свою очередь, обеспечило надежную идентификацию консервативных участков промотора, обогащенных регуляторными мотивами и формирующих характерные пики (см. рис. 2), которые, как показал приведенный анализ, точно совпадают с экспериментально доказанными. Таким образом, предложенный нами подход наглядно иллюстрирует суть филогенетического футпринтинга, выявляя наиболее консервативные участки выравниваний промоторов генов-ортологов.

Проведенный статистический анализ результатов сравнения *de novo* мотивов, выявленных во множественных запусках, с известными выборками ССТФ из БД SwissRegulon и БД PRODORIC позволил точно определить наиболее вероятные регуляторные факторы, связывающиеся с соответствующими участками анализируемого промотора. Предсказания метода MotifsOnFly полностью совпали с экспериментальными данными для пиков 1, 2, 3, определив ТФ FNR (Constantinidou et al., 2006; Myers et al., 2013; Xiao et al., 2016) как наиболее вероятный белок-регулятор для пиков 1 и 2 и ТФ CRP (Gaston et al., 1990; Ushida, Aiba, 1990; Xiao et al., 2016) – для пика 3 (см. рис. 3). Одновременно этот анализ показал, что не



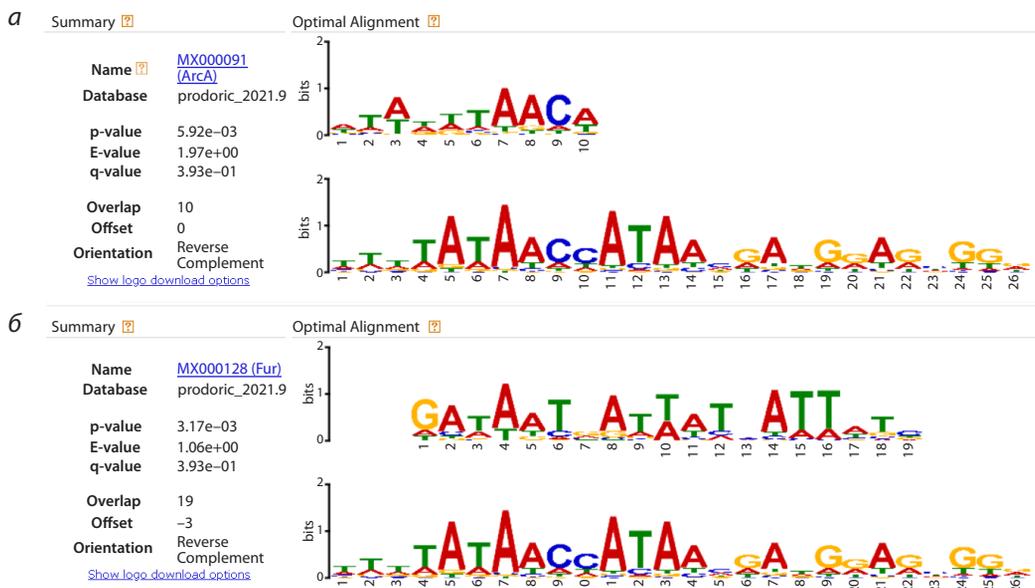
**Рис. 6.** Результаты сравнения выявленных объединенных выравниваний участков пиков 1–3.

Для каждого пика приведен скриншот лучшего совпадения с данными БД SwissRegulon из выдачи программы Tomtom. Показано значимое сходство этих выравниваний с ССТФ FNR для пиков 1 (а, E-value=3.13e-08) и 2 (б, E-value = 9.17e-08) и с ССТФ CRP для пика 3 (в, E-value = 1.80e-04), что полностью совпадает с экспериментальными данными.



**Рис. 7.** Результат сравнения выявленного объединенного выравнивания для участка пика 4.

Приведены скриншоты сравнения с данными БД SwissRegulon из выдачи программы Tomtom. Показано значимое сходство с выборкой ССТФ NanR (а) и незначимое сходство с выборкой ССТФ NarL (б).



**Рис. 8.** Результат сравнения выявленного объединенного выравнивания для участка пика 5.

Приведены скриншоты сравнения с данными БД PRODORIC из выдачи программы Tomtom. Показано незначительное сходство частей объединенного выравнивания для пика 5 с ССТФ ArcA (а) и Fur (б). Видно, что расположение выявленных сайтов выравнивания значительно пересекается.

всегда и далеко не в каждом однократном запуске имеется возможность правильной идентификации наиболее вероятного транскрипционного фактора, совпадающего с экспериментальными данными (см. рис. 3). Описанное выше продемонстрировало неустойчивость результатов при однократном запуске конвейера, тогда как многократный запуск, напротив, позволяет надежно выявлять наиболее вероятный ТФ не только с помощью анализа статистики многократных запусков (см. рис. 3), но и с помощью анализа объединенного выравнивания для соответствующего пика с последующим его сравнением с выбранными базами данных известных ССТФ (см. рис. 6). Полученная оценка значимости сходства объединенных выравниваний с эталонными выборками из БД SwissRegulon и PRODORIC, достигающая величины E-value ~3.1e-08, открывает возможность генерации качественных выборок для распознавания некоторых ССТФ, недостаточно представленных в базах данных. В рамках конвейера реализована процедура, позволяющая проследить наличие *de novo* мотивов, соответствующих каждому пику, в каждом из проанализированных промоторов генов-ортологов и отобразить эти данные на таксономическом дереве геномов, рассмотренных в процессе анализа (см. рис. 5). Эта функция может быть полезна при анализе эволюции и представленности отдельных регуляторных сигналов.

Интересен результат, полученный с помощью метода MotifsOnFly для пика 4 (см. рис. 4 и 7). Экспериментальные данные указывают на наличие в этом районе сайта связывания ТФ NarL (Tyson et al., 1993, 1994; Xiao et al., 2016). Однако анализ статистики множественных запусков конвейера (см. рис. 4) и объединенного выравнивания сходных консервативных участков ДНК (см. рис. 7) вы-

являет NanR как наиболее вероятный ТФ для связывания с данным сайтом. При этом ТФ NarL лишь второй по значимости кандидат на связывание, несмотря на сходство с консенсусной последовательностью для его сайта связывания “ТАСУУМТ” (где Y = C или T, а M = A или C) (Tyson et al., 1993) с объединенным выравниванием для пика 4 (см. рис. 7). В то же время, рассматривая этот результат, следует учесть, что последовательность GGTATA, выявленная нами как высокодостоверный сайт связывания ТФ NanR (см. рис. 7), может быть просто недостаточной для связывания активной формы этого фактора, который является димером. Более того, ранее экспериментально показано, что высокое сродство связывания ТФ NanR с ДНК обеспечивается кооперативным связыванием трех димеров ТФ NanR с тремя рядом расположенными повторами GGTATA (Kalivoda et al., 2013; Horne et al., 2021). Наличие только одной последовательности GGTATA в позиции пика 4 ставит под сомнение возможность участия NanR в регуляции экспрессии гена *ompW*. Кроме того, в настоящее время роль NanR в регуляции экспрессии генов *E. coli* считается минорной и ограничивается его участием в регуляции только четырех оперонов, кодирующих ферменты катаболизма сиаловой кислоты (Kalivoda et al., 2013; Shimada et al., 2018). Тем не менее возможность регуляции гена *ompW* или некоторого числа генов-ортологов через найденный нами ССТФ NanR заслуживает внимания и требует дополнительной экспериментальной проверки.

Что касается результатов анализа участка промотора, соответствующего пику 5, то в первую очередь следует отметить, что этот пик полностью совпадает с районом посадки РНК-полимеразы  $\sigma 70$ , включающим старт транскрипции и специфичные -35/-10 элементы. Этот участок промотора перед самым стартом транскрипции гена *ompW*

насыщен регуляторными сигналами. Так, согласно RegulonDB (Salgado et al., 2024), здесь сосредоточены сайты связывания ArcA (Park et al., 2013; Xiao et al., 2016) и Fur (Zhang et al., 2020). Известная пластичность архитектуры сайта связывания TF ArcA (Park et al., 2013) обеспечивает эффективный способ кодирования сайтов ArcA вместе с сайтами других факторов транскрипции в одном и том же узком пространстве последовательностей. Тот факт, что сайт связывания TF Fur в районе пика 5 оценивается как слабый (Zhang et al., 2020), может объясняться особенностями его конкурентных отношений с TF SoxS (Graham et al., 2012), обладающим способностью взаимодействовать с элементом  $-35$  (Taliaferro et al., 2012) в условиях окислительного стресса, провоцируемого активностью Fur в присутствии ионов железа. Таким образом, область промотора гена *OmpW* в районе пика 5 обладает значительным регуляторным потенциалом, однако по данным нашего анализа не содержит ни одного сайта со значительным сходством с известными ССТФ. Это можно интерпретировать как отсутствие сильных сайтов связывания каких-либо транскрипционных факторов, которые препятствовали бы посадке РНК-полимеразы  $\sigma 70$ . Одновременно сравнимое по значениям выявляемой достоверности сходство объединенного выравнивания для пика 5 (E-value) с большим числом известных ССТФ из БД PRODORIC можно интерпретировать как наличие множества слабых сайтов различных TF, включая пересекающиеся сайты ArcA и Fur (см. рис. 8), что находится в соответствии с приведенными выше литературными данными.

## Заключение

Предложенная в нашей работе модификация подхода филогенетического футпринтинга, реализованная в методе MotifsOnFly, является естественным развитием подхода В. Liu с коллегами (2016) в ответ на непрерывно увеличивающийся объем широкомасштабных геномных данных. MotifsOnFly позволяет производить множественный запуск конвейера на различающихся подвыборках промоторов генов-ортологов, получать уточненные данные по локализации ССТФ и проводить статистический анализ. Варианты такого статистического анализа не ограничиваются описанными в этой статье и могут быть расширены в соответствии с целями и задачами конкретных исследований. Метод MotifsOnFly, позволяющий получать, как продемонстрировано здесь, надежные и устойчивые результаты выявления ССТФ, может быть востребован широким кругом исследователей при аннотации и анализе регуляторных последовательностей бактериальных геномов.

## Список литературы / References

Мухин А.М., Ощепков Д.Ю., Лашин С.А. Вычислительный конвейер по распознаванию сайтов связывания транскрипционных факторов в бактериальных геномах *de novo*. *Проблемы информатики*. 2024;4(65):69-83. doi 10.24412/2073-0667-2024-4-69-83 [Mukhin A., Oschepkov D., Lashin S. A computational pipeline for *de novo* recognition of transcription factor binding sites in bacterial genomes. *Problems of Informatics*. 2024;4(65):69-83. doi 10.24412/2073-0667-2024-4-69-83 (in Russian)]

- Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., Noble W.S. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(W):W202-W208. doi 10.1093/nar/gkp335
- Blanchette M., Tompa M. Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Res.* 2002; 12(5):739-748. doi 10.1101/gr.6902
- Blanchette M., Tompa M. FootPrinter: a program designed for phylogenetic footprinting. *Nucleic Acids Res.* 2003;31(13):3840-3842. doi 10.1093/nar/gkg606
- Browning D.F., Busby S.J. The regulation of bacterial transcription initiation. *Nat Rev Microbiol.* 2004;2(1):57-65. doi 10.1038/nrmicro787
- Chen X., Guo L., Fan Z., Jiang T. W-AlignACE: an improved Gibbs sampling algorithm based on more accurate position weight matrices learned from sequence and gene expression/ChIP-chip data. *Bioinformatics.* 2008;24(9):1121-1128. doi 10.1093/bioinformatics/btn088
- Claeys M., Storms V., Sun H., Michoel T., Marchal K. MotifSuite: workflow for probabilistic motif detection and assessment. *Bioinformatics.* 2012;28(14):1931-1932. doi 10.1093/bioinformatics/bts293
- Constantinidou C., Hobman J.L., Griffiths L., Patel M.D., Penn C.W., Cole J.A., Overton T.W. A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *J Biol Chem.* 2006;281(8):4802-4815. doi 10.1074/jbc.M512312200
- Diesh C., Stevens G.J., Xie P., De Jesus Martinez T., Hershsberg E.A., Leung A., Guo E., ... Haw R., Cain S., Buels R.M., Stein L.D., Holmes I.H. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.* 2023;24(1):74. doi 10.1186/s13059-023-02914-z
- Dudek C.-A., Jahn D. PRODORIC: state-of-the-art database of prokaryotic gene regulation. *Nucleic Acids Res.* 2022;50(D1):D295-D302. doi 10.1093/nar/gkab1110
- Gaston K., Bell A., Kolb A., Buc H., Busby S. Stringent spacing requirements for transcription activation by CRP. *Cell.* 1990;62(4):733-743. doi 10.1016/0092-8674(90)90118-x
- Graham A.I., Sanguinetti G., Bramall N., McLeod C.W., Poole R.K. Dynamics of a starvation-to-surfeit shift: a transcriptomic and modelling analysis of the bacterial response to zinc reveals transient behaviour of the Fur and SoxS regulators. *Microbiology (Reading)*. 2012;158(Pt.1):284-292. doi 10.1099/mic.0.053843-0
- Gupta S., Stamatoyannopoulos J.A., Bailey T.L., Noble W.S. Quantifying similarity between motifs. *Genome Biol.* 2007;8(2):R24. doi 10.1186/gb-2007-8-2-r24
- Hertz G.Z., Stormo G.D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics.* 1999;15(7-8):563-577. doi 10.1093/bioinformatics/15.7.563
- Horne C.R., Venugopal H., Panjikar S., Wood D.M., Henrickson A., Brookes E., North R.A., Murphy J.M., Friemann R., Griffin M.D.W., Ramm G., Demeler B., Dobson R.C.J. Mechanism of NanR gene repression and allosteric induction of bacterial sialic acid metabolism. *Nat Commun.* 2021;12(1):1988. doi 10.1038/s41467-021-22253-6
- Huerta-Cepas J., Serra F., Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 2016;33(6): 1635-1638. doi 10.1093/molbev/msw046
- Hunter J.D. Matplotlib: a 2D graphics environment. *Comput Sci Eng.* 2007;9(3):90-95. doi 10.1109/MCSE.2007.55
- Kalivoda K.A., Steenbergen S.M., Vimr E.R. Control of the *Escherichia coli* sialoregulon by transcriptional repressor NanR. *J Bacteriol.* 2013;195(20):4689-4701. doi 10.1128/JB.00692-13
- Katara P., Grover A., Sharma V. Phylogenetic footprinting: a boost for microbial regulatory genomics. *Protoplasma.* 2012;249(4):901-907. doi 10.1007/s00709-011-0351-9
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lo-

- pez R., Thompson J.D., Gibson T.J., Higgins D.G. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947-2948. doi 10.1093/bioinformatics/btm404
- Levy S., Hannenhalli S., Workman C. Enrichment of regulatory signals in conserved non-coding genomic sequence. *Bioinformatics*. 2001; 17(10):871-877. doi 10.1093/bioinformatics/17.10.871
- Li G., Liu B., Ma Q., Xu Y. A new framework for identifying cis-regulatory motifs in prokaryotes. *Nucleic Acids Res*. 2011a;39(7):e42. doi 10.1093/nar/gkq948
- Li G., Ma Q., Mao X., Yin Y., Zhu X., Xu Y. Integration of sequence-similarity and functional association information can overcome intrinsic problems in orthology mapping across bacterial genomes. *Nucleic Acids Res*. 2011b;39(22):e150. doi 10.1093/nar/gkr766
- Liu B., Zhang H., Zhou C., Li G., Fennell A., Wang G., Kang Y., Liu Q., Ma Q. An integrative and applicable phylogenetic footprinting framework for cis-regulatory motifs identification in prokaryotic genomes. *BMC Genomics*. 2016;17:578. doi 10.1186/s12864-016-2982-x
- Liu X., Brutlag D.L., Liu J.S. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput*. 2001;6:127-138
- Liu X.S., Brutlag D.L., Liu J.S. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol*. 2002;20(8):835-839. doi 10.1038/nbt717
- Mao X., Ma Q., Zhou C., Chen X., Zhang H., Yang J., Mao F., Lai W., Xu Y. DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res*. 2014;42(D1): D654-D659. doi 10.1093/nar/gkt1048
- McCue L.A., Thompson W., Carmack C.S., Lawrence C.E. Factors influencing the identification of transcription factor binding sites by cross-species comparison. *Genome Res*. 2002;12(10):1523-1532. doi 10.1101/gr.323602
- Myers K.S., Yan H., Ong I.M., Chung D., Liang K., Tran F., Keleş S., Landick R., Kiley P.J. Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. *PLoS Genet*. 2013;9(6):e1003565. doi 10.1371/journal.pgen.1003565
- Olman V., Xu D., Xu Y. CUBIC: identification of regulatory binding sites through data clustering. *J Bioinform Comput Biol*. 2003;1(1): 21-40. doi 10.1142/s0219720003000162
- Pachkov M., Balwierz P.J., Arnold P., Ozonov E., van Nimwegen E. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Res*. 2013;41(D1):D214-D220. doi 10.1093/nar/gks1145
- Park D.M., Akhtar M.S., Ansari A.Z., Landick R., Kiley P.J. The bacterial response regulator ArcA uses a diverse binding site architecture to regulate carbon oxidation globally. *PLoS Genet*. 2013;9(10): e1003839. doi 10.1371/journal.pgen.1003839
- Peltek S., Bannikova S., Khlebodarova T.M., Uvarova Y., Mukhin A.M., Vasiliev G., Scheglov M., Shipova A., Vasilieva A., Oshchepkov D., Bryanskaya A., Popik V. The transcriptomic response of cells of the thermophilic bacterium *Geobacillus icigianus* to terahertz irradiation. *Int J Mol Sci*. 2024;25(22):12059. doi 10.3390/ijms252212059
- Salgado H., Gama-Castro S., Lara P., Mejia-Almonte C., Alarcón-Carranza G., López-Almazo A.G., Betancourt-Figueroa F., ... Hernández-Alvarez A.J., Santos-Zavaleta A., Capella-Gutiérrez S., Gelpi J.L., Collado-Vides J. RegulonDB v12.0: a comprehensive resource of transcriptional regulation in *E. coli* K-12. *Nucleic Acids Res*. 2024;52(D1):D255-D264. doi 10.1093/nar/gkad1072
- Sayers E.W., Beck J., Bolton E.E., Bourexis D., Brister J.R., Canese K., Comeau D.C., ... Wang J., Ye J., Trawick B.W., Pruitt K.D., Sherry S.T. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2021;49(D1):D10-D17. doi 10.1093/nar/gkaa892
- Shimada T., Ogasawara H., Ishihama A. Single-target regulators form a minor group of transcription factors in *Escherichia coli* K-12. *Nucleic Acids Res*. 2018;46(8):3921-3936. doi 10.1093/nar/gky138
- Tagle D.A., Koop B.F., Goodman M., Slightom J.L., Hess D.L., Jones R.T. Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol*. 1988;203(2):439-455. doi 10.1016/0022-2836(88)90011-3
- Taliaferro L.P., Keen E.F., Sanchez-Alberola N., Wolf R.E. Transcription activation by *Escherichia coli* Rob at class II promoters: protein-protein interactions between Rob's N-terminal domain and the  $\sigma^{70}$  subunit of RNA polymerase. *J Mol Biol*. 2012;419(3-4):139-157. doi 10.1016/j.jmb.2012.03.019
- Tompa M., Li N., Bailey T.L., Church G.M., De Moor B., Eskin E., Favorov A.V., ... Vandenbogaert M., Weng Z., Workman C., Ye C., Zhu Z. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*. 2005;23(1):137-144. doi 10.1038/nbt1053
- Tyson K.L., Bell A.I., Cole J.A., Busby S.J. Definition of nitrite and nitrate response elements at the anaerobically inducible *Escherichia coli* *nirB* promoter: interactions between FNR and NarL. *Mol Microbiol*. 1993;7(1):151-157. doi 10.1111/j.1365-2958.1993.tb01106.x
- Tyson K.L., Cole J.A., Busby S.J. Nitrite and nitrate regulation at the promoters of two *Escherichia coli* operons encoding nitrite reductase: identification of common target heptamers for both NarP- and NarL-dependent regulation. *Mol Microbiol*. 1994;13(6):1045-1055. doi 10.1111/j.1365-2958.1994.tb00495.x
- Ushida C., Aiba H. Helical phase dependent action of CRP: effect of the distance between the CRP site and the -35 region on promoter activity. *Nucleic Acids Res*. 1990;18(21):6325-6330. doi 10.1093/nar/18.21.6325
- Xiao M., Lai Y., Sun J., Chen G., Yan A. Transcriptional regulation of the outer membrane porin gene *ompW* reveals its physiological role during the transition from the aerobic to the anaerobic lifestyle of *Escherichia coli*. *Front Microbiol*. 2016;7:799. doi 10.3389/fmicb.2016.00799
- Zhang P., Ye Z., Ye C., Zou H., Gao Z., Pan J. OmpW is positively regulated by iron via Fur, and negatively regulated by SoxS contribution to oxidative stress resistance in *Escherichia coli*. *Microb Pathog*. 2020;138:103808. doi 10.1016/j.micpath.2019.103808

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 15.11.2025. После доработки 09.12.2025. Принята к публикации 10.12.2025.

doi 10.18699/vjgb-26-09

## Детекция колосков в колосе пшеницы на RGB-изображениях с использованием глубокого машинного обучения

М.А. Генаев <sup>1, 2</sup> , И.Д. Бусов<sup>1, 3</sup>, Ю.В. Кручинина <sup>1, 2</sup>, В.С. Коваль<sup>1, 2</sup>, Н.П. Гончаров<sup>1, 3</sup><sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия<sup>2</sup> Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия mag@bionet.nsc.ru

**Аннотация.** В работе рассматривается задача автоматизированного высокопроизводительного фенотипирования признаков колоса пшеницы с использованием современных методов компьютерного зрения и глубокого обучения. Точная оценка числа колосков является важным компонентом анализа продуктивности растения, однако традиционные методы ручной разметки и подсчета крайне трудоемки, плохо масштабируются и требуют значительных временных затрат. В исследовании предложен подход для эффективной детекции колосков, основанный на использовании упрощенной точечной разметки, при которой эксперт отмечает только центры колосков, без необходимости формировать трудоемкие пиксельные маски или ограничивающие рамки. Такая схема позволяет существенно снизить стоимость подготовки обучающей выборки и ускорить процесс аннотации. Для определения оптимального способа обработки упрощенной разметки были исследованы три метода: сегментация бинарных масок с помощью архитектуры U-Net, регрессия плотностных карт на основе двумерного нормального распределения и функции дивергенции Кульбака–Лейблера, а также детекция областей фиксированного размера с использованием модели YOLOv8. Проведено сравнение точности методов по количественным (MAE, MAPE) и пространственным метрикам (Precision, Recall, F1) на тестовых наборах изображений. Анализ результатов показал, что подходы, основанные на U-Net, обеспечивают высокую точность локализации и подсчета колосков при минимальных затратах на разметку данных, тогда как метод YOLOv8 менее устойчив к геометрической вариативности реальных объектов. Предложенный подход демонстрирует, что комбинация точечной разметки и современных моделей сегментации является эффективным инструментом для автоматизации фенотипирования, что может значительно ускорить селекционные исследования и расширить возможности высокопроизводительного анализа морфологических признаков растений.

**Ключевые слова:** компьютерное зрение; глубокое обучение; пшеница; колос; число колосков; фенотипирование; детекция объектов

**Для цитирования:** Генаев М.А., Бусов И.Д., Кручинина Ю.В., Коваль В.С., Гончаров Н.П. Детекция колосков в колосе пшеницы на RGB-изображениях с использованием глубокого машинного обучения. *Вавиловский журнал генетики и селекции*. 2026;30(1):27-35. doi 10.18699/vjgb-26-09

**Финансирование.** Подготовка данных, разработка и верификация алгоритма выполнены при поддержке Российского научного фонда, проект № 23-14-00150.

**Благодарности.** Расчеты выполнялись с использованием вычислительных ресурсов ЦКП «Биоинформатика» при поддержке бюджетного проекта № FWNR-2022-0020.

## Wheat spikelet detection on RGB images using deep machine learning

М.А. Genaev <sup>1, 2</sup> , I.D. Busov<sup>1, 3</sup>, Yu.V. Kruchinina <sup>1, 2</sup>, V.S. Koval<sup>1, 2</sup>, N.P. Goncharov<sup>1, 3</sup><sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia<sup>2</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia mag@bionet.nsc.ru

**Abstract.** This study addresses the challenge of automated high-throughput phenotyping of wheat spike characteristics using modern computer vision and deep learning methods. Accurate estimation of spikelet number is a key indicator of plant productivity, yet traditional manual counting approaches are labor-intensive, slow, and difficult to scale to large breeding datasets. To overcome these limitations, we propose a spikelet detection strategy based on simplified point annotations, where an expert marks only the centers of spikelets rather than drawing detailed segmentation masks or bounding boxes. This significantly reduces annotation time and lowers the overall cost of preparing training datasets for machine learning models. To determine the most effective way of utilizing such simplified annotations, three computational methods were explored: segmentation of binary masks using a U-Net architecture, density regression based on two-dimensional Gaussian distributions optimized via Kullback–Leibler divergence, and detection of fixed-size bounding regions using the YOLOv8 object detection framework. The models were evaluated on dedicated test datasets using both quantitative metrics (MAE,

MAPE) and spatial localization metrics (Precision, Recall, F1 score). The results demonstrate that U-Net-based approaches provide consistently high accuracy in spikelet localization and counting while maintaining robustness to annotation imperfections. In contrast, the YOLOv8-based method showed reduced performance, likely due to the geometric mismatch between fixed-size boxes and the natural elongated shape of spikelets. Overall, the proposed methodology highlights the effectiveness of combining minimalistic point-level annotation with advanced segmentation models for automating phenotyping workflows. This approach has the potential to accelerate breeding programs, enhance the efficiency of large-scale phenotypic data collection, and support further development of robust computer-vision tools for plant science applications.

**Key words:** computer vision; deep learning; wheat; spike; spikelets per spike; phenotyping; object detection

**For citation:** Genaev M.A., Busov I.D., Kruchinina Yu.V., Koval V.S., Goncharov N.P. Wheat spikelet detection on RGB images using deep machine learning. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed.* 2026;30(1):27-35. doi 10.18699/vjgb-26-09

## Введение

Рост населения и локальное и глобальное изменение климата обуславливают необходимость в ускоренной селекции сельскохозяйственных культур, таких как пшеница, для повышения урожайности и устойчивости к биотическим и абиотическим факторам внешней среды (Ефимов и др., 2024). Оценка признаков, в том числе фенотипирование параметров колоса, – один из ключевых этапов селекции. Среди них число колосков в колосе является наиболее важным показателем продуктивности растения пшеницы (Афонников и др., 2016; Скрипка и др., 2016; Маслова и др., 2018; Vahamidis et al., 2019; и др.).

Число колосков в колосе у пшеницы имеет сложный генетический контроль (Zhang B. et al., 2015) и часто связано с плотностью колоса – не менее важным селекционным и таксономическим признаком (Vavilova et al., 2017; Вавилова и др., 2019; Савин, 2019).

Традиционные методы фенотипирования параметров колоса, базирующиеся на визуальной оценке экспертами, являются медленными, дорогостоящими и субъективными (Konopatskaia et al., 2016). Это стимулирует разработку автоматизированных, использующих компьютерное зрение для высокопроизводительного анализа (Li et al., 2017; Liu et al., 2017; Genaev et al., 2019).

Методы компьютерного зрения на основе глубокого обучения доказали свою эффективность в автоматизации задач фенотипирования сельскохозяйственных растений (Artemenko et al., 2024). Существующие подходы к решению задачи подсчета колосков можно условно разделить на два типа. Первый опирается на детекцию объектов. Например, Ф. Хорошевская с коллегами (Khoroshevsky et al., 2021) использовали архитектуру RetinaNet для детекции и подсчета колосков непосредственно в полевых условиях; достигнуты средние значения процентной ошибки (MAPE) в диапазоне от 9.2 до 11.5 %. Л. Ши с соавторами (Shi et al., 2023) применили модель YOLOv5s для детекции числа колосков на изображениях колосьев, получив на тестовой выборке для зрелой пшеницы среднее значение абсолютной ошибки (MAE) числа колосков, равное 0.43. Второй тип подходов использует семантическую сегментацию. В работе (Misra et al., 2020) предложена архитектура SpikeSegNet на основе U-Net, которая достигла точности подсчета (accuracy) 95 %. Однако общим для этих методов является необходимость в трудоемкой и дорогостоящей полной разметке данных (использование ограничивающих

прямоугольников или пиксельных масок), что становится ключевым ограничением при масштабировании исследований.

В качестве альтернативы для снижения затрат на разметку предлагается использовать точечную аннотацию, когда эксперту требуется отметить лишь центр объекта. Ф. Чен с коллегами (Chen et al., 2021) продемонстрировали, что обучение с неполной точечной разметкой, где аннотировано лишь 50 % колосков, приводит к потере точности на 6.5 % (F1 падает с 84.15 до 78.65 %) по сравнению с полной разметкой. А уменьшение доли размеченных объектов до 10 % снижает точность на 16.5 %. Таким образом, уменьшение трудозатрат на разметку в 10 раз сохраняет 83.5 % исходной точности, что подтверждает перспективность этого направления. Другой путь упрощения подготовки данных – полуавтоматическая разметка (Alkhudaydi et al., 2019). Р. Куи с соавторами (Qiu et al., 2022) предложили метод, где первоначальная модель обучается на слабо размеченных данных, а затем используется для автоматической генерации разметки, на которой повторно обучается новая модель.

Эти исследования демонстрируют прогресс в области автоматизации фенотипирования параметров колоса, но также указывают на главную проблему – сложность и стоимость подготовки данных. Несмотря на прогресс, задача разработки точного и робастного алгоритма, эффективно работающего с упрощенной разметкой и сохраняющего высокую точность, остается актуальной.

В настоящей работе мы исследуем альтернативный подход, основанный на использовании упрощенной точечной разметки лишь центров колосков колоса, что значительно сокращает время и стоимость подготовки данных (Chen et al., 2021) и может стать эффективным и практичным решением для широкого применения в сельскохозяйственных исследованиях.

## Материалы и методы

### Биологический материал и набор данных

В работе использовались колосья растений пшеницы из коллекции Н.П. Гончарова. Выборка растений включала представителей различных видов ди- ( $2n = 2x = 14$ ), тетра- ( $2n = 4x = 14$ ) и гексаплоидных ( $2n = 6x = 42$ ) пшениц. Изображения колосьев были получены в лабораторных условиях согласно протоколу, описанному ранее (Генаев

и др., 2018). Для съемки использовалась цифровая камера Canon 350D, объектив EF-S 18–55 мм f/3.5–5.6. Параметры съемки: выдержка 1/160, диафрагма 11, ISO 100, фокусное расстояние 55 мм. Колос пшеницы располагался на синем фоне рядом с картой цветовой палитры X-Rite Mini ColorChecker Classic (<http://xritephoto.com/colorchecker-targets>). Использовались изображения колоса, полученные на прищепке, и колоса, расположенного на столе. Пример изображений, включенных в анализ, показан на рис. 1.

Всего набор данных включал 1745 цифровых изображений колосьев пшеницы (протокол «на столе» – 82 %, протокол «на прищепке» – 18 % изображений). Для оценки обобщающей способности моделей была взята отдельная отложенная выборка из 14 изображений, полученных по протоколу «на столе», которые не были включены нами ни в обучение, ни в тестирование алгоритмов.

### Разметка колосков на изображениях

В работе было использовано три типа разметки: на основе бинарных масок, гауссовых масок и разметки на основе ограничивающих квадратов. Первоначально разметка изображений осуществлялась в программе ImageJ (Schneider et al., 2012) вручную. Точками размечались центры колосков, их координаты сохранялись в отдельном файле для каждого изображения. На основе координат центров колосков применялись три варианта генерации разметок для машинного обучения (рис. 2).

**Разметка в виде бинарных масок** (см. рис. 2, б). Для каждого колоска генерировалась круговая область радиусом 2 мм. В этом случае мы использовали алгоритм сегментации изображений: нейросетевая модель обучалась предсказывать области на изображении, соответствующие бинарным маскам.

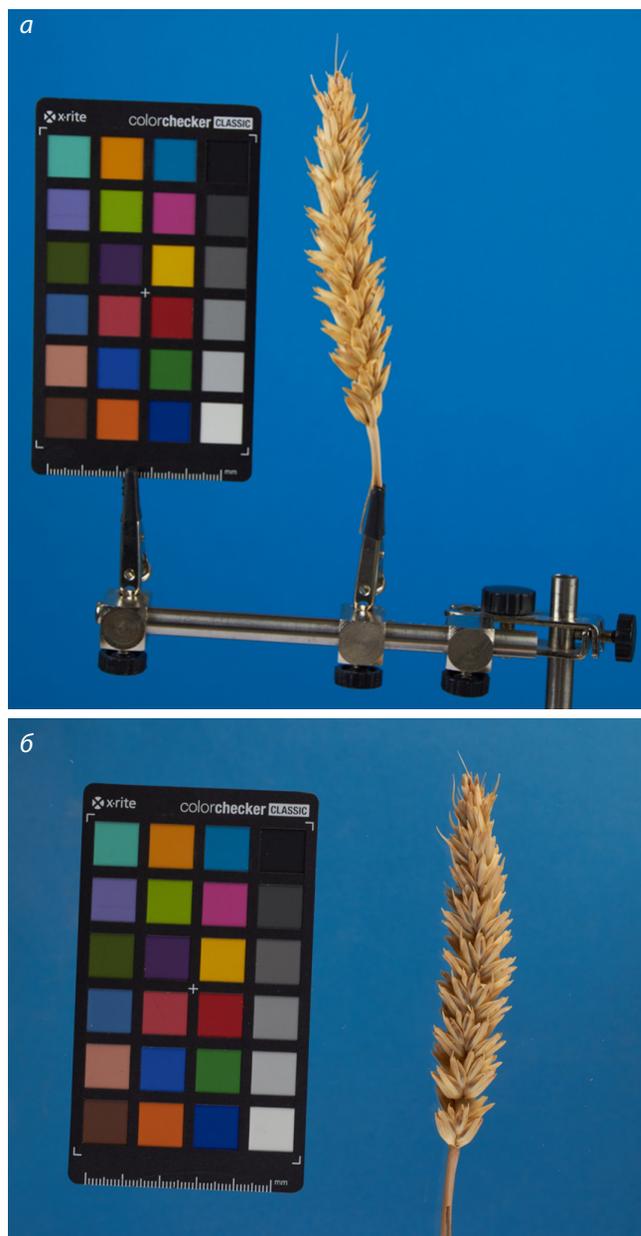
**Разметка в виде гауссовой маски** (см. рис. 2, в). Можно предположить, что центры колосков, размеченные вручную, не соответствуют в точности их геометрическим центрам, а отклоняются случайным образом. Мы предположили, что плотность этого случайного расположения меток вокруг центра имеет вид радиального распределения Гаусса. Поэтому в качестве второго варианта разметки генерировалось двумерное гауссово распределение, центр которого совпадал с размеченным центром колоска.

**Ограничивающие квадраты** (см. рис. 2, г). Вокруг каждого размеченного вручную центра колоска автоматически генерировался квадрат фиксированного размера (6×6 мм). Модель обучалась детектировать эти квадраты.

### Архитектуры моделей и обучение

Для обучения изображения были случайным образом разбиты на тренировочную (~60 %), валидационную (~20 %) и тестовую (~20 %) выборки.

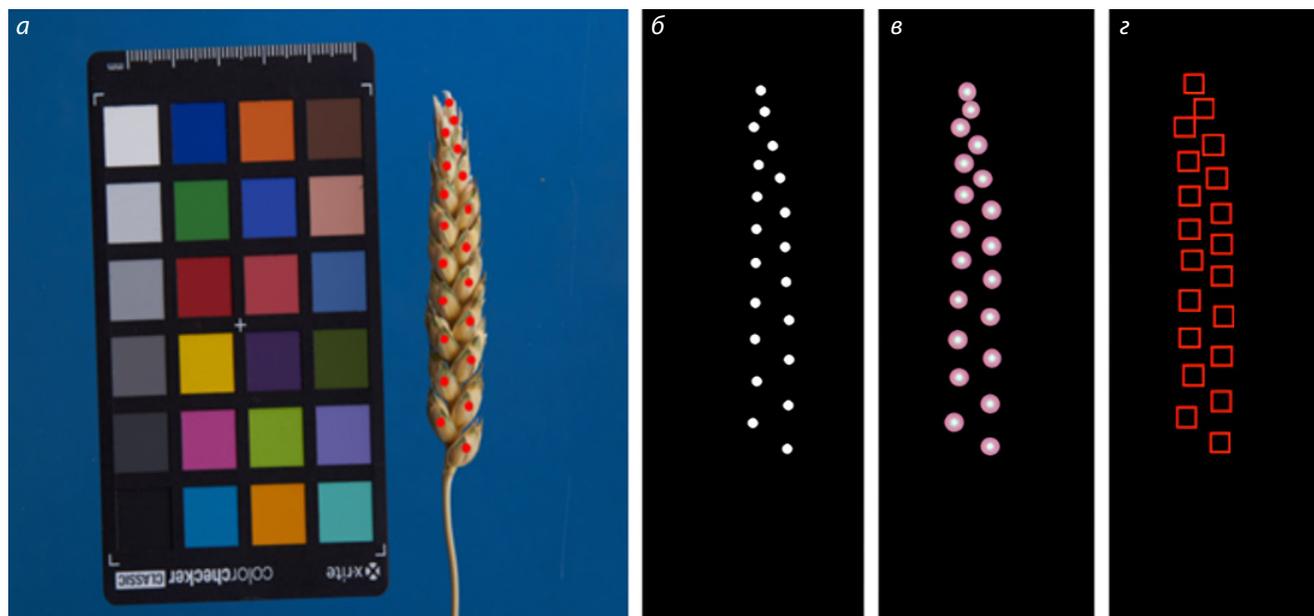
Были задействованы два подхода к определению центров колосков на изображении с помощью алгоритмов глубокого машинного обучения. Первый заключался в решении задачи сегментации пикселей изображения, согласно которому изображения делились на принадлежащие маске и не принадлежащие маске. Для этого применили модели семантической сегментации изображений на основе сети



**Рис. 1.** Примеры изображения колоса пшеницы, закрепленного на прищепке (а) и расположенного на столе (б).

U-Net (Ronneberger et al., 2015). Использовались два типа масок из описанных выше: бинарные и гауссовы. В случае бинарной маски предсказанными центрами колосков считались геометрические центры областей, соответствующих предсказанным маскам. В случае гауссовой маски центр колоска определялся как пиксель с максимальной вероятностью внутри области, удовлетворяющей условию: отношение вероятности каждого пикселя к максимальной вероятности в данной области должно быть не ниже порога  $C$ .

Сеть U-Net (Ronneberger et al., 2015) может использовать кодировщики слоев (энкодеры) различной архитектуры. В нашей работе были протестированы несколько вариантов энкодеров: efficientnet-b3, efficientnet-b4, mit\_b2, mit\_b1,



**Рис. 2.** Методы разметки колосков колоса на основе их центров, указанных вручную.

*а* – исходное изображение колоса с размеченными центрами колосков, показанными красными точками; *б* – бинарная маска для колоска в виде кругов, центры которых совпадают с разметкой центров колосков; *в* – гауссова маска, в которой центры распределения совпадают с центрами размеченных колосков, а плотность вероятности имеет радиальное распределение; *г* – набор ограничивающих квадратов, центры которых совпадают с центрами колосков.

timm-resnest26d, timm-regnetx\_032, timm-res2next50, timm-gernet\_m, timm-efficientnet-b4, timm-efficientnet-b3 (Tan, Le, 2019; Wightman, 2019; Radosavovic et al., 2020; Zhang H. et al., 2020; Gao et al., 2021; Xie et al., 2021).

Второй подход заключался в детекции областей изображения, ограниченных квадратами и соответствующих центральных областям колоска. Для этого использовалась архитектура сети YOLO в версии YOLOv8m, которая для каждого колоска на изображении определяет ограничивающий прямоугольник (Redmon et al., 2016).

Таким образом, мы применили для определения центров три метода: сегментацию с помощью сети U-Net для бинарных (далее обозначена как U-Net-BIN) и гауссовых (далее обозначена как U-Net-GAUSS) масок, ограничивающих положение центра и детекцию области центра колоска в виде квадрата с помощью сети архитектуры YOLO (далее обозначена как YOLOv8).

**Вид функции потерь.** Для алгоритма с бинарными масками использовалась функция бинарной кросс-энтропии (torch.nn.BCEWithLogitsLoss). Для гауссовых масок в качестве функции потерь использовалась дивергенция Кульбака–Лейблера (torch.nn.KLDivLoss).

**Организация процесса обучения.** Все три алгоритма обучались в течение 500 эпох. Веса моделей были инициализированы на основе предобученных параметров, полученных на наборе данных ImageNet (Deng et al., 2009). В качестве метода оптимизации параметров применили алгоритм Adam. В процессе обучения использовали набор аугментаций, реализованных в библиотеке albumentations (Buslaev et al., 2020). Изображение нормализовалось к размеру 512×224 (Resize), затем применялись: горизон-

тальное отражение с вероятностью 0.5 (HorizontalFlip (p=0.5)); вертикальное отражение с вероятностью 0.277 (VerticalFlip (p=0.277)); поворот на случайный угол в диапазоне  $-30...+30^\circ$  с вероятностью 0.735 (Rotate(limit=30, p=0.735)); гауссово размытие с размером ядра, случайно выбираемым в диапазоне от 1 до 3, с вероятностью 0.25 (GaussianBlur(blur\_limit=(1.3), p=0.25)); добавление гауссового шума с вероятностью 0.15 (GaussNoise(p=0.15)); случайная корректировка яркости и контраста с вероятностью 0.5 (RandomBrightnessContrast(p=0.5)); случайное изменение интенсивности каналов R, G и B в диапазоне  $\pm 15$  с вероятностью 0.5 (RGBShift(r\_shift\_limit=15, g\_shift\_limit=15, b\_shift\_limit=15, p=0.5)); цветовые преобразования – случайные изменения яркости, контраста, насыщенности и оттенка (ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.2, p=0.703)); преобразование изображения в градации серого с вероятностью 0.1 (ToGray (p=0.1)).

Исследованные нами модели нейронных сетей для идентификации центров колосков пшеницы зависели от целого ряда параметров. К ним относились: тип архитектуры энкодера для сети U-Net, радиус круговой области  $r$  для бинарных масок, порог  $C$  отношения вероятности пикселя принадлежать центру колоска к максимальной вероятности в случае гауссовой маски, параметры алгоритма оптимизации и некоторые другие. Их полный список и области значений приведены в табл. 1.

В процессе работы для алгоритмов требовалось подобрать наиболее оптимальные значения параметров из табл. 1. Для этого применен алгоритм байесовской оптимизации, реализованный в библиотеке Optuna (Akiba et

**Таблица 1.** Гиперпараметры, используемые в методах определения положения центра колосков на изображении

Метод	Гиперпараметры	Области значений
U-Net-BIN	Архитектура энкодера	efficientnet-b3, efficientnet-b4, mit_b2, mit_b1, timm-resnest26d, timm-regnetx_032, timm-res2next50, timm-gernet_m, timm-efficientnet-b4, timm-efficientnet-b3
	Коэффициент градиентного спуска на начальном этапе оптимизации	[0.0001, 0.001]
	Коэффициент градиентного спуска на финальном этапе оптимизации	[0.000001, 0.00005]
	Радиус бинарной маски (r, мм)	[0.5, 3]
U-Net-GAUSS	Архитектура энкодера	efficientnet-b3, efficientnet-b4, mit_b2, mit_b1, timm-resnest26d, timm-regnetx_032, timm-res2next50, timm-gernet_m, timm-efficientnet-b4, timm-efficientnet-b3
	Коэффициент градиентного спуска на начальном этапе оптимизации	[0.0001, 0.001]
	Коэффициент градиентного спуска на финальном этапе оптимизации	[0.000001, 0.00005]
	Параметр $\sigma$ для гауссова распределения	[0.000002, 0.0002]
	Порог вероятности для выбора области положения центра колоска (C)	[1/9, 1/2]
YOLOv8	Коэффициент градиентного спуска на начальном этапе оптимизации	[0.01, 0.00005]
	Коэффициент градиентного спуска на финальном этапе оптимизации	[0.000005, 0.000001]
	Параметр DropBlock	[0, 0.5]

al., 2019) с использованием изображений валидационной выборки, а в качестве целевого параметра оптимизации – метрика точности F1, описанная ниже.

**Метрики оценки точности алгоритмов**

Эффективность определения центров колосков анализировали при помощи двух типов метрик. Количественные метрики оценивали ошибку подсчета количества колосков в колосе на изображении. Среднюю абсолютную ошибку (MAE) и абсолютную процентную ошибку (MAPE) подсчета колосков определяли согласно:

$$MAE = \frac{1}{n} \sum_{i=1}^n |t_i - p_i|,$$

$$MAPE = \frac{100}{n} \% \sum_{i=0}^n \left| \frac{t_i - p_i}{t_i} \right|,$$

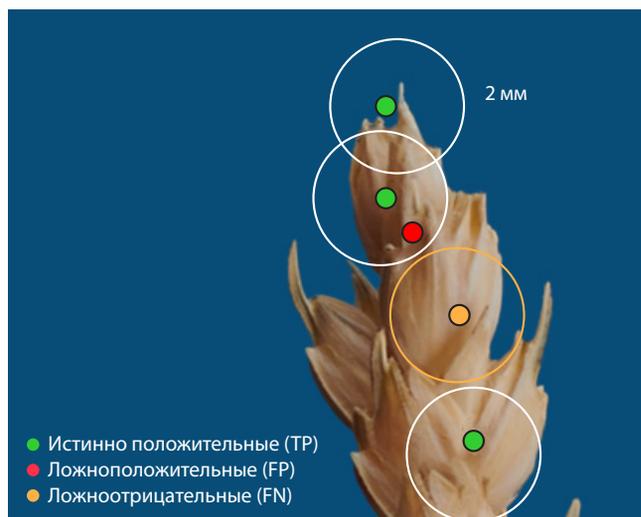
где  $t_i$  – истинное число колосков на изображении, подсчитанное вручную;  $p_i$  – число колосков, определенное на основе методов машинного обучения;  $n$  – число изображений в выборке, по которым оценивается точность.

Для оценки точности определения положения центра колосков на изображении подсчитывались метрики точности (Precision), полноты (Recall) и F1-мера:

$$Precision = \frac{TP}{TP + FP},$$

$$Recall = \frac{TP}{TP + FN},$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall},$$



**Рис. 3.** Визуализация истинно положительных, ложноположительных и ложноотрицательных предсказаний при вычислении пространственных метрик оценки качества детекции колосков.

где TP (true positive) – число верных положительных предсказаний положения колоска; FP (false positive) – число неверных предсказаний положения колоска; FN (false negative) – число ложных предсказаний положения колоска. Для определения параметров TP, FP, FN использовалось следующее правило (рис. 3): предсказанное положение

центра считалось истинно положительным (TP), если находилось в пределах радиуса 2 мм от истинного центра; ложноположительным (FP), если центр колоска предсказан неверно; ложноотрицательным (FN), если в радиусе 2 мм от истинного центра нет предсказанных центров.

Программы были реализованы на языке Python 3.11 с использованием библиотек PyTorch 2.3.0, OpenCV 4.9.0 и scikit-learn 1.5.0. Расчеты выполнялись на рабочей станции с процессором AMD Ryzen 9 7950X, 64 ГБ оперативной памяти и графическим ускорителем NVIDIA GeForce RTX 2080 Ti (11 ГБ).

## Результаты и обсуждение

Оптимальные значения гиперпараметров для трех моделей предсказания положения колосков на изображении, подобранные с помощью байесовской оптимизации, перечислены в табл. 2.

Наилучшей архитектурой энкодера для определения центра колоска на основе сегментации и бинарной маски оказалась MiT-B2 (Mix Transformer-B2) (Xie et al., 2021). Данная архитектура сочетает сверточные слои с механизмом самовнимания, что позволяет одновременно захватывать локальные особенности и глобальные контекстные зависимости на изображении. Это особенно важно для точного позиционирования мелких объектов, таких как центры колосков.

Для гауссовой маски наилучшие результаты показала архитектура EfficientNet-B4 (Tan, Le, 2019). Модель использует метод составного масштабирования (compound scaling), который сбалансированно увеличивает глубину, ширину и разрешение входных данных. Такой подход обеспечивает баланс между точностью и вычислительной эффективностью, что делает архитектуру хорошо подходящей для задач регрессии плотности распределения.

Результаты оценки на тестовой выборке (табл. 3) показали, что оба подхода на основе U-Net значительно превосходят метод на основе YOLOv8.

Как видно из табл. 3, оба подхода на основе архитектуры U-Net продемонстрировали высокую и сопоставимую точность как по количественным (MAE ~0.5), так и по пространственным (F1 > 0.96) метрикам на тестовой выборке. Незначимое превосходство модели U-Net-GAUSS по метрике MAE (0.502 против 0.512) и модели U-Net-BIN по метрике F1 (0.965 против 0.962) позволяет считать их одинаково эффективными для данных, распределение которых соответствует обучающей выборке. В то же время модель YOLOv8 дает значительно более низкие результаты (MAE = 3.641, F1 = 0.679), что указывает на ее непригодность для решения задачи в текущей конфигурации с автоматически сгенерированными квадратными ограничивающими рамками.

На отложенной выборке (табл. 4) модель U-Net-BIN обуславливает наилучшую точность, тогда как для модели U-Net-GAUSS точность оказалась ниже. Наибольшая ошибка наблюдается для модели YOLOv8.

Результаты на отложенной выборке, собранной в иное время и имеющей несколько иное распределение, демонстрируют более выраженные различия между методами. Модель U-Net-BIN сохранила высокую точность, показав лишь незначительный рост ошибки (MAE с 0.512 до 0.538), что свидетельствует о ее высокой обобщающей способности и надежности. В отличие от нее, точность модели U-Net-GAUSS на отложенной выборке заметно снизилась (MAE = 0.846 против 0.502 на тестовой выборке). Это позволяет предположить, что подход может быть более чувствителен к изменчивости данных. Модель YOLOv8, как и на тестовой выборке, показала наихудший результат (MAE = 3.0), дополнительно подтверждая вывод о том, что выбранный способ генерации ограничивающих квадратов не подходит для описания формы колосков.

Оба предложенных метода на основе U-Net позволили получить хорошие результаты на тестовой выборке. Однако модель, обученная на бинарных масках, продемон-

**Таблица 2.** Гиперпараметры для моделей нейронных сетей идентификации положения центров колосков на изображении, подобранные с помощью байесовской оптимизации

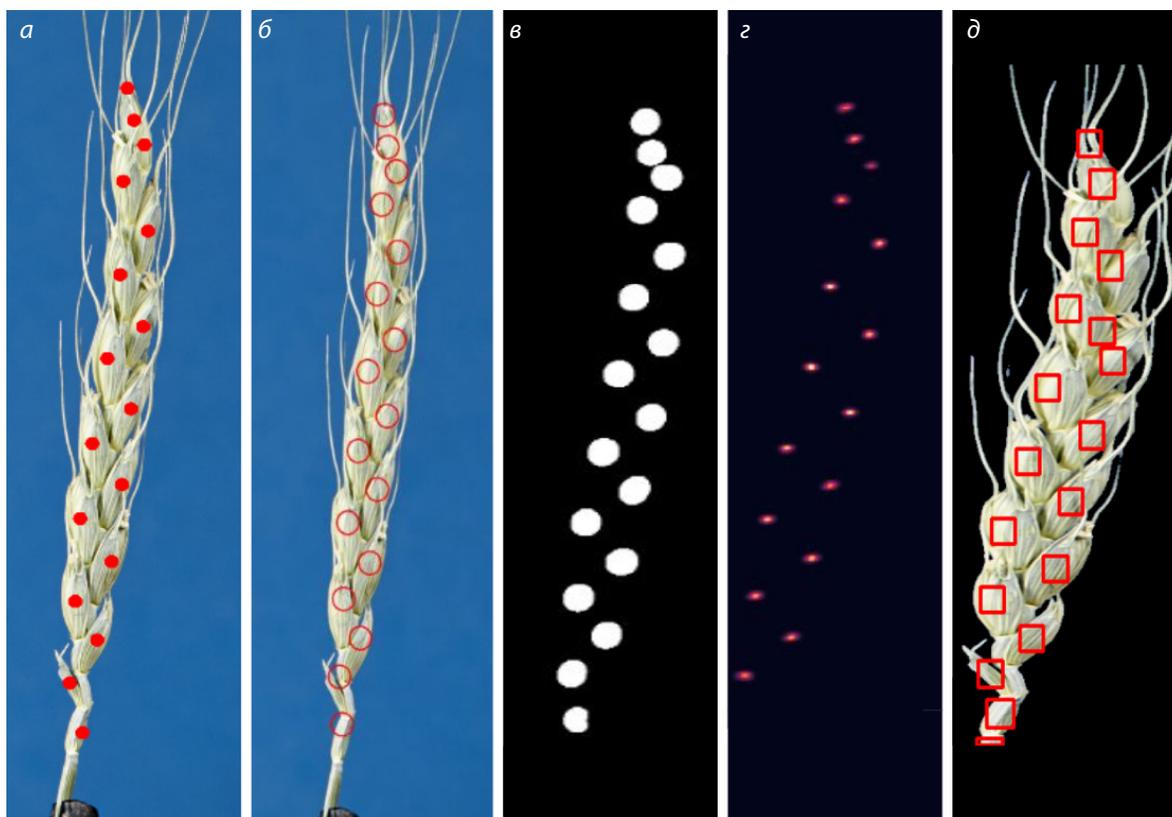
Модель	Гиперпараметры	Оптимальные значения
U-Net-BIN	Архитектура энкодера	mit_b2
	Коэффициент градиентного спуска на начальном этапе оптимизации	0.001
	Коэффициент градиентного спуска на финальном этапе оптимизации	0.000001
	Радиус бинарной маски (r, мм)	1.5999
U-Net-GAUSS	Архитектура энкодера	efficientnet-b4
	Коэффициент градиентного спуска на начальном этапе оптимизации	0.0001
	Коэффициент градиентного спуска на финальном этапе оптимизации	0.0000063723
	Параметр $\sigma$ для гауссова распределения	0.00001121
	Порог вероятности для выбора области положения центра колоска (C)	0.214564
YOLOv8	Коэффициент градиентного спуска на начальном этапе оптимизации	0.00042
	Коэффициент градиентного спуска на финальном этапе оптимизации	0.000001
	Параметр DropBlock	0.48

**Таблица 3.** Оценка эффективности методов определения положения центра колоска на тестовой выборке

Модель	MAE	MAPE	F1
U-Net-BIN	0.512	0.032	<b>0.965</b>
U-Net-GAUSS	<b>0.502</b>	0.032	0.962
YOLOv8	3.641	0.280	0.679

**Таблица 4.** Оценка эффективности методов определения положения центра колоска на отложенной выборке ( $n = 14$ )

Модель	MAE	MAPE	F1
U-Net-BIN	<b>0.538</b>	<b>0.035</b>	<b>0.972</b>
U-Net-GAUSS	0.846	0.062	0.942
YOLOv8	3.000	0.231	0.704



**Рис. 4.** Результаты работы моделей: а – размеченные вручную центры колосков в колосе; б – результат работы модели, обученной на бинарных масках, наложенный на реальное изображение колоса; в – результат работы модели, обученной на бинарных масках; з – результат работы модели, обученной на масках радиального распределения Гаусса; д – результат работы модели, обученной на ограничивающих квадратах.

стрировала значительно лучшую обобщающую способность на отложенной выборке, что говорит о ее большей надежности для работы с данными, имеющими иное распределение. Регрессионный анализ предсказанного и истинного числа колосков показал сильную линейную зависимость для моделей U-Net ( $R^2 > 0.95$  на тестовой выборке), подтверждая высокую точность подсчета.

Измерение скорости инференса на одном графическом ускорителе показало, что время обработки одного изображения для модели, обученной на гауссовых масках, состав-

ляет примерно 0.67 с, тогда как для модели на бинарных масках – около 0.64 с. Таким образом, подход U-Net-BIN демонстрирует не только более высокую точность, но и несколько большую скорость работы.

Низкая производительность YOLOv8, вероятно, связана с неоптимальностью автоматически сгенерированных квадратных рамок для описания продолговатой и изогнутой формы колосков (рис. 4). Это указывает на то, что для методов детекции требуется более тщательная и, возможно, ручная корректировка разметки.

## Заключение

Настоящее исследование демонстрирует, что использование упрощенной точечной разметки центров колосков в сочетании с современными архитектурами глубокого обучения позволяет достичь точности, сопоставимой с лучшими современными методами, использующими более сложную и дорогую разметку (Misra et al., 2020; Khoroshevsky et al., 2021; Zhou et al., 2021; Shi et al., 2023). Упрощенная схема разметки не только снижает стоимость экспериментов, но и открывает возможность быстрого расширения наборов данных. Предложенный подход может быть адаптирован для решения других задач подсчета морфологических признаков растений и интегрирован в системы автоматизированного фенотипирования для ускорения селекционных программ.

## Список литературы / References

- Афонников Д.А., Генаев М.А., Дорошков А.В., Комышев Е.Г., Пшеничникова Т.А. Методы высокопроизводительного фенотипирования растений для массовых селекционно-генетических экспериментов. *Генетика*. 2016;52(7):788-803. doi 10.7868/S001667581607002X  
[Afonnikov D.A., Genaev M.A., Doroshkov A.V., Komyshev E.G., Pshenichnikova T.A. Methods of high-throughput plant phenotyping for large-scale breeding and genetic experiments. *Russ J Genet*. 2016;52(7):688-701. doi 10.1134/S1022795416070024]
- Вавилова В.Ю., Конопатская И.Д., Блинов А.Г., Гончаров Н.П. Межвидовой полиморфизм генов *DEP1* и форма колоса у пшениц. *Генетика*. 2019;55(7):837-843. doi 10.1134/S0016675819070142  
[Vavilova V.Yu., Konopatskaia I.D., Blinov A.G., Goncharov N.P. Interspecific polymorphism of *DEP1* genes and the spike shape in wheats. *Russ J Genet*. 2019;55(7):908-913. doi 10.1134/S1022795419070147]
- Генаев М.А., Комышев Е.Г., Хао Ф., Коваль В.С., Гончаров Н.П., Афонников Д.А. SpikeDroidDB – информационная система для аннотации морфометрических характеристик колоса пшеницы. *Вавиловский журнал генетики и селекции*. 2018;22(1):132-140. doi 10.18699/VJ18.340  
[Genaev M.A., Komyshev E.G., Khao F., Koval V.S., Goncharov N.P., Afonnikov D.A. SpikeDroidDB: an information system for annotation of morphometric characteristics of wheat spike. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2018;22(1):132-140. doi 10.18699/VJ18.340 (in Russian)]
- Ефимов В.М., Речкин Д.В., Гончаров Н.П. Многомерный анализ многолетних климатических данных в связи с урожайностью, скороспелостью и проблемой глобального потепления. *Вавиловский журнал генетики и селекции*. 2024;28(2):155-165. doi 10.18699/vjgb-24-18  
[Efimov V.M., Rechkin D.V., Goncharov N.P. Multivariate analysis of long-term climate data in connection with yield, earliness and the problem of global warming. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2024;28(2):155-165. doi 10.18699/vjgb-24-18]
- Маслова Г.Я., Абдраев М.Р., Шарاپов И.И., Шарাপова Ю.А. Корреляционный анализ урожайности и элементов продуктивности сортов озимой мягкой пшеницы в засушливых условиях лесостепной зоны среднего Поволжья. *Известия Самарского научного центра Российской академии наук*. 2018;20(2-4):680-683. doi 10.24411/1990-5378-2018-00153  
[Maslova G., Abdryayev M.R., Sharapov I.I., Sharapova Ju.A. Correlation analysis of yield and elements of productivity of winter soft wheat varieties in the arid conditions of steppe zone of Middle Volga region. *Izvestiya of Samara Scientific Center of the Russian Academy of Sciences*. 2018;20(2-4):680-683. doi 10.24411/1990-5378-2018-00153 (in Russian)]
- Савин В.Ю. Определение усилий, необходимых для очеса колоса пшеницы. *Инженерные технологии и системы*. 2019;29(3):456-466. doi 10.15507/2658-4123.029.201903.456-466  
[Savin V.Yu. Determination of the force required for stripping the wheat ear. *Inzhenernyye Tekhnologii i Sistemy = Engineering Technologies and Systems*. 2019;29(3):456-466. doi 10.15507/2658-4123.029.201903.456-466 (in Russian)]
- Скрипка О.В., Самофалов А.П., Подгорный С.В., Громова С.Н. Урожайность и основные элементы продуктивности у сортов озимой пшеницы интенсивного типа селекции ВНИИЗК. *Достижения науки и техники АПК*. 2016;30(9):30-32.  
[Skripka O.V., Samofalov A.P., Podgorniy S.V., Gromova S.N. Productivity and main elements of productivity of winter wheat varieties of intensive type, bred in the All-Russian Research Institute of Grain Crops. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology in Agro-Industrial Complex*. 2016;30(9):30-32 (in Russian)]
- Akiba T., Sano S., Yanase T., Ohta T., Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA. ACM, 2019;2623-2631. doi 10.1145/3292500.3330701
- Alkhudaydi T., Zhou J., De La Iglesia B. SpikeletFCN: counting spikelets from infield wheat crop images using fully convolutional networks. In: Artificial Intelligence and Soft Computing. ICAISC 2019. Lecture Notes in Computer Science. Vol. 11508. Springer, Cham, 2019;3-13. doi 10.1007/978-3-030-20912-4\_1
- Artemenko N.V., Epifanov R.U., Genaev M.A., Komyshev E.G., Kruchinina Yu.V., Koval V.S., Goncharov N.P., Afonnikov D.A. Image-based classification of wheat spikes by glume pubescence using convolutional neural networks. *Front Plant Sci*. 2024;14:1336192. doi 10.3389/fpls.2023.1336192
- Buslaev A., Igloukov V.I., Khvedchenya E., Parinov A., Druzhinin M., Kalinin A. Albumentations: fast and flexible image augmentations. *Information*. 2020;11(2):125. doi 10.3390/info11020125
- Chen F., Pound M.P., French A.P. Learning to localise and count with incomplete dot-annotations. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada. IEEE, 2021;1612-1620. doi 10.1109/ICCVW54120.2021.00186
- Deng J., Dong W., Socher R., Li L.J., Li K., Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009, Miami. IEEE, 2009;248-255. doi 10.1109/CVPR.2009.5206848
- Gao S.-H., Cheng M.-M., Zhao K., Zhang X.-Y., Yang M.-H., Torr P. Res2Net: a new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021;43(2):652-662. doi 10.1109/TPAMI.2019.2938758
- Genaev M.A., Komyshev E.G., Smirnov N.V., Kruchinina Y.V., Goncharov N.P., Afonnikov D.A. Morphometry of the wheat spike by analyzing 2D images. *Agronomy*. 2019;9(7):390. doi 10.3390/agronomy9070390
- Khoroshevsky F., Khoroshevsky S., Bar-Hillel A. Parts-per-object count in agricultural images: solving phenotyping problems via a single deep neural network. *Remote Sens*. 2021;13(13):2496. doi 10.3390/rs13132496
- Konopatskaia I., Vavilova V., Blinov A., Goncharov N.P. Spike morphology genes in wheat species (*Triticum* L.). *Proc Latv Acad Sci B Nat Exact Appl Sci*. 2016;70(6):345-355. doi 10.1515/prolas-2016-0053
- Li Q., Cai J., Berger B., Okamoto M., Miklavcic S.J. Detecting spikes of wheat plants using neural networks with Laws texture energy. *Plant Methods*. 2017;13:83. doi 10.1186/s13007-017-0231-1
- Liu T., Chen W., Wang Y., Wu W., Sun C., Ding J., Guo W. Rice and wheat grain counting method and software development based on

- Android system. *Comput Electron Agric.* 2017;141:302-309. doi 10.1016/j.compag.2017.08.011
- Misra T., Arora A., Marwaha S., Chinnusamy V., Rao A., Jain R., Narayan R., Ray M., Kumar S., Raju D., Ranjan R., Nigam A., Goel S. SpikeSegNet – a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods.* 2020;16:40. doi 10.1186/s13007-020-00582-9
- Qiu R., He Y., Zhang M. Automatic detection and counting of wheat spikelet using semi-automatic labeling and deep learning. *Front Plant Sci.* 2022;13:872555. doi 10.3389/fpls.2022.872555
- Radosavovic I., Kosaraju R.P., Girshick R., He K., Dollár P. Designing network design spaces. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, Seattle, WA, USA. IEEE, 2020;10428-10436. doi 10.1109/CVPR42600.2020.01044
- Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, Las Vegas, NV, USA. IEEE, 2016;779-788. doi 10.1109/CVPR.2016.91
- Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science. Vol. 9351. Springer, 2015;234-241. doi 10.1007/978-3-319-24574-4\_28
- Schneider C.A., Rasband W.S., Eliceiri K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9(7):671-675. doi 10.1038/nmeth.2089
- Shi L., Sun J., Dang Y., Zhang S., Sun X., Xi L., Wang J. YOLOv5s-T: a lightweight small object detection method for wheat spikelet counting. *Agriculture.* 2023;13(4):872. doi 10.3390/agriculture13040872
- Tan M., Le Q.V. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, California, USA. ICML, 2019;6105-6114
- Vahamidis P., Karamanos A.J., Economou G. Grain number determination in durum wheat as affected by drought stress: an analysis at spike and spikelet level. *Ann Appl Biol.* 2019;174(2):190-208. doi 10.1111/aab.12487
- Vavilova V., Konopatskaia I., Kuznetsova A.E., Blinov A., Goncharov N.P. Genomic characterization of *DEP1* gene in wheats with normal and compact spike shape. *BMC Genetics.* 2017;18 (Suppl.1): 106. doi 10.1186/s12863-017-0583-6
- Wightman R. PyTorch Image Models (timm). *GitHub Repository.* 2019. doi 10.5281/zenodo.4414861
- Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J., Luo P. SegFormer: simple and efficient design for semantic segmentation with transformers. *arXiv.* 2021. doi 10.48550/arXiv.2105.15203
- Zhang B., Liu X., Xu W., Chang J., Li A., Mao X., Zhang X., Jing R. Novel function of a putative *MOC1* ortholog associated with spikelet number per spike in common wheat. *Sci Rep.* 2015;5(1):12211. doi 10.1038/srep12211
- Zhang H., Wu C., Zhang Z., Zhu Y., Zhang Z., Lin H., Sun Y., He T., Mueller J., Manmatha R., Li M., Smola A. ResNeSt: split-attention networks. *arXiv.* 2020. doi 10.48550/arXiv.2004.08955
- Zhou H., Riche A.B., Hawkesford M.J., Whalley W.R., Atkinson B.S., Sturrock C.J., Mooney S.J. Determination of wheat spike and spikelet architecture and grain traits using X-ray Computed Tomography imaging. *Plant Methods.* 2021;17(1):26. doi 10.1186/s13007-021-00726-5

---

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 10.09.2025. После доработки 23.10.2025. Принята к публикации 01.11.2025.

doi 10.18699/vjgb-26-14

# Different patterns of *P* transposon and *blood* retrotransposon distribution in Harwich and Canton-S sub-strains do not affect the manifestation of *Drosophila melanogaster* intraspecific PM hybrid dysgenesis

L.P. Zakharenko  , Y.Y. Ilinsky 

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

 zakharlp@bionet.nsc.ru

**Abstract.** Intraspecific infertility, the nature of which is not always understood, occurs in many eukaryotes. Intraspecific PM hybrid dysgenesis (PM HD) in *Drosophila melanogaster* manifests in one cross direction as offspring infertility and other genetic disorders due to incompatibility between the maternal cytoplasm and the paternal genome. PM HD is believed to result from a massive transposition of the *P*-element when the maternal cytoplasm lacks a repressor to block it. In this work, we have investigated the distribution of the *P* transposon and *blood* retrotransposon in the reference PM HD strains (Canton-S and Harwich), which have been maintained in different laboratories for several decades. *P*-element distribution patterns vary among Harwich sub-strains, indicating that the *P*-element was translocated in these genomes. The rate of movement of the *P*-element, which was not induced by crosses, is comparable to the rate of movement of other DNA transposons. The distribution pattern of the low-active *blood* retrotransposon in Harwich sub-strains is more stable than that of the *P*-element, indicating genetic relatedness between sub-strains. Derivatives of the *P*-element detected in some Canton-S sub-strains possibly indicate genetic contamination. The significant difference in the *blood* transposable element distribution pattern in Canton-S sub-strains also indicates genetic heterogeneity among them. Despite the complex genealogy of the studied sub-strains, including cases of possible genetic contamination, and differences in *P*-element distribution, the ability to express PM HD symptoms is preserved in the studied sub-strains.

**Key words:** PM hybrid dysgenesis; *Drosophila melanogaster*; *P*-element; *blood* retrotransposon

**For citation:** Zakharenko L.P., Ilinsky Y.Y. Different patterns of *P* transposon and *blood* retrotransposon distribution in Harwich and Canton-S sub-strains do not affect the manifestation of *Drosophila melanogaster* intraspecific PM hybrid dysgenesis. *Вавиловский журнал генетики и селекции*. 2026;30(1):36-42. doi 10.18699/vjgb-26-14

**Funding.** The work was supported by the Budget Project FWNR-2022-0019 from the Ministry of Science and Higher Education of the Russian Federation.

**Acknowledgements.** The authors thank the Shared Facility Center for Microscopic Analysis of Biological Objects SB RAS, Novosibirsk and Dr. O. Ignatenko for technical assistance and valuable comments. The authors thank the reviewers for their constructive comments.

## Паттерн распределения *P*-транспозона и *blood* ретротранспозона в отводках референсных линий Harwich и Canton-S из разных лабораторий не влияет на проявление внутривидового PM гибридного дисгенеза у *Drosophila melanogaster*

Л.П. Захаренко  , Ю.Ю. Илинский 

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

 zakharlp@bionet.nsc.ru

**Аннотация.** Внутривидовое бесплодие, природа которого не всегда понятна, встречается у многих эукариот. Внутривидовый PM гибридный дисгенез (PM ГД) у *Drosophila melanogaster* проявляется в одном из направлений скрещивания в бесплодии потомства и некоторых других генетических нарушениях в результате несовместимости между материнской цитоплазмой и отцовским геномом. Считается, что PM ГД является результатом массового перемещения *P*-элемента, если материнская цитоплазма не имеет репрессора, блокирующего его перемещение. В данной работе мы исследовали распределение *P*-элемента в отводках референсных для PM ГД линиях (Canton-S и Harwich), которые долгое время содержались в разных лабораториях. Паттерны распределения *P*-элемента различались в разных отводках Harwich. Скорость перемещения *P*-элемента, не

индуцированная скрещиваниями, сопоставима со скоростью перемещения других ДНК-транспозонов. Паттерн распределения малоактивного ретротранспозона *blood* в производных линии Harwich был более стабильным по сравнению с *P*-элементом, что свидетельствует о генетическом родстве между отводками. В некоторых отводках линии Canton-S были обнаружены следы *P*-элемента, что может указывать на генетическое загрязнение. Значительная разница в паттерне распределения транспозона *blood* в отводках линии Canton-S также говорит о генетической гетерогенности отводков. Несмотря на сложную генеалогию отводков исследованных линий, включающую случаи возможного генетического загрязнения, и различия в распределении *P*-элементов, способность проявлять симптомы РМ ГД у исследованных отводков сохраняется. Мы показали, что все изученные производные Canton-S имеют сильный М-цитотип, а все отводки Harwich – индуцирующий Р-цитотип.

**Ключевые слова:** РМ гибридный дисгенез; *Drosophila melanogaster*; *P*-элемент; ретротранспозон *blood*

## Introduction

Progeny of some *Drosophila melanogaster* suffer from the hybrid dysgenesis (HD) phenomenon that occurs in one cross direction (Kidwell et al., 1977) and results in female sterility, male recombination, high mutation rate and other genetic disorders (Bingham et al., 1982). It is commonly considered that HD is caused by massive *P*, *hobo* or *I* element transpositions in PM, HE or IR HD, respectively (Bingham et al., 1982; Bucheton et al., 1984; Blackman et al., 1987).

It is a commonly held opinion that *P*-elements appeared in the *D. melanogaster* genome by horizontal transfer in the middle of the last century. In American populations, *P*-elements were first identified in *D. melanogaster* obtained from nature in 1938, from Russian populations, in 1966, and from French populations, in 1967 (Anxolabehere et al., 1988). However, all *D. melanogaster* strains collected on different continents from 1969 contain *P*-elements in their genomes (Anxolabehere et al., 1988). Thus, the distribution of the *P*-elements throughout the world was extremely rapid despite possible harm.

PM is the most common variant of HD, manifesting with all symptoms listed above. It was postulated to occur in crosses between females with the M cytotype, which lack *P*-elements in their genome, and males that harbor a functional *P*-element (Bingham et al., 1982). However, it was later found that many M cytotype strains contained a complete *P*-element (Bingham et al., 1982), and the presence of a *P*-element in the genome did not guarantee the induction of HD (Itoh et al., 1999, 2001, 2007). For this reason, the nature of PM HD is studied mainly on the reference Canton-S (Maternal cytotype) and Harwich (Paternal cytotype) strains. Crossing these strains results in HD phenotypes with 100 % probability. Not every strain with a full-length *P*-element in its genome induces HD, and none of them consistently exhibits HD as strongly as the Harwich strain. This fact alone casts doubt on the hypothesis of the *P*-element's involvement in HD induction. A correlation was found between the strains' ability to induce PM HD and the presence of a highly truncated *P*-element variant named "Har-P" in the Harwich genome, which is the most frequent *de novo* insertion. However, genomic location may also influence host tolerance of *Har-P*s, as a significant rescue of viable pupae was observed in crosses with *Har-P*s located only on Chromosome 3, which has six *P*-elements, while no pupae survived with *Har-P*s located on Chromosome 2, which has nine *P*-elements (Srivastav et al., 2019). Thus, the reason for the uniqueness of the reference P strains remains not fully understood. According to a widely accepted hypothesis, the

*P*-element mobilization is repressed in P cytotype strains by small RNA (Brennecke et al., 2008; Khurana et al., 2011).

The Canton-S strain has been widely used among genetic scientists since 1925. However, *P*-elements have been identified in some Canton-S strains (Ignatenko et al., 2015), possibly as a result of genetic contamination during cultivation in laboratory. "A golden rule of *Drosophila* genetics is never trust a stock label, no matter how reputable the source from which it was obtained" (Ashburner, 1989).

Here, we analyze the pattern of *P* transposon and *blood* retrotransposon distribution in Canton-S and Harwich sub-strains from different laboratories and their ability to produce dysgenic progeny to describe accumulated genetic differences. We selected *blood* retrotransposon, as its distribution pattern remained stable for two decades in spite of the presence of a complete *blood* copy in the isogenic reference y; cn bw sp genome (Ignatenko et al., 2015). According to our data, the rate of movement of the *P*-element in Harwich sub-strains without dysgenic crosses is comparable to the rate of movement of other DNA transposons. All Canton-S derivatives have a strong reactive M cytotype, and all Harwich sub-strains have an inducer P cytotype.

## Methods

**Strains of *Drosophila melanogaster*.** Five Canton-S (CS) and three Harwich (H) sub-strains from different laboratories were used for the study (see the Table). The H2 sub-strain has a visible, spontaneously occurring mutation in the *sepia* gene (66 D5), and the H3 sub-strain has a mutation in the *white* gene (3 B6). There is no information in the literature about the artificial origin of these mutations. The LK-P (1A) strain containing one complete *P*-element per genome (from S. Ronsseray) was used to obtain a DNA probe of a full-sized *P*-element. Reference strain y; cn bw sp (No. 2057, Bloomington *Drosophila* collection) with a completely sequenced genome was used to obtain a *blood* probe (see the Table). The Harwich strain was taken from nature (Harwich, Massachusetts) in 1967 by M.L. Tracey (Kidwell et al., 1977). According to S. Ronsseray, the Harwich strain descended from two females (Ronsseray et al., 1984). Bloomington *Drosophila* Stock Center received the Harwich strain (stock number 4264) from M. Kidwell in 1997.

**Hybrid dysgenesis assays.** To test HD sensitivity, sub-strains were crossed to CS2 or H3, a cross between which results in high penetrance of HD (Ignatenko et al., 2015). Progeny from crosses between H3 males and females of each Canton-S sub-strain, as well as CS2 female and males of each Harwich sub-strains, were cultivated at 29 °C from the egg

Strains description

Strain	Resource	Year of strain appearance in ICG, Novosibirsk
CS1	Bloomington Drosophila Stock Center (BDSC) #ND	2013
CS2	ICG, I. Zakharov from Obninsk	1990
CS3	ICG, B. Chadov	1990
CS4	ICG, N. Gruntenko	2013
CS5	CNRS-Universite' Pierre et Marie Curie, France, S. Ronsseray	2013
Harwich-1	BDSC #4264	2013
Harwich-2	France, S. Ronsseray	2013
Harwich-3	ICG, I. Zakharov from L. Kaidanov	1996
LK-P (1A)	France, S. Ronsseray	2013
y; cn bw sp	BDSC #2057	2006

stage to adulthood. As a marker of PM HD, ovary morphology was analyzed in 50 females aged 4–5 days (Kidwell, Novy, 1979).

**Fluorescence *in situ* hybridization (FISH).** The *P*-element probe was obtained by PCR with a primer complementary to terminal inverted repeats 5'-TGATGAAATAACATAAGGTG GTCCCGTGC-3' (Takasu-Ishikawa et al., 1992; Lapie et al., 1993), and the *blood* probe, with primers 5'-CAGTGGCATAAC GCTCAAGA-3' and 5'-GGTTCGCGAAATACCAGTGT-3' (Ignatenko et al., 2015). The probes were labeled by nick-translation with digoxigenin-dUTP for the *P*-element, and with biotin-dUTP for *blood*. FISH analysis was performed for 3–5 larvae for each sub-strain as in Ignatenko et al. (2015). Squashed preparations of polytene chromosomes from the salivary glands of third-instar larvae were made according to standard procedures (Biémont et al., 1994). The hybridization sites were analyzed using an Axio Imager M1 (Carl Zeiss) microscope in the Shared Facility Center for Microscopic Analysis of Biological Objects SB RAS, Novosibirsk.

Results

To test sub-strains for PM HD, they were crossed to CS2 or H3, a cross between which results in high penetrance of HD (Ignatenko et al., 2015). Females of the CS2 sub-strain were crossed to H1, H2 or H3 males to check Harwich sub-strains induction ability. Females of five Canton-S sub-strains were crossed to H3 males to check the ability of Canton-S sub-strains to respond to induction. Indeed, abnormal gonad morphology was observed in progeny of all dysgenic crosses with 100 % penetrance, indicating a strong *P* cytotypic for Harwich and M cytotypic for Canton-S sub-strains (Table S1)<sup>1</sup>. Opposite crosses (female Harwich to Canton-S males) produced normal progeny (Table S1).

**Blood element pattern in Harwich and Canton-S sub-strains**

The pattern of the *blood* retrotransposon was studied to determine the level of similarity among the sub-strains. The reason for choosing the *blood* TE was its low-level mobility observed in the isogenic reference strain y; cn bw sp. The localization

<sup>1</sup> Supplementary Tables S1–S4 and Figs S1–S3 are available at: <https://vavilovj-icg.ru/download/pict-2026-30/appx2.pdf>

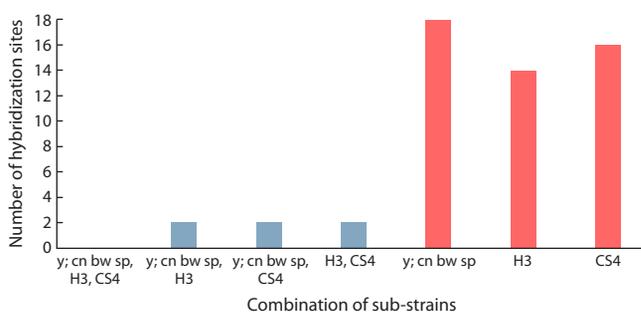


Fig. 1. Distribution of common and unique *blood* hybridization sites in unrelated strains.

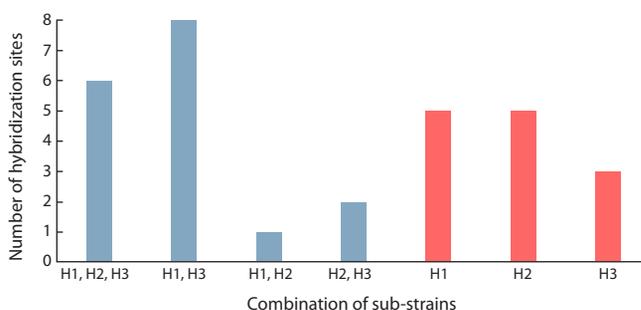
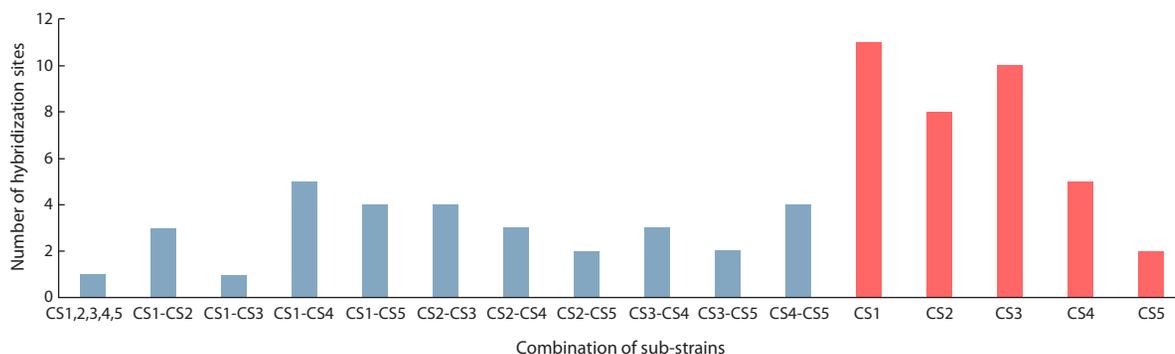


Fig. 2. Distribution of common and unique *blood* hybridization sites in Harwich sub-strains.

of the *blood* retrotransposon in the y; cn bw sp strain has not changed significantly over two decades (1992–2013) (Ignatenko et al., 2015). As expected, unrelated strains (H3, CS4 and y; cn bw sp) display mostly unique *blood* hybridization sites (89 %), and no hybridization sites were found to be common to all three strains. Pairwise matching sites were rare (Fig. 1, Tables S2, S4).

Thirty two *blood* hybridization sites were found in the genomes of Harwich sub-strains: 6 sites common for all three sub-strains, 15 unique sites, and 11 sites common for different pairs (Fig. 2). Thus, 20 % of *blood* hybridization sites are common to all three sub-strains (H1, H2, H3), and 47 % are unique.



**Fig. 3.** Distribution of common and unique *blood* hybridization sites in Canton-S sub-strains.

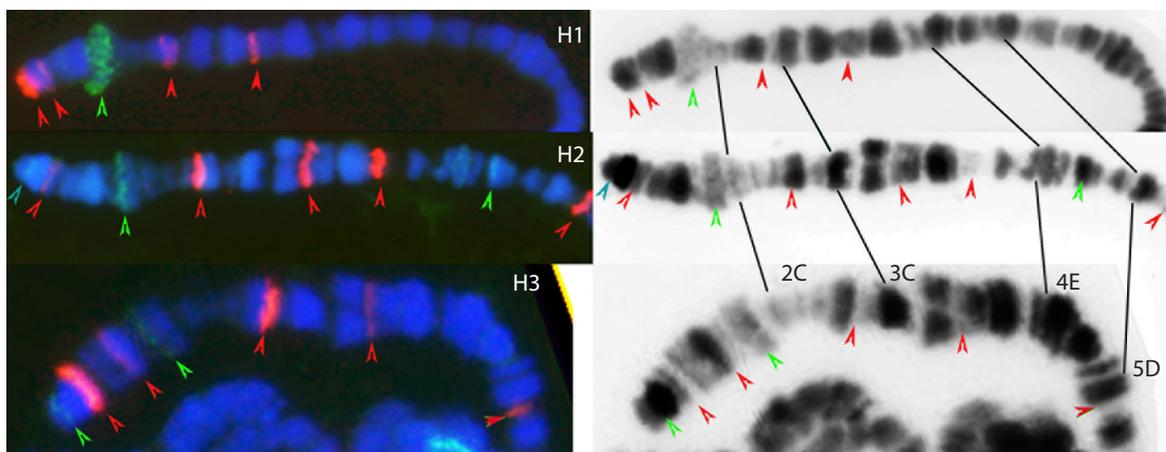
The H1 and H3 sub-strains displayed the greatest similarity, harboring 56 % (14 out of 25) common hybridization sites. It can thus be assumed that of the three investigated Harwich sub-strains, at least H1 and H3 sub-strains have common origin.

Strikingly, the *blood* distribution pattern shows extreme variability among Canton-S sub-strains. Of the total 60 hybridization sites, most are unique (60 %). One site is shared by all five sub-strains, two sites are shared by three sub-strains, and pairwise coincidences range from one to six common sites between different pairs (Fig. 3). Consequently, genetic contamination of at least some sub-strains cannot be excluded. Additional evidence for genetic contamination comes from the presence of traces of the *P*-element in some of these Canton-S sub-strains (Ignatenko et al., 2015) despite the fact that classically Canton-S serves as the gold standard for being *P*-element-negative. It is interesting that traces of the *P*-element (two sites per genome) were found only on the third chromosome and were localized in different regions of the CS1, CS2, CS4 genomes (Ignatenko et al., 2015). Notably, *P*-element appeared in the *D. melanogaster* genome later than the Canton-S strain was isolated from nature. Rahman with coauthors (2015) also reported introgression of *P*-element-containing lab strains into certain stocks of Oregon R, which should be free of *P*-elements originally (Robertson, Engels, 1989).

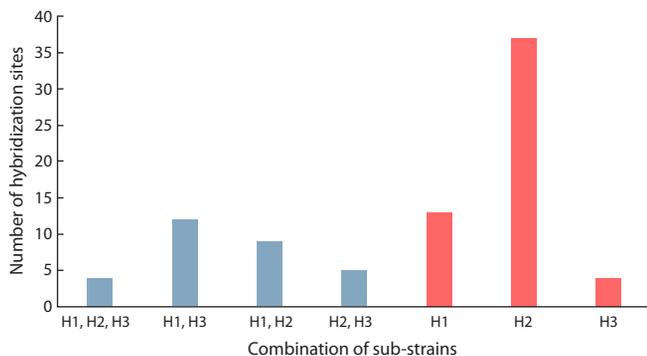
### *P*-element pattern in Harwich sub-strains

We identified 85 *P*-element hybridization sites in Harwich sub-strains, of which only 5 % were common for all three sub-strains (Fig. 4, Table S3, Figs S1–S3). The percentage of unique *P*-element hybridization sites is 33 % for H1, 69 % for H2 and 28 % for H3 (Fig. 5). This result shows much higher variability in *P*-element distribution compared to *blood* distribution among the Harwich sub-strains. Moreover, common *P*-element hybridization sites, for example in 3F regions, show weak signal in the H3 sub-strain, suggesting a possible difference in fine structure of *P*-elements in different sub-strains (Fig. 4). Although FISH is a semi-quantitative method, the use of a full-length copy of the *P*-element as a probe allows interpreting the intensity of the signal as a reflection of the size of the sequence with which the probe hybridizes on the chromosome. Thus, the difference between sub-strains in the distribution of hybridization sites may be greater than we assume.

The H2 sub-strain harbors the largest number of unique *P*-element sites (44 %), which is consistent with *blood* distribution in this sub-strain also being the most distinct from the other two Harwich sub-strains in pairwise comparisons. In turn, the H1 and H3 sub-strains, which show the greatest similarity in *blood* distribution, shared 35 % (16 out of 47)



**Fig. 4.** Localization of *P*-element (red) and *blood* TE (green) hybridization sites on the end of the X chromosomes of Harwich sub-strains from different laboratories. DAPI (blue).



**Fig. 5.** Distribution of common and unique *P*-element hybridization sites in Harwich sub-strains.

of *P*-element hybridization sites. Thus, although the Harwich sub-strains are more heterogeneous in *P*-element distribution compared to *blood* distribution, a closer relationship between the H1 and H3 sub-strains is suggested by distribution patterns of both retrotransposons. Besides, the H2 and H3 sub-strains have visible genetic markers that help to monitor the purity of the strains. H3 has a mutation in the *white* gene, H2, in the *sepia* gene. The origin of the *sepia* and *white* mutations has not been precisely determined. The H3 strain was obtained from L. Kaidanov in 1996. Similarities in *P*-element distributions at the end of the X chromosomes of different Harwich derivatives support the spontaneous occurrence of the *white* mutation at least (Fig. 4).

Three Canton-S sub-strains (CS1, CS2, CS4) contain *P*-element derivatives as a result of possible genetic contamination. Given that derivatives of *P*-element were found only on the third chromosome, these sub-strains may share a common origin and may have been distributed between laboratories after *P*-element contamination. In terms of *blood* distribution, the CS1 and CS4 sub-strains show the highest degree of similarity among all pairwise comparisons. CS3 and CS5 have no traces of the *P*-element in the genome; this circumstance makes them similar.

## Discussion

We found that all Canton-S sub-strains have strong M cytotypes and all Harwich sub-strains exhibit a strong P cytotype independent of possible genetic contamination and a different pattern of *P*-element distribution. A comparison of the Harwich sub-strains indicates that the *P*-element can change its position without crossing with M strains. This is similar to the genomic instability reported for *I*-elements, which also occurs in the absence of dysgenic crosses (Moschetti et al., 2010). The symptoms of IR (Inducer-Reactive) HD differ from those of PM HD: dysgenic females lay eggs, but the offspring die at the embryonic stage. In both cases, the symptoms are consistently reproduced with the same physiological manifestations typical of a given HD type and specific reference strains. All other things being equal, only a limited number of strains can induce HD in both cases.

The Harwich strain was separated from nature (Harwich, Massachusetts) in 1967 by M.L. Tracey (Kidwell et al., 1977).

The peak of activity in studying the PM HD phenomenon and the spread of the strain across laboratories occurred in the 90s of the last century (Anxolabehere et al., 1984, 1988; Wang et al., 1993; Simmons et al., 1996; Ronsseray et al., 1998). During this time, at least 300 generations of *Drosophila* have changed. The H1 and H3 sub-strains differ in 11 out of 25 common *blood* hybridization sites, and in 31 out of 47 *P*-element sites. In our case, the rate of transposition can be estimated as at least  $15 \times 10^{-4}$  (11/25/300) per site per genome per generation and  $22 \times 10^{-4}$  (31/47/300) for *blood* and for *P*-element, respectively.

According to the FISH analyses, the rate of TE transpositions in natural populations varies by several orders of magnitude, ranging from  $10^{-5}$  per copy per generation to  $10^{-2}$  per copy per generation (Pasyukova et al., 1998; Maside et al., 2000; Diaz-González et al., 2011). Sequencing analysis of parents and offspring from 18 families of full-sib *D. melanogaster* estimated the range of TE insertion rate from  $10^{-3}$  per copy per generation to  $10^{-5}$  per copy per generation (Wang et al., 2023). By comparing the parents and offspring in each family, authors identified 89 new TE insertions across the 89 samples, making it only one insertion per genome per generation on average (Wang et al., 2023). The rate of movement of the *P*-element in our experiment is in the range typical for the movement of other TEs.

Whole genome sequencing was used to estimate the rate of *P*-element movement in dysgenic ovaries and germ cells (Moon et al., 2018; Jansen et al., 2024).

Contrary to predictions based on the insertional mutagenesis model of hybrid dysgenesis, single-cell whole-genome sequencing analysis of DNA from dysgenic and non-dysgenic embryos at late embryonic stages (before dysgenic germ cell death) shows that dysgenic and non-dysgenic germ cells acquire unexpectedly similar, low numbers of new heterozygous *P*-element insertions (Jansen et al., 2024). The authors found double-strand breaks in generative cells and proposed that transposase excises the *P*-element, but does not integrate it into the genome (Jansen et al., 2024). To confirm their assumption, the authors simulate the excision of the *P*-element by using CRISPR Cas9. However, the gaps with blunt ends, induced by CRISPR Cas9, are usually completed by a non-homologous path (Quétier, 2016), whereas transposase recognizes the inverted terminal ends of the *P*-element and cuts them with the formation of sticky ends, which are successfully ligated or completed by repair enzymes using the homolog. The excision of the *P*-element can occur precisely, without loss of gene functionality (Weinert et al., 2005). Double-strand breaks in dysgenic germline cells can be induced not only by *P*-element translocation.

Moon with coauthors investigated the rate of *P*-element movement in the ovaries of dysgenic females raised at 18 °C, when symptoms of hybrid dysgenesis do not appear and the ovaries have normal morphology (Moon et al., 2018). Nevertheless, the authors found 527 new *P*-element insertion sites in dysgenic ovaries. However, only a single new *P*-element insertion was found in F1 and F2 dysgenic progeny (Eggleston et al., 1988). It means that cells with a huge number of new *P*-elements should be eliminated, but the ovaries of dysgenic females raised at 18 °C, paradoxically, have normal morpho-

logy. Thus, disputable results were obtained about the mass movement of *P*-elements during PM HD.

It should be noted that Moon with coauthors do not see the movement of other TE families in the dysgenic cross direction (Moon et al., 2018).

The Harwich genome contains 132 *P*-elements according to Moon, and 80 (half of which are heterozygous) according to Jansen with coauthors (Moon et al., 2018; Jansen et al., 2024). On the one hand, the annotation of sequenced genomes with repeats is still imperfect and the difference can be explained by a reading error. For example, Moon with coauthors (2018) admit that 34 new insertions per genome may be false positives. On the other hand, it is possible that the issue is not only in the accuracy of the sequencing method or the choice of tissue for sequencing but also that the Harwich sub-strains, although having the same origin, were cultivated in different laboratories for a long time. Jansen with coauthors used the Harwich strain from BDSC. According to Ronsseray, this strain descended from two females since 1967 (Ronsseray et al., 1984). The initial polymorphism of the strain may affect the heterogeneity of the sub-strains. Moon (Moon et al., 2018) received the Harwich strain from Khurana et al. (2011), Khurana (Khurana et al., 2011) received the Harwich strain from Ronsseray. Thus, sub-strains originate from the same source, but have been cultured independently for a long time.

According to our data, Harwich sub-strains from different laboratories also differ in the number and localization of *P*-element hybridization sites. The sub-strain obtained from Ronsseray has more *P*-element hybridization sites than the sub-strains from BDSC. This is less than that revealed by sequencing, because the sensitivity of the FISH method depends on the specific activity of the labeled probe. In addition, weak signals may be missed during FISH analysis. Nevertheless, according to our data, relationship between the Harwich sub-strains is undeniable, since the percentage of common hybridization sites between the Harwich sub-strains is higher than between obviously unrelated strains.

PM HD may result not only from possible massive *P*-element movement, but also from an incompatibility of temperature-dependent metabolic processes of sensitive to PM HD strains, as dysgenic flies, when reared at low temperatures (20 °C), do not display dysgenic symptoms (Engels, Preston, 1979; Dorogova et al., 2017). Additionally, there is a hormonal difference between dysgenic and non-dysgenic flies (Zakharenko et al., 2014). The reference PM HD Harwich and Canton-S strains also differ in a number of physiological characteristics: life expectancy, fertility, locomotor activity, development rate (Zakharenko et al., 2024). The asymmetry in the expression of the dysgenic traits, characteristic of interspecific hybrids, suggests a significant genetic distance between the reference PM HD strains not only in *P*-element absence/presence.

## Conclusion

The relationship between Harwich sub-strains is undeniable despite the difference in *P*-element distribution pattern. The rate of movement of the *P*-element in Harwich sub-strains without induction by crossing is in the range typical for the movement of other TEs. All Canton-S sub-strains have strong

M cytotypes and all Harwich sub-strains exhibit a strong P cytotype independently of possible genetic contamination and a different pattern of *P*-element distribution.

## References

- Anxolabehere D., Kai H., Nouaud D., Périquet G., Ronsseray S. The geographical distribution of *P*-*M* hybrid dysgenesis in *Drosophila melanogaster*. *Genet Sel Evol.* 1984;16(1):15-26. doi 10.1186/1297-9686-16-1-15
- Anxolabehere D., Kidwell M.G., Periquet G. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P*-elements. *Mol Biol Evol.* 1988;5(3):252-269. doi 10.1093/oxfordjournals.molbev.a040491
- Ashburner M. *Drosophila. A Laboratory Handbook.* New York: Cold Spring Harbor Laboratory Press, 1989
- Biémont C., Lemeunier F., Garcia Guerreiro M.P., Brookfield J.F., Gautier C., Aulard S., Pasyukova E.G. Population dynamics of the copia, mdg1, mdg3, gypsy, and P transposable elements in a natural population of *Drosophila melanogaster*. *Genet Res.* 1994;63(3):197-212. doi 10.1017/s0016672300032353
- Bingham P.M., Kidwell M.G., Rubin G.M. The molecular basis of P-M hybrid dysgenesis: the role of the *P*-element, a P-strain-specific transposon family. *Cell.* 1982;29(3):995-1004. doi 10.1016/0092-8674(82)90463-9
- Blackman R.K., Grimaila R., Koehler M.M., Gelbart W.M. Mobilization of hobo elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. *Cell.* 1987;49(4):497-505. doi 10.1016/0092-8674(87)90452-1
- Brennecke J., Malone C.D., Aravin A.A., Sachidanandam R., Stark A., Hannon G.J. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science.* 2008;322(5906):1387-1392. doi 10.1126/science.1165171
- Bucheton A., Paro R., Sang H.M., Pelisson A., Finnegan D.J. The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. *Cell.* 1984; 38(1):153-163. doi 10.1016/0092-8674(84)90536-1
- Díaz-González J., Vázquez J.F., Albornoz J., Domínguez A. Long-term evolution of the *roo* transposable element copy number in mutation accumulation lines of *Drosophila melanogaster*. *Genet Res.* 2011; 93(3):181-187. doi 10.1017/S0016672311000103
- Dorogova N.V., Bolobolova E.U., Zakharenko L.P. Cellular aspects of gonadal atrophy in *Drosophila* P-M hybrid dysgenesis. *Dev Biol.* 2017;424(2):105-112. doi 10.1016/j.ydbio.2017.02.020
- Eggleston W.B., Johnson-Schlitz D.M., Engels W.R. P-M hybrid dysgenesis does not mobilize other transposable element families in *D. melanogaster*. *Nature.* 1988;331(6154):368-370. doi 10.1038/331368a0
- Engels W.R., Preston C.R. Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility. *Genetics.* 1979; 92(1):161-174. doi 10.1093/genetics/92.1.161
- Ignatenko O.M., Zakharenko L.P., Dorogova N.V., Fedorova S.A. *P*-elements and the determinants of hybrid dysgenesis have different dynamics of propagation in *Drosophila melanogaster* populations. *Genetica.* 2015;143(6):751-759. doi 10.1007/s10709-015-9872-z
- Itoh M., Woodruff R.C., Leone M.A., Boussy I.A. Genomic *P*-elements and P-M characteristics of eastern Australian populations of *Drosophila melanogaster*. *Genetica.* 1999;106(3):231-245. doi 10.1023/a:1003918417012
- Itoh M., Sasai N., Inoue Y., Watada M. *P*-elements and P-M characteristics in natural populations of *Drosophila melanogaster* in the southernmost islands of Japan and in Taiwan. *Heredity (Edinb).* 2001;86:206-212. doi 10.1046/j.1365-2540.2001.00817.x
- Itoh M., Takeuchi N., Yamaguchi M., Yamamoto M.T., Boussy I.A. Prevalence of full-size *P* and *KP* elements in North American popu-

- lations of *Drosophila melanogaster*. *Genetica*. 2007;131(1):21-28. doi 10.1007/s10709-006-9109-2
- Jansen G., Gebert D., Kumar T.R., Simmons E., Murphy S., Teixeira F.K. Tolerance thresholds underlie responses to DNA damage during germline development. *Genes Dev*. 2024;38(13-14):631-654. doi 10.1101/gad.351701.124
- Khurana J.S., Wang J., Xu J., Koppetsch B.S., Thomson T.C., Nowosielska A., Li C., Zamore P.D., Weng Z., Theurkauf W.E. Adaptation to *P* element transposon invasion in *Drosophila melanogaster*. *Cell*. 2011;147(7):1551-1563. doi 10.1016/j.cell.2011.11.042
- Kidwell M.G., Novy J.B. Hybrid dysgenesis in *Drosophila melanogaster*: sterility resulting from gonadal dysgenesis in the P-M system. *Genetics*. 1979;92(4):1127-1140. doi 10.1093/genetics/92.4.1127
- Kidwell M.G., Kidwell J.F., Sved J.A. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics*. 1977;86(4):813-833. doi 10.1093/genetics/86.4.813
- Lapie P., Nasr F., Lepesant J.A., Deutsch J. Deletion scanning of the regulatory sequences of the *Fbp1* gene of *Drosophila melanogaster* using *P* transposase-induced deficiencies. *Genetics*. 1993;135(3):801-816. doi 10.1093/genetics/135.3.801
- Maside X., Assimakopoulos S., Charlesworth B. Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. *Genet Res*. 2000;75:275-284. doi 10.1017/S0016672399004474
- Moon S., Cassani M., Lin Y.A., Wang L., Dou K., Zhang Z.Z. A robust transposon-endogenizing response from germline stem cells. *Dev Cell*. 2018;47(5):660-671.e3. doi 10.1016/j.devcel.2018.10.011
- Moschetti R., Dimitri P., Caizzi R., Junakovic N. Genomic instability of I elements of *Drosophila melanogaster* in absence of dysgenic crosses. *PLoS One*. 2010;5(10):e13142. doi 10.1371/journal.pone.0013142
- Pasyukova E.G., Nuzhdin S.V., Filatov D.A. The relationship between the rate of transposition and transposable element copy number for *copia* and *Doc* retrotransposons of *Drosophila melanogaster*. *Genet Res*. 1998;72(1):1-11. doi 10.1017/S0016672398003358
- Quétier F. The CRISPR-Cas9 technology: closer to the ultimate toolkit for targeted genome editing. *Plant Sci*. 2016;242:65-76. doi 10.1016/j.plantsci.2015.09.003
- Rahman R., Chirn G.W., Kanodia A., Sytnikova Y.A., Brems B., Bergman C.M., Lau N.C. Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. *Nucleic Acids Res*. 2015;43(22):10655-10672. doi 10.1093/nar/gkv1193
- Robertson H.M., Engels W.R. Modified *P* elements that mimic the P cytotype in *Drosophila melanogaster*. *Genetics*. 1989;123(4):815-824. doi 10.1093/genetics/123.4.815
- Ronsseray S., Anxolabéhère D., Périquet G. Hybrid dysgenesis in *Drosophila melanogaster*: influence of temperature on cytotype determination in the P-M system. *Mol Gen Genet*. 1984;196(1):17-23. doi 10.1007/BF00334086
- Ronsseray S., Marin L., Lehmann M., Anxolabéhère D. Repression of hybrid dysgenesis in *Drosophila melanogaster* by combinations of telomeric *P*-element reporters and naturally occurring *P* elements. *Genetics*. 1998;149(4):1857-1866. doi 10.1093/genetics/149.4.1857
- Simmons M.J., Raymond J.D., Grimes C.D., Belinco C., Haake B.C., Jordan M., Lund C., Ojala T.A., Papermaster D. Repression of hybrid dysgenesis in *Drosophila melanogaster* by heat-shock-inducible sense and antisense *P*-element constructs. *Genetics*. 1996;144(4):1529-1544. doi 10.1093/genetics/144.4.1529
- Srivastav S.P., Rahman R., Ma Q., Pierre J., Bandyopadhyay S., Lau N.C. *Har-P*, a short *P*-element variant, weaponizes *P*-transposase to severely impair *Drosophila* development. *eLife*. 2019;8:e49948. doi 10.7554/eLife.49948
- Takasu-Ishikawa E., Yoshihara M., Hotta Y. Extra sequences found at *P*-element excision sites in *Drosophila melanogaster*. *Mol Gen Genet*. 1992;232:17-23. doi 10.1007/BF00299132
- Wang Y., Balter H., Levitan M., Margulies L. Mutability, sterility and suppression in P-M hybrid dysgenesis: the influence of P subline, cross, chromosome, sex and *P*-element structure. *Genet Res*. 1993;62(2):111-123. doi 10.1017/s0016672300031700
- Wang Y., McNeil P., Abdulazeez R., Pascual M., Johnston S.E., Keightley P.D., Obbard D.J. Variation in mutation, recombination, and transposition rates in *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res*. 2023;33(4):587-598. doi 10.1101/gr.277383.122
- Weinert B.T., Min B., Rio D.C. *P* element excision and repair by non-homologous end joining occurs in both G<sub>1</sub> and G<sub>2</sub> of the cell cycle. *DNA Repair (Amst)*. 2005;4(2):171-181. doi 10.1016/j.dnarep.2004.09.004
- Zakharenko L.P., Karpova E.K., Rauschenbach I.Y. P-M hybrid dysgenesis affects juvenile hormone metabolism in *Drosophila melanogaster* females. *Russ J Genet*. 2014;50:772-774. doi 10.1134/S1022795414060143
- Zakharenko L.P., Bobrovskikh M.A., Gruntenko N.E., Petrovskii D.V., Verevkin E.G., Putilov A.A. Two old wild-type strains of *Drosophila melanogaster* can serve as an animal model of faster and slower aging processes. *Insects*. 2024;15(5):329. doi 10.3390/insects15050329

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

Received July 24, 2025. Revised December 12, 2025. Accepted December 15, 2025.

doi 10.18699/vjgb-26-06

# Локус сои qDTF-7 как пример генетической гетерогенности, ассоциированной с продолжительностью цветения и созревания

Р.Н. Перфильев , М.И. Шматова<sup>1</sup>, А.Б. Щербань <sup>1, 2</sup>, Е.А. Салина <sup>1, 2</sup><sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия<sup>2</sup> Курчатовский геномный центр ИЦИГ СО РАН, Новосибирск, Россия PerfilyevRN@bionet.nsc.ru

**Аннотация.** В настоящее время полногеномный анализ ассоциаций (GWAS) стал стандартным подходом для идентификации локусов количественных признаков, ассоциированных с различными фенотипическими признаками. При последующем изучении локуса, т.е. поиске гена и мутации, которые влияют на признак, можно столкнуться с разными проблемами. Одна из проблем – генетическая (или локусная) гетерогенность, ситуация, когда аллели из разных близко расположенных генов в одном локусе влияют на один признак. Ранее с помощью GWAS на хромосоме 3 сои мы обнаружили локус qDTF-7, ассоциированный с продолжительностью цветения в условиях Новосибирска. Первоначально для данного локуса в качестве наиболее вероятного гена-кандидата мы определили *GmTOE1*, ортолог *TOE1* (*TARGET OF EAT 1*), известного регулятора цветения у арабидопсиса. У *GmTOE1* были обнаружены четыре основных гаплотипа, которые ассоциированы с цветением и созреванием сои и, вероятно, связаны с адаптацией сои к северным широтам. Однако этот ген демонстрировал очень слабую ассоциацию с цветением сои в Новосибирской области по сравнению с Орловской, что указывало на наличие в составе локуса другого гена, который может влиять на цветение сои. Повторно проанализировав гены в составе локуса qDTF-7, мы примерно на 21 т.п.н. выше гена *GmTOE1* обнаружили *GmRVE8c*, ортолог гена *RVE8* (*REVEILLE 8*), который вовлечен в процессы развития у арабидопсиса в качестве компонента циркадных ритмов. Изучив естественное нуклеотидное разнообразие *GmRVE8c*, мы выявили четыре основных гаплотипа, которые возникли из-за трех однонуклеотидных замен и одной делеции длиной 19 п.н., приводящей к сдвигу рамки считывания. Для определения трех гаплотипов *GmRVE8c<sup>hap1, 3, 4</sup>*, преобладающих в современных сортах сои, мы разработали ДНК-маркеры. С помощью этих маркеров мы генотипировали 129 сортообразцов сои, для которых ранее изучили продолжительность развития в условиях Новосибирской и Орловской областей. В результате с помощью наших сведений и данных из SoyOmics мы обнаружили ассоциацию гаплотипов *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* с поздним цветением и созреванием сои. Раннеспелый гаплотип *GmRVE8c<sup>hap1</sup>* преобладает в сортообразцах из северных регионов и, вероятно, связан с адаптацией сои к северным широтам. Гаплотип *GmRVE8c<sup>hap4</sup>* полностью сцеплен с раннеспелым аллелем *GmTOE1<sup>C</sup>*, а гаплотип *GmRVE8c<sup>hap3</sup>* сильно сцеплен с позднеспелым аллелем *GmTOE1<sup>T</sup>*. Кроме того, результат ANOVA показывает наличие взаимодействия *GmRVE8c* с главным регулятором цветения сои – геном *E1*. Данное взаимодействие выражается более сильным эффектом гаплотипов *GmRVE8c<sup>hap3, 4</sup>* на цветение и созревание на генетическом фоне аллеля *e1-as* в сравнении с *E1*. Все это формирует достаточно комплексный и интересный локус, который может выступать как возможный пример генетически гетерогенного локуса.

**Ключевые слова:** соя; цветение; *TOE1*; *RVE8*; генетическая гетерогенность

**Для цитирования:** Перфильев Р.Н., Шматова М.И., Щербань А.Б., Салина Е.А. Локус сои qDTF-7 как пример генетической гетерогенности, ассоциированной с продолжительностью цветения и созревания. *Вавиловский журнал генетики и селекции*. 2026;30(1):43-52. doi 10.18699/vjgb-26-06

**Финансирование.** Работа выполнена при поддержке РФФ (проект № 21-76-30003-П). Разработка маркеров и их верификация проведены в рамках бюджетного проекта FWNR-2025-0033.

## Soybean locus qDTF-7 as an example of genetic heterogeneity associated with flowering and maturity time

R.N. Perfil'ev , M.I. Shmatova<sup>1</sup>, A.B. Shcherban <sup>1, 2</sup>, E.A. Salina <sup>1, 2</sup><sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia<sup>2</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia PerfilyevRN@bionet.nsc.ru

**Abstract.** Genome-wide association studies (GWAS) have become a standard approach for identifying quantitative trait loci associated with diverse phenotypic traits. Further investigation of the locus – specifically, the search for the causal gene and mutation – may present various challenges. One of the challenges is genetic heterogeneity (or locus

heterogeneity), when alleles from different closely located genes can influence the same trait. Recently, using GWAS, we found the qDTF-7 locus on soybean chromosome 3, which is associated with flowering time under Novosibirsk conditions. Initially, we identified *GmTOE1*, an ortholog of *TOE1* (*TARGET OF EAT1*), a known flowering-time regulator in Arabidopsis, as the most likely candidate gene for this locus. Four major haplotypes were identified in *GmTOE1*, which are associated with soybean flowering and maturity and are likely to provide soybean adaptation to northern latitudes. However, this gene showed only a very weak association with soybean flowering in the Novosibirsk region compared to the Oryol region, suggesting the presence of another gene within the locus that influences flowering time. We therefore re-analyzed genes in the qDTF-7 locus and identified *GmRVE8c*, an Arabidopsis *RVE8* (*REVEILLE8*) ortholog, located ~21 kb upstream of *GmTOE1*; *RVE8* is a circadian clock component involved in plant development. After studying the natural variation of the *GmRVE8c* genes, we found four major haplotypes that arose due to three nonsynonymous substitutions and one 19-bp deletion leading to a frameshift. To identify three haplotypes, *GmRVE8c<sup>hap1,3,4</sup>*, which are predominant in improved soybean cultivars, we developed DNA markers. Using these markers, we genotyped 129 soybean accessions, the developmental time of which had been studied in the Novosibirsk and Oryol regions. Using our data and data from SoyOmics, we found the *GmRVE8c<sup>hap3</sup>* and *GmRVE8c<sup>hap4</sup>* haplotypes to be associated with late flowering and maturity in soybean. The early-maturing haplotype *GmRVE8c<sup>hap1</sup>* is predominant in cultivars from northern regions and is likely associated with the adaptation of soybean to high latitudes. The *GmRVE8c<sup>hap4</sup>* haplotype is in complete linkage with the early-maturing allele *GmTOE1<sup>C</sup>*, whereas the *GmRVE8c<sup>hap3</sup>* haplotype shows strong linkage with the late maturing allele *GmTOE1<sup>T</sup>*. Furthermore, the ANOVA results indicate an interaction between *GmRVE8c* and *E1*, the major regulator of flowering in soybean. This interaction is manifested as a stronger effect of the *GmRVE8c<sup>hap3,4</sup>* haplotypes on flowering and maturity in the genetic background of the *e1-as* allele compared with *E1*. Together, these findings define a complex and intriguing locus, which may serve as a possible example of a genetically heterogeneous locus.

**Key words:** soybean; flowering; *TOE1*; *RVE8*; genetic heterogeneity

**For citation:** Perfil'ev R.N., Shmatova M.I., Shcherban A.B., Salina E.A. Soybean locus qDTF-7 as an example of genetic heterogeneity associated with flowering and maturity time. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):43-52. doi 10.18699/vjgb-26-06

## Введение

Благодаря достижениям в методах массового фенотипирования и генотипирования растений и развитию математических моделей, позволяющих находить связи между фенотипом и генотипом, было открыто множество локусов количественных признаков у растений (Zhang et al., 2022). Продолжительность цветения и созревания сои не является исключением, и с помощью GWAS (genome-wide association studies) были обнаружены гены, контролирующие данный признак, например: *Tof5* (Dong et al., 2022), *Tof8* (Li H. et al., 2023a), *Tof13* (Li H. et al., 2023b), *Tof16* (Dong et al., 2021), *qFT13-3* (Li Y.F. et al., 2024), *GmAP1d* (Guo et al., 2024). Так, для локуса *Tof16* было установлено, что он кодирует ортолог арабидопсиса – ген *CCA1* (*CIRCADIAN CLOCK ASSOCIATED 1*), ключевой компонент циркадных ритмов и транскрипционный фактор с ДНК-связывающим доменом MYB (Dong et al., 2021). Рецессивные аллели *tof16*, у которых частично или полностью потеряна первоначальная функция кодируемого белка, задерживают цветение и созревание сои, что наиболее благоприятно для сортов, возделываемых в южных широтах (Dong et al., 2021).

В геноме сои было идентифицировано 54 *CCA1*-like белка (Bian et al., 2017), из них четыре попадают в одну филогенетическую кладу и показывают наибольшее сходство с генами арабидопсиса *RVE4* (*REVEILLE 4*) и *RVE8* (Bian et al., 2017; Shan et al., 2021; Bao et al., 2024). Потеря функции *RVE8* у арабидопсиса укорачивает время цветения в условиях короткого и длинного дня, тогда как сверхэкспрессия, наоборот, задерживает цветение (Rawat et al., 2011). Помимо уменьшения срока цветения, мутанты *rve4 6 8* и *rve3 4 5 6 8* арабидопсиса имеют ускоренное развитие и увеличенный размер клеток, что вызвано нарушением в экспрессии *PIF4* (*PHYTOCHROME*

*INTERACTING FACTOR 4*) и *PIF5* и сниженной протеасомной активностью (Gray et al., 2017; Scandola et al., 2022). Функциональная роль генов *RVE* в цветении и созревании у сои пока слабо охарактеризована (Shan et al., 2021). Недавно было установлено, что сверхэкспрессия *GmMYB133* (ген, наиболее близкий к *RVE5* арабидопсиса) и *GmRVE8a* в арабидопсисе приводит к задержке цветения (Shan et al., 2021; Bao et al., 2024). Интересно, что сверхэкспрессия *GmPIF4b* у сои ускоряет не только цветение, но и прохождение последующих фаз развития (Arya et al., 2021). У сои также недавно была предложена схема взаимодействия *GmMYB133-PRR5-PIF4*, поскольку *GmMYB133* в арабидопсисе может связываться с промотором *PRR5* (*PSEUDO-RESPONSE REGULATOR 5*), а *PRR5* – с *PIF4* (Shan et al., 2021).

В настоящее время лишь для относительно небольшой части локусов количественных признаков у растений были выявлены гены и мутации, которые влияют на признак. Есть много различных причин, мешающих точному картированию – определению гена и мутации в составе локуса, влияющей на формирование признака (Clauw et al., 2024). Одной из проблем является аллельная гетерогенность – ситуация, когда разные мутации в одном гене ведут к одинаковому или похожему фенотипу (Sasaki et al., 2021; Clauw et al., 2024; Liu H.-J. et al., 2024). Классический пример для растений – это множественные аллели у гена *FRI* (*FRIGIDA*) арабидопсиса, которые имеют доказанный эффект на экспрессию *FLC* (*FLOWERING LOCUS C*), но при этом не детектируются с помощью GWAS без добавления этих аллелей как ковариат в статистическую модель (Atwell et al., 2010). Другая, менее изученная проблема – генетическая (или локусная) гетерогенность – ситуация, когда аллели из разных генов в одном локусе влияют на один признак (Sasaki et al., 2021; Clauw et al., 2024;

Liu H.-J. et al., 2024). Примеров генетической гетерогенности известно значительно меньше. Недавно E. Sasaki с коллегами (2021) охарактеризовали подобный локус у арабидопсиса, показав, что ассоциация генов *AOP2/AOP3* (*ALKENYL HYDROXALKYL PRODUCING*) со временем цветения на самом деле может быть объяснена двумя соседними генами *NDX1* (*NODULIN HOMEODOMAIN GENES 1*) и *GAI* (*GIBBERELLIC ACID REQUIRING 1*). У сои тоже был найден подобный локус и показано, что устойчивость бобов к растрескиванию возникла в результате мутаций у двух близко расположенных генов, *Sh1* и *Pdh1* (Li S. et al., 2024).

В настоящей работе мы обсуждаем возможный пример генетически гетерогенного локуса у сои. Заново проанализировав гены в составе локуса qDTF-7, ассоциированного с цветением сои, мы обнаружили новый ген-кандидат *GmRVE8c*, который находится на расстоянии 21 т. п. н. от другого сильного гена-кандидата *GmTOE1*. У *GmRVE8c* были выявлены мутации, формирующие различные гаплотипы и ассоциированные с цветением и созреванием сои. Кроме того, эффект гаплотипов на время цветения и созревание *GmRVE8c* демонстрирует зависимость от аллельного состояния главного регулятора цветения сои – гена *E1*. Все это формирует достаточно интересный и комплексный локус, и наша работа может послужить теоретическим примером того, какие естественные мутации могут стоять за устанавливаемыми с помощью GWAS ассоциациями.

## Материалы и методы

**Растительный материал и фенотипы.** В качестве растительного материала мы использовали 129 сортообразцов сои (63 сорта и 66 селекционных линий), которые были описаны в наших предыдущих работах (Perfil'ev et al., 2023, 2024). Для этих сортообразцов в 2021 и 2022 гг. в условиях Новосибирска и Орла была изучена продолжительность от всходов до цветения (DTF, days from emergence to flowering) и созревания (DTM, days from emergence to maturity). Детали полевого эксперимента описаны в работе (Perfil'ev et al., 2023). В качестве фенотипов для данных 129 сортообразцов мы использовали значения BLUP (Best Linear Unbiased Prediction), которые были получены для каждого региона отдельно, как ранее (Perfil'ev et al., 2024).

Все доступные наблюдения по BBD (Beginning bloom date) и MD (Pod maturity date) были загружены из SoyOmics (<https://ngdc.cnbc.ac.cn/soyomics/index>) (Liu Y. et al., 2023). Для данных фенотипов были также рассчитаны значения BLUP с помощью функции “lmer” из пакета lme4 (Bates et

al., 2015) по формуле, где генотип, локация и год отмечены как случайный эффект. Полученные значения BLUP использовали в дальнейшем анализе.

**Характеристика локуса qDTF-7 и естественные мутации у гена *GmRVE8c*.** Список генов, расположенных в составе локуса qDTF-7, был взят из базы SoyBase (<https://www.soybase.org/>).

Информация о гаплотипах у гена *GmRVE8c/SoyZH13\_03G159500* в пределах кодирующей области была получена через модуль “NapSnap” с сайта SoyOmics. Настройки в “NapSnap” по умолчанию, за исключением параметров в “Variation type”, который не включал варианты “Synonymous” и “Unclassed”.

Для поиска основных белковых доменов у *GmRVE8c/Glyma.03G177300* использовали базу данных ModIDB (<https://mobidb.org/>) (Piovesan et al., 2025).

Колокализация qDTF-7 с ранее обнаруженными QTL была изучена с помощью модуля “BreedingTips” с сайта SoyOmics.

**Выделение ДНК, разработка ДНК-маркеров, генотипирование.** Геномную ДНК из 3–4-дневных проростков, выращенных в чашках Петри, выделяли с применением СТАВ буфера (Rogers, Bendich, 1985).

Для разработки ДНК-маркеров к гену *GmRVE8c/Glyma.03G177300* использовали нуклеотидную последовательность, полученную из базы данных Ensembl Plants (<https://plants.ensembl.org/index.html>). Подбор праймеров осуществляли в программе UGENE V 48.0 (Okonechnikov et al., 2012) (<http://ugene.net/ru/>). Нуклеотидные последовательности праймеров приведены в таблице.

Для ПЦР использовали БиоМастер HS-Тaq ПЦР-Color (2×) («Биолабмикс», Новосибирск, Россия). Смесь общим объемом 25 мкл содержала 10 mM Трис-НСl, pH 8.5, 50 mM KCl, 0.1 % Твин 20, 2 mM MgCl<sub>2</sub>, 0.25 mM каждого праймера, 50–100 нг ДНК и 1 е. а. Тaq ДНК-полимеразы. Условия ПЦР: 95 °C в течение 5 мин; далее 34 цикла: 95 °C 15 с, 55 °C 10 с, 72 °C 20 с; 72 °C в течение 1 мин.

Гидролиз ПЦР продуктов рестриктазой Mhl I (SibEnzyme, Новосибирск, Россия) проводили в реакционной смеси общим объемом 20 мкл, которая содержала: 8 мкл ПЦР продукта; 2 мкл 10x буфера, рекомендованного производителем; 1 е. а. фермента; 10 мкл ddH<sub>2</sub>O. Продукты рестрикции и амплификации разделяли в 2–2.5 % агарозном геле с этидиум бромидом. Результаты электрофореза фотографировали с помощью Gel Doc™ XR+ (BioRad, США). В качестве маркера молекулярных весов ДНК использовали “Step100+50” («Биолабмикс», Россия).

ДНК-маркеры, использованные для определения основных гаплотипов *GmRVE8c* (*Glyma.03G177300/SoyZH13\_03G159500*)

Ген	5'–3' направление	Последовательности праймеров	Гаплотипы	Ожидаемые продукты амплификации/рестрикции, п. н.	Рестрикционный фермент
<i>GmRVE8c</i>	F	AGATAGGATATGATACCATATATTG	hap1–3	173	–
	R	CGGACCAGCTCTCCCTTGAC	hap4	154	
<i>GmRVE8c</i>	F	AAGAACGGGGCAGTAGCACACG	hap2, 3	251	Mhl I
	R	TTCATCCCATGTGACATACCCA	hap1, 4	221+30	

Генотипы по гену *E1* и *GmTOE1* для 129 сортообразцов сои были установлены и опубликованы ранее (Perfil'ev et al., 2023, 2024). Для улучшенных сортов из базы данных SoyOmics генотипы по мутациям soy8699425 для гена *E1/SoyZH13\_06G195900* и soy4989324 для гена *GmTOE1/SoyZH13\_03G159700* были получены через модуль “HarSnap”, а для мутации soy8699425 дополнительно перекодированы как C → *e1-as*, G → *E1* аллель.

**Статистический анализ.** Ассоциация гаплотипов *GmRVE8c* с продолжительностью цветения и созревания была установлена с помощью базовой функции “aov” из R версии 4.4.2 (<https://www.r-project.org/>). Множественное сравнение гаплотипов проводилось с использованием функции “TukeyHSD”, построение диаграмм размаха и тепловых карт с результатами множественного сравнения – с использованием пакетов “ggplot2” (Wickham, 2016) и “ggcorplot” (Kassambara, 2023).

## Результаты

### Повторное исследование локуса qDTF-7 выявило новый ген-кандидат *GmRVE8c*

Ранее с помощью мультилокусной модели FarmCPU на хромосоме 3 сои мы обнаружили QTN (*Gm03\_40959110\_A\_G*), который был ассоциирован с цветением в условиях Новосибирска (Perfil'ev et al., 2024). Установлено, что данный QTN находится в LD блоке размером 553 т.п.н., этот участок был обозначен как локус qDTF-7 (QTL days from emergence to flowering 7) (Perfil'ev et al., 2024). В составе локуса находится 71 ген (табл. S1 Приложения)<sup>1</sup>, и в качестве наиболее вероятного гена-кандидата мы первоначально определили *Glyma.03G177500* (далее *GmTOE1*) (Perfil'ev et al., 2024). Однако данный ген демонстрировал очень слабую ассоциацию с цветением сои в Новосибирске по сравнению с Орлом (Perfil'ev et al., 2024). Поэтому мы решили заново более внимательно проанализировать гены в составе локуса и в результате обнаружили *Glyma.03G177300*. Этот ген кодирует один из четырех гомологов сои, которые наиболее родственны генам арабидопсиса *RVE8* (*REVEILLE 8*) и *RVE4* (Bao et al., 2024). В дальнейшем в рамках настоящей статьи ген *Glyma.03G177300* будет обозначен как *GmRVE8c*.

У арабидопсиса ген *RVE8* – известный компонент циркадных ритмов, который в том числе задерживает цветение (Rawat et al., 2011). Сверхэкспрессия в арабидопсисе одного из ортологов *RVE8* (*GmRVE8a*) приводит к задержке цветения (Bao et al., 2024). Все это делает *GmRVE8c* потенциальным геном-кандидатом в качестве регулятора цветения и созревания у сои.

### Ассоциация гаплотипов *GmRVE8c* с цветением и созреванием сои

С помощью базы данных SoyOmics мы изучили естественное нуклеотидное разнообразие в гене *GmRVE8c* в пределах его кодирующей области. В результате были обнаружены четыре гаплотипа, которые возникли из-за

одной делеции, приводящей к сдвигу рамки считывания, и трех несинонимичных замен (табл. S2). Данная делеция должна вызывать сдвиг рамки считывания после 4-й аминокислоты, что должно приводить к полной потере первоначальной функции у гена (см. табл. S2). Все выявленные несинонимичные замены находятся за пределами предсказанных ДНК-связывающих доменов MYB (рис. S1). Сложно сказать, оказывают ли эти замены влияние на структуру и функцию кодируемого белка.

Для трех наиболее распространенных гаплотипов мы изучили ассоциацию с BLUP\_BBD и BLUP\_MD у улучшенных сортов сои. По результатам анализа ассоциаций для BLUP\_BBD, возделываемые сорта с *GmRVE8c<sup>hap3</sup>* зацветают достоверно позже в сравнении с *GmRVE8c<sup>hap1</sup>* и *GmRVE8c<sup>hap4</sup>* (рис. 1, а). По результатам анализа ассоциаций для BLUP\_MD, сортообразцы с *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* созревают достоверно позже в сравнении с *GmRVE8c<sup>hap1</sup>* (см. рис. 1, а).

С целью идентификации этих трех гаплотипов мы разработали ДНК-маркеры для мутаций soy32494699 и soy4988787 (см. таблицу, рис. S2) и с их помощью генотипировали 129 сортообразцов сои из нашей коллекции, у которых локус qDTF-7 был первоначально идентифицирован. В результате у 109 сортообразцов был установлен *GmRVE8c<sup>hap1</sup>*, тогда как остальные 12 и 8 образцов несли гаплотипы *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* соответственно (табл. S3). Также *GmRVE8c* показывает достоверную ассоциацию с BLUP\_DTF и BLUP\_MD в условиях Новосибирской и Орловской областей, и сортообразцы, содержащие гаплотип *GmRVE8c<sup>hap4</sup>*, цветут и созревают значительно позже (см. рис. 1, б, в).

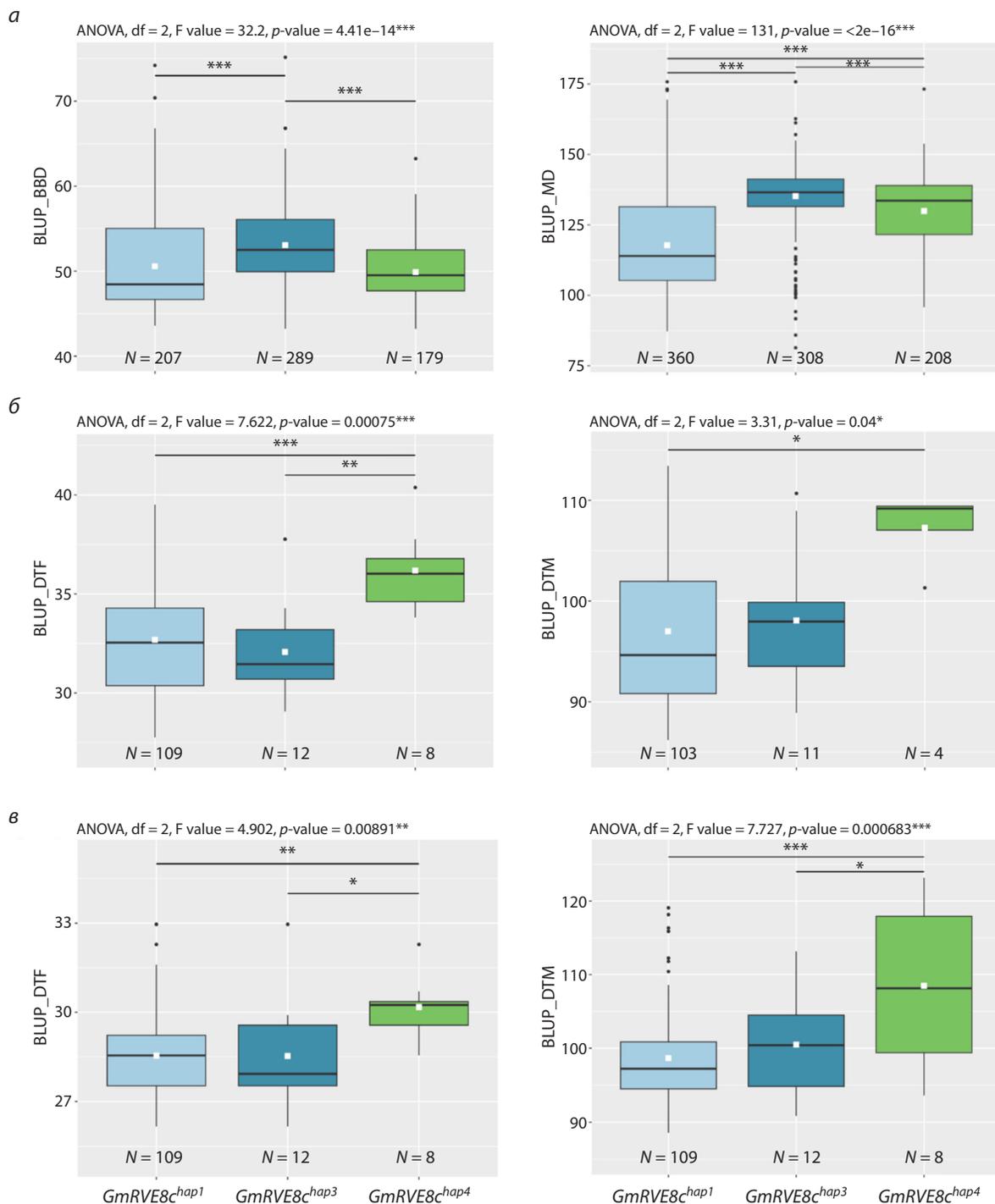
У сортов Грация и ВНИИС-1 мы ранее обнаружили аллель *GmTOE1<sup>T</sup>* (Perfil'ev et al., 2024), но в результате генотипирования по *GmRVE8c* у них был выявлен гаплотип *GmRVE8c<sup>hap4</sup>*, что необычно, поскольку *GmRVE8c<sup>hap4</sup>* сцеплен с аллелем *GmTOE1<sup>C</sup>*. Следует учитывать, что ген *GmRVE8c* физически очень близко расположен к *GmTOE1*. Это делает вероятность наличия кроссоверных генотипов *GmRVE8c<sup>hap4</sup> GmTOE1<sup>T</sup>* крайне низкой. К примеру, в SoyOmics мы не нашли ни одного такого генотипа. Поэтому мы заново выделили ДНК из данных сортообразцов, генотипировали их с помощью ранее разработанного ДНК-маркера на аллель *GmTOE1<sup>C</sup>* (Perfil'ev et al., 2024) и обнаружили у них аллель *GmTOE1<sup>C</sup>* (см. табл. S3).

### Ген *GmRVE8c* показывает эпистатическое взаимодействие с геном *E1*, а также расположен рядом с другим регулятором цветения *GmTOE1*

У сои гены *GmCCA1s* (*CIRCADIAN CLOCK ASSOCIATED 1*) активируют цветение через подавление экспрессии основного репрессора цветения – гена *E1* (Dong et al., 2021). Гены *RVE* относятся к семейству транскрипционных факторов с CCA1-like MYB доменами, поэтому мы предположили, что функция *GmRVE8c* в контроле цветения также может зависеть от гена *E1*.

Расстояние между *GmRVE8c* и *GmTOE1* составляет ~21 т.п.н., и *GmTOE1* практически следующий после *GmRVE8c* ген в 5'–3' направлении (см. табл. S1). У *GmTOE1*

<sup>1</sup> Табл. S1–S6 и рис. S1–S6 Приложения доступны по адресу: <https://vavilovj-icg.ru/download/pict-2026-30/appx3.zip>



**Рис. 1.** Ассоциация гаплотипов *GmRVE8c* с продолжительностью цветения и созревания.

*a* – ассоциация гаплотипов *GmRVE8c* с BLUP\_BBD и BLUP\_MD; *б* – ассоциация гаплотипов *GmRVE8c* с BLUP\_DTF и BLUP\_DTM в Новосибирске; *в* – ассоциация гаплотипов *GmRVE8c* с BLUP\_DTF и BLUP\_DTM в Орле. Белые квадраты на диаграммах – среднее значение. Значимые различия между сравниваемыми генотипами: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

есть естественные мутации, которые ассоциированы с продолжительностью цветения и созревания сои (Perfil'ev et al., 2024). Ранее мы обнаружили четыре основных гаплотипа, включая два раннеспелых и два позднеспелых (далее обозначены как аллель *GmTOE1C* и *GmTOE1T* соответственно), которые различаются между собой ну-

клеотидной заменой Т316С, приводящей к замене серина на пролин в 106-й аминокислоте (Perfil'ev et al., 2024).

Мы изучили ассоциацию гаплотипов *GmRVE8c* с учетом генетического фона *E1* и *GmTOE1* с помощью данных из SoyOmics и из нашего эксперимента (рис. S3–S6). Согласно результату ANOVA, между *GmRVE8c* и *E1* есть

достоверное взаимодействие на основании данных из SoyOmics (табл. S4, S5). Это косвенно указывает на наличие эпистатического взаимодействия между *GmRVE8c* и *E1*. Анализ установленных ассоциаций для BLUP\_BBD и BLUP\_MD показывает, что на фоне аллеля *e1-as* гаплотипы *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* имеют более сильный эффект на время цветения и созревания (рис. S3, a, b). В нашей коллекции растения с гаплотипом *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* несут аллель *e1-as* или *e1-nl*, и растения с *GmRVE8c<sup>hap4</sup>* демонстрируют позднее цветение и созревание относительно остальных генотипов (см. рис. S3, c, d).

По BLUP\_BBD растения с *GmRVE8c<sup>hap1</sup> GmTOE1<sup>T</sup>* и *GmRVE8c<sup>hap3</sup> GmTOE1<sup>T</sup>* зацветают достоверно позже в сравнении с *GmRVE8c<sup>hap4</sup> GmTOE1<sup>C</sup>* и *GmRVE8c<sup>hap1</sup> GmTOE1<sup>C</sup>* (рис. S4, a). По BLUP\_MD наблюдается похожая ситуация, за исключением генотипа *GmRVE8c<sup>hap4</sup> GmTOE1<sup>C</sup>*, который созревает позже в сравнении с *GmRVE8c<sup>hap1</sup> GmTOE1<sup>C</sup>* (см. рис. S4, a). По данным из нашей коллекции растения с *GmRVE8c<sup>hap4</sup> GmTOE1<sup>C</sup>* зацветают и созревают достоверно позже в сравнении с *GmRVE8c<sup>hap1</sup> GmTOE1<sup>C</sup>* в обоих регионах (см. рис. S4, b, c). По BLUP\_MD генотипы *GmRVE8c<sup>hap1</sup> GmTOE1<sup>C</sup>* созревают достоверно раньше в сравнении с *GmRVE8c<sup>hap1</sup> GmTOE1<sup>T</sup>* (см. рис. S4, a). По данным нашей коллекции генотипы *GmRVE8c<sup>hap1</sup> GmTOE1<sup>C</sup>* зацветают достоверно раньше в сравнении с *GmRVE8c<sup>hap1</sup> GmTOE1<sup>T</sup>* только в условиях Орла (см. рис. S4, b). Аллель *GmTOE1<sup>C</sup>* ассоциирован с ранним цветением и созреванием, как описано нами ранее, а *GmRVE8c<sup>hap4</sup>*, по-видимому, частично восстанавливает раннеспелый фенотип растений с *GmTOE1<sup>C</sup>*, делая его позднеспелым (см. рис. S4, a).

С учетом сразу трех генов для интерпретации результатов мы рассматривали генотипы, которые отличаются между собой по одному гену. По BLUP\_BBD и BLUP\_MD растения с генотипом *GmRVE8c<sup>hap1</sup> e1-as GmTOE1<sup>C</sup>* зацветают и созревают позже в сравнении с *GmRVE8c<sup>hap4</sup> e1-as GmTOE1<sup>C</sup>* (рис. S5). По BLUP\_BBD между генотипами *GmRVE8c<sup>hap1</sup> E1 GmTOE1<sup>C</sup>* и *GmRVE8c<sup>hap1</sup> E1 GmTOE1<sup>T</sup>* и генотипами *GmRVE8c<sup>hap1</sup> e1-as GmTOE1<sup>C</sup>* и *GmRVE8c<sup>hap1</sup> e1-as GmTOE1<sup>T</sup>* нет достоверных различий (см. рис. S5, a). По BLUP\_MD из этих двух генотипов достоверно различаются между собой только *GmRVE8c<sup>hap1</sup> E1 GmTOE1<sup>C</sup>* и *GmRVE8c<sup>hap1</sup> E1 GmTOE1<sup>T</sup>* (см. рис. S5, b). В нашей коллекции растения с *GmRVE8c<sup>hap1</sup> e1-nl GmTOE1<sup>C</sup>* зацветают достоверно раньше в сравнении с *GmRVE8c<sup>hap1</sup> e1-nl GmTOE1<sup>C</sup>* в условиях Орла (рис. S6, c). Также генотипы с *GmRVE8c<sup>hap4</sup> e1-as GmTOE1<sup>C</sup>* зацветают и созревают позже в сравнении с остальными генотипами (см. рис. S6). В целом этот результат согласуется с полученным ранее для двух генов.

Между генотипами *GmRVE8c<sup>hap1</sup> E1 GmTOE1<sup>C</sup>* и *GmRVE8c<sup>hap4</sup> E1 GmTOE1<sup>C</sup>* есть также достоверные различия по BLUP\_MD и BLUP\_BBD. Растения с *GmRVE8c<sup>hap4</sup>* созревают позже, но зацветают немного раньше (см. рис. S5). На фоне аллеля *e1-as* у гаплотипа *GmRVE8c<sup>hap4</sup>* наблюдается более сильный эффект на BLUP\_BBD (см. рис. S5, a). Для гаплотипа *GmRVE8c<sup>hap3</sup>* отмечается похожая ситуация, и между генотипами *GmRVE8c<sup>hap3</sup> E1*

*GmTOE1<sup>T</sup>* и *GmRVE8c<sup>hap1</sup> E1 GmTOE1<sup>T</sup>* нет достоверных различий по BLUP\_BBD и BLUP\_MD (см. рис. S5). Хотя между генотипами *GmRVE8c<sup>hap3</sup> e1-as GmTOE1<sup>T</sup>* и *GmRVE8c<sup>hap1</sup> e1-as GmTOE1<sup>T</sup>* достоверные различия по BLUP\_MD существуют (см. рис. S5, b). В нашей коллекции растения с гаплотипом *GmRVE8c<sup>hap4</sup>* даже на фоне нуль-аллеля *e1-nl* демонстрируют позднее цветение и созревание (рис. S5).

### Распространение гаплотипов *GmRVE8c* у сортообразцов сои различного происхождения

Для гена *GmRVE8c* мы изучили частоту гаплотипов у сортов сои из Китая в трех основных регионах возделывания сои (China I, China II, China III) и одном смешанном (China IV–VI), а также в нашей коллекции в трех основных группах: А – Западная Сибирь (Новосибирск и Омск); В – другие регионы России; С – другие страны (рис. 2).

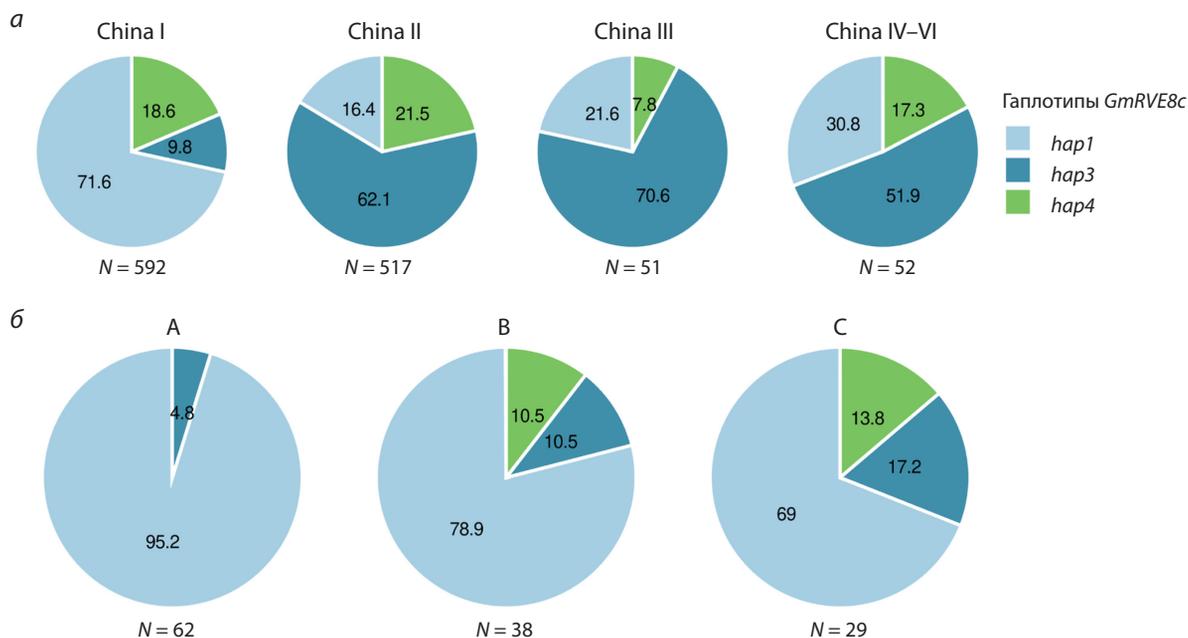
Гаплотип *GmRVE8c<sup>hap1</sup>* преобладает в сортах из северного Китая, тогда как в остальных регионах Китая доминирует *GmRVE8c<sup>hap3</sup>* (см. рис. 2, a). В нашей коллекции сортообразцы из Западной Сибири преимущественно несут гаплотип *GmRVE8c<sup>hap1</sup>*, позднеспелый гаплотип *GmRVE8c<sup>hap4</sup>* представлен только в сортообразцах из других регионов России и из других стран (см. рис. 2, б). Это показывает роль гаплотипов *GmRVE8c* в адаптации сои к различным регионам возделывания.

### Обсуждение

#### *GmRVE8c* – новый ген-кандидат в роли регулятора цветения и созревания сои

В качестве нового ген-кандидата для ранее обнаруженного локуса qDTF-7, ассоциированного с цветением сои (Perfil'ev et al., 2024), мы определили ген *Glyma.03G177300/GmRVE8c*. У *GmRVE8c* было выявлено четыре основных гаплотипа, которые возникли в результате трех однонуклеотидных замен и делеции 19 п. н. (см. табл. S2). Среди найденных у *GmRVE8c* мутаций замена soy32494699 (делеция 19 п. н.) у гаплотипа *GmRVE8c<sup>hap4</sup>*, возможно, ведет к полной потере первоначальной функции кодируемого белка из-за сдвига рамки считывания после 4-й аминокислоты (см. табл. S2). Другие три гаплотипа, *GmRVE8c<sup>hap1-3</sup>*, возникли из-за трех несинонимичных замен (см. табл. S2). И хотя все три замены находятся за пределами предсказанных ДНК-связывающих доменов (см. рис. S1), тем не менее они потенциально могут влиять на 3D структуру белка или на взаимодействие с другими белками. Так, две мутации находятся в предсказанных LIP (linear interacting peptides) регионах белка (см. рис. S1). Полная потеря функции *GmRVE8c*, по-видимому, ведет к задержке развития сои, поскольку растения с *GmRVE8c<sup>hap4</sup>* демонстрируют позднее цветение и созревание (см. рис. 1). Возможно, у гаплотипа *GmRVE8c<sup>hap3</sup>* тоже утрачена первоначальная функция белка, поскольку, согласно данным SoyOmics, образцы с этим гаплотипом также демонстрируют позднее цветение и созревание (см. рис. 1, a).

Частота гаплотипов *GmRVE8c* изменяется от дикой к культурной сое. В сравнении с дикой соей у местных и



**Рис. 2.** Распределение гаплотипов *GmRVE8c*: а – в улучшенных сортах сои из Китая (China I – северные провинции, China II, III, IV-VI – южные провинции); б – в сортообразцах сои из Западной Сибири (Новосибирская и Омская области) (А), из других регионов России (В), из других стран (С).

Числа на круговых диаграммах приведены в процентах; N – количество образцов.

возделываемых сортов почти не встречается гаплотип *GmRVE8c<sup>hap2</sup>* (см. табл. S2). В популяции дикой сои не обнаружен гаплотип *GmRVE8c<sup>hap4</sup>*, вероятно, он возник и был отобран после доместикации (см. табл. S2). Распределение гаплотипов *GmRVE8c* в различных регионах возделывания показывает, что раннеспелый гаплотип *GmRVE8c<sup>hap1</sup>* преобладает в сортах из северного Китая, тогда как в остальных регионах Китая доминирует *GmRVE8c<sup>hap3</sup>* (см. рис. 2, а). В нашей коллекции почти все сортообразцы из Западной Сибири тоже несут гаплотип *GmRVE8c<sup>hap1</sup>* (см. рис. 2, б). Это показывает роль гаплотипов *GmRVE8c* в адаптации к различным регионам возделывания сои. Из-за позднего цветения и созревания гаплотип *GmRVE8c<sup>hap4</sup>* является неблагоприятным аллелем для северных широт, однако для южных широт он, наоборот, благоприятный. Мы полагаем, что *GmRVE8c<sup>hap4</sup>* может найти применение в селекции сортов сои, пригодных к возделыванию в южных широтах.

### Ген *GmRVE8c* показывает эпистатическое взаимодействие с геном *E1*

Гены семейства *RVE* кодируют транскрипционные факторы с CCA1-like MYB доменами и близкородственны генам *CCA1* (*CIRCADIAN CLOCK ASSOCIATED 1*) и *LHY* (*LATE ELONGATED HYPOCOTYL*) (Rawat et al., 2009). У сои гены *CCA1/LHY* – известные регуляторы цветения сои, в условиях короткого дня *GmLHY* напрямую связываются с промотором *E1* и подавляют его экспрессию, что ускоряет цветение сои, а потеря функции *GmLHY* задерживает ее цветение (Dong et al., 2021). На основании этого мы предположили, что функция *GmRVE8c* в контроле цветения

также может зависеть от гена *E1*. Результаты ANOVA на данных SoyOmics демонстрирует наличие достоверного взаимодействия между *E1* и *GmRVE8c*, поэтому мы предполагаем, что *GmRVE8c* тоже является репрессором гена *E1* (см. табл. S3, S4). Анализ ассоциаций этих данных показывает, что гаплотипы *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* гораздо сильнее задерживают цветение и созревание на генетическом фоне полупункционального аллеля *e1-as* (см. рис. S4, а, б). Это довольно неожиданно, поскольку, если предполагать, что *GmRVE8c* может подавлять экспрессию гена *E1*, то на фоне слабо функционального аллеля *e1-as* эффект от потери функции *GmRVE8c* не должен приводить к изменению фенотипа или, по крайней мере, должен быть слабее. Так, эффект на время цветения локусов/генов *Tof16/LHYa*, *Tof8/FKF1b* и *J/ELF3* на фоне аллеля *e1-as* полностью пропал или становился меньше (Lu S. et al., 2017; Dong et al., 2021; Li H. et al., 2023a). Анализ фенотипов в нашей коллекции показывает эффект *GmRVE8c<sup>hap4</sup>* на цветение и созревание даже на фоне *e1-nl*, аллеля с делецией всего гена *E1* (см. рис. S3, с, d). Возможно, в наших условиях *GmRVE8c* может участвовать в транскрипционном контроле всего семейства генов *E1* (*E1*, *E1La*, *E1Lb*), или *GmRVE8c* регулирует время цветения через независимый от гена *E1* механизм.

У арабидопсиса белки LNK1 (NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED 1) и LNK2 рекрутируют RVE4 и RVE8, чтобы активировать транскрипцию генов *TOC1* (*TIMING OF CAB EXPRESSION 1*) и *PRR5* (*PSEUDO-RESPONSE REGULATOR 5*), и потеря функции генов *LNK1-4* приводит к задержке цветения в условиях длинного дня (Xie et al., 2014; De Leone et al., 2018). Но-

каут четырех ортологов *GmLNK2* у сои ускоряет цветение в условиях длинного дня из-за уменьшения экспрессии гена *E1* (Li Z. et al., 2021). Характерная особенность генов *LNK* – отсутствие собственных ДНК-связывающих доменов, для чего они рекрутируют *CCA1*, *LHY*, *RVE4* и *RVE8* и переключают их с репрессоров на активаторов транскрипции (Xie et al., 2014). Дальнейшие исследования покажут, зависит ли функция *GmRVE8c* в контроле цветения от семейства *E1* и рекрутируют ли *GmLNK2s* белки *GmRVE8s* для контроля транскрипции гена *E1*.

### Локус qDTF-7 как возможный пример гетерогенного локуса у сои

Локусная, или генетическая, гетерогенность – ситуация, когда аллели из разных генов в одном локусе влияют на один признак. Для локуса qDTF-7 мы ранее предложили достаточно сильный ген-кандидат *GmTOE1* (Perfil'ev et al., 2024), у которого ранее была показана роль в контроле цветения у сои (Wang T. et al., 2016). В рамках настоящей работы на ~21 т. п. н. выше *GmTOE1* мы обнаружили новый ген-кандидат *GmRVE8c*, который потенциально тоже может быть вовлечен в контроль цветения и созревания сои. Оба гена имеют нуклеотидные замены, ассоциированные с этими признаками, что делает qDTF-7 возможным примером генетически гетерогенного локуса у сои. С помощью доступных нам данных мы попытались проанализировать, как мутации в генах *GmRVE8c* и *GmTOE1* соотносятся между собой. На фоне аллеля *GmTOE1<sup>C</sup>*, ассоциированного с ранним цветением и созреванием сои, возник позднеспелый гаплотип *GmRVE8c<sup>hap4</sup>*. В базе данных SoyOmics и в нашей коллекции мы не нашли ни одного генотипа *GmRVE8c<sup>hap4</sup> GmTOE1<sup>T</sup>* (см. рис. S4). При этом, судя по фенотипам из SoyOmics, эффект от гаплотипа *GmRVE8c<sup>hap4</sup>* может частично компенсировать эффект от раннеспелого аллеля *GmTOE1<sup>C</sup>* на время цветения и созревания (см. рис. S4, а, рис. S5).

В нашей коллекции наблюдается похожая ситуация, но у гаплотипа *GmRVE8c<sup>hap4</sup>* эффект на цветение и созревание гораздо сильнее, чем у *GmTOE1<sup>C</sup>* и *GmTOE1<sup>T</sup>* (см. рис. S4, b, c, рис. S6, a, b). Аллели *GmTOE1* имеют минорный эффект на цветение и созревание в наших условиях. Так, генотип *GmRVE8c<sup>hap1</sup> GmTOE1<sup>C</sup>* достоверно отличается от *GmRVE8c<sup>hap1</sup> GmTOE1<sup>T</sup>* только по BLUP\_DTF в условиях Орла (см. рис. S4, b, c, рис. S6). Интересно, что другой гаплотип – *GmRVE8c<sup>hap3</sup>* – значительно преобладает на фоне аллеля *GmTOE1<sup>T</sup>*; к примеру, в базе данных SoyOmics и в нашей коллекции мы обнаружили только два сортообразца с генотипом *GmRVE8c<sup>hap3</sup> GmTOE1<sup>C</sup>* (см. рис. S4). Поэтому сложно сказать, какой из генов – *GmRVE8c<sup>hap3</sup>* или *GmTOE1<sup>T</sup>* – задерживает цветение и созревание, хотя по данным SoyOmics оба гена оказывают эффект на данные признаки (см. рис. S4, а, рис. S5). Однако анализ фенотипов в нашей коллекции не выявил какого-либо достоверного эффекта *GmRVE8c<sup>hap3</sup>* на цветение и созревание (см. рис. S4, b, c, рис. S6, a, b).

Следует сказать, что обнаруженные мутации в генах *GmRVE8c* и *GmTOE1* требуют функциональной валидации для подтверждения того, что они действительно влияют на

фенотип. Если это окажется действительно так, то локус qDTF-7 может стать интересным примером того, каким образом естественные мутации формируют многообразие фенотипов у растений, и в частности у сои.

### Влияние гена *GmRVE8c* на другие признаки сои

Раннеспелый аллель *GmTOE1<sup>C</sup>*, вероятно, связан с адаптацией сои к северным широтам из-за сокращения времени цветения (Perfil'ev et al., 2024). Гаплотип *GmRVE8c<sup>hap4</sup>*, напротив, неблагоприятен для северных широт из-за позднего цветения и созревания; в нашей коллекции все сортообразцы из Западной Сибири несут раннеспелый гаплотип *GmRVE8c<sup>hap1</sup>* (см. рис. 2). Как говорилось выше, *GmRVE8c<sup>hap4</sup>*, по-видимому, восстанавливает раннеспелый фенотип *GmTOE1<sup>C</sup>*, хотя возникает вопрос: почему на фоне раннеспелого аллеля *GmTOE1<sup>C</sup>* возникла мутация *GmRVE8c<sup>hap4</sup>*? Возможное объяснение этого факта – плейотропный эффект *GmRVE8c* на другие признаки сои. Интересно, что на хромосоме 3 с помощью двуродительского QTL-картирования было обнаружено несколько локусов, ассоциированных с засухоустойчивостью сои (табл. S6). В границы данных локусов попадают и гены *GmRVE8c* и *GmTOE1*. Однако надо отметить, что в составе этих локусов находится также известный регулятор засухоустойчивости *GmSALT3* (Shi et al., 2018; Guan et al., 2021).

Сверхэкспрессия *GmRVE8a* в арабидопсисе повышает его засухо- и солеустойчивость, и *GmRVE8a* может взаимодействовать с белком GmNAC019 (Bao et al., 2024), который связан с АВА-опосредованной реакцией на засуху (Hoang et al., 2019). Гены *GmLHY* тоже участвуют в АВА-опосредованной реакции сои на засуху, нокаут четырех *GmLHY* увеличивает засухоустойчивость сои (Wang K. et al., 2021). В работе (Li C. et al., 2024) показано, что компонент циркадных ритмов сои *GmPRR3b* напрямую подавляет экспрессию *GmABF3* (*ABSCISIC ACID-RESPONSIVE ELEMENT-BINDING FACTOR 3*), сверхэкспрессия которого улучшает засухоустойчивость сои. Мы предполагаем, что потеря функции *GmRVE8c* также может положительно влиять на засухоустойчивость сои. Интересно, что у арабидопсиса для генов *TOE1* и *TOE2* недавно была показана ключевая роль в избегании засухи; потеря функции *TOE1/2* ускоряет цветение, но снижает устойчивость растений к засухе (Chen et al., 2025). Для раннеспелого аллеля *GmTOE1<sup>C</sup>* мы предполагаем по крайней мере частичную потерю функции кодируемого гена (Perfil'ev et al., 2024). Возможно, гаплотип *GmRVE8c<sup>hap4</sup>* возник, чтобы восстановить засухоустойчивость у растений с *GmTOE1<sup>C</sup>*. На наш взгляд, это интересная гипотеза, которая требует дальнейшего экспериментального подтверждения.

Согласно данным табл. S6, ген *GmRVE8c* колокализуется с другими локусами, которые ассоциированы с высотой, массой растений, содержанием масла и изофлавонов, толщиной семян, устойчивостью к соевой цистообразующей нематоде. Теоретически *GmRVE8c* может влиять на некоторые из этих признаков, в частности на высоту и массу растений через регуляцию *PRR5-PIF4* модуля, как было предложено для *GmMYB133* (Shan et al., 2021). Также *GmMYB133* является положительным регулятором

биосинтеза изофлавонов у сои (Bian et al., 2018). У арабидопсиса мутанты *rve 4 6 8* демонстрируют измененный метаболизм углеводов, жирных кислот и липидов и сниженную активность протеасом, что и вызывает увеличение размера клеток (Scandola et al., 2022). Для различных генов циркадных ритмов изучается и показана роль в защите растений от патогенов (Lu H. et al., 2017). В частности, у сои недавно продемонстрирована роль *GmCCA1-1* в ответе на заражение нематодой *Heterodera glycines* (Niraula et al., 2022). Все это делает *GmRVE8c* интересным геном для дальнейшего изучения в контексте не только цветения и созревания сои, но и других сельскохозяйственно ценных признаков.

## Заключение

В настоящей работе мы обнаружили новый ген-кандидат *GmRVE8c*, гаплотипы *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* которого ассоциированы с поздним цветением и созреванием сои. Для определения данных гаплотипов мы разработали ДНК-маркеры, которые могут быть использованы в маркер-ориентированной селекции сои. Ген *GmRVE8c* расположен очень близко к другому регулятору цветения сои, гену *GmTOE1*, который тоже имеет мутации, ассоциированные с продолжительностью цветения и созревания. Предположительно *GmTOE1* и *GmRVE8c* вместе образуют генетически гетерогенный локус, т.е. локус, в котором аллели из разных близко расположенных генов могут влиять на один и тот же признак. Влияние гаплотипов *GmRVE8c<sup>hap3</sup>* и *GmRVE8c<sup>hap4</sup>* на время цветения и созревания свидетельствует также о частичной зависимости от аллельного состояния гена *E1*. Все это делает *qDTF-7* очень комплексным и интересным локусом, и в дальнейшем мы планируем валидировать мутации у *GmRVE8c* и *GmTOE1* с помощью гибридных популяций сои.

## Список литературы / References

- Arya H., Singh M.B., Bhalla P.L. Overexpression of *PIF4* affects plant morphology and accelerates reproductive phase transitions in soybean. *Food Energy Secur.* 2021;10(3):e291. doi 10.1002/fes3.291
- Atwell S., Huang Y.S., Vilhjálmsson B.J., Willems G., Horton M., Li Y., Meng D., ... Weigel D., Marjoram P., Borevitz J.O., Bergelson J., Nordborg M. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature.* 2010;465(7298):627-631. doi 10.1038/nature08800
- Bao G., Sun G., Wang J., Shi T., Xu X., Zhai L., Bian S., Li X. Soybean *RVE8a* confers salt and drought tolerance in *Arabidopsis*. *Biochem Biophys Res Commun.* 2024;704:149660. doi 10.1016/j.bbrc.2024.149660
- Bates D., Mächler M., Bolker B., Walker S. Fitting linear mixed-effects models using lme4. *J Stat Software.* 2015;67(1):1-48. doi 10.18637/jss.v067.i01
- Bian S., Jin D., Li R., Xie X., Gao G., Sun W., Li Y., Zhai L., Li X. Genome-wide analysis of CCA1-like proteins in soybean and functional characterization of *GmMYB138a*. *Int J Mol Sci.* 2017;18(10):2040. doi 10.3390/ijms18102040
- Bian S., Li R., Xia S., Liu Y., Jin D., Xie X., Dhaubhadel S., Zhai L., Wang J., Li X. Soybean *CCA1-like MYB* transcription factor *GmMYB133* modulates isoflavonoid biosynthesis. *Biochem Biophys Res Commun.* 2018;507(1-4):324-329. doi 10.1016/j.bbrc.2018.11.033
- Chen W., Wang T., Li X., Feng J., Liu Q., Xu Z., You Q., Yang L., Liu L., Chen S., Yue Z., Wang H., Yu D. *Arabidopsis* RGLG1/2 regulate flowering time under different soil moisture conditions by affecting the protein stability of TOE1/2. *New Phytol.* 2025;246(4):1609-1626. doi 10.1111/nph.70073
- Clauw P., Ellis T.J., Liu H.-J., Sasaki E. Beyond the standard GWAS – a guide for plant biologists. *Plant Cell Physiol.* 2025;66(4):431-443. doi 10.1093/pcp/pcae079
- De Leone M.J., Hernando C.E., Romanowski A., Garcia-Hourquet M., Careno D., Casal J., Rugnone M., Mora-García S., Yanovsky M.J. The *LNK* gene family: at the crossroad between light signaling and the circadian clock. *Genes.* 2019;10(1):2. doi 10.3390/genes10010002
- Dong L., Fang C., Cheng Q., Su T., Kou K., Kong L., Zhang C., ... Yuan X., Weller J.L., Lu S., Kong F., Liu B. Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. *Nat Commun.* 2021;12:5445. doi 10.1038/s41467-021-25800-3
- Dong L., Cheng Q., Fang C., Kong L., Yang H., Hou Z., Li Y., ... Tang Y., Zhao X., Lu S., Liu B., Kong F. Parallel selection of distinct *Tof5* alleles drove the adaptation of cultivated and wild soybean to high latitudes. *Mol Plant.* 2022;15(2):308-321. doi 10.1016/j.molp.2021.10.004
- Gray J.A., Shalit-Kaneh A., Chu D.N., Hsu P.Y., Harmer S.L. The REVEILLE clock genes inhibit growth of juvenile and adult plants by control of cell size. *Plant Physiol.* 2017;173(4):2308-2322. doi 10.1104/pp.17.00109
- Guan R., Yu L., Liu X., Li M., Chang R., Gilliam M., Qiu L. Selection of the salt tolerance gene *GmSALT3* during six decades of soybean breeding in China. *Front Plant Sci.* 2021;12:794241. doi 10.3389/fpls.2021.794241
- Guo S., Li Y., Qiu H., Hu G., Zhao C., Wang R., Zhang H., Tian Y., Li X., Liu B., Li Ying-hui, Qiu L. *GmAP1d* regulates flowering time under long-day photoperiods in soybean. *Crop J.* 2024;12(3):845-855. doi 10.1016/j.cj.2024.03.004
- Hoang X., Nguyen N., Nguyen Y.-N., Watanabe Y., Tran L.-S. The soybean *GmNAC019* transcription factor mediates drought tolerance in *Arabidopsis* in an abscisic acid-dependent manner. *Int J Mol Sci.* 2019;21(1):286. doi 10.3390/ijms21010286
- Kassambara A. ggcorrplot: Visualization of a Correlation Matrix Using “ggplot2”. 2023. R package version 0.1.4.999. Available: <https://github.com/kassambara/ggcorrplot>
- Li C., Chen Y., Hu Q., Yang X., Zhao Y., Lin Y., Yuan J., Gu J., Li Y., He J., Wang D., Liu B., Wang Z.-Y. PSEUDO-RESPONSE REGULATOR 3b and transcription factor ABF3 modulate abscisic acid-dependent drought stress response in soybean. *Plant Physiol.* 2024;195(4):3053-3071. doi 10.1093/plphys/kiac269
- Li H., Du H., He M., Wang J., Wang F., Yuan W., Huang Z., ... Liu B., Kong F., Fang C., Zhao X., Yu D. Natural variation of *FKF1* controls flowering and adaptation during soybean domestication and improvement. *New Phytol.* 2023a;238(4):1671-1684. doi 10.1111/nph.18826
- Li H., Du H., Huang Z., He M., Kong L., Fang C., Chen L., Yang H., Zhang Y., Liu B., Kong F., Zhao X. The AP2/ERF transcription factor *TOE4b* regulates photoperiodic flowering and grain yield per plant in soybean. *Plant Biotechnol J.* 2023b;21(8):1682-1694. doi 10.1111/pbi.14069
- Li S., Wang W., Sun L., Zhu H., Hou R., Zhang H., Tang X., Clark C.B., Swarm S.A., Nelson R.L., Ma J. Artificial selection of mutations in two nearby genes gave rise to shattering resistance in soybean. *Nat Commun.* 2024;15:7588. doi 10.1038/s41467-024-52044-8
- Li Y.F., Zhang L., Wang J., Wang X., Guo S., Xu Z.J., Li D., Liu Z., Li Y.H., Liu B., Qiu L.J. Flowering time regulator *qFT13-3* involved in soybean adaptation to high latitudes. *Plant Biotechnol J.* 2024;22(5):1164-1176. doi 10.1111/pbi.14254
- Li Z., Cheng Q., Gan Z., Hou Z., Zhang Y., Li Y., Li H., ... Kou K., Wang L., Kong F., Liu B., Dong L. Multiplex CRISPR/Cas9-mediated knockout of soybean *LNK2* advances flowering time. *Crop J.* 2021;9(4):767-776. doi 10.1016/j.cj.2020.09.005
- Liu H.-J., Swarts K., Xu S., Yan J., Nordborg M. On the contribution of genetic heterogeneity to complex traits. *bioRxiv.* 2024. doi 10.1101/2024.03.27.586967

- Liu Y., Zhang Y., Liu X., Shen Y., Tian D., Yang X., Liu S., Ni L., Zhang Z., Song S., Tian Z. SoyOmics: a deeply integrated database on soybean multi-omics. *Mol Plant*. 2023;16(5):794-797. doi 10.1016/j.molp.2023.03.011
- Lu H., McClung C.R., Zhang C. Tick Tock: circadian regulation of plant innate immunity. *Annu Rev Phytopathol*. 2017;55:287-311. doi 10.1146/annurev-phyto-080516-035451
- Liu S., Zhao X., Hu Y., Liu S., Nan H., Li X., Fang C., ... Weller J.L., Liu B., Hou X., Tian Z., Kong F. Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nat Genet*. 2017;49(5):773-779. doi 10.1038/ng.3819
- Niraula P.M., McNeece B.T., Sharma K., Alkharouf N.W., Lawrence K.S., Klink V.P. The central circadian regulator *CCA1* functions in *Glycine max* during defense to a root pathogen, regulating the expression of genes acting in effector triggered immunity (ETI) and cell wall metabolism. *Plant Physiol Biochem*. 2022;185:198-220. doi 10.1016/j.plaphy.2022.05.028
- Okonechnikov K., Golosova O., Fursov M., UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012; 28(8):1166-1167. doi 10.1093/bioinformatics/bts091
- Perfil'ev R., Shcherban A., Potapov D., Maksimenko K., Kiryukhin S., Gurinovich S., Panarina V., Polyudina R., Salina E. Impact of allelic variation in maturity genes *E1-E4* on soybean adaptation to Central and West Siberian regions of Russia. *Agriculture*. 2023;13(6):1251. doi 10.3390/agriculture13061251
- Perfil'ev R., Shcherban A., Potapov D., Maksimenko K., Kiryukhin S., Gurinovich S., Panarina V., Polyudina R., Salina E. Genome-wide association study revealed some new candidate genes associated with flowering and maturity time of soybean in Central and West Siberian regions of Russia. *Front Plant Sci*. 2024;15:1463121. doi 10.3389/fpls.2024.1463121
- Piovesan D., Del Conte A., Mehdiabadi M., Aspromonte M.C., Blum M., Tesei G., von Bülow S., Lindorff-Larsen K., Tosatto S.C.E. MOBIDB in 2025: integrating ensemble properties and function annotations for intrinsically disordered proteins. *Nucleic Acids Res*. 2025;53(D1):D495-D503. doi 10.1093/nar/gkae969
- Rawat R., Schwartz J., Jones M.A., Sairanen I., Cheng Y., Andersson C.R., Zhao Y., Ljung K., Harmer S.L. REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. *Proc Natl Acad Sci USA*. 2009;106(39):16883-16888. doi 10.1073/pnas.0813035106
- Rawat R., Takahashi N., Hsu P.Y., Jones M.A., Schwartz J., Salemi M.R., Phinney B.S., Harmer S.L. REVEILLE8 and PSEUDO-REPONSE REGULATOR5 form a negative feedback loop within the Arabidopsis circadian clock. *PLoS Genet*. 2011;7(3):e1001350. doi 10.1371/journal.pgen.1001350
- Rogers S.O., Bendich A.J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol Biol*. 1985;5(2):69-76. doi 10.1007/BF00020088
- Sasaki E., Köcher T., Filaault D.L., Nordborg M. Revisiting a GWAS peak in *Arabidopsis thaliana* reveals possible confounding by genetic heterogeneity. *Heredity*. 2021;127(3):245-252. doi 10.1038/s41437-021-00456-3
- Scandola S., Mehta D., Li Q., Rodriguez Gallo M.C., Castillo B., Uhrig R.G. Multi-omic analysis shows *REVEILLE* clock genes are involved in carbohydrate metabolism and proteasome function. *Plant Physiol*. 2022;190(2):1005-1023. doi 10.1093/plphys/kiac269
- Shan B., Wang W., Cao J., Xia S., Li R., Bian S., Li X. Soybean *GmMYB133* inhibits hypocotyl elongation and confers salt tolerance in *Arabidopsis*. *Front Plant Sci*. 2021;12:764074. doi 10.3389/fpls.2021.764074
- Shi X., Yan L., Yang C., Yan W., Moseley D.O., Wang T., Liu B., Di R., Chen P., Zhang M. Identification of a major quantitative trait locus underlying salt tolerance in 'Jidou 12' soybean cultivar. *BMC Res Notes*. 2018;11:95. doi 10.1186/s13104-018-3202-3
- Wang K., Bu T., Cheng Q., Dong L., Su T., Chen Z., Kong F., Gong Z., Liu B., Li M. Two homologous *LHY* pairs negatively control soybean drought tolerance by repressing the abscisic acid responses. *New Phytol*. 2021;229(5):2660-2675. doi 10.1111/nph.17019
- Wang T., Sun M.-Y., Wang X.-S., Li W.-B., Li Y.-G. Over-expression of GmGla-regulated soybean *miR172a* confers early flowering in transgenic *Arabidopsis thaliana*. *Int J Mol Sci*. 2016;17(5):645. doi 10.3390/ijms17050645
- Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer, 2016. doi 10.1007/978-3-319-24277-4
- Xie Q., Wang P., Liu X., Yuan L., Wang L., Zhang C., Li Y., Xing H., Zhi L., Yue Z., Zhao C., McClung C.R., Xu X. LNK1 and LNK2 are transcriptional coactivators in the *Arabidopsis* circadian oscillator. *Plant Cell*. 2014;26(7):2843-2857. doi 10.1105/tpc.114.126573
- Zhang M., Liu S., Wang Z., Yuan Y., Zhang Z., Liang Q., Yang X., Duan Z., Liu Y., Kong F., Liu B., Ren B., Tian Z. Progress in soybean functional genomics over the past decade. *Plant Biotechnol J*. 2022;20(2):256-282. doi 10.1111/pbi.13682

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 14.08.2025. После доработки 25.11.2025. Принята к публикации 23.12.2025.

doi 10.18699/vjgb-26-05

## A histochemical assay for polyphenolic profiling in cereal grains

S.R. Mursalimov  , O.Yu. Shoeva 

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

 mursalimovsr@gmail.com

**Abstract.** In different cell layers, cereal grains may accumulate various economically important polyphenols such as colored anthocyanins and melanins and colorless proanthocyanidins. To effectively create new cultivars with different combinations of these compounds, a simple, fast, and precise screening method is required. Here, a histochemical assay that includes a combination of hot ethanolic, acidic, alkaline, and ammoniacal silver treatments of grain cryosections followed by microscopy was successfully applied to distinguish these substances in cereal grains. Barley lines previously characterized chemically for the presence of anthocyanins, proanthocyanidins, and melanins in grains were used as a model. In black barley grains, this approach allowed to visually distinguish insoluble melanins that do not react to a pH change from anthocyanins, which can be insoluble or soluble but always react to changing pH. For the first time, ammoniacal silver staining commonly used for melanin identification in human and animal tissues was adapted for melanin identification in plant tissues. Along with melanins, this reagent stains other polyphenols thereby helping to detect colorless polyphenols including proanthocyanidins in the testa of barley grains as confirmed by *p*-dimethylaminocinnamaldehyde (DMACA) staining. The applicability of this assay to polyphenol profiling was demonstrated not only in the barley grain but also in wheat and common vetch grains. The proposed histochemical assay allows rapid polyphenol screening using a single grain, making it a practical and efficient alternative to time-consuming chromatographic methods for preliminary selection from large sample sets prior to detailed quantitative and qualitative chemical analysis.

**Key words:** anthocyanin; analytical technique; DMACA; proanthocyanidins; melanins; Fontana–Masson

**For citation:** Mursalimov S.R., Shoeva O.Yu. A histochemical assay for polyphenolic profiling in cereal grains. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2026;30(1):53-60. doi 10.18699/vjgb-26-05

**Funding.** The study was supported by RSF grant No. 25-16-20101, <https://rscf.ru/project/25-16-20101/> and the Ministry of Science and Innovation Policy of the Novosibirsk region (the agreement No. 30-2025-000848).

**Acknowledgements.** We are very grateful for Dr. A. Zukin (SPBGU) for providing us with the DMACA reagent. Plants for gathering developing seeds were grown using resources of the Greenhouse Core Facility supported by ICG project number FWNR-2026-0029.

## Гистохимический тест для определения профиля полифенольных соединений в зерне злаков

С.Р. Мурсалимов  , О.Ю. Шоева 

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

 mursalimovsr@gmail.com

**Аннотация.** Зерна злаков в некоторых типах клеток могут накапливать разнообразно экономически важные полифенольные соединения, такие как окрашенные антоцианы и меланины, а также бесцветные проантоцианидины. Для эффективного создания новых сортов с различными комбинациями этих соединений в зернах необходим простой, быстрый и точный метод определения их полифенольного профиля. В данной статье для идентификации указанных веществ предложено использовать гистохимический анализ, включающий комбинацию обработок криосрезом зерен горячим этанолом, кислотой, щелочью и аммиачным раствором серебра с последующей микроскопией. В качестве модели использовались линии ячменя, зерна которых ранее были охарактеризованы с помощью аналитических методов на наличие антоцианов, проантоцианидинов и меланинов. В черных зернах ячменя данный подход позволил отличить нерастворимые меланины, не реагирующие на изменение pH, от антоцианов, которые могут быть растворимыми или нерастворимыми, но всегда меняют окраску при изменении pH. Впервые обработка аммиачным серебром, широко применяемая для идентификации меланина в тканях человека и животных, была адаптирована для использования в растительных тканях. Наряду с меланинами, этот реагент окрашивает и другие полифенолы, что позволяет выявлять в том числе бесцветные соединения. С помощью этого подхода были выявлены проантоцианидины в оболочке зерна ячменя, что подтверждено окрашиванием 4-(диметиламино)циннамальдегидом (DMACA). Разработанный протокол успешно применен для определения полифенольного профиля зерен ячменя, пшеницы и вики посевной. Метод позволяет проводить быструю оценку полифенольного профиля с использованием единичных зерен, что является эффективной альтернативой трудоемким хроматографическим методам на этапе предварительного отбора коллекционного материала перед его детальным химическим анализом.

**Ключевые слова:** антоцианы; аналитическая техника; DMACA; проантоцианидины; меланины; реакция Фонтана–Массона

## Introduction

In different cell layers, cereal grains may accumulate different polyphenols such as colored anthocyanins and melanins and colorless proanthocyanidins. Melanins and anthocyanins are known to participate in protection of grain against severe environments and threats by predators and pathogens (Winkel-Shirley, 1998; Glagoleva et al., 2020), while proanthocyanidins are conducive to dormancy and longevity of seeds (Debeaujon et al., 2000).

The aforementioned polyphenols are industrially important; they affect final application of grains as a raw material. For example, due to their protein-binding capacity leading to chill haze (reducing beer quality), proanthocyanidins synthesized in the testa (i. e., the seed coat) of barley under the control of the *Ant28* gene are undesirable in malt cultivars (von Wettstein, 2007). Anthocyanins that accumulate either in the aleurone layer or in the pericarp under the control of genes *Bxl1–5* or *Ant1* and *Ant2* cause the grain color to be blue and purple, respectively, and melanins that accumulate in husks and the pericarp under the control of *Blp1* cause the grain color to be black (Shoeva et al., 2018). These anthocyanins and melanins are promising functional food ingredients and are desirable in cultivars for human consumption (Matseychik et al., 2020; Loskutov, Khlestkina, 2021). Currently, the generation of new cultivars with distinct compositions of polyphenolic pigments in grain is an advanced task for plant breeders. To reach this goal, analysis of many specimens for certain compounds to choose adequate donors carrying the desired genes and screen the resulting hybrids is required.

Currently available methods of chemical profiling of polyphenols such as high-performance liquid chromatography (HPLC) or thin layer chromatography (TLC) allow to separate a large number of individual compounds (Vermerris, Nicholson, 2008), but they are unnecessarily time-consuming for preliminary screening of many specimens. In addition, they do not show histological patterns of polyphenol accumulation, which are known to be unevenly distributed within cereal grain and should be taken into account while processing grain for food (Barron et al., 2017). As an alternative to these methods, microscopy techniques can be used (Panato et al., 2017).

Among polyphenols, anthocyanins can be easily detected by microscopy owing to their color, which may vary from red to blue and their shades and depends on the chemical structure of the molecules, the presence of copigments, and metal ions (Davies, 2009). Change of color at different pH levels is a well-known characteristic feature of anthocyanins. In a strongly acidic medium (pH < 1), anthocyanins are in the cationic form, having a red color, whereas between pH 4 and 6, the cation loses two protons and turns blue (Davies, 2009). This pH-dependent color change of anthocyanins has been adapted to the detection and quantification of these compounds (Lee et al., 2005; Wrolstad et al., 2005).

Melanins are dark brown to black pigments formed in plants by oxidation of diverse phenolic precursors among which catechol, caffeic, chlorogenic, protocatechuic, *p*-coumaric and gallic acid have been reported (Bell, Wheeler, 1986; Solano, 2014; Varga et al., 2016). The polymeric nature and poor solubility of melanins constrains research on their chemical struc-

ture. In plants, melanins are the least studied group of pigments unlike melanins of the other kingdoms of living organisms, such as animals, bacteria, and fungi melanins, which were surveyed in details at the biochemical and molecular genetic levels (Britton, 1983; Solano, 2014; Glagoleva et al., 2020). It has been shown that melanins in barley grains accumulate in specific plastids called melanoplasts (Shoeva et al., 2020; Mursalimov et al., 2022).

Proanthocyanidins are colorless and require histochemical staining to be detected (Gardner, 1975; Treutter, 1989). An aromatic aldehyde *p*-dimethylaminocinnamaldehyde (DMACA) that binds to meta-oriented dihydroxy- or trihydroxy-substituted benzene rings is routinely used to detect proanthocyanidins and their immediate precursor molecules, namely, flavan-3,4-diols and flavan-3-ols, in plant tissues, including grains of cereals (Aastrup et al., 1984; Abeynayake et al., 2011; Kohyama et al., 2017; Zykina et al., 2020).

In barley, it can often be challenging to distinguish visually dark melanins from anthocyanins owing to a brownish hue they acquire during grain maturation (Glagoleva et al., 2022b). Furthermore, anthocyanins and melanins can accumulate in one grain simultaneously (Glagoleva et al., 2022a). Melanins are usually identified by a series of solubility tests, and spectroscopy techniques such as Fourier transform infrared (FT-IR), nuclear magnetic resonance (NMR), and the electron paramagnetic resonance (EPR) spectroscopy (Glagoleva et al., 2020). They dissolve in alkali solutions, discolor under the influence of strong oxidizing agents, react with FeCl<sub>3</sub> and are stable in common organic solvents that have been reported as hallmarks of melanin (Sava et al., 2001). Histochemical analysis of the melanin pigments in plant tissues is not common. For detection of melanins in human and animal tissues, the Fontana–Masson (FM) protocol with ammoniacal silver staining is commonly used. The protocol is based on the reduction of ammoniacal silver to metallic silver by phenolic substances. The product of the reaction is an insoluble black precipitate, which can be identified on tissue sections by light microscopy (Wildi, 1951; Bancroft, Gamble, 2008).

Here, we use the barley lines, previously characterized chemically for the presence of anthocyanins, proanthocyanins, and melanins, to demonstrate the ability of a new histochemical assay to differentiate these substances in grain tissue. Among the lines used in the study are Bowman (hereafter: Bw) back-cross-derived near-isogenic lines (NILs) developed for the presence of anthocyanin (*i*:*BwAnt1Ant2*, hereafter: PLP) and melanin (*i*:*BwBlp1*, hereafter: BLP) compounds chemically identified by HPLC, a series of solubility tests, and Fourier transform infrared spectroscopy (Shoeva et al., 2020; Glagoleva et al., 2022c). The lines have been created by introgression of recombinant genetic fragments from colored donor lines into the genetic background of the parental cultivar Bowman, which lacks pigmentation in the grain (Druka et al., 2011). Another line used in the study is the mutant line *ant25.264*, which lacks proanthocyanidins in the grain. This line was induced by chemical mutagenesis of the parental cultivar Secobra 18193, which retained its ability to accumulate proanthocyanidins in the grain (Jende-Strid, 1993). NILs and mutant lines, which have relatively few genetic differences from their parental

cultivars, and are used extensively in genetic studies aiming to discover the association of phenotypic differences between the lines with specific chromosome regions or nucleotide polymorphisms. The use of these lines in studies of the genetic control of the synthesis of different classes of polyphenolic compounds in barley grains has been accompanied by their chemical profiling (Jende-Strid, 1993; Shoeva et al., 2020; Glagoleva et al., 2022c).

Here, taking advantage of the detailed characterization of these lines, a histochemical assay was developed to identify the different classes of polyphenolic compounds in barley grains. The described protocol may be employed for fast screening of a large number of samples and requires only a single grain for every sample. The method could be adapted to the grains/seeds/fruits of a wide variety of plant species. Here, as an example, besides barley, we successfully applied the assay to polyphenolic pigment profiling of black grains of wheat line *s:S29Ba14Th(4D)Pp3<sup>P</sup>Pp-DI<sup>PF</sup>* (hereafter: S29BLACK) accumulating anthocyanins in the pericarp and aleurone layers simultaneously (Gordeeva et al., 2022), and common vetch cultivar Obskaya 16, having a black color of grain due to accumulation of anthocyanins in the macrosclereids (Goncharova, 2020; Mursalimov et al., 2021).

## Materials and methods

**Plant material and growing conditions.** The plant genetic resources utilized in the current study and their origin are listed in Table S1<sup>1</sup>. To obtain seeds for developing the protocol for histochemical polyphenolic profiling, cereal plants were grown in a greenhouse of the ICG SB RAS (Novosibirsk, Russia) under a 16 h photoperiod at a temperature range of 20–25 °C, whereas vetch was grown in an experimental field of the ICG SB RAS (Novosibirsk, Russia) in 2021.

**The histochemical assay.** The grains were sampled at the hard dough developmental stage (BBCH-87) from each of the aforementioned barley and wheat lines and at the brown pod stage for vetch grains. All of them were snap-frozen in liquid nitrogen and stored at –70 °C until analysis. Prior to sectioning, the frozen grains were kept at –20 °C for 30 min, mounted, and embedded in the Tissue-Tek O.C.T.TM compound medium (Sakura Finetek Europe B.V., Netherlands). The sectioning was carried out at –20 °C using an HM 505N cryostat microtome (Microm, Germany). Sections of 20 µm thickness were mounted on a poly-L-lysine slide (Thermo Fisher Scientific, Germany) and fixed by the addition of 1 mL of a precooled fixative (4 °C) onto the slide. Then, the slides were air dried at room temperature. For sections prepared for hot ethanol treatment, 80 % ethanol in water was utilized as a fixative, and 8 % formaldehyde (Sigma-Aldrich, USA) solution in phosphate-buffered saline (pH 7.4) was used as a fixative for the rest of the samples. All the slides were next rinsed for 15 min in distilled water to remove the fixative and embedding medium. After that, for hot ethanol treatment, the slides were transferred into 80 % (v/v) ethanol in water at 80 °C for 60 min incubation. The slides were then rinsed in distilled water and mounted in glycerol. For the pH test, the slides were immersed

for a few seconds either in 1 M NaOH or 1 M HCl, quickly rinsed in distilled water and mounted in glycerol.

For ammoniacal silver staining, the FM protocol (Bancroft, Gamble, 2008; Kiernan, 2008) was performed. A stock solution was prepared by adding ammonium hydroxide solution drop by drop to a 10 % (w/v) silver nitrate water solution, until the solution yielded a precipitate and cleared again. After that, a 0.01 % working solution was prepared by dilution with distilled water. Next, the slides with the sections were incubated in the working solution for 15 min at 60 °C. The slides were then rinsed in distilled water and incubated in a 5 % (w/v) sodium thiosulfate water solution for 15 min, rinsed in distilled water again, and mounted in glycerol. The sections were analyzed under an Axio Observer Z1 microscope equipped with an AxioCam HRc camera (Zeiss, Germany). The ZEN software was used for image analysis and processing (Zeiss, Germany). The microscopic analysis was carried out at the Multi-Access Center for Microscopy of Biological Objects at the Institute of Cytology and Genetics, SB RAS (Novosibirsk, Russia). For each variety, at least three independent grains were analyzed, and a minimum of ten sections were examined for each grain.

**Dimethylaminocinnamaldehyde (DMACA) staining.** The sections fixed as described above were incubated overnight in a fresh solution of 0.5 % (w/v) DMACA (Sigma-Aldrich, Germany, kindly provided by Dr. Pavel Zykin from St. Petersburg State University, Russia) in 6 N HCl. The microscopic analysis was carried out as described above. For each variety, at least three independent grains were analyzed, and a minimum of ten sections were examined for each grain.

## Results

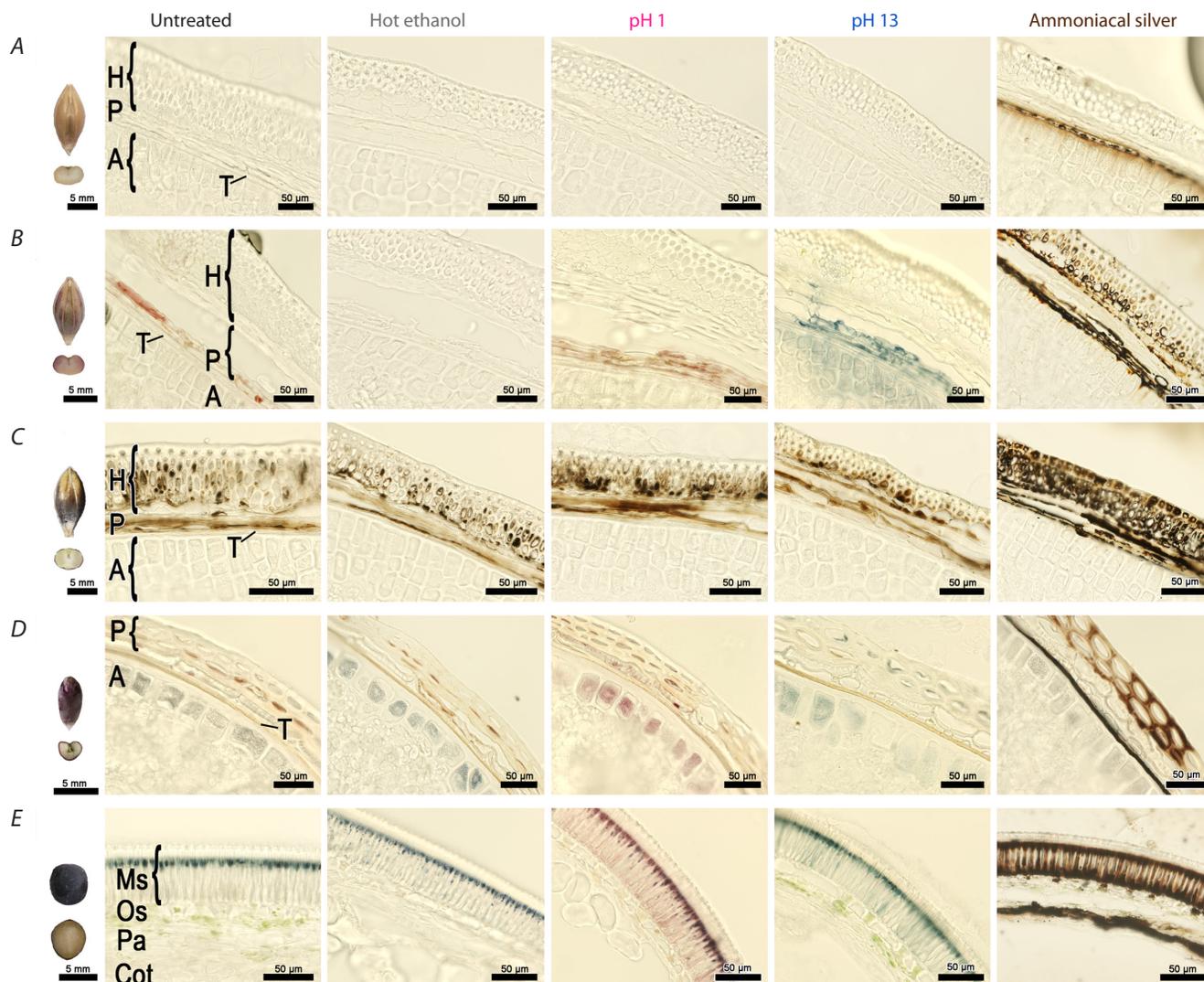
Grain sections of three barley lines were analyzed by the proposed assay: the control unpigmented Bowman line, the PLP line with purple grains, and the BLP line with black grains. Five groups of sections were obtained for every grain: untreated or treated with hot ethanol, HCl, NaOH, or ammoniacal silver. All the resulting specimens were examined by brightfield microscopy (Fig. 1).

Grain tissues on sections of the control Bowman line did not have any visible pigmentation at the hard dough stage (Fig. 1A). There was no visible staining after hot ethanol treatment as well as after the pH change. The only noticeable dark staining appeared after ammoniacal silver treatment of the testa.

Untreated sections of the PLP line are characterized by accumulation of purple pigments in cells of the pericarp (Fig. 1B). The other cell types of this line do not have any pigmentation. The purple color disappeared totally after hot ethanol treatment, and the whole slice became colorless. At pH 1, the purple color persisted, whereas at pH 13, it turned bluish-green and notably leached into surrounding tissues. Intensive dark staining was observed in the pericarp, testa, and husk after ammoniacal silver treatment.

In the BLP line, brownish-black pigments were observed in untreated sections in the pericarp and husk (Fig. 1C). This pigmentation did not change after hot ethanol treatment or after the pH changes. After ammoniacal silver treatment, the

<sup>1</sup> Supplementary Table S1 is available at:  
<https://vavilovj-icg.ru/download/pict-2026-30/appx4.pdf>



**Fig. 1.** Seeds and cross-sections of the seeds of barley, wheat and vetch after corresponding treatment (brightfield microscopy). Horizontal lines: A – Bowman, B – PLP, C – BLP, D – S29BLACK, E – vetch. A: aleurone layer, H: husk, P: pericarp, T: testa, Cot: cotyledon, Ms: macrosclereids, Os: osteosclereids, Pa: parenchyma.

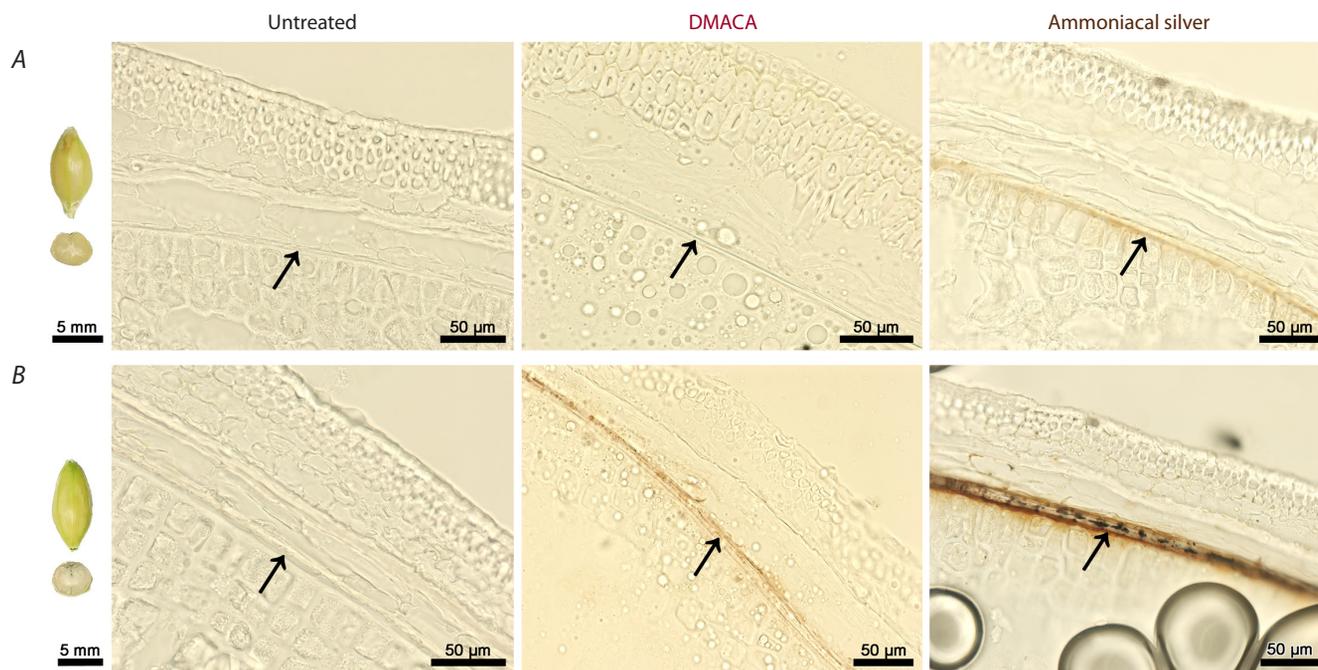
brownish-black pigmentation became even more intense in the pericarp and husk, where it was present before, and it emerged in the previously colorless testa and upper epidermal cells of the husk.

The sections prepared from wheat and vetch grains were assayed in the same manner. In black grains of wheat line S29BLACK, pale bluish-gray and purple pigments were revealed in aleurone and pericarp cells, respectively (Fig. 1D). The pigmentation did not disappear after hot ethanol treatment, and the bluish-gray color turned blue. At pH 1, blue aleurone cells became purple, while the purple pericarp cells did not change their color. At pH 13, the purple color of pericarp cells became blue, while the bluish-gray color of aleurone cells changed to bluish-green. Ammoniacal silver treatment gave rise to gray pigmentation in the aleurone, a brownish-black color in pericarp cells, and an intensive black color in the testa. It should be pointed out that only in wheat grains was the brownish-black color detected in cell walls of the pericarp.

In black vetch grains, blue pigmentation was observed in macrosclereids (Fig. 1E), becoming more intense after hot ethanol treatment and failing to be washed out. The blue pigments turned purple at pH 1 and became bluish-green at pH 13. In macrosclereids, ammoniacal silver stained the blue pigments brownish-black. In addition, the brownish-black pigmentation emerged in previously colorless parts of macrosclereids as well as in osteosclereids and cells of the parenchyma.

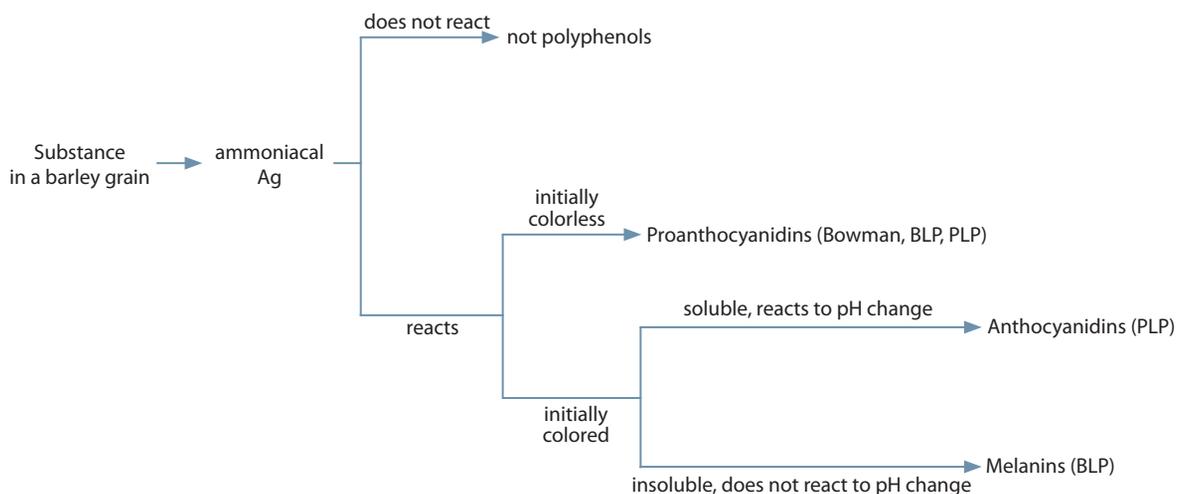
A comparison between ammoniacal silver staining and DMACA staining for revealing colorless polyphenols was performed on proanthocyanidin-less barley mutant line *ant25.264* and its parental cultivar Secobra 18193 containing proanthocyanidins in grains (Fig. 2).

There was no noticeable staining after the DMACA treatment of grains of *ant25.264*, whereas pale brown staining appeared in testa after ammoniacal silver treatment (Fig. 2). The brownish-purple staining was observed in the testa of the parental Secobra 18193 grains after DMACA treatment, and



**Fig. 2.** Seeds and cross-sections of the seeds of anthocyanin-free (A) and anthocyanin-rich (B) grains of barley after DMACA and ammoniacal silver staining (brightfield microscopy).

Horizontal lines: A – *ant25.264*, B – *Secobra 18193*. Arrows indicate testa.



**Fig. 3.** The scheme for identification of polyphenols in barley grains.

in the same tissue, intense black staining was registered after ammoniacal silver treatment (Fig. 2).

After summarizing the results of all the above-mentioned tests, we built a scheme for identifying anthocyanins, melanins, and proanthocyanidins in barley grains (Fig. 3).

### Discussion

The analysis of polyphenolic profiles of crops is quite important for both the elucidation of fundamental aspects of polyphenol biosynthesis and for industrial applications of the raw material. Some attempts have been made to find a fast and simple way to perform polyphenolic profiling of a large

number of samples, such as high-throughput phenotyping approaches based on image analysis of color characteristics (ElMasry et al., 2019; Komyshev et al., 2023). Nevertheless, histochemical techniques combined with microscopy seem to be the only approach nowadays allowing a fast and precise polyphenolic profiling on a limited amount of the experimental material and providing information about the cell type where these polyphenols accumulate. Our assay involving a series of standard chemical tests applied to cryosections followed by microscopic examination successfully showed the possibility of identifying polyphenolic pigments in barley, wheat, and vetch grains.

Hot ethanol treatment, which is a common test for solubility (Takahashi, Hara, 2014), allowed us to easily distinguish soluble anthocyanins in PLP grains from insoluble melanins in BLP grains. It is known that brown-pigment melanins necessitate a special harsh treatment to become soluble and are insoluble under the hot ethanol conditions we used (Sava et al., 2001). Nevertheless, not all anthocyanins are soluble. Melanins can be distinguished from insoluble anthocyanins by the absence of a color transition at changing pH, which is a well-known characteristic feature of anthocyanins (Davies, 2009).

Thus, in the PLP line, the solubility and pH tests enabled us to identify the purple pigments – that completely disappeared during hot ethanol treatment and changed color to blue in the presence of sodium hydroxide – as anthocyanins, and to identify the BLP line's black pigments – that did not react to hot ethanol treatment and changes in pH – as melanins. The same results have been previously obtained by HPLC, a series of solubility tests, and Fourier transform infrared spectroscopy, when extracts from PLP and BLP grains have been assayed (Shoeva et al., 2020; Glagoleva et al., 2022c).

As a first step, a FM protocol was included in our histochemical assay to confirm the presence of melanins, which do not react to hot ethanol treatment and a pH change. The FM protocol is commonly used for the detection of melanins in human and animal tissues (Bancroft, Gamble, 2008; Kiernan, 2008) and has not been used previously for the detection of melanins in plant tissues. It turned out that the FM protocol allows the detection of not only melanins by staining them deep black but also of other polyphenols in a grain, including colorless ones, owing to their common ability to reduce ammoniacal silver to metallic one. It is known that silver nitrate reacts with aromatic compounds that contain two or more phenolic hydroxyl groups (Wildi, 1951).

In the barley grain, black melanins, purple anthocyanins, and colorless proanthocyanidins can be detected by ammoniacal silver staining. In grains of the BLP line, the melanin-containing tissues (pericarp and husk) were expectedly stained by ammoniacal silver; however, metallic silver was also noted in the testa, implying the presence of hidden polyphenols. Ammoniacal silver staining in the BLP line was the strongest in comparison to lines Bowman and PLP, thus pointing to the presence of a much greater amount of polyphenols in BLP. The higher accumulation of phenolic compounds in the BLP line has also been observed elsewhere by HPLC analysis (Shoeva et al., 2016). Moreover, for the black-grained BLP line, the highest antioxidant activity has been documented (Glagoleva et al., 2017). The ability of melanins and colorless polyphenols to reduce ammoniacal silver may be explained by their well-known antioxidant properties (Quideau et al., 2011; Panzella et al., 2012). In the PLP line, ammoniacal silver strongly stained both visible anthocyanin pigments and previously colorless tissues, but less strongly than in the BLP line. In all three barley lines, ammoniacal silver stained colorless polyphenols in the testa. It was especially noticeable in noncolored control Bowman grains, which do not contain any visible pigments and do not react to hot ethanol and to changes in pH but show prominent black granules after ammoniacal silver staining in

only one tissue: the testa. It is in this tissue of the barley grain that colorless proanthocyanidins are synthesized (Jende-Strid, 1993).

To evaluate the efficiency of ammoniacal silver staining at detecting proanthocyanidins, it was compared with the DMACA staining. Both methods were applied to barley grain sections enriched with and low in proanthocyanidins. We observed the better sensitivity of ammoniacal silver staining: DMACA did not detect any proanthocyanidins in proanthocyanidin-less *ant25.264*, while the ammoniacal silver reaction in its turn was weak but positive. It is also important to state that DMACA staining is performed in 6 M HCl, and HCl negatively affects tissue preservation. Moreover, owing to the low pH of the DMACA solution, this reagent must not be used in the presence of anthocyanins because they could change their color under such conditions and yield a false positive result. Due to the aforementioned problem, we could not apply DMACA staining to the detection of proanthocyanidins in grains of PLP barley, S29BLACK wheat, and vetch. The results allowed us to conclude that ammoniacal silver staining has some advantages against DMACA staining and could replace it to locate proanthocyanidins in grains.

The observations made on the barley grain sections treated with hot ethanol, HCl, NaOH, and ammoniacal silver helped us to develop a protocol for the identification of polyphenolic pigments in barley grains; it is presented as a scheme in Figure 3. The relevance of the protocol was confirmed by chromatography and independent chemical tests reported previously for the same barley lines (Shoeva et al., 2020; Glagoleva et al., 2022c). It is worth noting that this protocol was devised for barley. For polyphenolic profiling of grains of other plant species in the same manner, it is desirable to verify the reliability of the proposed histochemical techniques by precise analytical chemistry approaches. Nevertheless, for a preliminary analysis, the proposed assay could be used in its current state even for different species. Here we extended the application of this assay to polyphenol profiling of darkly pigmented grains of wheat and vetch. Wheat line S29BLACK was constructed by marker-assisted selection for the presence of genes *ThMyc4E* and *Pp*, which control the synthesis of blue and purple anthocyanins in the aleurone and pericarp (Gordeeva et al., 2022). The black color of grains of the vetch line we analyzed in the current work has been ascribed to the blue pigments accumulating in macrosclereids; however, the precise chemical composition has not been determined (Mursalimov et al., 2021).

Our assay revealed a sufficient amount of polyphenols in testa and pericarps of the tested wheat grains, which yielded positive staining with ammoniacal silver. This approach allowed us to detect deposition of the polyphenols in cell walls of the wheat grains pericarp with the bound phenolic acids among which dehydrodiferulates have been reported as predominant (Parker et al., 2005). Abundant colorless polyphenols were also detected by ammoniacal silver staining in seed envelopes of vetch; this feature may affect feeding properties of this crop and deserves deeper research too. Unlike purple pigments in the barley pericarp, the purple pigments in the wheat pericarp are insoluble and are not washed out by hot ethanol; however,

they alter their color after a pH change, thereby enabling us to identify these purple pigments as anthocyanins. Previously, cyanidin-based compounds have been identified in the wheat pericarp as major pigments, among which cyanidin-3-glucoside, cyanidin-3-rutinoside, and peonidin-3-glucoside are predominant (Abdel-Aal et al., 2006; Ficco et al., 2014). Insoluble blue pigments have also been observed in the aleurone layer of wheat and in macrosclereids of vetch. They have been identified as anthocyanins owing to their color transition from blue (original color) to red in an acidic medium. In blue-grained wheat, delphinidin-derived anthocyanins – delphinidin-3-glucoside and delphinidin-3-rutinoside – have been identified as predominant anthocyanins (Knievel et al., 2009; Abdel-Aal et al., 2016), whereas in vetch, the precise chemical structure of the blue pigments has not been determined yet. Judging by the results of our test, we can predict that vetch grains accumulate similar blue delphinidin-derived anthocyanins.

It also should be noted that differential behavior of anthocyanins during hot ethanol treatment may be attributed not only to their distinct chemical structure but also to the type of cells and organelles accumulating these pigments. For example, in wheat, blue anthocyanins accumulate in aleurone cells' vacuoles. There are spherical particles called aleurone granules in aleurone cells, which are either phytate inclusions (type 1: composed of phytic acid minerals) or niacin inclusions (type 2: composed of niacin and proteins), each granule being surrounded by a fine layer of lipidic droplets (Morrison et al., 1975; Yu et al., 2021). The anthocyanins accumulated in aleurone cells may form complexes with metal ions or interact with the proteins; both these properties are attributed to anthocyanins and reportedly improve stability of the molecules (Li et al., 2021).

## Conclusion

In the current work, the histochemical assay consisting of a series of standard chemical tests applied to cryosections followed by microscopic examination successfully showed the possibility of distinguishing polyphenolic pigments and uncolored polyphenols in barley, wheat, and vetch grains. However, the extrapolation of results obtained on barley (and confirmed by analytical chemistry approaches) to other species requires caution, and further chemical analyses are necessary for exact identification of polyphenols. Nevertheless, the proposed histochemical assay has a high potential for polyphenolic profiling as part of preliminary screening of grains of barley and many other species where only a single grain is required. After the proposed preliminary screening, the selected lines/varieties could be propagated and analyzed in more detail for the presence of certain chemicals by more precise quantitative and qualitative methods.

## References

Aastrup S., Outtrup H., Erdal K. Location of the proanthocyanidins in the barley grain. *Carlsberg Res Commun.* 1984;49:105-109. doi 10.1007/BF02913969

Abdel-Aal el-S.M., Young J.C., Rabalski I. Anthocyanin composition in black, blue, pink, purple, and red cereal grains. *J Agric Food Chem.* 2006;54(13):4696-4704. doi 10.1021/jf0606609

Abdel-Aal el-S.M., Hucl P., Shipp J., Rabalski I. Compositional differences in anthocyanins from blue- and purple-grained spring wheat grown in four environments in central Saskatchewan. *Cereal Chem J.* 2016;93(1):32-38. doi 10.1094/CCHEM-03-15-0058-R

Abeynayake S.W., Panter S., Mouradov A., Spangenberg G. A high-resolution method for the localization of proanthocyanidins in plant tissues. *Plant Methods.* 2011;7(1):13. doi 10.1186/1746-4811-7-13

Bancroft J.D., Gamble M. (Eds) Theory and Practice of Histological Techniques. Edinburgh: Churchill Livingstone, 2008. doi 10.1016/B978-0-443-10279-0.50001-4

Barron C., Holopainen-Mantila U., Sahlstrom S., Hotekjolen A.K., Lullien-Pellerin V. Assessment of biochemical markers identified in wheat for monitoring barley grain tissue. *J Cereal Sci.* 2017;74: 11-18. doi 10.1016/j.jcs.2017.01.004

Bell A.A., Wheeler M.H. Biosynthesis and functions of fungal melanins. *Annu Rev Phytopathol.* 1986;24:411-451. doi 10.1146/annurev.py.24.090186.002211

Britton G. The Biochemistry of Natural Pigments. Cambridge Univ. Press, 1983

Davies K.M. Modifying anthocyanin production in flowers. In: Winefield C., Davies K., Gould K. (Eds) Anthocyanins: Biosynthesis, Functions, and Applications. Springer, 2009;49-80. doi 10.1007/978-0-387-77335-3\_3

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):403-414. doi 10.1104/pp.122.2.403

Druka A., Franckowiak J., Lundqvist U., Bonar N., Alexander J., Houston K., Radovic S., Shahinnia F., Vendramin V., Morgante M., Stein N., Waugh R. Genetic dissection of barley morphology and development. *Plant Physiol.* 2011;155(2):617-627. doi 10.1104/pp.110.166249

ElMasry G., Mandour N., Al-Rejaie S., Belin E., Rousseau D. Recent applications of multispectral imaging in seed phenotyping and quality monitoring – an overview. *Sensors.* 2019;19(5):1090. doi 10.3390/s19051090

Ficco D.B.M., De Simone V., Colecchia S.A., Pecorella I., Platani C., Nigro F., Finocchiaro F., Papa R., De Vita P. Genetic variability in anthocyanin composition and nutritional properties of blue, purple, and red bread (*Triticum aestivum* L.) and durum (*Triticum turgidum* L. ssp. *turgidum* convar. *durum*) wheats. *J Agric Food Chem.* 2014;62(34):8686-8695. doi 10.1021/jf5003683

Gardner R.O. Vanillin-hydrochloric acid as a histochemical test for tannin. *Stain Technol.* 1975;50(5):315-317. doi 10.3109/10520297509117081

Glagoleva A.Y., Shmakov N.A., Shoeva O.Y., Vasiliev G.V., Shatskaya N.V., Bömer A., Afonnikov D.A., Khlestkina E.K. Metabolic pathways and genes identified by RNA-seq analysis of barley near-isogenic lines differing by allelic state of the *Black lemma and pericarp* (*Blp*) gene. *BMC Plant Biol.* 2017;17:182. doi 10.1186/s12870-017-1124-1

Glagoleva A.Y., Shoeva O.Y., Khlestkina E.K. Melanin pigment in plants: current knowledge and future perspectives. *Front Plant Sci.* 2020;11:770. doi 10.3389/fpls.2020.00770

Glagoleva A., Kukoeva T., Mursalimov S., Khlestkina E., Shoeva O. Effects of combining the genes controlling anthocyanin and melanin synthesis in the barley grain on pigment accumulation and plant development. *Agronomy.* 2022a;12(1):112. doi 10.3390/agronomy12010112

Glagoleva A.Y., Novokreschenov L.A., Shoeva O.Y., Kovaleva O.N., Khlestkina E.K. Studying grain color diversity in the barley collection of VIR. *Proc Appl Bot Genet Breed.* 2022b;183(3):76-84. doi 10.30901/2227-8834-2022-3-76-84 (in Russian)

Glagoleva A.Y., Vikhorev A.V., Shmakov N.A., Morozov S.V., Chernyak E.I., Vasiliev G.V., Shatskaya N.V., Khlestkina E.K., Shoeva O.Y. Features of activity of the phenylpropanoid biosynthesis pathway in melanin-accumulating barley grains. *Front Plant Sci.* 2022c;13:923717. doi 10.3389/fpls.2022.923717

- Goncharova A.V. Spring common vetch sowing cultivar Obskaya 16. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Lett Vavilov J Genet Breed.* 2020;6(1):15-17. doi 10.18699/Letters2020-6-03 (in Russian)
- Gordeeva E., Shoeva O., Mursalimov S., Adonina I., Khlestkina E. Fine points of marker-assisted pyramiding of anthocyanin biosynthesis regulatory genes for the creation of black-grained bread wheat (*Triticum aestivum* L.) lines. *Agronomy.* 2022;12:2934. doi 10.3390/agronomy12122934
- 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
- Kiernan J. *Histological and Histochemical Methods: Theory and Practice.* Cold Spring Harbor, 2008
- Kniewel D.C., Abdel-Aal el-S.M., Rabalski I., Nakamura T., Hucl P. Grain color development and the inheritance of high anthocyanin blue aleurone and purple pericarp in spring wheat (*Triticum aestivum* L.). *J Cereal Sci.* 2009;50(1):113-120. doi 10.1016/j.jcs.2009.03.007
- Kohyama N., Chono M., Nakagawa H., Matsuo Y., Ono H., Matsumura H. Flavonoid compounds related to seed coat color of wheat. *Biosci Biotechnol Biochem.* 2017;81(11):2112-2118. doi 10.1080/09168451.2017.1373589
- Komyshv E.G., Genaev M.A., Busov I.D., Kozhekin M.V., Artemenko N.V., Glagoleva A.Y., Koval V.S., Afonnikov D.A. Determination of the melanin and anthocyanin content in barley grains by digital image analysis using machine learning methods. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2023;27(7):859-868. doi 10.18699/VJGB-23-99
- Lee J., Durst R.W., Wrolstad R.E. Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method. *J AOAC Int.* 2005; 88(5):1269-1278. doi 10.1093/jaoac/88.5.1269
- Li J., Wang B., He Y., Wen L., Nan H., Zheng F., Liu H., Lu S., Wu M., Zhang H. A review of the interaction between anthocyanins and proteins. *Food Sci Technol Int.* 2021;27(5):470-482. doi 10.1177/1082013220962613
- Loskutov I.G., Khlestkina E.K. Wheat, barley, and oat breeding for health benefit components in grain. *Plants.* 2021;10(1):86. doi 10.3390/plants10010086
- Matseychik I.V., Korpacheva S.M., Lomovsky I.O., Serasutdinova K.R. Prospects for the use of products of processing of buckwheat as functional ingredients. *Technol Merchandising Innovative Foodstuff.* 2020;61(2):53-57. doi 10.33979/2219-8466-2020-61-2-53-57 (in Russian)
- Morrison I.N., Kuo J., O'Brien T.P. Histochemistry and fine structure of developing wheat aleurone cells. *Planta.* 1975;123(2):105-116. doi 10.1007/BF00383859
- Mursalimov S.R., Goncharova A.V., Glagoleva A.Yu., Shoeva O.Yu. Black seed color of the spring common vetch (*Vicia sativa* L.) cultivar Obskaya 16 is caused by blue anthocyanins accumulating in macrosclereids. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Lett Vavilov J Genet Breed.* 2021;7(2):91-95. doi 10.18699/Letters-VJ2021-7-10
- Mursalimov S., Glagoleva A., Khlestkina E., Shoeva O. Chlorophyll deficiency delays but does not prevent melanogenesis in barley seed melanoplasts. *Protoplasma.* 2022;259(2):317-326. doi 10.1007/s00709-021-01669-3
- Panato A., Antonini E., Bortolotti F., Ninfali P. The histology of grain caryopses for nutrient location: a comparative study of six cereals. *Int J Food Sci Technol.* 2017;52(5):1238-1245. doi 10.1111/ijfs.13390
- Panzella L., Eidenberger T., Napolitano A., d'Ischia M. Black sesame pigment: DPPH assay-guided purification, antioxidant/antinitrosating properties, and identification of a degradative structural marker. *J Agric Food Chem.* 2012;60(36):8895-8901. doi 10.1021/jf2053096
- Parker M.L., Ng A., Waldron K.W. The phenolic acid and polysaccharide composition of cell walls of bran layers of mature wheat (*Triticum aestivum* L. cv. Avalon) grains. *J Sci Food Agric.* 2005;85(15): 2539-2547. doi 10.1002/jsfa.2304
- Quideau S., Deffieux D., Douat-Casassus C., Pouységu L. Plant polyphenols: chemical properties, biological activities, and synthesis. *Angew Chem Int.* 2011;50(3):586-621. doi 10.1002/anie.201000044
- Sava V.M., Yang S.-M., Hong M.-Y., Yang P.-C., Huang G.S. Isolation and characterization of melanic pigments derived from tea and tea polyphenols. *Food Chem.* 2001;73(2):177-184. doi 10.1016/S0308-8146(00)00258-2
- Shoeva O.Y., Mock H.-P., Kukoeva T.V., Börner A., Khlestkina E.K. Regulation of the flavonoid biosynthesis pathway genes in purple and black grains of *Hordeum vulgare*. *PLoS One.* 2016;11(10): e0163782. doi 10.1371/journal.pone.0163782
- Shoeva O.Yu., Strygina K.V., Khlestkina E.K. Genes determining the synthesis of flavonoid and melanin pigments in barley. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2018;22(3): 333-342. doi 10.18699/VJ18.369 (in Russian)
- Shoeva O.Y., Mursalimov S.R., Gracheva N.V., Glagoleva A.Y., Börner A., Khlestkina E.K. Melanin formation in barley grain occurs within plastids of pericarp and husk cells. *Sci Rep.* 2020;10:179. doi 10.1038/s41598-019-56982-y
- Solano F. Melanins: skin pigments and much more – types, structural models, biological functions, and formation routes. *New J Sci.* 2014; 5:1-28. doi 10.1155/2014/498276
- Takahashi I., Hara M. Enhancement of starch accumulation in plants by exogenously applied methyl jasmonate. *Plant Biotechnol Rep.* 2014;8:143-149. doi 10.1007/s11816-013-0304-1
- Treutter D. Chemical reaction detection of catechins and proanthocyanidins with 4-dimethylaminocinnamaldehyde. *J Chromatogr.* 1989; 467(1):185-193. doi 10.1016/S0021-9673(01)93963-9
- Varga M., Berkesi O., Darula Z., May N.V., Palágyi A. Structural characterization of allomelanin from black oat. *Phytochemistry.* 2016; 130:313-320. doi 10.1016/j.phytochem.2016.07.002
- Vermeris W., Nicholson R. Phenolic compounds and their effects on human health. In: *Phenolic Compound Biochemistry.* Springer, 2008;235-255. doi 10.1007/978-1-4020-5164-7\_7
- 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
- Wildi B.S. Silver nitrate as a test for ortho and para dihydric phenols. *Science.* 1951;113(2929):188. doi 10.1126/science.113.2929.188
- Winkel-Shirley B. Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. *Seed Sci Res.* 1998;8(4):415-422. doi 10.1017/S0960258500004372
- Wrolstad R.E., Durst R.W., Lee J. Tracking color and pigment changes in anthocyanin products. *Trends Food Sci Technol.* 2005;16(9): 423-428. doi 10.1016/j.tifs.2005.03.019
- Yu R., Wu X., Liu J., Howitt C.A., Bird A.R., Liu C.M., Larkin P.J. Rice with multilayer aleurone: a larger sink for multiple micronutrients. *Rice.* 2021;14(1):102. doi 10.1186/s12284-021-00543-3
- Zykin P.A., Andreeva E.A., Tsvetkova N.V., Voylovkov A.V. Anatomical and image analysis of grain coloration in rye. *Preprints.* 2020. doi 10.20944/preprints202011.0530.v1

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

Received May 25, 2025. Revised September 1, 2025. Accepted September 24, 2025.

doi 10.18699/vjgb-26-03

# Modification of the BphP1-QPAS1 optogenetic system for gene expression regulation in *Nicotiana benthamiana* tobacco leaves using near-infrared light

E.S. Surkova <sup>1, 2, §</sup>, Y.A. Galimova <sup>1, §</sup>, N.V. Battulina <sup>1</sup>, D.M. Motorina <sup>1</sup>, E.S. Omelina <sup>1</sup> <sup>1</sup> Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia omelina@mcb.nsc.ru

**Abstract.** In plants, the regulation of transgene transcription is typically achieved using chemical agents. A safe alternative to chemically induced systems may be optogenetic systems. The BphP1-QPAS1 system has distinct advantages over other optogenetic systems, as it is activated by near-infrared (NIR, 780 nm) light, which is beyond the spectrum of plant photoreceptors. This system is based on the use of a split transcription factor (TF), consisting of the DNA-binding and dimerization domains of the yeast TF Gal4, fused to the QPAS1 component, along with the transactivation domain VP16 fused to BphP1. Under NIR light, BphP1 interacts with QPAS1, leading to the formation of the functional TF Gal4-VP16. A primary obstacle to using optogenetic systems in plants is their undesired activation under white light, which is vital for normal plant growth. A potential solution to this issue is temporarily removing one component of the split TF from the nucleus under white light. We modified the BphP1-QPAS1 system to activate reporter gene expression in *Nicotiana benthamiana* leaves using NIR light. We combined BphP1-QPAS1 with several variants of LOV domain-containing proteins activated by blue light (460–480 nm). The best results were achieved by combining the BphP1-QPAS1 system with the AsLOV2 domain, which carries the degron sequence RRRG at the C-terminal  $\alpha$  helix and initiates the degradation of the chimeric protein NES-Gal4-QPAS1-AsLOV2-RRRG under white light. This modification induced the BphP1-QPAS1 system in tobacco leaves only under NIR light, but not in the dark or under white light. We believe that, in the future, the BphP1-QPAS1 system could be applied to enhance plant resistance to adverse environmental conditions, pests, and viral diseases.

**Key words:** optogenetic system; BphP1-QPAS1; VVD; AsLOV2; *Nicotiana benthamiana*; gene expression regulation; Gal4/UAS

**For citation:** Surkova E.S., Galimova Y.A., Battulina N.V., Motorina D.M., Omelina E.S. Modification of the BphP1-QPAS1 optogenetic system for gene expression regulation in *Nicotiana benthamiana* tobacco leaves using near-infrared light. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2026;30(1):61-71. doi 10.18699/vjgb-26-03

**Funding.** This research was funded by the Russian Science Foundation, grant No. 22-74-10118.

## Модификация оптогенетической системы BphP1-QPAS1 для регуляции экспрессии генов в листьях табака *Nicotiana benthamiana* с помощью ближнего инфракрасного света

Э.С. Суркова <sup>1, 2, §</sup>, Ю.А. Галимова <sup>1, §</sup>, Н.В. Баттулина <sup>1</sup>, Д.М. Моторина <sup>1</sup>, Е.С. Омелина <sup>1</sup> <sup>1</sup> Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия omelina@mcb.nsc.ru

**Аннотация.** В растениях регуляция транскрипции трансгенов обычно осуществляется с помощью химических агентов. Безопасной альтернативой химически индуцируемым системам могут служить оптогенетические (т. е. светоиндуцируемые) системы. Бактериальная система BphP1-QPAS1 выгодно отличается от других оптогенетических систем, поскольку активируется ближним инфракрасным светом (БИК, 780 нм), выходящим за пределы спектра восприятия растительных фоторецепторов. Система BphP1-QPAS1 основана на использовании «расщепленного» транскрипционного фактора (ТФ), состоящего из ДНК-связывающего и димеризующего доменов дрожжевого ТФ Gal4, слитого с компонентом QPAS1, и трансактивационного домена VP16, слитого с BphP1. Под действием БИК света BphP1 взаимодействует с QPAS1, что приводит к сближению Gal4 с VP16 и запуску экспрессии репортерного гена. Основным препятствием для широкого использования оптогенетических систем в растениях является нежелательная активация таких систем при воздействии дневного (белого) света, который включает широкий спектр длин волн и жизненно важен для нормального роста растений. Решением проблемы нежелательной активации может стать временное удаление из ядра клетки одного из компонентов

«расщепленного» ТФ при воздействии белого света. В данной работе мы модифицировали систему BphP1-QPAS1 для активации экспрессии репортерного гена в листьях табака *Nicotiana benthamiana* с помощью БИК света. Для этого мы комбинировали систему BphP1-QPAS1 с несколькими вариантами LOV доменов-содержащих белков, активируемых синим светом (460–480 нм). Наилучшие результаты были достигнуты при комбинации компонентов системы BphP1-QPAS1 с доменом AsLOV2 из белка фототропин 1 *Avena sativa*, несущим на C-концевой  $\alpha$  спирали последовательность дегрона RRRG и запускающим деградацию химерного белка NES-Gal4-QPAS1-AsLOV2-RRRG при воздействии белого света. Такая модификация активировала систему BphP1-QPAS1 в листьях табака только при воздействии БИК света, но не в темноте либо при белом свете. Мы полагаем, что в будущем система BphP1-QPAS1 может быть применена для повышения устойчивости растений к неблагоприятным условиям среды, вредителям, вирусным заболеваниям.

**Ключевые слова:** оптогенетическая система; BphP1-QPAS1; VVD; AsLOV2; *Nicotiana benthamiana*; регуляция экспрессии генов; Gal4/UAS

## Introduction

The use of non-channel light-sensitive proteins, which can change conformation or form homo- or heterodimeric complexes under specific wavelengths of light, has enabled the application of optogenetic systems for precise temporal transcription regulation (Shimizu-Sato et al., 2002; Wang et al., 2012; Konermann et al., 2013; Ochoa-Fernandez et al., 2016; Gligorovski et al., 2023). Light control may serve as a valuable alternative to chemically inducible gene expression systems. The advantages of optogenetic systems over chemical ones include the following: (1) the ability to focus light on a specific small area of an organism or cell; (2) control over the intensity and duration of light exposure; (3) activation or deactivation of the system by switching the light On or Off, respectively; (4) independence from the diffusion rate of a chemical agent; and (5) absence of Off-target effects caused by chemical inducers (Shimizu-Sato et al., 2002; Omelina et al., 2022).

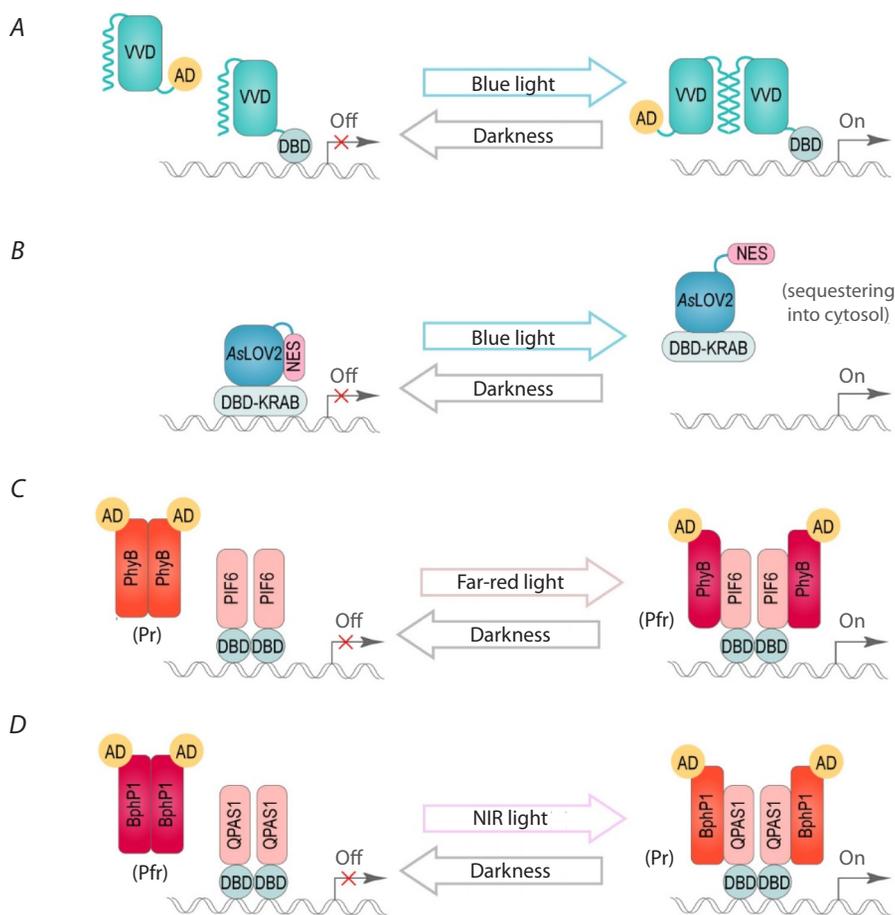
One of the strategies for regulating expression by light is based on photoreversible conformational changes (Fig. 1A, B). For instance, light illumination may induce the unfolding of one of the  $\alpha$ -helices in the LOV domain-containing proteins (Crosson, Moffat, 2001; Zayner et al., 2012), which can be used to expose an engineered peptide that controls intracellular protein localization or protein-protein interactions (Krueger et al., 2019). Another strategy is the use of split transcription factors (TFs), where one component is a DNA-binding domain of a TF fused to a light-sensitive protein, while its partner is fused to an activator or repressor domain. When exposed to light, heterodimers are formed, resulting in the generation of an active transcription factor (Fig. 1C, D) (Pathak et al., 2017; Noda, Ozawa, 2018; Omelina, Pindyurin, 2019; Hernández-Candia, Tucker, 2020).

One of the most popular systems for the activation and suppression of gene transcription is the LOV domain-containing proteins. LOV domains belong to the Per-Arnt-Sim (PAS) protein family (Herrou, Crosson, 2011). The small size (~11–15 kDa) and the presence of flavin chromophores in most, if not all, cell types are key advantages of these domains for the design of optogenetic tools (Wu Y.I. et al., 2009; Losi, Gärtner, 2011; Christie et al., 2012; Renicke et al., 2013; Niopek et al., 2016; Redchuk et al., 2017). One of the strategies for developing LOV-derived optogenetic tools is based on light-induced homodimerization. This approach utilizes the LOV domain-containing protein Vivid (VVD) from the filamentous fungus *Neurospora crassa* (Wang et al., 2012). VVD has an N-terminal helix that is docked on the core in the dark state.

Blue light (460–480 nm) leads to the dissociation of this helix from the VVD core and interaction with the N-terminal helix of another VVD, resulting in the formation of a VVD homodimer (Fig. 1A) (Zoltowski et al., 2007; Zoltowski, Crane, 2008). The VVD-based strategy was used to develop the LightON system, providing blue-light-induced transcription of target transgenes in mammalian cells and in mice (Wang et al., 2012). VVD was fused with the DNA-binding domain of the yeast TF Gal4 and the p65 transcription activation domain. It formed homodimers under blue light, resulting in the generation of the active dimer Gal4-p65 and the transcriptional activation of the UAS reporter genes. Dark relaxation caused dimer dissociation, meaning the system returned to the non-active state.

Another LOV domain is derived from the LOV2 domain of *Avena sativa* phototropin 1 (AsLOV2). In LOV2, illumination with blue light leads to the unwinding of the caged C-terminal  $\alpha$  helix and the exposure of the fused C-terminal peptide, thus making it available for protein-protein interactions (Peter et al., 2010; Lungu et al., 2012; Strickland et al., 2012; Yumerefendi et al., 2015; Kim et al., 2024; Kaya et al., 2025). This property of the LOV2 domain has been utilized to create a light-inducible nuclear export system (LEXY) (Niopek et al., 2016). In this system, blue light releases the photocaged nuclear export signal (NES) from the AsLOV2 core, thereby inducing the export of the protein of interest from the nucleus (Fig. 1B). To create this system, a library of 33 different NESs was generated, providing varying efficiencies of nuclear export for the chimeric protein NLS-mCherry-LEXY from the nucleus to the cytoplasm under blue light. This system has been tested in various mammalian cell cultures, and it has been demonstrated that during the dark relaxation period, the protein NLS-mCherry-LEXY returns to the nucleus, while the export of the protein from the nucleus can be repeatedly induced under blue light. An interesting application of LEXY is the spatiotemporal control of transcriptional repressors by sequestering them into the cytosol, which could be of particular value for controlling reporter gene activity (Fig. 1B) (Niopek et al., 2016).

Phytochromes are also a widely used optogenetic tool for the activation and suppression of gene transcription. These photoreceptors are found in plants, bacteria, cyanobacteria, and fungi, playing essential roles in light-adaptive processes. They share common domains in a photosensory core module and phytochrome-specific domains (Wagner et al., 2007; Yang et al., 2008; Chernov et al., 2017). Bacterial and fungal phytochromes incorporate a tetrapyrrole biliverdin IX $\alpha$  (BV)



**Fig. 1.** Different strategies for regulating gene expression using light.

**A,** A schematic representation of the formation of homodimers due to conformational changes in the LOV domain of the VVD protein induced by blue light illumination. The N-terminal helix dissociates from the VVD core and interacts with the N-terminal helix of another VVD under blue light. As a result, the DBD and AD of the split TF associate into a functional TF, leading to the transcription of the reporter gene. **B,** A schematic representation of the unfolding of the C-terminal  $\alpha$  helix from the AsLOV2 domain and the exposure of the fused NES signal, leading to the relocalization of the transcriptional repressor (DBD-KRAB) from the nucleus into the cytoplasm to control gene activity. **C, D,** A schematic representation of the formation of heterodimers PhyB-PIF6 under far-red light (**C**) or BphP1-QPAS1 under NIR light (**D**), and their dissociation in darkness. Upon light induction, the DBD interacts with the AD, resulting in the transcription of the reporter gene. The Pr and Pfr states refer to two interconvertible forms of phytochromes. AD, activating domain; DBD, DNA-binding domain; KRAB, Kruppel associated box; NES, nuclear export signal.

as a chromophore. In plants and cyanobacteria, BV is further enzymatically reduced to phytochromobilin (PФВ) or phycocyanobilin (PCB), which bind to plant and cyanobacterial phytochromes, respectively. Phytochromes exist in one of two interconvertible states: either the Pr state or the Pfr state (Fig. 1C, D). The Pr state absorbs light at 660–700 nm, whereas the Pfr state absorbs light at 740–780 nm (Li J. et al., 2011; Chernov et al., 2017). Typically, the ground (inactive) state of phytochromes is Pr, which converts to the active Pfr state upon illumination with 660–680 nm light (Fig. 1C) (Shcherbakova et al., 2015). A specific group of phytochromes, termed bathy, adopts Pfr as the ground state (Fig. 1D). Many photoreversible dimerization tools have been developed based on the light-regulated interaction of plant phytochrome with basic helix–loop–helix proteins known as phytochrome-interacting factors (PIFs). The plant phytochromes PhyA and PhyB from *Arabidopsis thaliana* are readily activated by far-red (640–680 nm) light and deactivated by near-infrared (NIR)

(740–760 nm) light or in darkness. In the active Pfr state, PhyA and PhyB interact with PIF3 or PIF6 (Fig. 1C) (Ni et al., 1999; Zhu et al., 2000). A far-red light-inducible gene expression system in yeast was developed by fusing the components of the PhyB-PIF3 system to the DNA-binding and transactivation domains of the Gal4 TF (Shimizu-Sato et al., 2002). The interaction of PhyA and PhyB with PIF3 was exploited to develop a light-switchable gene expression system in mammalian cells (Müller et al., 2013). The PhyB-PIF6 system for activating transcription in plants has also been reported (Ochoa-Fernandez et al., 2020). PhyB is known to be very light-sensitive and can be easily activated even in dim white light. To prevent this, the authors combined the PhyB-PIF6 protein pair with a LOV-based optogenetic system. The system was successfully tested in tobacco and *Arabidopsis*.

Among phytochromes, a subclass of bacterial phytochrome photoreceptors (BphPs) that incorporate BV should be emphasized (Giraud, Verméglio, 2008; Aldridge, Forest, 2011;

Piatkevich et al., 2013). BV has the most red-shifted absorbance relative to other tetrapyrroles, including PΦB and PCB. The first BphP1-PpsR2 optogenetic system for transcriptional activation in eukaryotic cells was developed based on the bacterial bathy phytochrome BphP1 from the nonsulphur purple bacterium *Rhodospseudomonas palustris*. Under NIR light, BphP1 interacts with its natural partner protein PpsR2. This property was utilized to create a transcription activation system in mammalian cell cultures and mice (Kaberniuk et al., 2016). The original BphP1-PpsR2 system was optimized by reducing PpsR2 to a single-domain binding partner, QPAS1, which is three times smaller than full-length PpsR2, includes only the  $\alpha$ -helical Q-linker and PAS1 domain and lacks oligomerization (Redchuk et al., 2017). The pair of proteins BphP1 and QPAS1 was used to create NIR light-induced systems for the activation (Fig. 1D) and suppression of gene expression, as well as for the control of protein localization. These systems were tested in various mammalian cell cultures (Redchuk et al., 2017, 2018).

Previously, optogenetic systems for the activation of expression in plants utilized the plant phytochrome PhyB (Müller et al., 2014; Ochoa-Fernandez et al., 2016, 2020) or the bacterial photoreceptor CarH (Chatelle et al., 2018) activated by far-red (640–680 nm) or green light (525 nm), respectively (Omelina et al., 2022). A significant drawback of these systems is the potential effect on endogenous plant photoreceptors (Müller et al., 2014; Battle, Jones, 2020; Christie, Zurbriggen, 2021). In this study, for the regulation of the reporter gene transcription in plants, we applied the bacterial BphP1-QPAS1 system activated by NIR light (780 nm), which is outside the spectra of plant photoreceptors' excitation and differs from the wavelength required for their reversion. BphP1 is known to incorporate BV as a chromophore, which is synthesized in chloroplasts as a precursor of PΦB, suggesting the usability of BV-required tools in plants (Shikata, Denninger, 2022). However, the application of optogenetic systems in plant research is known to have a significant obstacle. Plants are phototrophic organisms and require exposure to white light, which may non-specifically activate optogenetic systems (Andres et al., 2019). To block the activity of the BphP1-QPAS1 system under white light, we combined the BphP1-QPAS1 protein pair with blue light-controlled systems (the LOV domain-containing VVD protein or *AsLOV2* domain fused to several variants of the NES and degron sequences). The best results were observed for the combination of the BphP1-QPAS1 system with the *AsLOV2* domain fused to a degron sequence. In the leaves of these transiently transformed tobacco plants, we observed strong fluorescence of the plant EGFP (pEGFP) reporter under NIR light and found no pEGFP signal when the plants were incubated in standard day/night conditions.

We believe that in the future, the optogenetic system BphP1-QPAS1 will enable the development of several approaches for creating safe and highly productive agricultural plants that are resistant to stress conditions, pests, and viral diseases, which are common among agricultural crops. Unlike methods that use various chemicals to combat pests and infectious diseases, optogenetic biotechnologies are non-toxic and facilitate safe and high-quality food production.

## Materials and methods

**Design of plasmids.** All plasmids generated in this study are presented in the Table.

*Control reporter plasmids without the activator (R1, R2).* Two variants of reporter plasmids R1 and R2 (Fig. 2A, B; Supplementary Figs. 1, 2)<sup>1</sup> were obtained using the Gibson assembly method based on the pBI121 vector (Chen et al., 2003), which was hydrolyzed at the SbfI/ClaI restriction sites. For that, the UAS sequence was amplified from the pUASTattB vector (GenBank EF362409.1) (Bischof et al., 2007). The pEGFP coding sequence, fused with the NOS terminator, and the sequence of the minimal promoter from the Cauliflower mosaic virus (mini35S) were amplified from the pCAMBIA\_pGFP vector (Gayatri, Basu, 2020).

*Control reporter plasmids with the activator (R1A, R2A).* The sequences of the full-length promoter of the Cauliflower mosaic virus (CaMV35S) and light-independent activator Gal4-VP16 were added to the R1 and R2 plasmids to obtain the R1A and R2A plasmids. The CaMV35S promoter sequence was amplified from the pBI121 vector. The sequence of the activator Gal4-VP16 (Fig. 2C; Suppl. Fig. 3) was amplified from the plasmid pGal4-VP16 (Redchuk et al., 2017). The NOS terminator sequence was amplified from the pCAMBIA\_pGFP vector. All fragments were inserted into the reporter plasmids R1 or R2 using the Gibson assembly method after being hydrolyzed at the PmeI/ClaI restriction sites.

*Control reporter plasmids without the full-length CaMV35S promoter from the pBI121 plasmid.* To eliminate the influence of the CaMV35S promoter from the pBI121 plasmid on the expression of pEGFP, this fragment was excised from R1, R2, R1A and R2A plasmids at the SbfI/BamHI sites. Subsequently, the vectors were treated with a Klenow fragment (NEB), followed by self-ligation. These constructs were named R1d, R2d, R1Ad and R2Ad, respectively.

*Plasmid carrying the coding sequences of the BphP1-VP16, NLS-Gal4-QPAS1-VVD and VVD-CAAX chimeric proteins (BQV).* To generate the BQV plasmid (Suppl. Fig. 4), the components of the BphP1-QPAS1 system were amplified from the plasmid pQP-T2A (Addgene #102583 (Redchuk et al., 2017)). The coding sequence of the VVD protein was amplified from the GAVPO construct (Wang et al., 2012). The NLS and CAAX sequences were included in the primers. These fragments were inserted into the reporter plasmid R2d using the Gibson assembly method after being hydrolyzed at the PmeI/ClaI restriction sites.

*Plasmids carrying the coding sequences of the BphP1-VP16 and NES-Gal4-QPAS1-AsLOV2-NES21 or NES-Gal4-QPAS1-AsLOV2-NES27 or NES-Gal4-QPAS1-AsLOV2-RRRG chimeric proteins (BQL1, BQL2, BQL3, respectively).* To generate the BQL1, BQL2 and BQL3 plasmids (Suppl. Fig. 5), the sequences of the NOS and CaMV35S promoters, the NOS terminator, and the NeoR/KanR sequence were amplified from the pBI121 vector. The components of the BphP1-QPAS1 system were amplified from the plasmid pQP-T2A. The Cauliflower mosaic virus polyadenylation signal sequence was amplified from the pCAMBIA\_pGFP vector. The *AsLOV2*

<sup>1</sup> Supplementary Figures 1–5 are available at:  
<https://vavilovj-icg.ru/download/pict-2026-30/appx5.zip>

sequence was amplified from the plasmid pKM546 (Müller et al., 2015). The N-terminal NES and C-terminal NES21, NES27, and degron sequences were included in the primers. These fragments were inserted into the BQV plasmid using the Gibson assembly method after being hydrolyzed at the PmeI/HindIII restriction sites.

**Preparation of chemically competent *Agrobacterium tumefaciens* cells.** An overnight culture (3 ml) of *A. tumefaciens* strain GV3101 was grown in sterile YEP medium (10 g/l yeast extract, 10 g/l Bacto Peptone, 5 g/l NaCl) for 16–20 hours at 28 °C. Subsequently, 2 ml of this culture was transferred to a flask containing 50 ml of sterile YEP medium and incubated at 28 °C until an optical density at 600 nm (OD<sub>600</sub>) of 0.5–1.0 was reached (typically several hours to overnight). The culture was then cooled on ice for 30 minutes. Cells were harvested by centrifugation at 5,000g at 4 °C for 8 minutes. The supernatant was discarded, and the cell pellet was gently resuspended in 2 ml of ice-cold 20 mM CaCl<sub>2</sub>. The suspension was incubated on ice for an additional 30 minutes. Aliquots of 100 µl were dispensed into pre-chilled microcentrifuge tubes, flash-frozen in liquid nitrogen, and stored at –70 °C.

***A. tumefaciens* transformation.** Chemically competent *A. tumefaciens* strain GV3101 was transformed with the plasmids of interest (see the Table). Specifically, 1 µg of plasmid DNA was added to 100 µl of competent cells. These cells were thawed for 5 minutes at 37 °C, after which 1 ml of YEP medium supplemented with 50 µg/ml kanamycin and 100 µg/ml rifampicin was added. The mixture was incubated with shaking for 3 to 5.5 hours at 28 °C. The cells were then centrifuged at 1,000g for 5 minutes, after which the pellet was resuspended in 100 µl of YEP medium and plated on YEP agar plates containing kanamycin and rifampicin. Plates were incubated in an inverted position at 28 °C for 72 hours until individual colonies appeared. Colonies were confirmed by colony PCR using Taq DNA polymerase (Biolabmix).

**Transient transformation of *Nicotiana benthamiana* plants.** *A. tumefaciens* cultures were adjusted to an optical density measured at a wavelength of 600 nm (OD<sub>600</sub> = 0.5) in infiltration medium (10 mM MgCl<sub>2</sub>, 10 mM MES in H<sub>2</sub>O). The cultures were incubated for 1–3 hours at 28 °C prior to infiltration through the adaxial surface of the leaves of 4- to 5-week-old *N. benthamiana*.

**Illumination conditions and microscopy.** To test the Gal4/UAS system (plasmids #1–8 in the Table) in tobacco leaves, the plants were kept in a plant incubator with fluorescent tubes (cool white light, OSRAM), employing alternating cycles of light and darkness (16 hours of light and 8 hours of darkness) to simulate standard day/night conditions for 48 hours.

To analyze the modified BphP1-QPAS1 system variants, *N. benthamiana* plants transiently transformed with the plasmids BQV, BQL1, BQL2, or BQL3 (see the Table) were grown for 48 hours in simulated day/night conditions (16 hours of white light and 8 hours of darkness). Thereafter, half of the plants were incubated under NIR light for 24 hours to induce pEGFP expression, while the other half continued to grow under the same day/night conditions.

NIR illumination was performed using custom-assembled LED arrays with a wavelength of 780/20 nm. For microscopic

analysis, we conducted at least three independent experiments, taking 1–3 samples from different areas of the tobacco leaves and using a Zeiss Axio Imager M2 (Carl Zeiss).

## Results

### Testing of light-independent expression of the control reporter constructs in tobacco leaves

To test the functioning of the optogenetic system BphP1-QPAS1 in plants, we first selected an optimal structure for the regulatory region of the UAS enhancer-based reporter construct. This structure should provide minimal reporter activity in the absence of the activator and a high level of reporter gene expression in the presence of the activator. The creation of different variants of the reporter construct for use in *N. benthamiana* is necessary due to the lack of a single version of the UAS regulatory regions of reporter genes in plants, as their structure depends on the plant species (Wu C. et al., 2003; Johnson A.A.T. et al., 2005; Sakvarelidze et al., 2007; Yun et al., 2023). Additionally, the DNA-binding activity of Gal4 is known to be sensitive to the methylation of its binding site in plant chromatin, which can result in weak expression or even the loss of Gal4-mediated expression (Gälweiler et al., 2000). To develop these control reporter constructs, we employed the widely used binary *Agrobacterium* vector pBI121, which carries a GUS (*β-glucuronidase*) reporter gene for plant transformation (Chen et al., 2003). Based on the pBI121 vector backbone, we designed two variants of the control reporter plasmids, R1 (Fig. 2A; the Table; Suppl. Fig. 1) and R2 (Fig. 2B; the Table; Suppl. Fig. 2). Both plasmids contain five UAS repeats for the binding of the TF Gal4-VP16 and the reporter gene *pEGFP*. The R2 plasmid additionally contains the sequence of the minimal promoter from the Cauliflower mosaic virus (CaMV35S) – mini35S.

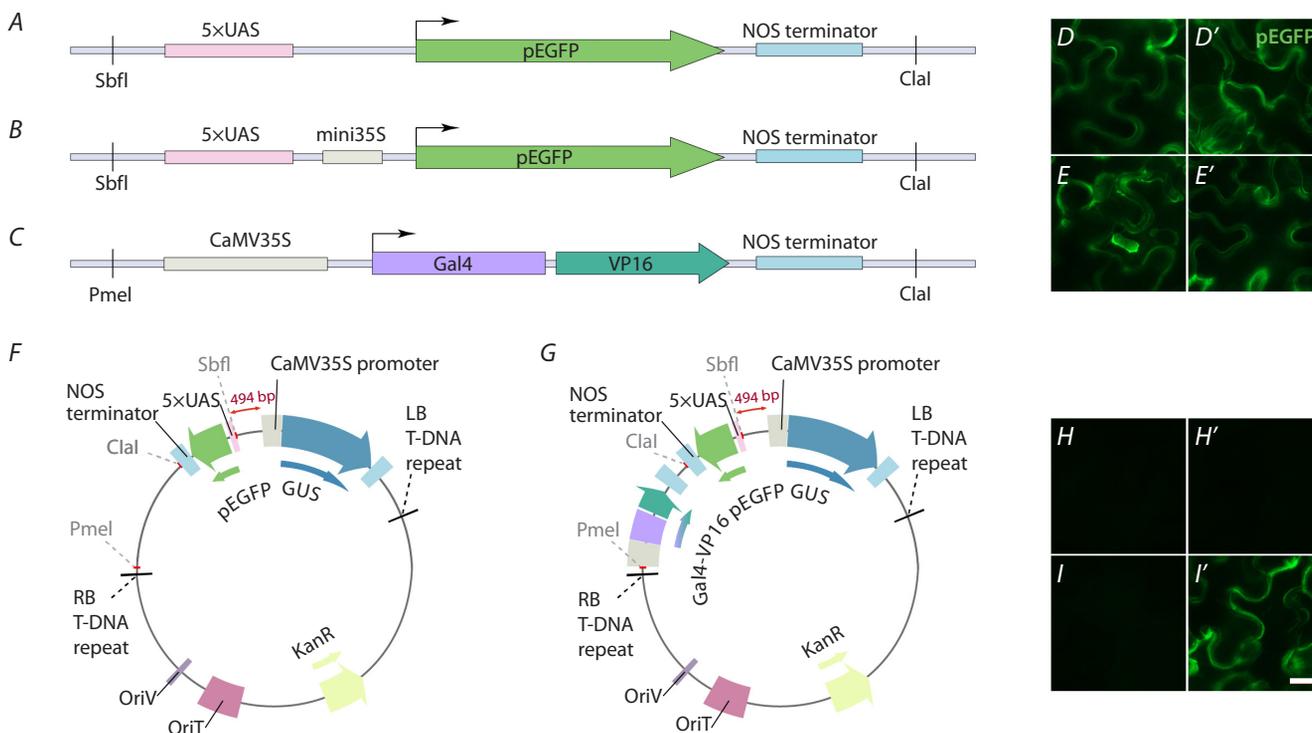
The sequence of the light-insensitive Gal4-VP16 activator was added to the reporter plasmids R1 and R2, resulting in the creation of the R1A and R2A plasmids (see the Table). The Gal4-VP16 activator consists of the DNA-binding and dimerization domains of the yeast TF Gal4 (148 a. a.) fused with the transactivation domain VP16 (Fig. 2C; Suppl. Fig. 3). For testing in plants, the plasmids R1, R2, R1A, and R2A were used to transform *A. tumefaciens* cells, which were subsequently used to infiltrate the leaves of the tobacco *N. benthamiana*. A microscopic analysis of pEGFP protein expression in the transiently transformed tobacco leaves was conducted 48 hours after infiltration. We anticipated observing pEGFP expression in the tobacco leaves transformed with one of the variants of the reporter plasmids in the presence of the Gal4-VP16 activator (R1A or R2A), but not in its absence (R1 or R2) (see the Table). However, all the analyzed constructs exhibited strong pEGFP expression (Fig. 2D–E').

As we mentioned above, we cloned the reporter plasmids based on the pBI121 vector backbone. This vector carries the GUS coding sequence under the control of the full-length CaMV35S promoter located 494 bp from the UAS repeats in the reverse orientation (Fig. 2F, G). We did not initially remove the GUS sequence and its promoter from the vector, assuming that its expression would not interfere with the analysis of

Plasmids for testing the Gal4/UAS (#1–8) and optogenetic (#9–12) systems in tobacco leaves

#	Name	Key element(s) (reporter, activator, components of the optogenetic system)	Bb	Cloning method, restriction sites	Presence of the CaMV35S promoter*
1	R1	5xUAS-pEGFP	pBI121	GA, SbfI/Clal	+
2	R2	5xUAS-mini35S-pEGFP	pBI121	GA, SbfI/Clal	+
3	R1A	5xUAS-pEGFP, CaMV35S:Gal4-VP16	R1	GA, PmeI/Clal	+
4	R2A	5xUAS-mini35S-pEGFP, CaMV35S:Gal4-VP16	R2	GA, PmeI/Clal	+
5	R1d	5xUAS-pEGFP	R1	Restriction and self-ligation, SbfI/BamHI	-
6	R2d	5xUAS-mini35S-pEGFP	R2	Restriction and self-ligation, SbfI/BamHI	-
7	R1Ad	5xUAS-pEGFP, CaMV35S:Gal4-VP16	R1A	Restriction and self-ligation, SbfI/BamHI	-
8	R2Ad	5xUAS-mini35S-pEGFP, CaMV35S:Gal4-VP16	R2A	Restriction and self-ligation, SbfI/BamHI	-
9	BQV	5xUAS-mini35S-pEGFP, CaMV35S:BphP1-VP16-T2A-VVD-CAAX, NOS:NLS-Gal4-QPAS1-VVD	R2d	GA, PmeI/Clal	-
10, 11, 12	BQL1, BQL2, BQL3	5xUAS-mini35S-pEGFP, CaMV35S:BphP1-VP16-T2A-NeoR/KanR, NOS:NES-Gal4-QPAS1-AsLOV2-NES21/NES27/degroN	BQV	GA, PmeI/HindIII	-

Note. Bb, backbone; GA, Gibson assembly method. \* The full-length promoter CaMV35S from the pBI121 vector backbone, which controls expression of the GUS reporter gene.



**Fig. 2.** Structure of the control reporter plasmids for testing the Gal4/UAS system in transiently transformed *N. benthamiana* leaves.

A, Reporter construct R1 without a minimal promoter. B, Reporter construct R2, containing the sequence of the minimal CaMV35S promoter – mini35S. C, Scheme of the light-independent Gal4-VP16 activator, incorporated into the reporter plasmids R1 and R2 for the activation of the reporter gene. It consists of the DNA-binding and dimerization domains of the yeast TF Gal4 (148 a.a.) fused with the transactivation domain VP16. The plasmid elements are not shown to scale. D–E', Microscopy images of *N. benthamiana* leaves transiently transformed with the constructs R1 (D), R1A (D'), R2 (E), R2A (E'). F, A scheme of the reporter construct R1 (analogous to R2) based on the pBI121 vector backbone. G, A scheme of the construct R1A (analogous to R2A) with the Gal4-VP16 activator based on the pBI121 vector backbone. F, G, The full-length CaMV35S promoter, which regulates the activity of the GUS gene, is located 494 bp from the UAS repeats in the reverse orientation. H–I', Microscopy images of *N. benthamiana* leaves transiently transformed with the constructs R1d (H), R1Ad (H'), R2d (I), R2Ad (I'). Plants were incubated in simulated day/night conditions for 48 hours (D–E', H–I'). Scale, 20 µm.

the pEGFP fluorescence. However, a thorough review of the literature indicated that the full-length CaMV35S promoter could alter the activity level of adjacent genes located approximately 2 kb downstream of the promoter (Zheng et al., 2007). Furthermore, this promoter can act as an enhancer in either orientation and lead to an increase in the expression level of neighboring genes (Zheng et al., 2007). To eliminate the presumed influence of the CaMV35S promoter on the expression of the pEGFP reporter, we removed its sequence from plasmids R1, R2, R1A and R2A. As a result, we obtained two new variants of the reporter constructs (R1d and R2d) and two constructs that carry both the reporter gene pEGFP and the activator Gal4-VP16 (R1Ad and R2Ad) (see the Table). Analysis of the activity of these constructs in tobacco leaves showed that pEGFP protein expression was only observed in the leaves transformed with the R2Ad construct with the reporter containing the mini35S sequence in the presence of the Gal4-VP16 activator (Fig. 2I'). In the case of the other constructs, the level of the pEGFP signal corresponded to autofluorescence (Fig. 2H–I). Therefore, we subsequently used the R2d reporter plasmid for the application of the optogenetic system BphP1-QPAS1 in plants.

### Optogenetic control of gene expression with the BphP1-QPAS1 system in transiently transformed *N. benthamiana* leaves

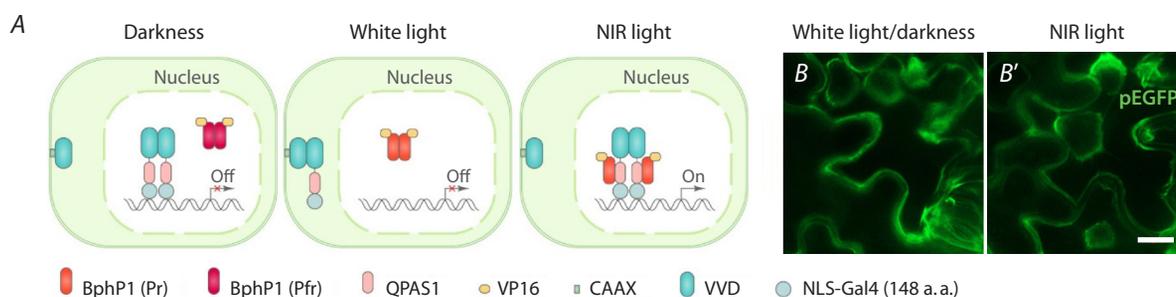
The application of optogenetic systems in plant research is known to face a significant obstacle in the form of undesired activity under ambient white light, which is vital for normal plant growth and development (Braguy, Zurbriggen, 2016). The original BphP1-QPAS1 system is based on the split TF Gal4-VP16, which is assembled by exposure to NIR light. The DNA-binding domain of Gal4 is fused to QPAS1 (Gal4-QPAS1) and BphP1 is fused with the VP16 transactivation domain (BphP1-VP16) (Fig. 1D) (Redchuk et al., 2017). To address the problem of undesired activation, we fused the Gal4-QPAS1 chimeric protein with several variants of blue

light-sensitive LOV domains. This approach might provide the removal of the Gal4-QPAS1 component from the cell nucleus under white light.

We combined the components of the BphP1-QPAS1 system with the VVD protein (see the BQV plasmid in the Table), which forms homodimers under blue light (or white light covering a broad spectrum of different wavelengths, including blue light at 460–480 nm). To achieve this, we fused the DNA-binding protein NLS-Gal4-QPAS1 with VVD, resulting in the chimeric protein NLS-Gal4-QPAS1-VVD (Fig. 3A). We also added to the BQV plasmid an additional VVD-containing construct, VVD-CAAX, which facilitates the attachment of the VVD protein to the plasma membrane via the CAAX peptide (Fig. 3A; Suppl. Fig. 4) (Redchuk et al., 2017; Willumsen et al., 1984a, b). We hypothesized that in this modified variant of the optogenetic system, BphP1-VP16 would predominantly reside in the nucleus of the cells (Fig. 3A). In darkness, the system exists in a state of relaxation. Under white light, the chimeric proteins NLS-Gal4-QPAS1-VVD and VVD-CAAX form dimers due to VVD homodimerization, resulting in the relocation of the chimeric protein NLS-Gal4-QPAS1-VVD to the membrane. Under NIR light, BphP1 interacts with QPAS1, leading to the activation of the reporter gene (Fig. 3A).

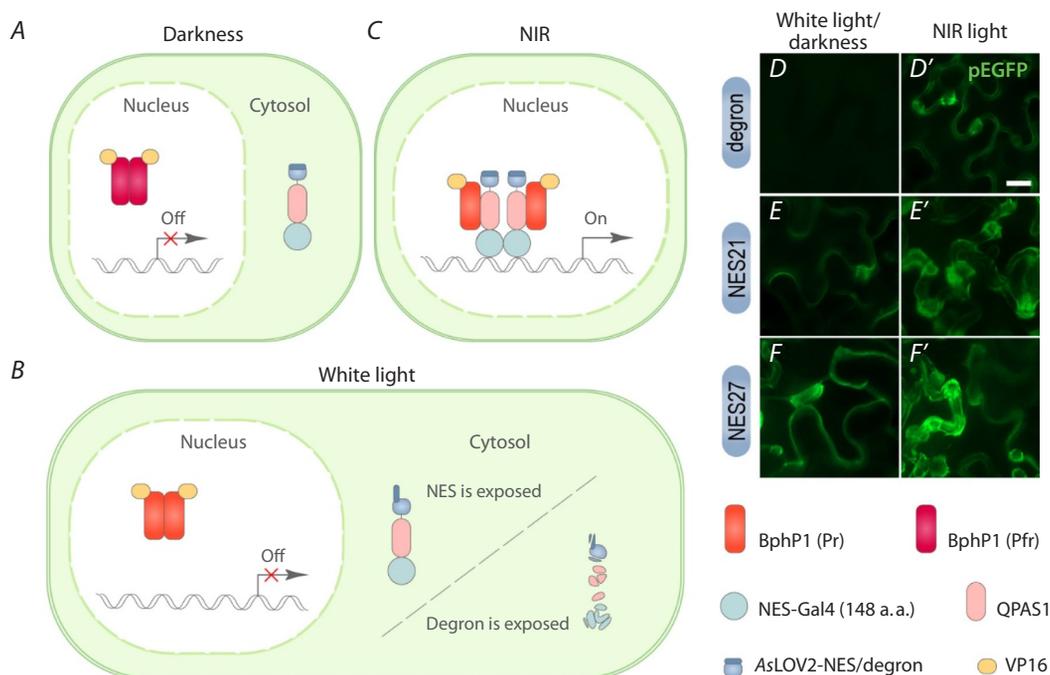
To test the functionality of the BphP1-QPAS1 system combined with VVD, we transformed *Agrobacterium* with the BQV plasmid (see the Table), followed by infiltration of the leaves of tobacco plants. The plants were incubated in simulated day/night conditions (16 hours of light and 8 hours of darkness) for 48 hours. Thereafter, half of the plants were incubated under NIR light for 24 hours, while the other plants continued to grow in simulated day/night conditions. As a result, we detected pEGFP fluorescence in the transiently transformed tobacco leaves incubated both under NIR light and in white light/dark conditions (Fig. 3B, B').

It should be noted that a standard practice for testing optogenetic systems involves incubating samples under different light conditions: in darkness and under pulsed light (e. g.,



**Fig. 3.** NIR light-controlled transcription activation of the reporter gene using the BphP1-QPAS1 system in combination with the VVD protein.

A, A scheme of the proposed functioning of the BphP1-QPAS1 system under different light conditions. In darkness, BphP1 exists in the inactive Pfr state and cannot interact with QPAS1. Under white light, BphP1 converts to the active Pr state. However, it cannot interact with QPAS1, as the NLS-Gal4-QPAS1-VVD chimeric protein relocates to the plasma membrane due to dimer formation with VVD-CAAX. Exposure to NIR light leads to the interaction of the BphP1 and QPAS1 components in the nucleus and activation of the reporter gene expression. B, B', Microscopy images of *N. benthamiana* leaves transiently transformed with the plasmid BQV. The plants were incubated in simulated day/night conditions for 72 hours (B) or in simulated day/night conditions for 48 hours, followed by incubation under NIR light for 24 hours (B'). Scale, 20  $\mu$ m.



**Fig. 4.** Two strategies of the BphP1-QPAS1 system functioning to avoid undesired activation in simulated day/night conditions.

A, A schematic representation of the proposed functioning of the modified BphP1-QPAS1 system in darkness. BphP1-VP16 exists in the ground Pfr state and predominantly resides in the nucleus of the cells. The NES-Gal4-QPAS1-AsLOV2-NES/degron localizes in the cytoplasm due to a weak N-terminal NES signal. B, A scheme of the proposed functioning of the modified BphP1-QPAS1 system under white light. BphP1-VP16 possibly exists in the active Pr state. However, it cannot interact with QPAS1, as under white light, the NES or degron signal unwinds from the AsLOV2 core. This results in the relocalization from the nucleus to the cytoplasm or degradation of the NES-Gal4-QPAS1-AsLOV2-NES/degron chimeric protein. C, A scheme of the proposed functioning of the modified BphP1-QPAS1 system under NIR light. BphP1 exists in the active Pr state and interacts with QPAS1, leading to the formation of the functional TF Gal4-VP16 and the activation of the reporter gene expression. D–F', Microscopy images of *N. benthamiana* leaves transiently transformed with plasmids carrying the components of the BphP1-QPAS1 system in combination with AsLOV2 fused to different peptides (degron in D, D'; NES21 in E, E'; NES27 in F, F'). The plants were incubated in simulated day/night conditions for 72 hours (D, E, F) or in simulated day/night conditions for 48 hours, followed by incubation under NIR light for 24 hours (D', E', F'). Scale bar, 20 μm.

30 seconds On and 180 seconds Off). Initially, after the transient transformation of the tobacco leaves with the BQV plasmid, we employed various light conditions. We kept the plants either in darkness or under continuous blue (480 nm), NIR (780 nm), or white light for 48 hours. However, these conditions were not optimal for the normal growth of the plants, and they exhibited wilting. Therefore, in all subsequent experiments, we decided to simulate the most natural lighting conditions with a day/night cycle during the first 48 hours after transformation. Thereafter, half of the plants were incubated under NIR light for 24 hours to induce pEGFP expression, while the remaining plants continued to grow in simulated day/night conditions.

Thus, the combination of BphP1-QPAS1 with the used variant of the VVD protein appears to be ineffective in plants, as the addition of VVD did not eliminate the undesired activity of the BphP1-QPAS1 system under white light (and possibly in darkness). To address this issue, we first replaced the N-terminal nuclear localization signal (NLS) with a weak NES signal in the Gal4-QPAS1 component in the BQV plasmid. It may relocalize part of this chimeric protein from the cell nucleus, which could potentially reduce the level of undesired activation in darkness. Secondly, we replaced VVD with the

blue-light-sensitive *AsLOV2* domain, which carries a specific peptide on its C-terminal  $\alpha$  helix. Utilizing the different peptides allowed for two strategies: (1) white light-induced nuclear export of the chimeric protein NES-Gal4-QPAS1-*AsLOV2*-NES; (2) white light-induced degradation of the chimeric protein NES-Gal4-QPAS1-*AsLOV2*-degron (Fig. 4A–C). To relocalize the NES-Gal4-QPAS1-*AsLOV2*-NES chimeric protein from the nucleus to the cytosol, we used two different NES signals, presented in plasmids BQL1 and BQL2 (see the Table; Suppl. Fig. 5). These signals (NES21 and NES27, respectively) were taken from a library of NES signals, which, in fusion with *AsLOV2*, have shown different efficiencies of nucleocytoplasmic translocation under blue light (Niopek et al., 2016). For light-inducible degradation of the NES-Gal4-QPAS1-*AsLOV2*-degron protein, the plasmid BQL3 was generated (see the Table; Suppl. Fig. 5). For this strategy, we utilized the previously described B-LID system carrying the degron variant RRRG (Bonger et al., 2014).

We transformed *Agrobacterium* with the plasmids BQL1, BQL2, and BQL3, followed by the infiltration of tobacco leaves. The plants were incubated for 48 hours in simulated day/night conditions. Thereafter, half the plants were incubated under NIR light for 24 hours, while the others continued to

grow in simulated day/night conditions for a further 24 hours. The strategy based on white light-sensitive degradation enabled the avoidance of undesired activation of the BphP1-QPAS1 system under white light and in darkness. We observed a strong pEGFP signal under NIR light and did not detect any pEGFP fluorescence in plants grown under standard day/night conditions (Fig. 4D, D'). The strategy based on white light-induced nucleocytoplasmic translocation was much less effective. Both variants of the NES signals displayed pEGFP fluorescence under NIR light as well as in alternating cycles of white light and darkness (Fig. 4E–F').

## Discussion

The undesired activation of optogenetic systems under ambient white light is known to be a significant obstacle to using these systems in plant research, as white light is vital for normal plant growth. This explains the relatively low number of articles on the application of optogenetic systems in plants compared to studies employing chemically inducible systems (Omelina et al., 2022). However, optogenetic systems offer several undeniable advantages over chemical systems, making their application in plants very appealing. To address the problem of undesired activation, optogenetic systems may be combined with other light-sensitive proteins that have higher light sensitivity to block undesired activity under white light.

In this study, we first applied the BphP1-QPAS1 system in the leaves of *N. benthamiana* to induce the expression of a reporter gene using NIR light. To avoid undesired activity of the BphP1-QPAS1 system under white light in plants, we combined it with the blue light-sensitive VVD protein (Wang et al., 2012). Previous studies in mice have demonstrated that VVD is more light-sensitive compared to the BphP1-QPAS1 protein pair (Kaberniuk et al., 2016). However, in plants, the combination of these systems did not yield the desired result. This may be related to the previously described dependence of the photoinducible dimerization properties of VVD on temperature (Qian et al., 2023). In this study, we used a variant of the VVD protein bearing the mutations N56K and C71V, which have previously shown a high On/Off ratio at 37 °C and a low On/Off ratio at 25 °C in HEK293 cells (Qian et al., 2023). The combination of the BphP1-QPAS1 system with the VVD protein is very appealing for use in plants, as all components are orthogonal to plant cells. Using another mutant version of the VVD protein or other localization signals in the future might enable avoiding undesired activation of the BphP1-QPAS1 system in combination with VVD under white light. We replaced the VVD protein with the plant-derived *AsLOV2* domain. As a result, a variant of the *AsLOV2* domain fused with the degron sequence has shown very good results. We observed a high level of pEGFP fluorescence after NIR light illumination and did not detect the pEGFP signal in plants grown in simulated day/night conditions.

It is important to note that light-sensitive proteins of plant origin have been applied in plants before. For instance, the PULSE system, which was previously successfully tested in plants, is based on the use of the plant phytochrome PhyB (Müller et al., 2014; Ochoa-Fernandez et al., 2020). In this system, the problem of undesired activation under white light

was solved by the addition of the bacterial LOV-domain protein EL222 fused with the repressor domain SRDX and binding to the same promoter of the reporter gene as the PhyB-PIF6-based split TF. We propose that the BphP1-QPAS1 system has several advantages compared to the PhyB-based system, as its activation occurs at NIR light (780 nm) rather than at 640–680 nm. NIR light is preferable not only because it is not toxic to eukaryotic cells, including plant cells, but also because it does not affect the excitation or relaxation of endogenous phytochromes (Johnson M.P., 2016; Guidi et al., 2017). Additionally, passive activation of degradation by white light (but not blue light) does not affect the activation of endogenous phototropins and cryptochromes. The chromophores required for photoactivation of BphP1 and *AsLOV2* (BV and flavin chromophore, respectively) are plant metabolites and are present in sufficient quantities in plant cells. In contrast to BphP1-QPAS1, the functioning of other optogenetic systems may lead to the potential activation of endogenous photoreceptors, which play a crucial role in plant development. Moreover, it should also be noted that the BphP1-based system is significantly less sensitive to ambient light compared to PhyB-derived optogenetic tools (Kaberniuk et al., 2016). Furthermore, the PhyB protein is present in all seed plants (Mathews, 2010; Li F.W. et al., 2015), so the interaction of the components of PhyB-based systems with endogenous proteins cannot be ruled out. These points favourably distinguish the BphP1-QPAS1 system from previously used systems for light-inducible regulation of gene expression in plants. Thus, in the near future, the BphP1-QPAS1 system has all the potential for wide application in plant research and biotechnology, for example, for plant protection from adverse climatic and chemical conditions (such as drought, toxins, soil salinity, and radiation), plant diseases (including moulds, bacteria, and viruses), and pests (such as insects and rodents).

## References

- Andres J., Blomeier T., Zurbriggen M.D. Synthetic switches and regulatory circuits in plants. *Plant Physiol.* 2019;179(3):862-884. doi 10.1104/pp.18.01362
- Auldridge M.E., Forest K.T. Bacterial phytochromes: more than meets the light. *Crit Rev Biochem Mol Biol.* 2011;46(1):67-88. doi 10.3109/10409238.2010.546389
- Battle M.W., Jones M.A. Cryptochromes integrate green light signals into the circadian system. *Plant Cell Environ.* 2020;43(1):16-27. doi 10.1111/pce.13643
- Bischof J., Maeda R.K., Hediger M., Karch F., Basler K. An optimized transgenesis system for *Drosophila* using germ-line-specific  $\phi$ C31 integrases. *Proc Natl Acad Sci USA.* 2007;104(9):3312-3317. doi 10.1073/pnas.0611511104
- Bonger K.M., Rakhit R., Payumo A.Y., Chen J.K., Wandless T.J. General method for regulating protein stability with light. *ACS Chem Biol.* 2014;9(1):111-115. doi 10.1021/cb400755b
- Braguy J., Zurbriggen M.D. Synthetic strategies for plant signalling studies: molecular toolbox and orthogonal platforms. *Plant J.* 2016; 87(1):118-138. doi 10.1111/tj.13218
- Chatelle C., Ochoa-Fernandez R., Engesser R., Schneider N., Beyer H.M., Jones A.R., Timmer J., Zurbriggen M.D., Weber W. A green-light-responsive system for the control of transgene expression in mammalian and plant cells. *ACS Synth Biol.* 2018;7(5): 1349-1358. doi 10.1021/acssynbio.7b00450
- Chen P.-Y., Wang C.-K., Soong S.-C., To K.-Y. Complete sequence of the binary vector pBI121 and its application in cloning T-DNA in-

- sertion from transgenic plants. *Mol Breed.* 2003;11(4):287-293. doi 10.1023/A:1023475710642
- Chernov K.G., Redchuk T.A., Omelina E.S., Verkhusha V.V. Near-infrared fluorescent proteins, biosensors, and optogenetic tools engineered from phytochromes. *Chem Rev.* 2017;117(9):6423-6446. doi 10.1021/acs.chemrev.6b00700
- Christie J.M., Zurbriggen M.D. Optogenetics in plants. *New Phytol.* 2021;229(6):3108-3115. doi 10.1111/nph.17008
- Christie J.M., Gawthorne J., Young G., Fraser N.J., Roe A.J. LOV to BLUF: flavoprotein contributions to the optogenetic toolkit. *Mol Plant.* 2012;5(3):533-544. doi 10.1093/mp/sss020
- Crosson S., Moffat K. Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction. *Proc Natl Acad Sci USA.* 2001;98(6):2995-3000. doi 10.1073/pnas.051520298
- Gälweiler L., Conlan R.S., Mader P., Palme K., Moore I. Technical advance: the DNA-binding activity of Gal4 is inhibited by methylation of the Gal4 binding site in plant chromatin. *Plant J.* 2000;23(1):143-157. doi 10.1046/j.1365-313x.2000.00805.x
- Gayatri T., Basu A. Development of reproducible regeneration and transformation system for *Sesamum indicum*. *Plant Cell Tiss Organ Cult.* 2020;143(2):441-456. doi 10.1007/s11240-020-01931-1
- Giraud E., Verméglio A. Bacteriophytochromes in anoxygenic photosynthetic bacteria. *Photosynth Res.* 2008;97(2):141-153. doi 10.1007/s11120-008-9323-0
- Gligorovski V., Sadeghi A., Rahi S.J. Multidimensional characterization of inducible promoters and a highly light-sensitive LOV-transcription factor. *Nat Commun.* 2023;14(1):3810. doi 10.1038/s41467-023-38959-8
- Guidi L., Tattini M., Landi M. How does chloroplast protect chlorophyll against excessive light? In: Chlorophyll. InTech, 2017. doi 10.5772/67887
- Hernández-Candia C.N., Tucker C.L. Optogenetic control of gene expression using cryptochrome 2 and a light-activated degron. *Methods Mol Biol.* 2020;2173:151-158. doi 10.1007/978-1-0716-0755-8\_10
- Herrou J., Crosson S. Function, structure and mechanism of bacterial photosensory LOV proteins. *Nat Rev Microbiol.* 2011;9(10):713-723. doi 10.1038/nrmicro2622
- Johnson A.A.T., Hibberd J.M., Gay C., Essah P.A., Haseloff J., Tester M., Guiderdoni E. Spatial control of transgene expression in rice (*Oryza sativa* L.) using the GAL4 enhancer trapping system. *Plant J.* 2005;41(5):779-789. doi 10.1111/j.1365-313X.2005.02339.x
- Johnson M.P. Photosynthesis. *Essays Biochem.* 2016;60(3):255-273. doi 10.1042/EBC20160016
- Kaberniuk A.A., Shemetov A.A., Verkhusha V.V. A bacterial phytochrome-based optogenetic system controllable with near-infrared light. *Nat Methods.* 2016;13(7):591-597. doi 10.1038/nmeth.3864
- Kaya S.G., Hovan A., Fraaije M.W. Engineering of LOV-domains for their use as protein tags. *Arch Biochem Biophys.* 2025;763:110228. doi 10.1016/j.abb.2024.110228
- Kim C., Yun S.R., Lee S.J., Kim S.O., Lee H., Choi J., Kim J.G., Kim T.W., You S., Kosheleva I., Noh T., Baek J., Ihee H. Structural dynamics of protein-protein association involved in the light-induced transition of *Avena sativa* LOV2 protein. *Nat Commun.* 2024;15(1):6991. doi 10.1038/s41467-024-51461-z
- Konermann S., Brigham M.D., Trevino A.E., Hsu P.D., Heidenreich M., Cong L., Platt R.J., Scott D.A., Church G.M., Zhang F. Optical control of mammalian endogenous transcription and epigenetic states. *Nature.* 2013;500(7463):472-476. doi 10.1038/nature12466
- Krueger D., Izquierdo E., Viswanathan R., Hartmann J., Pallares Cartes C., De Renzis S. Principles and applications of optogenetics in developmental biology. *Development.* 2019;146(20):dev175067. doi 10.1242/dev.175067
- Li F.W., Melkonian M., Rothfels C.J., Villarreal J.C., Stevenson D.W., Graham S.W., Wong G.K., Pryer K.M., Mathews S. Phytochrome diversity in green plants and the origin of canonical plant phytochromes. *Nat Commun.* 2015;6(1):7852. doi 10.1038/ncomms8852
- Li J., Li G., Wang H., Wang Deng X. Phytochrome signaling mechanisms. *Arabidopsis Book.* 2011;9:e0148. doi 10.1199/tab.0148
- Losi A., Gärtner W. Old chromophores, new photoactivation paradigms, trendy applications: flavins in blue light-sensing photoreceptors. *Photochem Photobiol.* 2011;87(3):491-510. doi 10.1111/j.1751-1097.2011.00913.x
- Lungu O.I., Hallett R.A., Choi E.J., Aiken M.J., Hahn K.M., Kuhlman B. Designing photoswitchable peptides using the AsLOV2 domain. *Chem Biol.* 2012;19(4):507-517. doi 10.1016/j.chembiol.2012.02.006
- Mathews S. Evolutionary studies illuminate the structural-functional model of plant phytochromes. *Plant Cell.* 2010;22(1):4-16. doi 10.1105/tpc.109.072280
- Müller K., Engesser R., Metzger S., Schulz S., Kämpf M.M., Busacker M., Steinberg T., Tomakidi P., Ehrbar M., Nagy F., Timmer J., Zurbriggen M.D., Weber W. A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. *Nucleic Acids Res.* 2013;41(7):e77. doi 10.1093/nar/gkt002
- Müller K., Siegel D., Rodriguez Jahnke F., Gerrer K., Wend S., Decker E.L., Reski R., Weber W., Zurbriggen M.D. A red light-controlled synthetic gene expression switch for plant systems. *Mol Biosyst.* 2014;10(7):1679-1688. doi 10.1039/c3mb70579j
- Müller K., Zurbriggen M.D., Weber W. An optogenetic upgrade for the Tet-OFF system. *Biotechnol Bioeng.* 2015;112(7):1483-1487. doi 10.1002/bit.25562
- Ni M., Tepperman J.M., Quail P.H. Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature.* 1999;400(6746):781-784. doi 10.1038/23500
- Niopek D., Wehler P., Roensch J., Eils R., Di Ventura B. Optogenetic control of nuclear protein export. *Nat Commun.* 2016;7:10624. doi 10.1038/ncomms10624
- Noda N., Ozawa T. Light-controllable transcription system by nucleocytoplasmic shuttling of a truncated phytochrome B. *Photochem Photobiol.* 2018;94(5):1071-1076. doi 10.1111/php.12955
- Ochoa-Fernandez R., Samodelov S.L., Brandl S.M., Wehinger E., Müller K., Weber W., Zurbriggen M.D. Optogenetics in plants: red/far-red light control of gene expression. *Methods Mol Biol.* 2016;1408:125-139. doi 10.1007/978-1-4939-3512-3\_9
- Ochoa-Fernandez R., Abel N.B., Wieland F.-G., Schlegel J., Koch L.A., Miller J.B., Engesser R., Giuriani G., Brandl S.M., Timmer J., Weber W., Ott T., Simon R., Zurbriggen M.D. Optogenetic control of gene expression in plants in the presence of ambient white light. *Nat Methods.* 2020;17(7):717-725. doi 10.1038/s41592-020-0868-y
- Omelina E.S., Pindyurin A.V. Optogenetic regulation of endogenous gene transcription in mammals. *Vavilovskii Zhurnal Genetiki i Selektzii = Vavilov J Genet Breed.* 2019;23(2):219-225. doi 10.18699/VJ19.485
- Omelina E.S., Yushkova A.A., Motorina D.M., Volegov G.A., Kozhevnikova E.N., Pindyurin A.V. Optogenetic and chemical induction systems for regulation of transgene expression in plants: use in basic and applied research. *Int J Mol Sci.* 2022;23(3):1737. doi 10.3390/ijms23031737
- Pathak G.P., Spiltoir J.I., Höglund C., Polstein L.R., Heine-Koskinen S., Gersbach C.A., Rossi J., Tucker C.L. Bidirectional approaches for optogenetic regulation of gene expression in mammalian cells using *Arabidopsis* cryptochrome 2. *Nucleic Acids Res.* 2017;45(20):e167. doi 10.1093/nar/gkx260
- Peter E., Dick B., Baeurle S.A. Mechanism of signal transduction of the LOV2-Ja photosensor from *Avena sativa*. *Nat Commun.* 2010;1:122. doi 10.1038/ncomms1121
- Piatkevich K.D., Subach F.V., Verkhusha V.V. Engineering of bacterial phytochromes for near-infrared imaging, sensing, and light-control in mammals. *Chem Soc Rev.* 2013;42(8):3441-3452. doi 10.1039/c3cs35458j
- Qian Y., Li T., Zhou S., Chen X., Yang Y. A single-component optogenetic Gal4-UAS system allows stringent control of gene expression in zebrafish and Drosophila. *ACS Synth Biol.* 2023;12(3):664-671. doi 10.1021/acssynbio.2c00410

- Redchuk T.A., Omelina E.S., Chernov K.G., Verkhusha V.V. Near-infrared optogenetic pair for protein regulation and spectral multiplexing. *Nat Chem Biol.* 2017;13(6):633-639. doi 10.1038/nchembio.2343
- Redchuk T.A., Karasev M.M., Omelina E.S., Verkhusha V.V. Near-infrared light-controlled gene expression and protein targeting in neurons and non-neuronal cells. *ChemBiochem.* 2018;19(12):1334-1340. doi 10.1002/cbic.201700642
- Renicke C., Schuster D., Usherenko S., Essen L.O., Taxis C. A LOV2 domain-based optogenetic tool to control protein degradation and cellular function. *Chem Biol.* 2013;20(4):619-626. doi 10.1016/j.chembiol.2013.03.005
- Sakvarelidze L., Tao Z., Bush M., Roberts G.R., Leader D.J., Doonan J.H., Rawsthorne S. Coupling the GAL4 UAS system with *alcR* for versatile cell type-specific chemically inducible gene expression in *Arabidopsis*. *Plant Biotechnol J.* 2007;5(4):465-476. doi 10.1111/j.1467-7652.2007.00254.x
- Shcherbakova D.M., Shemetov A.A., Kaberniuk A.A., Verkhusha V.V. Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools. *Annu Rev Biochem.* 2015;84(1):519-550. doi 10.1146/annurev-biochem-060614-034411
- Shikata H., Denninger P. Plant optogenetics: applications and perspectives. *Curr Opin Plant Biol.* 2022;68:102256. doi 10.1016/j.pbi.2022.102256
- Shimizu-Sato S., Huq E., Tepperman J.M., Quail P.H. A light-switchable gene promoter system. *Nat Biotechnol.* 2002;20(10):1041-1044. doi 10.1038/nbt734
- Strickland D., Lin Y., Wagner E., Hope C.M., Zayner J., Antoniou C., Sosnick T.R., Weiss E.L., Glotzer M. TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nat Methods.* 2012;9(4):379-384. doi 10.1038/nmeth.1904
- Wagner J.R., Zhang J., Brunzelle J.S., Vierstra R.D., Forest K.T. High resolution structure of *Deinococcus* bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J Biol Chem.* 2007;282(16):12298-12309. doi 10.1074/jbc.M611824200
- Wang X., Chen X., Yang Y. Spatiotemporal control of gene expression by a light-switchable transgene system. *Nat Methods.* 2012;9(3):266-269. doi 10.1038/nmeth.1892
- Willumsen B.M., Christensen A., Hubbert N.L., Papageorge A.G., Lowy D.R. The p21 *ras* C-terminus is required for transformation and membrane association. *Nature.* 1984a;310(5978):583-586. doi 10.1038/310583a0
- Willumsen B.M., Norris K., Papageorge A.G., Hubbert N.L., Lowy D.R. Harvey murine sarcoma virus p21 *ras* protein: biological and biochemical significance of the cysteine nearest the carboxy terminus. *EMBO J.* 1984b;3(11):2581-2585. doi 10.1002/j.1460-2075.1984.tb02177.x
- Wu C., Li X., Yuan W., Chen G., Kilian A., Li J., Xu C., Li X., Zhou D.X., Wang S., Zhang Q. Development of enhancer trap lines for functional analysis of the rice genome. *Plant J.* 2003;35(3):418-427. doi 10.1046/j.1365-313x.2003.01808.x
- Wu Y.I., Frey D., Lungu O.I., Jaehrig A., Schlichting I., Kuhlman B., Hahn K.M. A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature.* 2009;461(7260):104-108. doi 10.1038/nature08241
- Yang X., Kuk J., Moffat K. Crystal structure of *Pseudomonas aeruginosa* bacteriophytochrome: photoconversion and signal transduction. *Proc Natl Acad Sci USA.* 2008;105(38):14715-14720. doi 10.1073/pnas.0806718105
- Yumerefendi H., Dickinson D.J., Wang H., Zimmerman S.P., Bear J.E., Goldstein B., Hahn K., Kuhlman B. Control of protein activity and cell fate specification via light-mediated nuclear translocation. *PLoS One.* 2015;10(6):e0128443. doi 10.1371/journal.pone.0128443
- Yun A., Kang J., Lee J., Song S.J., Hwang I. Design of an artificial transcriptional system for production of high levels of recombinant proteins in tobacco (*Nicotiana benthamiana*). *Front Plant Sci.* 2023;14:1138089. doi 10.3389/fpls.2023.1138089
- Zayner J.P., Antoniou C., Sosnick T.R. The amino-terminal helix modulates light-activated conformational changes in *AsLOV2*. *J Mol Biol.* 2012;419(1-2):61-74. doi 10.1016/j.jmb.2012.02.037
- Zheng X., Deng W., Luo K., Duan H., Chen Y., McAvoy R., Song S., Pei Y., Li Y. The cauliflower mosaic virus (CaMV) 35S promoter sequence alters the level and patterns of activity of adjacent tissue- and organ-specific gene promoters. *Plant Cell Rep.* 2007;26(8):1195-1203. doi 10.1007/s00299-007-0307-x
- Zhu Y., Tepperman J.M., Fairchild C.D., Quail P.H. Phytochrome B binds with greater apparent affinity than phytochrome A to the basic helix-loop-helix factor PIF3 in a reaction requiring the PAS domain of PIF3. *Proc Natl Acad Sci USA.* 2000;97(24):13419-13424. doi 10.1073/pnas.230433797
- Zoltowski B.D., Crane B.R. Light activation of the LOV protein *vivid* generates a rapidly exchanging dimer. *Biochemistry.* 2008;47(27):7012-7019. doi 10.1021/bi8007017
- Zoltowski B.D., Schwerdtfeger C., Widom J., Loros J.J., Bilwes A.M., Dunlap J.C., Crane B.R. Conformational switching in the fungal light sensor *Vivid*. *Science.* 2007;316(5827):1054-1057. doi 10.1126/science.1137128

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

Received April 8, 2025. Revised April 30, 2025. Accepted May 28, 2025.

doi 10.18699/vjgb-26-04

## Редактирование оснований в гене *AUTS2* и высокопроизводительное NGS-генотипирование клонов: стратегия создания клеточной модели

А.П. Ян <sup>1, 2, 3</sup> , П.А. Сальников <sup>1, 2</sup>, А.А. Буздин<sup>4, 5, 6</sup>, В.А. Ковальская <sup>7</sup>, Е.В. Мусатова<sup>8</sup>, П.С. Орлов <sup>1, 9</sup>, О.П. Рыжкова<sup>7</sup>, А.И. Субботовская <sup>9</sup>, М.В. Сунцова<sup>4, 5</sup>, А.Ю. Христинченко <sup>5</sup>, А.А. Хабарова <sup>1</sup> 

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

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>3</sup> Научно-технологический университет «Сириус», федеральная территория «Сириус», Краснодарский край, Россия

<sup>4</sup> Первый Московский государственный медицинский университет им. И.М. Сеченова Министерства здравоохранения Российской Федерации, Москва, Россия

<sup>5</sup> Национальный медицинский исследовательский центр эндокринологии им. академика И.И. Дедова, Москва, Россия

<sup>6</sup> Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук, Москва, Россия

<sup>7</sup> Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, Россия

<sup>8</sup> Центр генетики и репродуктивной медицины «Genetico», Москва, Россия

<sup>9</sup> Федеральный исследовательский центр фундаментальной и трансляционной медицины, Новосибирск, Россия

 a.yan@alumni.nsu.ru, khabarova@bionet.nsc.ru

**Аннотация.** Изучение молекулярных механизмов, лежащих в основе расстройств аутистического спектра (РАС), требует создания клеточных моделей, способных отражать цис-регуляторные эффекты и аллель-специфичную экспрессию генов. В настоящем исследовании мы представляем подход к получению индуцированных плюрипотентных стволовых клеток (ИПСК), модифицированных с использованием аденинового редактора оснований (ABE), для введения синонимичных однонуклеотидных замен в ген *AUTS2* – кандидата на участие в патогенезе РАС. Эти замены позволяют маркировать аллели и отслеживать различия в экспрессии нормального и реорганизованного аллелей в цис-контексте. Мы разработали стратегию высокоэффективного генотипирования клонов с использованием секвенирования продуктов ПЦР (ампликонов) на платформе нового поколения (NGS). Анализ более 100 субклонов показал, что предложенный подход превосходит секвенирование по Сэнгеру по масштабируемости, чувствительности и экономичности. Мы отобрали клоны с целевыми гетерозиготными заменами, оценили уровень мозаицизма и провели фазирование с герминальными гетерозиготными вариантами, позволяющее убедиться в моноклональном происхождении клеточной линии или идентифицировать аллель, ассоциированный с мутацией. Полученные линии ИПСК маркируют разные аллели гена *AUTS2*, что открывает перспективу анализа влияния цис-регуляторных элементов на экспрессию гена в различных типах клеток. Результаты работы подчеркивают практическую значимость редакторов оснований и целевого NGS-генотипирования при создании клеточных моделей с однонуклеотидными заменами для фундаментальных и прикладных исследований.

**Ключевые слова:** индуцированные плюрипотентные стволовые клетки (ИПСК); CRISPR/Cas9; однонуклеотидные замены; аллель-специфичная оценка экспрессии; хромосомная перестройка

**Для цитирования:** Ян А.П., Сальников П.А., Буздин А.А., Ковальская В.А., Мусатова Е.В., Орлов П.С., Рыжкова О.П., Субботовская А.И., Сунцова М.В., Христинченко А.Ю., Хабарова А.А. Редактирование оснований в гене *AUTS2* и высокопроизводительное NGS-генотипирование клонов: стратегия создания клеточной модели. *Вавиловский журнал генетики и селекции*. 2026;30(1):72-84. doi 10.18699/vjgb-26-04

**Финансирование.** Работа поддержана грантом РФФИ 24-25-00152.

**Благодарности.** Секвенирование библиотек ДНК поддержано Министерством науки и высшего образования Российской Федерации (Соглашение № 075-15-2022-310 от 20 апреля 2022 г.). Культивирование клеток выполнено на базе ЦКП «Коллекция плюрипотентных культур клеток человека и млекопитающих общебиологического и биомедицинского направления» ИЦиГ СО РАН. Биоинформационный анализ данных проведен при финансовой поддержке бюджетного проекта (№ FWNR-2025-0017). Работа А.А. Буздина, М.В. Сунцовой и А.Ю. Христинченко поддержана проектом МЗ РФ «Разработка системы жидкостной биопсии для мониторинга ответа на терапию и прогрессии рака легкого».

**Соблюдение этических стандартов.** Настоящая статья не содержит описания каких-либо исследований с участием людей или животных в качестве объектов.

**Индивидуальный вклад авторов.** А.А. Хабарова разработала концепцию исследования и руководила проектом. А.А. Хабарова, А.П. Ян, П.А. Сальников, А.А. Буздин, В.А. Ковальская, О.П. Рыжкова, М.В. Сунцова, А.Ю. Христинченко, Е.В. Мусатова, П.С. Орлов, А.И. Субботовская провели эксперименты. П.А. Сальников выполнил биоинформационную обработку результатов секвенирования. А.А. Хабарова, П.А. Сальников, А.П. Ян проанализировали результаты и подготовили текст рукописи. Все авторы редактировали рукопись и одобрили ее окончательный вариант.

## Base editing in the *AUTS2* gene and high-throughput NGS genotyping of clones: a strategy for generating a cellular model

A.P. Yan <sup>1,2,3</sup> , P.A. Salnikov <sup>1,2</sup>, A.A. Buzdin<sup>4,5,6</sup>, V.A. Kovalskaia <sup>7</sup>, E.V. Musatova<sup>8</sup>, P.S. Orlov <sup>1,9</sup>, O.P. Ryzhkova<sup>7</sup>, A.I. Subbotovskaia <sup>9</sup>, M.V. Suntsova<sup>4,5</sup>, A.U. Khristichenko <sup>5</sup>, A.A. Khabarova <sup>1</sup> 

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russia

<sup>4</sup> I.M. Sechenov First Moscow State Medical University of the Ministry of Healthcare of the Russian Federation, Moscow, Russia

<sup>5</sup> National Medical Research Center for Endocrinology named after Academician I.I. Dedov, Moscow, Russia

<sup>6</sup> M.M. Shemyakin–Yu.A. Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Moscow, Russia

<sup>7</sup> Research Centre for Medical Genetics, Moscow, Russia

<sup>8</sup> Center of Genetics and Reproductive Medicine “Genetico”, Moscow, Russia

<sup>9</sup> Federal Research Center of Fundamental and Translational Medicine, Novosibirsk, Russia

 a.yan@alumni.nsu.ru, khabarova@bionet.nsc.ru

**Abstract.** Studying the molecular mechanisms underlying autism spectrum disorders (ASD) requires cellular models capable of capturing *cis*-regulatory effects and allele-specific gene expression. In this study, we present an approach for generating induced pluripotent stem cells (iPSCs) modified using an adenine base editor (ABE) to introduce synonymous single-nucleotide substitutions in the *AUTS2* gene – a candidate involved in ASD pathogenesis. These substitutions serve as allele-specific markers, enabling the tracking of expression differences between normal and rearranged alleles in a *cis*-regulatory context. We developed a high-efficiency strategy for genotyping clones using amplicon-based next-generation sequencing (NGS). Analysis of over 100 subclones demonstrated that this approach surpasses Sanger sequencing in scalability, sensitivity, and cost-effectiveness. We selected clones with targeted heterozygous substitutions, assessed mosaicism levels, and performed phasing with germline heterozygous variants to confirm monoclonal origin and identify the allele carrying the chromosomal rearrangement. The resulting iPSC lines mark distinct *AUTS2* alleles, providing a foundation for analyzing the impact of *cis*-regulatory elements on gene expression across different cell types. Our findings highlight the practical value of base editors and targeted NGS genotyping in creating cellular models with single-nucleotide substitutions for both basic and applied research.

**Key words:** induced pluripotent stem cells (iPSCs); CRISPR/Cas9; single-nucleotide substitutions; allele-specific expression analysis; chromosomal rearrangement

**For citation:** Yan A.P., Salnikov P.A., Buzdin A.A., Kovalskaia V.A., Musatova E.V., Orlov P.S., Ryzhkova O.P., Subbotovskaia A.I., Suntsova M.V., Khristichenko A.U., Khabarova A.A. Base editing in the *AUTS2* gene and high-throughput NGS genotyping of clones: a strategy for generating a cellular model. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):72-84. doi 10.18699/vjgb-26-04

### Введение

По данным Global Burden of Disease, в 2021 г. с расстройствами аутистического спектра жили примерно 61.8 млн человек по всему миру, а совокупные потери лет жизни, скорректированных по нетрудоспособности, составили 11.5 млн – т.е. 147 лет жизни на каждые 100 000 человек (Global Burden of Disease Study 2021..., 2025). Выяснение генетических причин аутизма так же, как и других врожденных неврологических заболеваний, зачастую требует исследования механизмов патологических процессов, происходящих в головном мозге. С учетом малой доступности этого органа для прямого получения клеточного материала особенно важно создание клеточных моделей, которые позволят отслеживать изменения в экспрессии интересующих генов. Незаменимым инструментом для этого являются индуцированные плюрипотентные стволовые клетки (ИПСК), которые получают из большого спектра клеточного материала пациента. Такие клетки можно направленно дифференцировать в различные клеточные типы, что открывает большую свободу моделирования процессов, происходящих в разных тканях организма (Rowe, Daley, 2019; De Masi et al., 2020).

Для реконструкции патогенных мутаций создания в клеточных моделях широко применяется система геномного редактирования CRISPR/Cas9. Однако, несмотря на ее распространенность, эта система имеет ограничения, включающие неспецифическую активность и неточность репарации двуцепочечных разрывов, что может приводить к нежелательным мутациям (Смирнов и др., 2016; Uddin et al., 2020). В настоящее время развивается ряд более точных методов, основанных на технологии CRISPR/Cas9, лишенных основных недостатков исходной технологии. Один из таких методов – редакторы оснований, позволяющие специфично заменять один нуклеотид на другой (A→G или C→T) без индукции двуцепочечных разрывов (Gaudelli et al., 2017).

Внесение модификаций в геном с одонуклеотидной точностью не только открывает возможность моделирования генетических заболеваний, вызванных точечными мутациями (Lu, Huang, 2018; Geurts et al., 2023), но и позволяет маркировать аллели, что дает уникальную возможность проследить влияние *cis*-регуляторных вариантов на экспрессию и функциональную активность генов на моноаллельном уровне. Сравнение уровней экспрессии

аллелей повторяет дизайн гибридологического подхода, позволяя оценить цис-регуляторные эффекты при нивелировании любых транс-влияний (Salnikov et al., 2025). Кроме того, редакторы оснований активно используются для разработки генно-терапевтических стратегий для лечения заболеваний, связанных с вариантами в ядерной и митохондриальной ДНК (Billon et al., 2017; Liang et al., 2023).

В системе редактирования оснований используется гибридный белок, состоящий из лишённого нуклеазной активности белка Cas9 и фермента нуклеотид-дезаминазы. Такой химерный белок позволяет вносить замены A→G или C→T без создания двуцепочечных разрывов ДНК. В случае аденинового редактора оснований (Adenin base editor, ABE) дезаминирование происходит на одной цепи ДНК в узком окне активности дезаминазы (4 нуклеотида), а одноцепочечный разрыв на другой цепи стимулирует систему репарации восстанавливать последовательность по модифицированной цепи (Rees, Liu, 2018). Система ABE индуцирует конверсию A→G путем дезаминирования аденина с превращением его в гипоксантин, который затем распознается ДНК-полимеразой как гуанин (Gaudelli et al., 2017; Chen et al., 2023).

Таким образом, редакторы оснований катализируют прямое химическое преобразование нуклеотидов без создания двуцепочечных разрывов, что минимизирует риск хромосомных нарушений и исключает необходимость в доставке экзогенной ДНК как матрицы для гомологичной репарации. Это существенно снижает частоту нецелевых инсерций, делеций и хромосомных перестроек по сравнению с классическим Cas9.

Несмотря на ряд очевидных преимуществ данной системы редактирования, редакторы оснований в своем применении имеют ряд ограничений. Ключевой недостаток – каталитическая специфичность дезаминазных ферментов, позволяющая осуществлять только транзиции (замены C→T или A→G (замена пурина на пурин или пиримидина на пиримидин), но не трансверсии (замена пурина на пиримидин или наоборот), что значительно сужает спектр корректируемых мутаций (Komor et al., 2016; Gaudelli et al., 2017; Rees, Liu, 2018). Другую серьезную проблему представляет нецелевая активность. Она проявляется как в редактировании не полностью комплементарных сайтов в ДНК в пределах «окна редактирования», так и, что особенно важно, в неспецифической модификации клеточной РНК. Дезаминазы, особенно цитозиновые, могут случайно дезаминировать цитозины в РНК, потенциально нарушая транскриптом и вызывая токсические эффекты, причем эта активность часто не зависит от направляющей РНК (Grünwald et al., 2019; Jin et al., 2019; Zuo et al., 2019; Yu et al., 2020).

Настоящая статья посвящена использованию системы редактирования оснований ABE и инновационной стратегии генотипирования клеточных клонов при создании клеточных моделей на основе ИПСК. При анализе кариотипа пациента с расстройством аутистического спектра была обнаружена хромосомная перестройка на одном из аллелей, граница которой находится в межгенной области

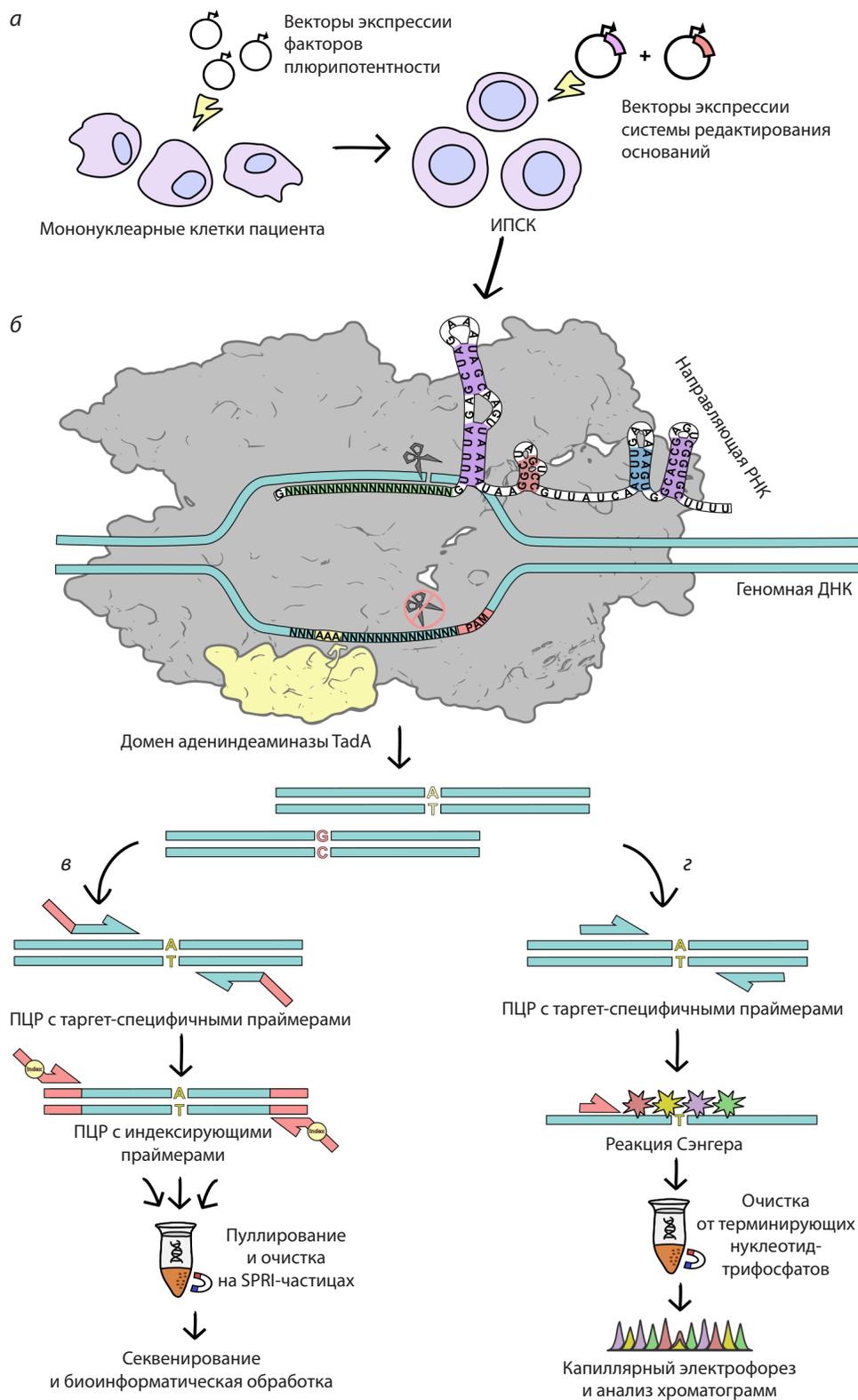
вблизи гена *AUTS2* (Gridina et al., 2025). Мы полагаем, что эта перестройка может нарушать регуляцию экспрессии гена *AUTS2*, что, в свою очередь, может быть связано с наблюдаемым фенотипом пациента. Для экспериментальной проверки этой гипотезы необходимо создать модель, позволяющую оценивать различия в экспрессии нормального аллеля и находящегося в цис-положении с хромосомной перестройкой в разных типах клеток. Однако для этого пациента не найдено однонуклеотидных замен в кодирующей области гена *AUTS2*, которые бы позволили дискриминировать транскрипты с разных аллелей. Из ранее полученных ИПСК данного пациента с использованием системы ABE мы создали модельные линии, несущие гетерозиготные однонуклеотидные замены в 10-м экзоне гена *AUTS2*. Эти линии дают возможность получать нейроны/нейральные предшественники *in vitro* и количественно оценивать аллель-специфичную транскрипцию *AUTS2* с перестроенной и интактной хромосомами.

Генотипирование клонов и оценку эффективности редактирования рутинно оценивают секвенированием по Сэнгеру ампликонов целевого локуса. Наиболее удобный и точный подход – секвенирование ампликона, содержащего целевую замену. Для секвенирования ампликонов применимы два подхода: метод Сэнгера и секвенирование нового поколения (NGS) (рис. 1, в, з). При секвенировании ампликонов выбор между методом Сэнгера и NGS определяется конкретными задачами исследования. Метод Сэнгера остается оптимальным для анализа отдельных продуктов ПЦР при небольшом количестве образцов. Его ограниченная пропускная способность делает его экономически нецелесообразным при работе с большим количеством образцов. В отличие от этого, NGS-технологии позволяют проводить одновременное секвенирование множества образцов, что значительно снижает стоимость анализа на один образец и делает метод незаменимым для массового анализа. Однако следует учитывать, что NGS требует биоинформатического анализа. Таким образом, для рутинного анализа единичных ампликонов метод Сэнгера сохраняет свои преимущества, тогда как при необходимости масштабировать анализ для большого количества образцов NGS демонстрирует явное превосходство.

В рамках этого исследования с помощью NGS проанализировано 117 субклонов и отобраны линии, подходящие для дифференцировки и анализа экспрессии гена *AUTS2*. Мы подробно описываем протокол пробоподготовки для NGS-генотипирования и биоинформационного анализа полученных данных, который может быть адаптирован для других задач генотипирования.

## Материалы и методы

**Создание вектора MLM3636-BstV2I.** В качестве основы для вектора мы взяли плазмиду MLM3636 (Addgene #43860), содержащую последовательность направляющей РНК (нРНК) с сайтом клонирования спейсера, рассчитанного на использование фермента рестрикции BsmBI (New England Biolabs, США). Были задействованы следующие последовательности протоспейсеров: 5'-CAA AAGTTGACCCATTCTAC-3' для H044 и 5'-TCACACT



**Рис. 1.** Общая схема эксперимента по получению однонуклеотидных замен в ИПСК человека и сравнение этапов протоколов генотипирования клонов методами NGS и секвенирования по Сэнгеру.

*а* – электропорация моноцитарных клеток крови пациента векторами экспрессии факторов репрограммирования, получение ИПСК и затем электропорация ИПСК плазмидами, кодирующими систему редактирования оснований: целевые направляющие РНК и модифицированный белок Cas9, слитый с адениндеаминазой TadA; *б* – схематичное взаимодействие комплекса модифицированного белка Cas9 (серый цвет), слитого с доменом адениндеаминазы TadA (светло-желтый), и направляющей РНК с последовательностью протоспейсера в геномной ДНК. Аденины в окне редактирования выделены желтым; *в* – схема приготовления NGS-библиотеки для генотипирования клонов путем добавления адаптерных последовательностей и индексирования с помощью двух раундов ПЦР; *з* – схема стандартного протокола секвенирования по Сэнгеру.

GTGCCGGTAGAAT-3' для H068. Данный фермент не производится на территории РФ, поэтому для упрощения дальнейших манипуляций мы заменили оригинальный сайт клонирования на аналогичный сайт из плазмиды pSpCas9(BB)-2A-GFP (PX458) (Addgene #48138), распознаваемый ферментом BstV2I, выпускаемым российским производителем ООО «СибЭнзим».

Для этого мы амплифицировали фрагмент, содержащий сайт клонирования pSpCas9(BB)-2A-GFP (PX458) с праймеров pSpCas9\_clonF и pSpCas9\_clonR, и почти всю последовательность вектора MLM3636 с применением праймеров MLM3636\_clonF и MLM3636\_clonR. Полученный продукт обработали 1 ед. фермента DpnI (New England Biolabs, США) и очистили с помощью суспензии SPRI-частиц VAHTS DNA Clean Beads (Vazyme, Китай). Сборку конструкции проводили с помощью NEBuilder HiFi (New England Biolabs) в соответствии с протоколом производителя.

Для получения векторов, экспрессирующих нРНК, 1 мкг плазмиды MLM3636-BstV2I линейризовали 25 ед. фермента BstV2I в 50 мкл реакционной смеси при 55 °С в течение ночи. После рестрикции ДНК очищали 0.8-кратным объемом суспензии SPRI-частиц VAHTS DNA Clean Beads (Vazyme). Однопочечные олигонуклеотиды H044\_F и H044\_R (по 2 мкМ каждого) фосфорилировали в 1× буфере для T4-ДНК-лигазы с 100 нМ АТФ и 10 ед. T4-полинуклеотидкиназы («СибЭнзим»). Для пары H068 применяли аналогичную процедуру. Затем олигонуклеотиды нагревали до 95 °С и гибридизовали медленным охлаждением со скоростью 0.1 °С/с до 4 °С.

Полученный раствор дц-олигонуклеотида H044 (H068) F+R на следующем этапе лигировали в линейризованную векторную конструкцию MLM3636-BstV2I. Для этого смешивали на льду следующие компоненты: 100 нМ раствора дц-олигонуклеотида до 20 нг линейризованной ранее плазмиды, 100 нМ АТФ, 1× SE-буфер T4 ДНК лигазы, 20 ед. T4 ДНК лигазы. Реакционную смесь инкубировали в течение ночи при 4 °С. Очищали продукт 0.8 объемом суспензии SPRI-частиц VAHTS DNA Clean Beads (Vazyme), элюировали в 5 мкл H<sub>2</sub>O. Для последующей трансформации бактериальных клеток использовали половину полученного объема продукта.

Трансформацию *Escherichia coli* проводили с помощью метода электропорации, плазмидную ДНК выделяли набором Midi-prep («Евроген», Россия) и дополнительно очищали 0.8 объемом суспензии VAHTS DNA Clean Beads (Vazyme). Элюцию проводили в минимальном объеме воды, чтобы достигнуть концентрации ДНК не менее 1 мкг/мкл. Для селекции клеток с применением пурамицина использовали плазмиду pURC\_puro (предоставлена сотрудником отдела молекулярных механизмов онтогенеза ИЦиГ СО РАН А. Нурисламовым). Данная конструкция экспрессирует встроенный в плазмиду под промотором CMV ген устойчивости к пурамицину, что позволяет отобрать такие клетки и обогатить популяцию клеток событиями редактирования.

**Культивирование и электропорация клеток с помощью системы Neon.** ИПСК культивировали на пласти-

ковых планшетах, дно которых предварительно покрывали раствором культурального матрикса Matrigel (Corning, США), с использованием среды mTesr1 (StemCell, Канада) с добавлением смеси антибиотиков пенициллина и стрептомицина («ПанЭко», Россия). Для дезагрегации клеток применяли реагент TripLE (Thermo Fisher, США), рассаживали в соотношении 1:3–1:5 с добавлением ингибитора Y-27632 (10 мкМ, ингибитор Rho-киназы) (StemCell, Канада). За день до электропорации линии ИПСК пересаживали так, чтобы клеточная культура покрывала дно планшета не больше чем на 70 % от площади лунки. Перед переносом трансформированных клеток дно культуральных планшетов покрывали матриксом Matrigel (1 мл на 10 см<sup>2</sup>) и инкубировали в термостате при 37 °С в течение 20–30 мин. За 1 ч до трансформации к клеткам добавляли 10 мкМ Y-27632.

Клеточную культуру дезагрегировали с помощью 700 мкл TripLE (Thermo Fisher). Затем добавляли в каждую лунку по 1 мл промывочной среды, мягко ресуспендировали клетки и собирали суспензию в центрифужную пробирку. Подсчет клеток осуществляли в камере Горяева. Суспензию центрифугировали 5 мин при 300 g. После подсчета необходимое количество клеточной суспензии (150–200 тыс. клеток на одну реакцию электропорации) отмывали один раз в фосфатном буфере. В пробирке на 1500 мкл смешивали три плазмиды (каждой по 0.5 мкг): плазмиды экспрессии нРНК (MLM3636-BstV2I), редактора оснований (addgene #108382) и pURC\_puro (устойчивость к пурамицину) и добавляли 150–250 тыс. клеток, ресуспендированных после отмывания в фосфатном буфере в буфере R из набора для электропорации Neon. Суммарный объем суспензии клеток и смеси плазмид не должен превышать 10 мкл.

Электропорацию проводили с помощью системы Neon при условиях 1100 V – 30 мс – 1 импульс и переносили клетки в лунку культурального планшета (6 лунок) в 3 мл среды mTesr1, не содержащей антибиотиков. Через сутки в среду добавляли антибиотики (1× смесь пенициллина и стрептомицина) и также добавляли селективный антибиотик пурамицин в концентрации 1 мкг/мл на 48 ч (Gibco, США). Среду меняли полностью через 1 день. Через 7–10 дней колонии клеток индивидуально переносили в отдельные лунки 24-луночного планшета (снимали около 30 колоний). Как только колонии достигали размеров 1/4 от площади лунки, их дезагрегировали с помощью TripLE, затем после центрифугирования часть клеточной суспензии оставляли для выделения ДНК и последующего генотипирования, а часть суспензии замораживали в KSR (Gibco) с добавлением 10 % диметилсульфоксида.

**Генотипирование клонов с помощью NGS-секвенирования ампликонов.** Для постановки ПЦР использовали набор реактивов от компании «МБС-Технология» (Россия). Общий объем 50 мкл содержал: 50 нг ДНК, 1× БР MBUision, 0.2 мМ каждого dNTP; 0.2 мкМ праймера AUTS\_gen\_F; 0.2 мкМ праймера AUTS\_gen\_R (табл. 1); 0.4 мкл MBUision полимеразы («МБС-Технология»); H<sub>2</sub>O до 50 мкл. ПЦР проводили при следующих условиях: 95 °С 3 мин, 20 циклов: 95 °С 15 с, 60 °С 30 с, 72 °С 30 с, и затем

**Таблица 1.** Праймеры и олигонуклеотиды

Название	Последовательность 5'–3'
MLM_genF	TCCGGCAGGAAGAGGGCCTATTTTC
MLM_genR	CCTCGAGCGGCCCAAGCTTAAAAA
H044_F	CACCGCAAAAGTTGACCCATTCTAC
H044_R	AAACGTAGAATGGGTCAACTTTTGC
H068_F	CACCGTCACACTGTGCCGGTAGAAT
H068_R	AAACATTCTACCGGCACAGTGTGAC
AUTS_gen_F	CTGGAGTTCAGACGTGTGCTCTCCGATCTTGACAGCTTAATGACAGGGAAGC
AUTS_gen_R	TCTTTCCCTACACGACGCTCTCCGATCTCTGATCTGTGGCAAGAGCACAA
P5_IP{index_ID}	AATGATACGGCGACCACCGAGATCTACAC{8bp_Index}ACACTCTTTCCCTACACGAC
P7_IP{index_ID}	CAAGCAGAAGACGGCATACGAGAT{8bp_Index}GTGACTGGAGTTCAGACGTGT
pSpCas9_clonF	GGCCTATTTCCCATGATTCCT
pSpCas9_clonR	CGACTCGGTGCCACTTTT
MLM3636_clonF	AGGAATCATGGGAAATAGGCC
MLM3636_clonR	AAAAGTGGCACCGAGTCG
AUTS2_phase_F1	CTGGAGTTCAGACGTGTGCTCTCCGATCTACTAGCTTTTGCTTTGATCCC
AUTS2_phase_R1	TCTTTCCCTACACGACGCTCTCCGATCTGCCTGACTGCCACTAAAGAG
AUTS2_phase_F2	CTGGAGTTCAGACGTGTGCTCTCCGATCTCCACTGTAGCAGTGAACAAA
AUTS2_phase_R2	TCTTTCCCTACACGACGCTCTCCGATCTATGCCTGACTGCCACTAAAG

72 °C 1 мин, 4 °C. Полученные продукты ПЦР разводили в 100 раз и использовали 2 мкл данного раствора в качестве матрицы для последующей ПЦР, которую проводили с индексированием и включением полных последовательностей, необходимых для секвенирования, P5\_IP {index\_ID} и P7\_IP {index\_ID} (см. табл. 1).

Аmplification осуществляли на данных условиях: 95 °C 3 мин; 8 циклов: 95 °C 15 с; 60 °C 30 с; 72 °C 30 с и затем 72 °C 1 мин, 4 °C. Приблизительную концентрацию продуктов ПЦР оценивали с помощью их визуализации на электрофореze в агарозном геле, затем все образцы смешивали в примерно равных количествах и очищали одним объемом суспензии VAHNTS DNA Clean Beads (Vazyme), затем элюировали в 20 мкл H<sub>2</sub>O.

Секвенирование выполняли в режиме парных прочтений по 150 п. н. (в среднем ~10000 прочтений на образец). Из полученных данных убрали последовательности адаптеров (cutadapt), после чего выровняли на индексированную референс-последовательность целевого региона инструментом bowtie2, параметры по умолчанию. Полученные данные в формате BAM визуализировали с помощью Integrative Genomics Viewer (IGV, США).

Для идентификации вариантов формировали VCF-файлы командой bcftools mpileup с записью глубины аллелей (флаг AD). Далее собственным скриптом на Python рассчитывали частоту представленности мутантного аллеля по каждой позиции, агрегировали результаты в единую матрицу и строили линейные графики для групп образцов, что позволяло оценить распределение аллельных частот для замен по всем возможным позициям. Все использо-

ванные скрипты с примерами доступны на GitHub <https://github.com/Somatich/NGS-genotyping-Yan-et-al.-2025>.

**Фазирование внесенных замен с герминальным вариантом.** Под фазированием однонуклеотидных замен мы подразумеваем определение того, находятся ли два варианта разных замен в цис-конфигурации или в транс-конфигурации. Мы подобрали пару праймеров, фланкирующих как сайт индуцированного мутагенеза, так и герминальный вариант (chr7:70768282 hg38, rs3829006 G/A), с ожидаемой длиной продукта ПЦР длиной 363 п. н.; к 5'-концам праймеров были добавлены технические последовательности для NGS. Данный герминальный вариант находился на расстоянии 238 п. н. от сайта индуцируемого мутагенеза и был выбран при анализе данных секвенирования экзона пациента (Gridina et al., 2025). Удаление на расстояние не более 200–300 нуклеотидов позволяет получить продукт ПЦР, длина которого оптимальна для секвенирования с помощью системы NGS Illumina (секвенирование образца с длиной фрагмента ДНК более 500 п. н. значительно снижало бы его эффективность). Были использованы две разные пары праймеров: AUTS2\_phase\_F1/R1 и AUTS2\_phase\_F2/R2 (см. табл. 1).

Для постановки ПЦР задействовали набор реактивов от компании «МБС-Технология». Индексирование ПЦР и очистку проводили так же, как указано выше. Библиотеки секвенировали в режиме парных прочтений по 150 п. н., получив около 5000 пар прочтений на образец. Чтения картировали утилитой Bowtie2 с настройками по умолчанию. Фазирование SNP выполняли вручную, анализируя выравнивания в геномном браузере IGV.

## Результаты

### Дизайн нРНК и оценка эффективности редактирования в 10-м экзоне гена *AUTS2*

Целью данного исследования является создание клеточной модели на основе ИПСК, содержащих гетерозиготную однонуклеотидную синонимичную замену в 10-м экзоне гена *AUTS2*, с применением аденинового редактирования оснований (ABE) (см. рис. 1, а). Выбор 10-го экзона был обусловлен тем, что он присутствует в большинстве описанных и предсказанных транскриптов. Кроме этого, именно рядом с 10-м экзоном обнаружен герминальный однонуклеотидный вариант, позволяющий нам фазировать внесенную замену относительно аллелей.

На этапе дизайна нРНК была проанализирована нуклеотидная последовательность 10-го экзона гена *AUTS2*. В результате этого анализа найдены позиции аденина, расположенные на третьем месте в составе кодонов и попадающие в рабочее окно ABE (нуклеотиды 4–7 протоспейсера). Таким образом, замена таких аденинов на гуанин является синонимичной заменой. В 10-м экзоне идентифицировано два таких аденина: chr7:70768044 A→G и chr7:70768068 T→C (hg38). Замена аденина на позиции chr7:70768044 приводит к синонимичной замене аденина в триплете AAA на AAG (p.Lys570Lys). Данный вариант встречается в популяции с вероятностью 0.000005596, не предсказан как влияющий на сплайсинг или создающий новый сайт сплайсинга (при анализе последовательности в базах данных SpliceAI и MobiDetails). Замена на позиции chr7:70768068 T→C – синонимичная замена AGT на AGC, данный вариант не описан в базе данных gnomAD и также не предсказан как вариант, влияющий на сплайсинг или создающий новый сайт сплайсинга (SpliceAI и MobiDetails) (Reese et al., 1997; Jaganathan et al., 2019; Vaux et al., 2021).

Для обоих этих вариантов сконструированы две направляющие РНК: Н044 и Н068 (см. табл. 1). Оценка эффективности направляющих РНК в условиях *in vivo* проведена на культуре ИПСК с использованием плазмиды, экспрессирующей белок Cas9 (addgene 62988), и тоже показала примерно одинаковую эффективность обеих нРНК: для позиции Н044 (chr7:70768044) составила 9.2 %, для Н068 (chr7:70768068) – 7.6 %.

Анализ выполняли с помощью инструмента Cas-Analyser (Crisper Rgen) (Hwang et al., 2018). Важно подчеркнуть, что эффективность геномного редактирования определяется множеством других трудно контролируемых факторов, таких как локальное состояние хроматина конкретной линии, поэтому не следует полностью полагаться на расчетное предсказание активности и специфичности направляющей РНК, но, чтобы повысить вероятность получения требуемой мутации, целесообразно параллельно использовать несколько нРНК с разными протоспейсерами. Такой подход повышает вероятность получения требуемой мутации. Поэтому в дальнейших экспериментах мы применяли обе нРНК.

В качестве пробного эксперимента на популяции клеток (после трансфекции для анализа были взяты все клет-

ки, без отбора тех, в которые действительно попали конструкции), была оценена эффективность обеих направляющих РНК в комбинации с редактором оснований Cas9(ABE7.10). После электропорации клеток генетическими конструкциями, экспрессирующими белок Cas9 и направляющую РНК, на третий день из клеток выделяли ДНК и проводили пробоподготовку образцов для NGS. Анализ осуществляли с помощью инструмента BE-Analyser (Crisper Rgen, Южная Корея) (Hwang et al., 2018). Эффективность внесения замены аденина с помощью редактора оснований для Н044 (chr7:70768044) составила 5.1 %.

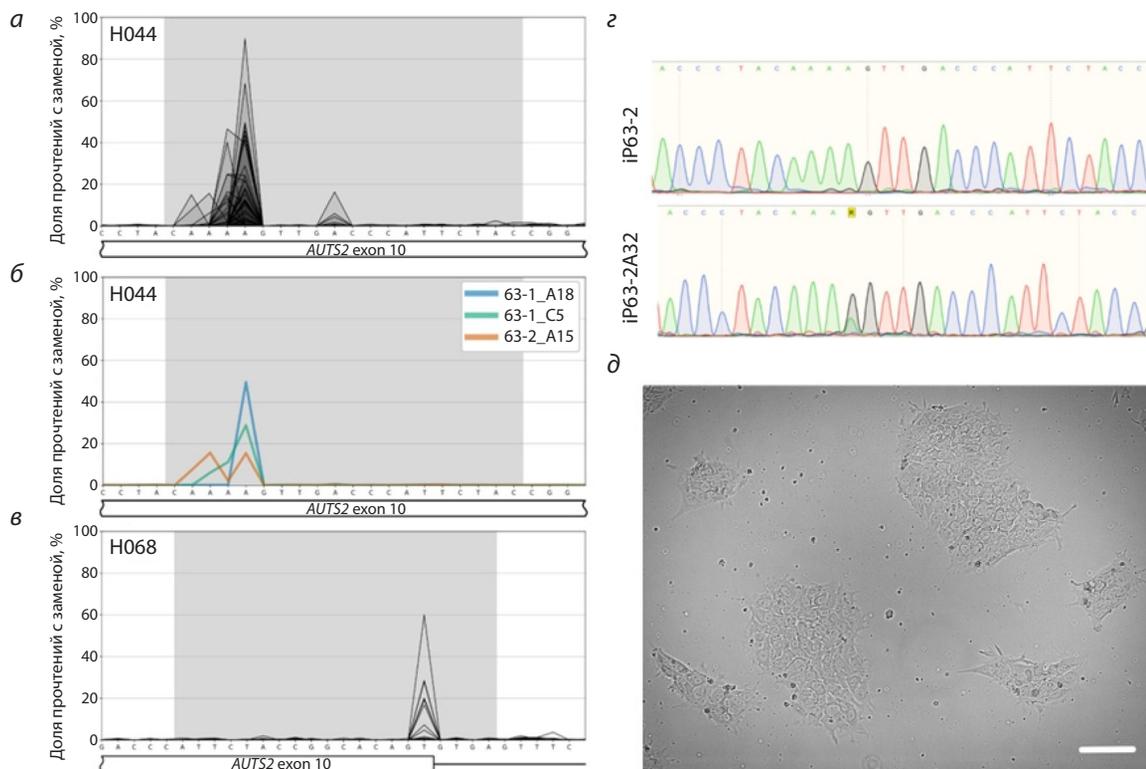
На соседних позициях (chr7:70768041–70768043) с левым аденином (chr7:70768044) находятся четыре аденина. Мы оценили вероятность редактирования редактором оснований: так, для аденина в позиции chr7:70768043 она составила 3 %, частота внесения замены аденина еще в двух соседних позициях (chr7:70768041–70768042) была 0.4 и 0.7 % соответственно. Для Н068 в позиции chr7:70768068 эффективность составила 3 %. Полученные данные свидетельствуют о том, что, несмотря на наличие кластера из четырех аденинов (chr7:70768041–70768044) вблизи целевой позиции, протоспейсер Н044 обеспечивает более высокую эффективность редактирования в заданной позиции по сравнению с Н068. Достаточно высокая частота нецелевой замены аденина в позиции chr7:70768043 в нашем случае не являлась критическим фактором для использования данной нРНК, но вынуждает задействовать большую выборку модифицированных клонов для поиска нужного нам генотипа с заменой в необходимой позиции.

Также для повышения доли клеток с необходимой генетической модификацией в популяции мы решили применить методы обогащения, такие как котрансфекция клеток с плазмидой, несущей селективный маркер (например, устойчивости к пурамицину), что позволит эффективно отобрать клетки, успешно трансфицированные генетическими конструкциями, и значительно повысит выход модифицированных клеток. Таким образом, результаты пилотного эксперимента подтверждают работоспособность системы редакторов оснований в локусе гена *AUTS2*; нРНК Н044 демонстрирует более высокую эффективность редактирования целевой позиции по сравнению с нРНК Н068, несмотря на потенциальные проблемы, связанные с кластером аденинов.

### Получение и отбор клонов ИПСК

#### с внесенной заменой после редактирования оснований

Для внесения целевой замены в 10-м экзоне гена *AUTS2* применили электропорацию с использованием системы Neon™. Электропорацию двух исходных линий человеческих ИПСК проводили смесью трех плазмид, кодирующих направляющую РНК, адениновый редактор оснований и маркер устойчивости к пурамицину. После электропорации клетки рассаживали в низкой плотности на матриксе и культивировали на среде mTesr1, содержащей селективный антибиотик в течение трех суток. На 7–10-й день после трансфекции были визуально идентифицированы и механически отобраны индивидуальные колонии, де-



**Рис. 2.** Примеры результатов представленного протокола.

*а* – визуализация суммарных результатов генотипирования для 69 клеточных клонов, модифицированных с помощью направляющей РНК Н044; *б* – пример визуализации результатов генотипирования для трех избранных образцов. По оси абсцисс – геномная последовательность chr7:70768037–70768062, hg38 в нуклеотидах, по оси ординат – доля прочтений секвенирования, содержащих замену; *в* – визуализация суммарных результатов генотипирования для 48 клеточных клонов, модифицированных с помощью направляющей РНК Н068. По оси абсцисс – геномная последовательность chr7:70768048–70768077, hg38 в нуклеотидах. Серая область соответствует последовательности направляющей РНК, использованной в эксперименте; *з* – секвеннограммы образцов iP63-2 (исходная линия ИПСК) и iP63-2A32 (линия ИПСК с синонимичной заменой триплета AAA/AAG) для сайта мутагенеза; *д* – морфология клеточной культуры ИПСК в проходящем свете; белая линия – 100 мкм.

монстрирующие характерную для ИПСК морфологию (рис. 2, *д*). В результате проведенного отбора для последующей детальной характеристики было сохранено 117 клонов, которые показали стабильный рост и поддерживали типичный для плюрипотентных клеток морфологический фенотип.

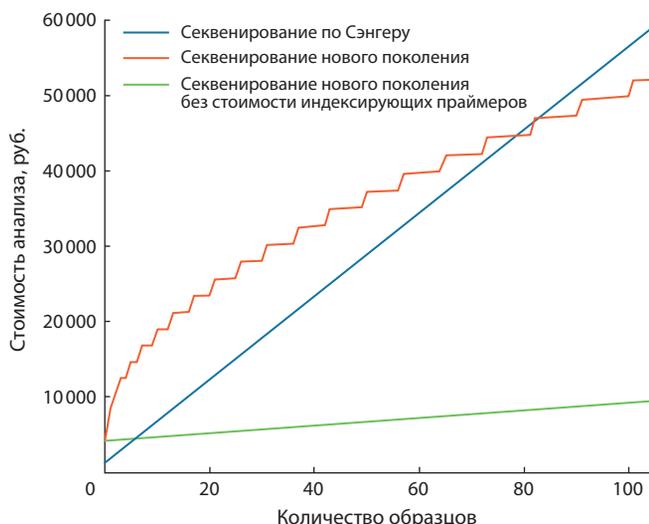
На следующем этапе было необходимо решить экспериментальную задачу по генотипированию более чем 100 образцов. Для оценки общей эффективности и результатов геномного редактирования традиционно используют секвенирование по Сэнгеру ампликонов редактируемого сайта (см. рис. 1, *в*, *з*). Однако этот метод трудозатратен, а при анализе большого количества образцов (больше 80) становится нерентабельным (рис. 3). Мы предлагаем альтернативное решение – массовое генотипирование клеточных клонов с использованием секвенирования нового поколения. Этот подход обеспечивает высокую масштабируемость, существенно ускоряет процесс анализа, позволяя одновременно получить информацию как о составе аллелей, так и об их частотах в популяции клеток.

Для анализа NGS целевой фрагмент, содержащий внесенную замену, амплифицировали с помощью сайт-специфичных праймеров, фланкирующих сайт мутагенеза.

Ключевым требованием для успешного генотипирования (при использовании парных прочтений длиной 150 пар оснований) является расположение праймеров на расстоянии от сайта мутагенеза в пределах 50–130 пар оснований до 5'-концов праймеров (см. рис. 1, *в*). При таком дизайне эксперимента обеспечивается покрытие внесенной замены обоими прочтениями в рамках одной пары. Для интеграции в NGS, на 5'-концы сайт-специфичных праймеров необходимо добавить технические последовательности, представляющие собой 3'-регионы ДНК-адаптеров для секвенирования. Для завершения последовательностей адаптеров проводится второй раунд ПЦР с набором универсальных праймеров, содержащих индексные последовательности для дальнейшего пулирования и демультиплексирования библиотек.

#### Оценка эффективности редактирования и мозаичности колоний

После электропорации получили и проанализировали в общей сложности 117 клонов: 69 клонов с использованием нРНК Н044 (см. рис. 2, *а*) и 48 для Н068 (см. рис. 2, *в*). Общая эффективность редактирования для направляющей РНК Н044, рассчитанная как доля клонов с любыми об-



**Рис. 3.** Зависимость стоимости секвенирования методом секвенирования по Сэнгеру (синий цвет) и методом NGS с учетом (оранжевый) и без учета (зеленый) стоимости синтеза индексирующих праймеров.

наруженными заменами с частотой встречаемости аллеля более 2 %, составила 54 % (37/69). Поскольку для дальнейших исследований нам требовалось получить линии, несущие замены в гетерозиготном состоянии, мы провели фильтрацию и отобрали клоны с долей отредактированного аллеля  $\geq 30$  %. При таком пороге эффективность редактирования составила 13 % (9/69), при этом два клон демонстрировали частоты внесенных замен 68 и 89 %, что говорит об их потенциальной гомозиготности (доля отредактированного аллеля  $\geq 30$  %). Для направляющей РНК Н068 суммарная эффективность составила 31 % (15/48) при пороге 2 % и 4 % (2/48) при пороге 30 %.

Видно, что при теоретически возможных только гомо- и гетерозиготных состояниях локуса наблюдаемые частоты встречаемости альтернативного аллеля могут существенно отклоняться от ожидаемых 0, 50 и 100 %. Часть наблюдаемых расхождений, вероятно, связана с особенностями подготовки NGS-библиотек (многочисленные циклы амплификации и риск кросс-контаминации). Однако ключевую роль, по-видимому, играет вероятность кросс-контаминации культур на этапе субклонирования после электропорации. Исключить ее полностью затруднительно, поскольку клетки не всегда диссоциируют до одиночных, могут повторно агрегировать и мигрировать в ходе совместного культивирования. Дополнительную вариабельность вносит асинхронность процесса редактирования: между трансфекцией и началом экспрессии и работы компонентов системы редактирования оснований проходит время, за которое некоторые клетки успевают пройти цикл деления. В результате в разных клетках одной и той же колонии могут появиться различные мутации или не появиться вовсе, что формирует мозаичную по внесенным заменам культуру. Таким образом, независимо от исходной эффективности мутагенеза, необходимо

исследовать значительное количество клонов, чтобы исключить мозаичные колонии.

В окно редактирования (4–6-й нуклеотиды спейсера) попадает сразу несколько аденинов. Результаты свидетельствуют о том, что замены происходят по всем «валентным» позициям, но с различной эффективностью: наибольшей – по пятой позиции (chr7:70768044, около 45 % позитивных клонов при пороге в 2 % частоты аллеля) и вдвое реже по четвертой позиции (chr7:70768043, около 17 % при том же пороге) (см. рис. 2, а). Шестая позиция представлена гуанином и не поддается оценке эффективности замен, хотя и попадает в теоретическое окно редактирования. Частота редактирования по остальным аденинам региона не превышала 5 %. Примечательно, что случаев одновременных замен двух нуклеотидов в пределах одного и того же аллеля не зафиксировано. Также не зафиксировано случаев внесения замены на нецелевой цепи (конверсии T→C).

Для наглядности на рис. 2 приведен предлагаемый способ визуализации результатов (см. рис. 2, а–в). Представленный метод визуализации позволяет быстро сравнивать образцы между собой, оценивать уровень шума и отбирать позитивные образцы для получения модельных клеточных линий для дальнейшей работы.

Мы отобрали пять колоний с подходящими заменами в потенциально гетерозиготном и генетически гомогенном состоянии. Для оценки того, какой аллель маркирует внесенная замена, а также для оценки мозаичности, необходимо фазировать внесенные замены с герминальными гетерозиготными вариантами в геномном окружении. Так, если наблюдается всего два комплаунд-аллеля, то можно считать, что полученные гетерозиготные линии моноклональны. Мы фазировали полученные замены для клонов с однонуклеотидным вариантом (SNV chr7:70768282 hg38, rs3829006 G/A), существование которого было обнаружено при полногеномном секвенировании пациента, маркирующий аллель, находящийся в цис-окружении с хромосомной аберрацией. Было отобрано три немозаичных клонов (с частотой встречаемости мозаичных аллелей менее 5 %), два из которых несли синонимичную замену в гетерозиготном состоянии, при этом маркирующую разные аллели у разных клонов (табл. 2).

**Таблица 2.** Фазирование и оценка мозаичности модифицированных по гену *AUTS2* клонов

Клон	Позиция замены, hg38	Частота замен, %	Аллель	Мозаицизм
63-2_A32	chr7:70768044	41.5	SNV	–
63-1_C16	chr7:70768043	40.2	WT	–
63-1_C21	chr7:70768044	48.4	WT	–
63-2_A30	chr7:70768043	46.5	WT	+
63-2_A30	chr7:70768044	39.5	SNV	+
63-2_C19	chr7:70768064	28.60	SNV	+

Примечание. WT – дикий тип; SNV – вариант с заменой.

**Таблица 3.** Сравнение генотипирования с помощью секвенирования по Сэнгеру и с помощью NGS

Критерий	Секвенирование по Сэнгеру	NGS-генотипирование
Масштабируемость	Позволяет секвенировать по восемь образцов одновременно, а анализировать результаты по одному образцу	Позволяет секвенировать и анализировать десятки-сотни образцов одновременно
Трудоемкость пробоподготовки	Один раунд ПЦР, реакция секвенирования и очистка реакции секвенирования. Требуется индивидуальная пробоподготовка (ПЦР и очистка образцов)	Два раунда ПЦР с целевыми и индексированными праймерами индивидуально для каждого образца. После индексирования в результате второго раунда ПЦР их можно объединить в один и очистить с помощью магнитных SPRI-частиц
Скорость анализа результатов	Индивидуальный анализ хроматограмм. Требуется ручная интерпретация хроматограмм	Быстрый и автоматизированный анализ
Чувствительность	Анализирует усредненный сигнал; сложно интерпретировать при наличии нескольких аллелей. Порог обнаружения – около 15–20 %; низкочастотные мутации не выявляются	Позволяет различать и количественно оценивать частоту отдельных аллелей (аллель-специфичный анализ). До 0.1 % – позволяет обнаруживать низкочастотные варианты и мозаицизм
Качество секвенирования последовательности	Крайние участки последовательности (до 100 п. о.), близкие к местам посадки праймеров, секвенируются в низком качестве и часто не пригодны для анализа	Крайние участки секвенируемой последовательности хорошего качества
Требования к праймерам	Обычные	Длинные праймеры, чувствительны к ошибкам, требуют очистки на ПААГ
Выбор клонов	Не позволяет надежно различать мозаичные или субклональные события	Позволяет выбирать клоны с заданной частотой аллеля; выявляет субклоны и исключает мозаицизм
Применимость метода	Подходит для проверки отдельных мутаций с высокой частотой в образце. Низкую эффективность редактирования (менее ~20 %) не удастся обнаружить	Подходит для широкого круга задач, включая количественную оценку редактирования, отслеживание субпопуляций и др. Обнаруживает даже 5–9 % редактирования в популяции клеток
Стоимость	Около 700 рублей на образец	Около 100 рублей на образец

После анализа колоний на мозаицизм выбранные клеточные линии культивировались для дальнейшего анализа аллельной экспрессии гена *AUTS2*. Для избежания вероятных ошибок в ходе массового культивирования клеточных линий мы повторно проанализировали наличие гетерозиготных замен, результаты анализа представлены на секвенограмме (см. рис. 2, з). Таким образом, мы получили необходимые линии и инструмент для анализа влияния хромосомной перестройки пациента на экспрессию гена *AUTS2*.

### Обсуждение

В данной работе мы получили три генетически модифицированные линии ИПСК пациента, несущие однонуклеотидные замены в экзоне 10 гена *AUTS2*, а также продемонстрировали эффективность и работоспособность стратегии отбора генетически модифицированных колоний с использованием генотипирования с помощью NGS секвенирования ампликонов. Эти линии характеризуются генетической гомогенностью, получены из двух разных линий ИПСК одного пациента, две из них несут синонимичные замены и маркируют разные аллели. Таким образом, они полностью подходят для дальнейшего изучения влияния хромосомной перестройки пациента на экспрессию гена *AUTS2* и ее дальнейшего функционального влияния.

Генетическая маркировка аллелей позволяет оценить различия в их цис-регуляции в одной и той же клетке. Такая экспериментальная схема минимизирует влияние

трудно контролируемых факторов на биологическую модель. Поскольку оба аллеля находятся под действием одного и того же набора транс-регуляторов, отражающих клеточное состояние и внешние условия, обнаруживаемые различия в экспрессии обусловлены исключительно их цис-окружением. По этой причине в анализ можно включать даже линии с несинонимичными заменами: если изменение аминокислотной последовательности влияет на транскрипцию через механизмы обратной связи, этот эффект одинаково влияет на оба аллеля и, таким образом, не искажает оценку различий в цис-регуляции.

Применение NGS для генотипирования значительно сокращает время лабораторных манипуляций и упрощает процесс анализа по сравнению с секвенированием первого поколения. В отличие от секвенирования отдельных образцов методом Сэнгера, NGS позволяет масштабировать процесс и анализировать более 100 образцов за один раз (табл. 3). Для приготовления NGS-библиотеки необходимы рутинное выделение ДНК и проведение двух раундов ПЦР с использованием праймеров, специфичных к целевой последовательности нужного района генома и вносящих необходимые технические последовательности для секвенирования.

Интеграция NGS с биоинформатическими инструментами обеспечивает автоматизированную идентификацию нуклеотидных вариантов и устраняет ошибки, обусловленные качеством хроматограмм в методе Сэнгера. Помимо этого, применение NGS при анализе результатов

**Таблица 4.** Оценка стоимости генотипирования образцов при использовании секвенирования методом Сэнгера и с помощью NGS

Этап	Наименование (реагент/процесс)	Количество	Цена за шт. (руб.)	Итого (руб.)			
<b>Секвенирование по Сэнгеру</b>							
На сайт	Сайт-специфичные праймеры	Синтез олигонуклеотидов 20 оснований	2*	540	1080		
На образец	ПЦР	MBUision полимераза	1	20	20	555***	
	Реакция Сэнгера	GenSeq-1000	2	159.5	319		
	Очистка	Переосаждение спиртом		2	0		0
		Sephadex G-50		2 на 20 мг	4		8
		SeqMag-1000		2	33		66
		VAHTS DNA Clean Beads		2	55		110
		Секвенирование	Генетический анализатор «НАНОФОР 05»	2	104**		208
<b>NGS</b>							
На сайт	Сайт-специфичные праймеры	Синтез олигонуклеотидов 55 оснований	2	2035	4070		
Зависит от количества образцов в пуле	Индексирующие праймеры*	Синтез олигонуклеотидов 55 оснований	$2\sqrt{N}$ , где $N$ – кол-во образцов	2035	$4060\sqrt{N}$ , где $N$ – кол-во образцов		
На образец	ПЦР	MBUision полимераза	2	20	40	49.744	
	Секвенирование	GenoLab MV2.0 (ячейка FCM, 300 циклов, 250M прочтений)	0.01M прочтений	243 600	9.744		
На пул	Очистка	VAHTS DNA Clean Beads	1	110	110		

\* В идеале следует использовать систему из четырех праймеров: два – для наработки продукта с помощью ПЦР и еще два – для секвенирования в обоих направлениях.

\*\* Стоимость анализа одного образца в пересчете на расход полимера в 35-сантиметровом капилляре.

\*\*\* С учетом очистки на спин-колонках с сорбентом Sephadex G-50.

редактирования генома расширяет экспериментальный арсенал, позволяя решать задачи, недостижимые при традиционном секвенировании по Сэнгеру. Если метод Сэнгера фиксирует усредненный сигнал по каждой позиции и способен выявлять мутации лишь при их доле  $\geq 20\%$  (по визуальному анализу хроматограмм) (Davidson et al., 2012; Bennett et al., 2020), то NGS предоставляет информацию по каждому аллелю, обеспечивая точность детекции до 2% и тем самым исключая двусмысленность интерпретации. Используя этот подход, мы оценили эффективность направляющей РНК в клеточной популяции.

Анализ данных NGS показал низкую частоту модификации ( $\approx 5\%$  от общей популяции), что обусловило корректировку стратегии получения мутантных линий: была выбрана конструкция с селективным маркером для обогащения клеток, в которые успешно интегрирован вектор.

На следующих этапах данные NGS-анализа позволили нам выявить клеточные линии с нужной мутацией и без мозаичности, обеспечивая чистоту генетического материала, или выбрать клоны с нужной частотой аллеля, что критически важно для получения стабильных генетических линий. Благодаря высокой глубине покрытия (секвенирование ампликона обычно выполняли с покрытием 10000 прочтений на фрагмент около 300 п. н.) можно об-

наружить даже низкочастотные варианты, что невозможно при использовании менее чувствительных методов.

Распространено мнение, что NGS секвенирование – слишком дорогое и тем самым нивелирует свои преимущества. Однако важно понимать, что в нашем случае используется целевое секвенирование ампликонов, что снижает стоимость затрат на само секвенирование до  $< 10$  руб. за образец. Основная статья затрат – индексирующие праймеры, число которых растет пропорционально  $\sqrt{N}$ , где  $N$  – количество образцов. Несмотря на это, данные праймеры универсальны: их можно применять как в амплификационном протоколе подготовки библиотек, описанном в работе, так и в стандартных схемах с лигированием универсальных адаптеров. Также возможно использование готовых коммерческих наборов индексирующих праймеров, однако зачастую это дороже индивидуального синтеза. В конечном итоге при учете цены заказа индексирующих праймеров, NGS-генотипирование начинает превосходить метод Сэнгера по стоимости при объеме  $> 80$  образцов; без учета этой статьи (при уже имеющемся наборе индексирующих праймеров) экономия и вовсе достигается уже при анализе  $> 7$  образцов (см. рис. 3), что делает предложенный метод перспективным для широкого круга исследователей (табл. 4).

NGS-секвенирование ампликонов может быть применено и при анализе эффективности других систем редактирования, основанных на системе CRISPR/Cas9. Если индуцируемая мутация, например, при получении протяженных делеций или инверсий, превышает возможности покрытия NGS (максимальная длина пары ридов 300 пар оснований), достаточно секвенировать только границы мутации, чтобы проверить целостность прилегающих участков и выявить нежелательные события мутагенеза. Метод применим, пока мутагенез не затронул сайты праймеров; тем же ограничением обладает и любой другой способ ПЦР-генотипирования.

Однако следует принимать во внимание ограничения и недостатки предлагаемого метода. Так как происходит большое количество циклов ПЦР, в данных секвенирования зачастую появляются амплификационные артефакты. Также из-за большого количества манипуляций с ампликонами до индексирования и значительной чувствительности метода высока чувствительность и к контаминации. Следует обязательно придерживаться правил правильной работы с пост-амплификационным материалом и исключать любые его контакты с местами приготовления смесей ПЦР и выделения ДНК. Кроме того, использование праймеров, содержащих длинные некомplementарные геномной ДНК последовательности на 5'-конце, приводит к общему снижению эффективности амплификации, что требует индивидуальной оптимизации условий ПЦР для каждой пары.

## Заключение

В настоящей статье представлен метод получения модельных линий на основе ИПСК с маркированным синонимичной мутацией аллелем гена *AUTS2* с применением технологии редакторов оснований. Полученные линии ИПСК будут использованы для изучения влияния хромосомных aberrаций на экспрессию *AUTS2* и для исследования функциональных последствий этого нарушения.

Предложенный подход к генотипированию большого количества клонов с помощью NGS-секвенирования ампликонов демонстрирует высокую эффективность и масштабируемость при снижении стоимости анализа в сравнении с традиционными методами. Большая пропускная способность метода позволила провести анализ значительного количества образцов, а также оценить мозаицизм, состав и частоты аллелей в клеточной популяции.

Таким образом, представленная стратегия дает возможность эффективно получать клеточные модели для исследования влияния цис-регуляторных вариантов на транскрипционную активность генов, при этом контролируя генетическую гомогенность модифицированных клеточных моделей, что может быть востребовано как в фундаментальных исследованиях, так и в решении прикладных задач.

## Список литературы / References

Смирнов А.В., Юнусова А.М., Лукьянчикова В.А., Баттулин Н.Р. Система CRISPR/Cas9 – универсальный инструмент геномной инженерии. *Вавиловский журнал генетики и селекции*. 2016; 20(4):493-510. doi 10.18699/VJ16.175

- [Smirnov A.V., Yunusova A.M., Lukyanchikova V.A., Battulin N.R. CRISPR/Cas9, a universal tool for genomic engineering. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2016;20(4): 493-510. doi 10.18699/VJ16.175 (in Russian)]
- Baux D., Van Goethem C., Ardouin O., Guignard T., Bergougnot A., Koenig M., Roux A.-F. MobiDetails: online DNA variants interpretation. *Eur J Hum Genet*. 2021;29(2):356-360. doi 10.1038/s41431-020-00755-z
- Bennett E.P., Petersen B.L., Johansen I.E., Niu Y., Yang Z., Chamberlain C.A., Met Ö., Wandall H.H., Frödin M. INDEL detection, the 'Achilles heel' of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Res*. 2020;48(21):11958-11981. doi 10.1093/nar/gkaa975
- Billon P., Bryant E.E., Joseph S.A., Nambiar T.S., Hayward S.B., Rothstein R., Ciccio A. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol Cell*. 2017;67(6):1068-1079.e4. doi 10.1016/j.molcel.2017.08.008
- Chen X., McAndrew M.J., Lapinaite A. Unlocking the secrets of ABEs: the molecular mechanism behind their specificity. *Biochem Soc Trans*. 2023;51(4):1635-1646. doi 10.1042/BST20221508
- Davidson C.J., Zeringer E., Champion K.J., Gauthier M.-P., Wang F., Boonyaratankornkit J., Jones J.R., Schreiber E. Improving the limit of detection for Sanger sequencing: a comparison of methodologies for *KRAS* variant detection. *BioTechniques*. 2012;53(3):182-188. doi 10.2144/000113913
- De Masi C., Spitalieri P., Murdocca M., Novelli G., Sangiuolo F. Application of CRISPR/Cas9 to human-induced pluripotent stem cells: from gene editing to drug discovery. *Hum Genomics*. 2020; 14(1):25. doi 10.1186/s40246-020-00276-2
- Gaudelli N.M., Komor A.C., Rees H.A., Packer M.S., Badran A.H., Bryson D.I., Liu D.R. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*. 2017;551(7681): 464-471. doi 10.1038/nature24644
- Geurts M.H., Gandhi S., Boretto M.G., Akkerman N., Derks L.L.M., Van Son G., Celotti M., ... Andersson-Rolf A., Chuvá De Sousa Lopes S.M., Van Es J.H., Van Boxtel R., Clevers H. One-step generation of tumor models by base editor multiplexing in adult stem cell-derived organoids. *Nat Commun*. 2023;14(1):4998. doi 10.1038/s41467-023-40701-3
- Global Burden of Disease Study 2021 Autism Spectrum Collaborators. The global epidemiology and health burden of the autism spectrum: findings from the Global Burden of Disease Study 2021. *Lancet Psychiatry*. 2025;12(2):111-121. doi 10.1016/S2215-0366(24) 00363-8
- Gridina M., Lagunov T., Belokopytova P., Torgunakov N., Nuriddinov M., Nurislamov A., Nazarenko L.P., ... Filipenko M., Rogaev E., Shilova N.V., Lebedev I.N., Fishman V. Combining chromosome conformation capture and exome sequencing for simultaneous detection of structural and single-nucleotide variants. *Genome Med*. 2025;17(1):47. doi 10.1186/s13073-025-01471-3
- Grünwald J., Zhou R., Garcia S.P., Iyer S., Lareau C.A., Aryee M.J., Joung J.K. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*. 2019;569(7756):433-437. doi 10.1038/s41586-019-1161-z
- Hwang G.-H., Park J., Lim K., Kim S., Yu J., Yu E., Kim S.-T., Eils R., Kim J.-S., Bae S. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics*. 2018;19(1):542. doi 10.1186/s12859-018-2585-4
- Jaganathan K., Kyriazopoulou Panagiotopoulou S., McRae J.F., Fazl Darbandi S., Knowles D., Li Y.I., Kosmicki J.A., ... Gao H., Kia A., Batzoglou S., Sanders S.J., Farh K.K.-H. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535-548.e24. doi 10.1016/j.cell.2018.12.015
- Jin S., Zong Y., Gao Q., Zhu Z., Wang Y., Qin P., Liang C., Wang D., Qiu J.-L., Zhang F., Gao C. Cytosine, but not ade-

- nine, base editors induce genome-wide off-target mutations in rice. *Science*. 2019; 364(6437):292-295. doi 10.1126/science.aaw7166
- Komor A.C., Kim Y.B., Packer M.S., Zuris J.A., Liu D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533(7603):420-424. doi 10.1038/nature17946
- Liang Y., Chen F., Wang K., Lai L. Base editors: development and applications in biomedicine. *Front Med*. 2023;17(3):359-387. doi 10.1007/s11684-023-1013-y
- Lu Z., Huang X. Base editors: a powerful tool for generating animal models of human diseases. *Cell Stress*. 2018;2(10):242-245. doi 10.15698/cst2018.10.156
- Rees H.A., Liu D.R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*. 2018;19(12):770-788. doi 10.1038/s41576-018-0059-1
- Reese M.G., Eeckman F.H., Kulp D., Haussler D. Improved splice site detection in Genie. In: Proceedings of the First Annual International Conference on Computational Molecular Biology (RECOMB '97). 1997;232-240. doi 10.1145/267521.267766
- Rowe R.G., Daley G.Q. Induced pluripotent stem cells in disease modelling and drug discovery. *Nat Rev Genet*. 2019;20(7):377-388. doi 10.1038/s41576-019-0100-z
- Salnikov P., Belokopytova P., Yan A., Viesná E., Korablev A., Serova I., Lukyanchikova V., Stepanchuk Y., Torgunakov N., Tikhomirov S., Fishman V. Direction and modality of transcription changes caused by TAD boundary disruption in *Slc29a3/Unc5b* locus depends on tissue-specific epigenetic context. *Epigenetics Chromatin*. 2025; 18(1):55. doi 10.1186/s13072-025-00618-1
- Uddin F., Rudin C.M., Sen T. CRISPR gene therapy: applications, limitations, and implications for the future. *Front Oncol*. 2020;10:1387. doi 10.3389/fonc.2020.01387
- Yu Y., Leete T.C., Born D.A., Young L., Barrera L.A., Lee S.-J., Rees H.A., Ciaramella G., Gaudelli N.M. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat Commun*. 2020;11(1):2052. doi 10.1038/s41467-020-15887-5
- Zuo E., Sun Y., Wei W., Yuan T., Ying W., Sun H., Yuan L., Steinmetz L.M., Li Y., Yang H. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*. 2019;364(6437):289-292. doi 10.1126/science.aav9973

---

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 14.06.2025. После доработки 05.11.2025. Принята к публикации 11.11.2025.

doi 10.18699/vjgb-26-02

# Биоинформатический анализ механизмов жизнеспособности линий опухолевых клеток при делеции генов-супермишеней

Д.А. Четверина <sup>1</sup>, Н.Я. Козельчук <sup>1</sup>, Д.В. Ломаев <sup>1</sup>, А.А. Штиль <sup>2</sup>, М.М. Ерохин <sup>1</sup> <sup>1</sup> Институт биологии гена Российской академии наук, Москва, Россия<sup>2</sup> Национальный медицинский исследовательский центр онкологии им. Н.Н. Блохина, Москва, Россия yermmaxbio@yandex.ru

**Аннотация.** Мутации генов и изменения эпигенетической регуляции экспрессии генов являются характерными признаками злокачественных новообразований. Сочетания данных нарушений формируют биологические особенности индивидуальных опухолей на молекулярном уровне. Разработка стратегий персонализированного лечения требует глубокого понимания молекулярных «портретов» отдельных опухолей. В рамках крупномасштабного проекта Dependency Map (DepMap) обширные панели линий опухолевых клеток человека тестируются на чувствительность к инактивации отдельных генов. Ранее на основе данных DepMap нами охарактеризованы гены, получившие название «супермишени», при делеции которых существенно снижена жизнеспособность клеток конкретного тканевого происхождения при минимальном нарушении жизнеспособности других линий. В настоящем исследовании определены факторы низкой жизнеспособности (ингибирования пролиферации или гибели) клеточных линий при инактивации генов-супермишеней. В результате установлено, что в 79 % случаев низкая жизнеспособность может быть вызвана эпигенетическими изменениями в экспрессии генов. В остальных случаях (21 %) она вызвана нарушениями структуры генов. Исходя из полученных данных, можно выделить три группы, содержащие разного типа нарушения экспрессии генов. В первой группе низкая жизнеспособность клеток коррелирует с повышением экспрессии гена-супермишени (например, *SOX10* и *HNF1B*). Во второй группе детектируется гиперэкспрессия гена, отличного от делетируемой супермишени (пары генов *FOXA1* и *SPDEF*, *TP63* и *SERPINB13* и др.). Третья группа характеризуется корреляциями между пониженной экспрессией определенных генов и чувствительностью опухолевых клеток (пары генов *FAM126A* и *FAM126B*, *SMARCA2* и *SMARCA4* и др.). Наблюдаемые генетические изменения включали GOF-мутации (*KRAS*, *BRAF* и др.), LOF-мутации (*STAG1*, *SMARCA2* и др.), слияние генов (*BCR-ABL1*, *PAX3-FOXO1* и др.) и амплификации (*CPM*, *BEST3* и др.). Таким образом, разные молекулярные механизмы выступают предикторами ответа опухолевых клеток на ингибирование генов-супермишеней.

**Ключевые слова:** опухоли; рак; онкомаркеры; Dependency Map; DepMap; транскрипция; мутации; *SMARCA2*; *SMARCA4*; *APP*; *FOXA1*; *ATP6V0A2*; *ATP6V0A1*; биоинформатика; анализ баз данных; персонализированная медицина

**Для цитирования:** Четверина Д.А., Козельчук Н.Я., Ломаев Д.В., Штиль А.А., Ерохин М.М. Биоинформатический анализ механизмов жизнеспособности линий опухолевых клеток при делеции генов-супермишеней. *Вавиловский журнал генетики и селекции*. 2026;30(1):85-93. doi 10.18699/vjgb-26-02

**Финансирование.** Работа выполнена при поддержке гранта Российского научного фонда № 20-74-10099.

**Благодарности.** Работа проводилась с использованием приборной базы ИБГ РАН, поддержанной Министерством науки и высшего образования РФ.

## The computational analysis of tumor cell sensitivity to supertarget deletion

D.A. Chetverina <sup>1</sup>, N.Y. Kozelchuk <sup>1</sup>, D.V. Lomaev <sup>1</sup>, A.A. Shtil <sup>2</sup>, M.M. Erokhin <sup>1</sup> <sup>1</sup> Institute of Gene Biology Russian Academy of Sciences, Moscow, Russia<sup>2</sup> N.N. Blokhin National Medical Research Center of Oncology, Moscow, Russia yermmaxbio@yandex.ru

**Abstract.** Gene mutations and altered epigenetic regulation of gene expression are characteristic features of malignant neoplasms. Combinations of these abnormalities form molecular features of individual tumors. In the large-scale Dependency Map (DepMap) project, the broad panels of human tumor cell lines are being tested for sensitivity to single gene inactivation. Using DepMap data, we have previously identified a set of genes termed supertargets, the deletion of which significantly reduced the survival of cells of a particular tissue origin while minimally impairing the unrelated cell lines. In the present study, we determined the factors of viability (inhibition of proliferation or death) of cell lines in which the supertarget genes have been deleted. We found that, in 79 % of cases, the reduced survival may be caused by epigenetic changes of gene expression. In the remaining 21 % of cases, it is associated with altered gene structure. Three groups containing different types of gene expression alterations can be distinguished. In the first

group, the reduced cell survival correlated with a higher expression of the supertarget gene (e.g., *SOX10* and *HNF1B*). In the second group, a gene different from the deleted supertarget was overexpressed (gene pairs: *FOXA1* and *SPDEF*, *TP63* and *SERPINB13*, etc.). The third group was characterized by correlations between low expression of a certain gene and tumor cell sensitivity (e.g., *FAM126A* and *FAM126B*, *SMARCA2* and *SMARCA4*). The genetic changes included GOF mutations (*KRAS*, *BRAF* genes, etc.), LOF mutations (*STAG1*, *SMARCA2* genes, etc.), gene fusions (*BCR-ABL1*, *PAX3-FOXO1*, etc.), and amplification (*CPM*, *BEST3*, etc.). Therefore, many different molecular mechanisms act as predictors of tumor cell response to inhibition of supertarget genes.

**Key words:** tumors; cancer; oncomarkers; Dependency Map; DepMap; transcription; mutations; *SMARCA2*; *SMARCA4*; *APP*; *FOXA1*; *ATP6V0A2*; *ATP6V0A1*; bioinformatics; database analysis; personalized medicine

**For citation:** Chetverina D.A., Kozelchuk N.Y., Lomaev D.V., Shtil A.A., Erokhin M.M. The computational analysis of tumor cell sensitivity to supertarget deletion. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):85-93. doi 10.18699/vjgb-26-02

## Введение

Современный подход к терапии опухолей предусматривает выявление молекулярных механизмов, специфичных для конкретного типа опухоли. Принцип мишень-направленной терапии – инактивация таких факторов для достижения антипролиферативного эффекта и/или гибели клеток при наименьшем повреждении других клеток, в первую очередь неопухолевых (Verma, 2012; Pfohl et al., 2021). Поиск опухолеспецифичных мишеней включает скрининговые исследования на широких панелях линий клеток различного тканевого происхождения. Наиболее масштабным является проект Dependency Map (DepMap, <https://depmap.org/>), в котором использована инактивация обширного набора генов с помощью технологий RNAi и CRISPR/Cas9 (Tsherniak et al., 2017; Arafteh et al., 2025).

Ранее с помощью ресурса DepMap мы проанализировали 27 гистологических типов опухолей и для каждого типа определили пять генов, CRISPR/Cas9-опосредованный нокаут которых обуславливал сниженную жизнеспособность (ингибирование пролиферации или гибель) клеток именно данного типа (Chetverina et al., 2023). Такие гены, названные «супермишенями», могут рассматриваться как перспективные мишени для персонализированной терапии.

В настоящем исследовании мы с помощью ресурса DepMap изучили корреляции сниженной жизнеспособности клеток с генетическими и эпигенетическими изменениями при делеции супермишеней. Выявленные факторы могут быть задействованы для прогноза чувствительности клеток отдельных типов опухолей к инактивации тех или иных молекулярных механизмов.

## Материалы и методы

Анализ жизнеспособности клеток опухолевых линий выполняли на основе базы DepMap (<https://depmap.org/portal/>, Tsherniak et al., 2017), версия Public 22Q4+Score, Chronos (<https://doi.org/10.25452/figshare.plus.27993248.v1>). Уровень жизнеспособности клеточных линий анализировали с помощью параметра “Gene effect”. Для отнесения признака к факторам низкой жизнеспособности применены два критерия: значение параметра значимости (Importance)  $\geq 10\%$  и степень достоверности (критерий Стьюдента)  $< 0.01$ . Уровни экспрессии генов проанализированы с использованием версии DepMap Expression Public 22Q4 ([https://depmap.org/portal/data\\_](https://depmap.org/portal/data_)

[page/?tab=allData&releasename= DepMap%20Public%2022Q4&filename=OmicsExpressionProteinCodingGenesTPMLogp1.csv](https://depmap.org/portal/data_page/?tab=allData&releasename=DepMap%20Public%2022Q4&filename=OmicsExpressionProteinCodingGenesTPMLogp1.csv)). Копийность генов определяли с помощью версии DepMap Copy Number Public 22Q4 ([https://depmap.org/portal/data\\_page/?tab=allData&releasename=DepMap%20Public%2022Q4&filename=OmicsCNGene.csv](https://depmap.org/portal/data_page/?tab=allData&releasename=DepMap%20Public%2022Q4&filename=OmicsCNGene.csv)).

Для анализа белок-белковых взаимодействий использовали базу данных STRING (<https://string-db.org/>, версия 12.0). Учитывали только взаимодействия, подтвержденные биохимическими методами.

## Результаты и обсуждение

### Низкая жизнеспособность линий опухолевых клеток при делеции генов, кодирующих супермишени

Ранее для 27 гистологических типов опухолей нами определено по пять генов (всего идентифицировано 124 уникальных гена-супермишени, 9 генов дважды повторялись в разных типах опухолей и один трижды), нокаут которых приводит к наибольшему ингибированию пролиферации или к гибели клеток (Chetverina et al., 2023). Для выявления генетических и/или эпигенетических изменений, коррелирующих с чувствительностью клеток к делеции супермишеней, мы использовали параметр значимости (Importance) проекта DepMap. При поиске корреляций исследован общий пул клеточных линий из базы DepMap без разделения на гистологические типы.

Всего обнаружено 167 корреляций, ассоциированных с низкой жизнеспособностью опухолевых клеток (т.е. ингибирование пролиферации/или гибелью) при делеции генов-супермишеней. Последние проанализированы индивидуально; для ряда генов выявлено не менее двух факторов низкой жизнеспособности; для 23 из 124 супермишеней достоверных корреляций не обнаружено. Среди факторов низкой жизнеспособности большинство (132 случая) связано с изменениями экспрессии генов, а 35 случаев – с изменениями структуры генов (мутациями).

### Нарушения экспрессии генов, коррелирующие с низкой жизнеспособностью клеток при делеции генов-супермишеней

**Повышение экспрессии супермишени.** Детальный анализ выявил три группы, содержащие разного типа нарушения экспрессии генов при делеции супермишеней.

**Таблица 1.** Гены-супермишени, при делеции которых наблюдается корреляция между их гиперэкспрессией и низкой жизнеспособностью клеток

№ п/п	Супермишень	Importance, %	Pearson	p-value	№ п/п	Супермишень	Importance, %	Pearson	p-value
1	<i>SOX10</i>	81.8	-0.84	2.99E-268	26	<i>SOX17</i>	39.6	-0.446	2.90E-50
2	<i>IRF4</i>	78.9	-0.813	7.16E-238	27	<i>BATF3</i>	38.7	-0.347	9.19E-30
3	<i>POU2AF1</i>	73.3	-0.67	7.79E-132	28	<i>DUSP4</i>	38.3	-0.416	2.64E-43
4	<i>HNF1B</i>	72.7	-0.751	1.76E-182	29	<i>PHOX2A</i>	37.3	-0.455	2.36E-52
5	<i>PAX8</i>	71.9	-0.677	1.38E-135	30	<i>MYOG</i>	37.0	-0.554	1.03E-81
6	<i>MYB</i>	71.4	-0.671	5.92E-132	31	<i>TP63</i>	36.4	-0.702	1.19E-149
7	<i>PAX5</i>	67.6	-0.645	3.21E-119	32	<i>LEF1</i>	35.7	-0.278	3.05E-19
8	<i>SPDEF</i>	66.7	-0.446	4.28E-50	33	<i>MECOM</i>	34.5	-0.363	1.38E-32
9	<i>MYCN</i>	62.9	-0.647	5.57E-120	34	<i>FLI1</i>	33.9	-0.312	4.86E-24
10	<i>MEF2C</i>	62.1	-0.464	1.27E-54	35	<i>SPI1</i>	31.8	-0.523	2.14E-71
11	<i>FOXA1</i>	61.0	-0.433	3.48E-47	36	<i>CCNE1</i>	31.5	-0.451	2.29E-51
12	<i>ESR1</i>	59.0	-0.481	3.55E-59	37	<i>PPARG</i>	31.3	-0.385	1.05E-36
13	<i>SATB2</i>	58.5	-0.278	3.06E-19	38	<i>PRDM1</i>	29.4	-0.286	2.27E-20
14	<i>WT1</i>	52.4	-0.209	2.34E-11	39	<i>ISL1</i>	27.5	-0.456	1.42E-52
15	<i>GATA3</i>	52.1	-0.522	3.84E-71	40	<i>HAND2</i>	27.4	-0.469	6.63E-56
16	<i>KLF5</i>	51.8	-0.609	8.22E-103	41	<i>ZBTB18</i>	24.5	-0.247	2.16E-15
17	<i>MITF</i>	51.4	-0.472	9.45E-57	42	<i>KLB</i>	20.8	-0.383	1.80E-36
18	<i>MYOD1</i>	51.3	-0.711	3.27E-155	43	<i>TEAD3</i>	20.2	-0.355	3.67E-31
19	<i>GFI1</i>	49.2	-0.423	8.75E-45	44	<i>NFIA</i>	15.1	-0.155	8.37E-07
20	<i>TRPS1</i>	48.0	-0.458	4.83E-53	45	<i>DOCK5</i>	14.9	-0.312	4.96E-24
21	<i>BCL6</i>	44.7	-0.383	2.03E-36	46	<i>PAX3</i>	13.2	-0.367	2.26E-33
22	<i>NKX2-1</i>	41.9	-0.391	5.39E-38	47	<i>PARD3</i>	12.3	-0.243	6.06E-15
23	<i>PHOX2B</i>	41.4	-0.458	4.49E-53	48	<i>SOX4</i>	12.3	-0.163	2.09E-07
24	<i>IGF2BP1</i>	40.3	-0.348	5.62E-30	49	<i>CDX2</i>	11.8	-0.399	1.09E-39
25	<i>IKZF1</i>	40.2	-0.583	1.58E-92	50	<i>IRS1</i>	10.0	-0.292	3.81E-21

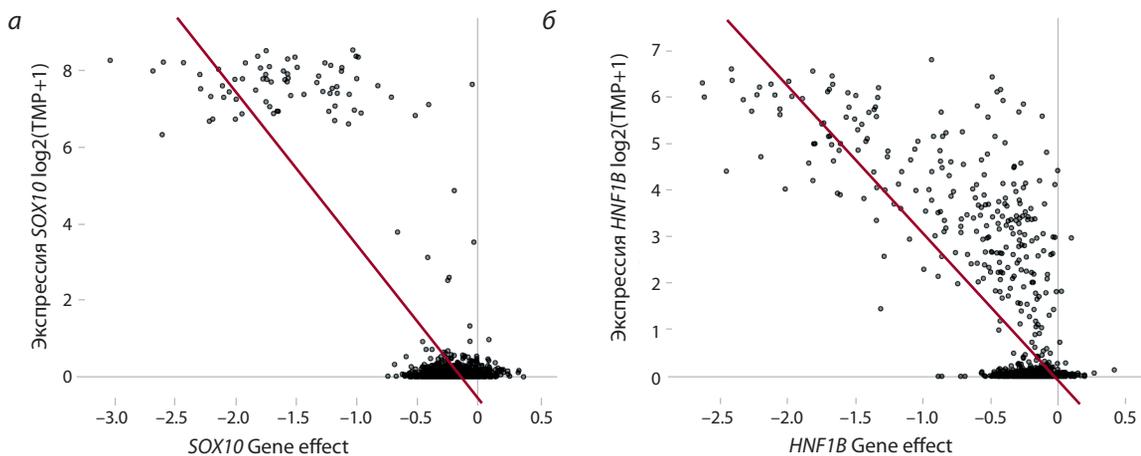
В первой группе (50 случаев) низкая жизнеспособность клеток коррелирует с повышенной экспрессией супермишени (табл. 1).

На рис. 1 представлено два характерных примера. Ген *SOX10* (SRY-box transcription factor 10, см. рис. 1, а) был идентифицирован нами ранее как супермишень для клеточных линий, происходящих из меланомы. Этот ген кодирует транскрипционный фактор, важный для дифференцировки клеток в различных тканях, включая центральную и периферическую нервную систему, меланоциты и хондроциты. Дерегуляция *SOX10* связана с канцерогенезом (Bahmad et al., 2023).

Ген *HNF1B* (HNF1 homeobox B, см. рис. 1, б) выявлен в качестве супермишени для линий рака почки. Этот ген кодирует гомеобокс-содержащий транскрипционный фактор, важный в эмбриональном развитии печени, почек и поджелудочной железы (Chandra et al., 2021). Нарушения функций *HNF1B*, включая герминальные и соматические мутации, обнаружены в солидных опухолях различного тканевого происхождения (Bártů et al., 2018).

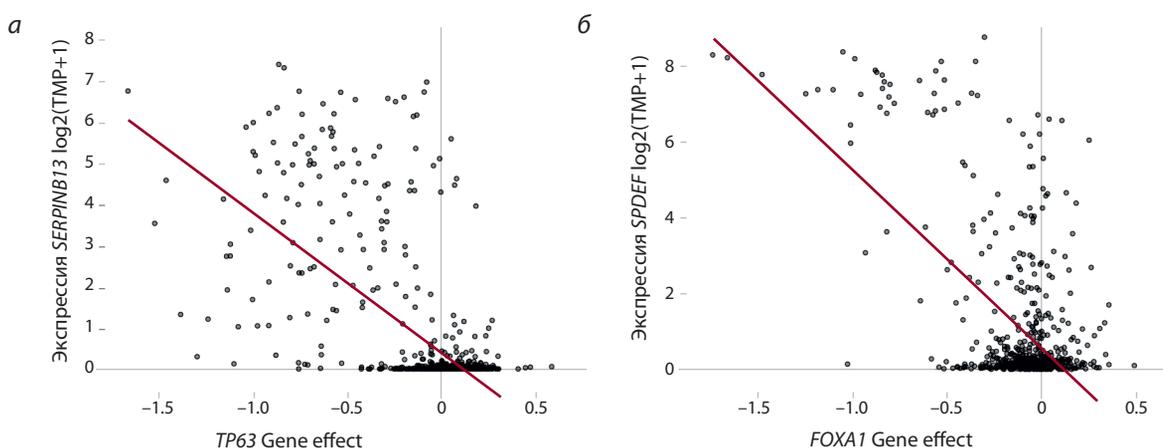
**Гиперэкспрессия гена, отличающегося от делетируемой супермишени.** Вторая группа корреляций содержит 71 случай, когда низкая жизнеспособность клеток при делеции супермишени коррелирует с гиперэкспрессией гена, отличного от нокаутированной супермишени (табл. 2).

В 65 случаях выявлены корреляции между низкой жизнеспособностью клеток при нокауте гена-супермишени и повышенным уровнем экспрессии гена, не входящего в список супермишеней. Примером такой зависимости может служить пара генов *TP63* и *SERPINB13* (рис. 2, а). Ген *TP63*, описанный нами как супермишень для линий рака слизистой оболочки верхних дыхательных и пищеварительных путей (upper aerodigestive cancer), кодирует белок p63 – транскрипционный регулятор, важный, как и его гомологи p53 и p73, в биологии опухолей (Sadu Murari et al., 2025). Для *SERPINB13* данных меньше; известно, что этот ген кодирует ингибитор сериновых протеаз, подавляющий активность катепсинов K и L (Jayakumar et al., 2003; Welss et al., 2003). Недавние исследования показали, что SERPINB13 может выступать в роли потенциального



**Рис. 1.** Повышенная экспрессия гена-супермишени и низкая жизнеспособность опухолевых клеток.

Корреляции между экспрессией генов и низкой жизнеспособностью клеток при делеции генов *SOX10* (а) и *HNF1B* (б). По оси Y – уровень экспрессии соответствующих генов (значения представлены в виде параметра  $\log_2(\text{TPM}+1)$ ), по оси X – оценочный параметр Gene effect, отражающий жизнеспособность клеток при делеции гена-супермишени. Чем ниже параметр Gene effect, тем ниже жизнеспособность и выше ингибирование пролиферации и/или гибель клеток данной клеточной линии при инактивации соответствующего гена. Здесь и на рис. 2–4: черные точки – значения для отдельных линий; диагональная прямая – линия регрессии.



**Рис. 2.** Гиперэкспрессия гена, не относящегося к супермишеням, и низкая жизнеспособность клеток при делеции супермишени.

а – корреляция между экспрессией *SERPINB13* и жизнеспособностью клеток при делеции *TP63*; б – корреляция между экспрессией *SPDEF* и жизнеспособностью клеток при делеции *FOXA1*.

онкогена в клетках плоскоклеточного рака легкого (Zhang et al., 2025). Однако, насколько известно, данных о кооперативности p63 и SERPINB13 нет.

В шести случаях мы наблюдали корреляцию между низкой жизнеспособностью клеток при делеции гена-супермишени и уровнем экспрессии другого гена, но так же, как и делелируемый, классифицированного как супермишень в предыдущем скрининге: пары генов *FOXA1–SPDEF*, *MYOD1–MYOG*, *TCF7L2–HNF4A*, *RXRA–PPARG*, *HAND2–PHOX2A*, *SMARCA1–MYOD1*. Так, для пары *FOXA1* (forkhead box A1) и *SPDEF* (SAM pointed domain containing ETS transcription factor, см. рис. 2, б) оба гена ранее идентифицированы нами как супермишени в линиях рака молочной железы (Chetverina et al., 2023). Продукт гена *FOXA1* активирует ген *SPDEF*, а белок SPDEF – тран-

скрипцию *FOXA1* (Buchwalter et al., 2013; Paranjarye et al., 2020). Хотя данные о непосредственных взаимодействиях белков FOXA1 и SPDEF отсутствуют, их гены находятся в одном регуляторном кластере, включающем также *ER* и *GATA3*, важные в патогенезе рака молочной железы (Lemieux et al., 2017).

Обнаруженные зависимости могут быть обусловлены физическими взаимодействиями между белками, кодируемыми исследованными генами, и участием в одних и тех же регуляторных путях. Для проверки такой возможности мы протестировали наличие физических контактов между продуктами выявленных пар генов с использованием базы данных STRING. Выяснилось, что для 12 пар генов есть данные о прямых белок-белковых взаимодействиях (см. табл. 2). Можно предположить, что в остальных случаях

**Таблица 2.** Гены-супермишени, при делеции которых низкая жизнеспособность клеток коррелирует с повышенной экспрессией другого гена

№ п/п	Супер-мишень	Ген чувствительности	Importance, %	Pearson	p-value	№ п/п	Супер-мишень	Ген чувствительности	Importance, %	Pearson	p-value
1	<i>EBF1</i>	<i>CD79A</i>	83.5	-0.76	1.57E-189	37	<i>SOS1</i>	<i>ITGA2B</i>	20.6	-0.333	2.37E-27
2	<i>EBF1</i>	<i>VPREB3</i>	39.8	-0.758	4.90E-188	38	<i>TRIM8</i>	<i>TPTE2</i>	20.6	-0.536	8.63E-76
3	<i>MEF2B</i>	<i>BORCS8-MEF2B</i>	73.8	-0.575	2.46E-89	39	<i>TRIM8</i>	<i>CYP4F22</i>	15.3	-0.501	7.02E-65
4	<i>MEF2B</i>	<i>ELL3</i>	39.2	-0.375	7.15E-35	40	<i>MECOM</i>	<i>WNT7A</i>	18.6	-0.355	3.26E-31
5	<i>TP63</i>	<i>SERPINB13</i>	61.1	-0.659	8.31E-126	41	<i>CRTC2</i>	<i>AVPI1</i>	18.5	-0.284	4.89E-20
6	<i>FOXA1</i>	<i>SPDEF</i>	57.6	-0.596	1.86E-97	42	<i>CRTC2</i>	<i>SIK1</i>	16.3	-0.194	5.41E-10
7	<i>CTNNB1*</i>	<i>AXIN2*</i>	55.9	-0.619	2.43E-107	43	<i>BCL6</i>	<i>ELL3</i>	18.3	-0.221	1.31E-12
8	<i>PAX5</i>	<i>CD19</i>	45.4	-0.701	5.36E-149	44	<i>EPAS1</i>	<i>SEC14L6</i>	18.3	-0.27	3.30E-18
9	<i>PAX5</i>	<i>VPREB3</i>	14.9	-0.706	2.08E-152	45	<i>EPAS1*</i>	<i>EGLN3*</i>	11.8	-0.123	8.80E-05
10	<i>PAX5</i>	<i>CD79A</i>	11.8	-0.682	6.22E-138	46	<i>RUNX1*</i>	<i>TAL1*</i>	17.5	-0.435	1.87E-47
11	<i>TCF7L2*</i>	<i>RNF43*</i>	42.6	-0.5	1.40E-64	47	<i>RUNX1</i>	<i>GMFG</i>	10.9	-0.497	9.38E-64
12	<i>TCF7L2</i>	<i>HNF4A</i>	14.6	-0.604	8.80E-101	48	<i>HAND2</i>	<i>PHOX2A</i>	15.7	-0.509	3.09E-67
13	<i>PAX3</i>	<i>SOX8</i>	40.4	-0.382	3.72E-36	49	<i>POU2AF1</i>	<i>JCHAIN</i>	14.7	-0.676	9.14E-135
14	<i>NFE2L2</i>	<i>AKR1C1</i>	38.9	-0.545	1.27E-78	50	<i>ISL1</i>	<i>STMN2</i>	14.1	-0.539	8.01E-77
15	<i>NFE2L2*</i>	<i>MAFG*</i>	16.2	-0.267	7.06E-18	51	<i>EGFR</i>	<i>ITGB4</i>	13.9	-0.491	5.21E-62
16	<i>NFE2L2</i>	<i>NQO1</i>	16.1	-0.304	7.02E-23	52	<i>EGFR*</i>	<i>AREG*</i>	12.3	-0.504	1.01E-65
17	<i>LMO2</i>	<i>LYL1</i>	36.1	-0.496	2.06E-63	53	<i>EGFR</i>	<i>C1orf116</i>	10.5	-0.53	9.47E-74
18	<i>LMO2</i>	<i>NFE2</i>	30.0	-0.525	5.43E-72	54	<i>EGFR</i>	<i>GJB3</i>	10.5	-0.53	8.17E-74
19	<i>LMO2</i>	<i>SIGLEC8</i>	11.1	-0.599	1.14E-98	55	<i>HS2ST1*</i>	<i>CKAP4*</i>	13.7	-0.355	4.24E-31
20	<i>LMO2</i>	<i>GATA1</i>	10.7	-0.612	4.03E-104	56	<i>PARD3*</i>	<i>PARD6B*</i>	13.7	-0.347	9.26E-30
21	<i>PRDM1</i>	<i>TXNDC5</i>	35.2	-0.281	1.12E-19	57	<i>PARD3</i>	<i>KCNJ16</i>	13.0	-0.44	9.86E-49
22	<i>CDX2</i>	<i>CDX1</i>	34.3	-0.495	2.98E-63	58	<i>DUSP4</i>	<i>MIA</i>	13.0	-0.509	3.01E-67
23	<i>MAPK1</i>	<i>S100B</i>	31.5	-0.489	2.32E-61	59	<i>STAT5B</i>	<i>PRSS57</i>	12.9	-0.453	5.73E-52
24	<i>MAPK1</i>	<i>CPN1</i>	16.8	-0.548	1.52E-79	60	<i>HMGA2</i>	<i>OS9</i>	12.8	-0.282	9.55E-20
25	<i>IL13RA1</i>	<i>IL5RA</i>	28.0	-0.345	2.30E-29	61	<i>EDF1</i>	<i>HCAR2</i>	12.4	-0.347	8.23E-30
26	<i>JAK2*</i>	<i>CSF2RB*</i>	26.8	-0.434	2.75E-47	62	<i>FZD5</i>	<i>FER1L6</i>	12.2	-0.32	3.01E-25
27	<i>JAK2*</i>	<i>MPL*</i>	20.8	-0.404	1.42E-40	63	<i>PRKAR1A*</i>	<i>PRKACA*</i>	12.2	-0.333	1.95E-27
28	<i>JUN</i>	<i>SERPINE1</i>	24.8	-0.502	5.32E-65	64	<i>PRKAR1A</i>	<i>TICAM2</i>	11.5	-0.45	2.94E-51
29	<i>JUN</i>	<i>PDCD1LG2</i>	12.8	-0.544	2.63E-78	65	<i>PRKAR1A</i>	<i>RTL3</i>	10.4	-0.351	2.13E-30
30	<i>MYOD1</i>	<i>MYOG</i>	23.7	-0.724	1.95E-163	66	<i>PHOX2A</i>	<i>SLC6A2</i>	11.7	-0.412	2.60E-42
31	<i>SPI1</i>	<i>TYROBP</i>	22.9	-0.509	2.52E-67	67	<i>MITF</i>	<i>MLANA</i>	11.6	-0.579	8.32E-91
32	<i>FGFR1</i>	<i>FGF2</i>	22.3	-0.432	8.11E-47	68	<i>SMARCAL1</i>	<i>MYOD1</i>	11.1	-0.298	5.12E-22
33	<i>ITGB1</i>	<i>GJB5</i>	22.3	-0.488	2.82E-61	69	<i>BATF3</i>	<i>MYBPC2</i>	10.8	-0.439	1.37E-48
34	<i>ITGB1</i>	<i>LAMA3</i>	11.1	-0.39	6.99E-38	70	<i>TRPS1</i>	<i>CREB3L4</i>	10.2	-0.303	8.35E-23
35	<i>RXRA*</i>	<i>PPARG*</i>	20.8	-0.371	3.87E-34	71	<i>RAB10</i>	<i>BICDL2</i>	10.0	-0.387	2.86E-37
36	<i>ETV6</i>	<i>PTGER3</i>	20.7	-0.411	4.64E-42						

Примечание. Символом «\*» отмечены пары белков, для которых биохимическими методами показаны прямые взаимодействия (база данных STRING). Оттенками серого цвета отмечены идущие подряд строки с одинаковыми генами.

имеет место опосредованное влияние, или данные о физических взаимодействиях между белками еще предстоит установить.

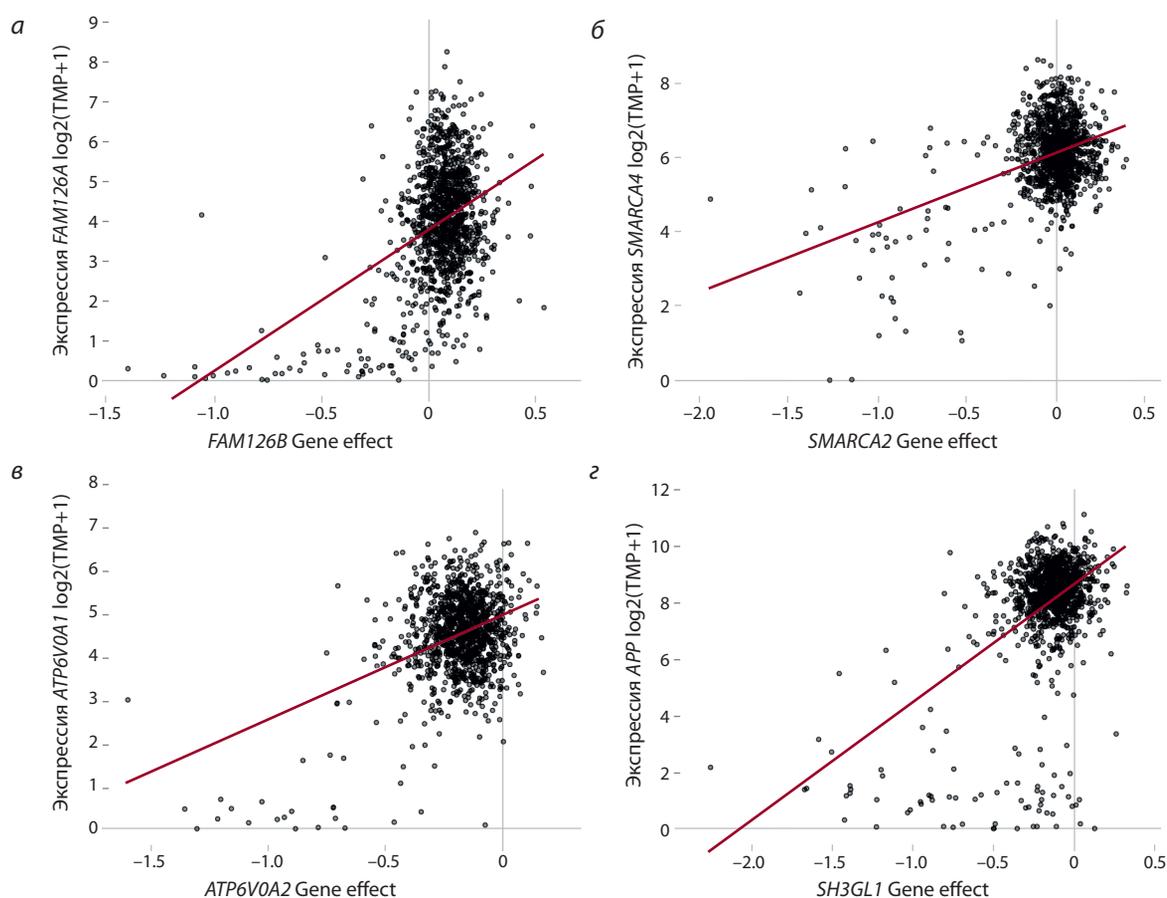
**Пониженная экспрессия гена, не относящегося к супермишеням.** В третьей группе корреляций представлено 11 случаев, когда низкая жизнеспособность клеток при

делеции супермишени коррелирует с низкой экспрессией определенного гена. Наиболее часто (9 случаев) такая ситуация наблюдается при наличии близкого гомолога гена-супермишени (табл. 3).

На рис. 3, а, б представлены характерные примеры. В первом случае низкая жизнеспособность клеток при де-

**Таблица 3.** Гены-супермишени, при делеции которых низкая жизнеспособность клеток коррелирует с пониженной экспрессией другого гена

№ п/п	Супермишень	Ген чувствительности	Importance, %	Pearson	p-value	Гены являются гомологами
1	<i>RPP25L</i>	<i>RPP25</i>	75.0	0.743	1.05E-176	Да
2	<i>FAM126B</i>	<i>FAM126A</i>	71.6	0.408	1.64E-41	Да
3	<i>ATP6V0E1</i>	<i>ATP6V0E2</i>	60.3	0.564	2.13E-85	Да
4	<i>SMARCA2</i>	<i>SMARCA4</i>	55.0	0.39	9.62E-38	Да
5	<i>ATP6V0A2</i>	<i>ATP6V0A1</i>	51.6	0.412	2.61E-42	Да
6	<i>STAG1</i>	<i>STAG2</i>	28.9	0.253	3.61E-16	Да
7	<i>SH3GL1</i>	<i>APP</i>	26.0	0.535	1.86E-75	Нет
8	<i>TIMM17A</i>	<i>TIMM17B</i>	24.7	0.399	1.04E-39	Да
9	<i>ARHGEF7</i>	<i>ARHGEF6</i>	23.0	0.457	6.96E-53	Да
10	<i>VRK1</i>	<i>VRK2</i>	12.2	0.357	1.66E-31	Да
11	<i>STAT5B</i>	<i>NPHP1</i>	10.8	0.198	2.33E-10	Нет



**Рис. 3.** Пониженная экспрессия гена, не относящегося к супермишеням, коррелирует с низкой жизнеспособностью клеток при делеции супермишени.

*a* – корреляция экспрессии *FAM126A* и чувствительности клеток к делеции *FAM126B*; *б* – корреляция экспрессии *SMARCA4* и жизнеспособности клеток при делеции *SMARCA2*; *в* – корреляция экспрессии *ATP6V0A1* и чувствительности клеток к делеции *ATP6V0A2*; *з* – корреляция экспрессии *APP* и чувствительности клеток к делеции *SH3GL1*.

лации гена *FAM126B* зависит от низкого уровня экспрессии его гомолога *FAM126A* (см. рис. 3, *a*). Гены *FAM126A* и *FAM126B* (известные также как *HYCC1* и *HYCC2*, *hуccin* PI4KA lipid kinase complex subunit) кодируют белки, являющиеся компонентами комплекса протеинкиназы PI4KIIIα/

PI4KA, регулирующего липидный состав и асимметрию плазматической мембраны (Suresh et al., 2024). Пример, показанный на рис. 3, *б*, демонстрирует зависимость низкого показателя Gene effect при делеции *SMARCA2* от низкого уровня экспрессии его гомолога *SMARCA4*. *SMARCA2* и

SMARCA4 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2) являются компонентами комплекса SWI/SNF, ремоделирующего хроматин. Нарушения структуры и функций комплекса SWI/SNF обнаруживаются в опухолях (Nguyen et al., 2023; Reddy et al., 2023).

В случае нокаута гена *ATP6V0A2* (ATPase H<sup>+</sup> transporting V0 subunit a2) наблюдается зависимость низкой жизнеспособности клеток от низкого уровня экспрессии его гомолога *ATP6V0A1* (см. рис. 3, в). Гены *ATP6V0A2* и *ATP6V0A1* кодируют компонент протонного канала V-АТФазы, поддерживающей электрохимический градиент протонов по обе стороны плазматической мембраны. Однако V-АТФаза, помимо основной функции, участвует в сигнальном пути Notch/Wnt (Eaton et al., 2021). Детали функционирования *ATP6V0A1/2* в опухолевых клетках еще предстоит установить.

При нокауте гена *SH3GL1* (SH3 domain containing GRB2 like 1) низкая жизнеспособность клеток ассоциирована с пониженной экспрессией гена *APP* (рис. 4, з). Ген *SH3GL1* кодирует эндофилин А2, важный для динамики внутриклеточных мембран, в частности для эндоцитоза (Yang et al., 2023). Роль *APP* (amyloid precursor protein) детально изучена при болезни Альцгеймера: фрагмент *APP* (бета-

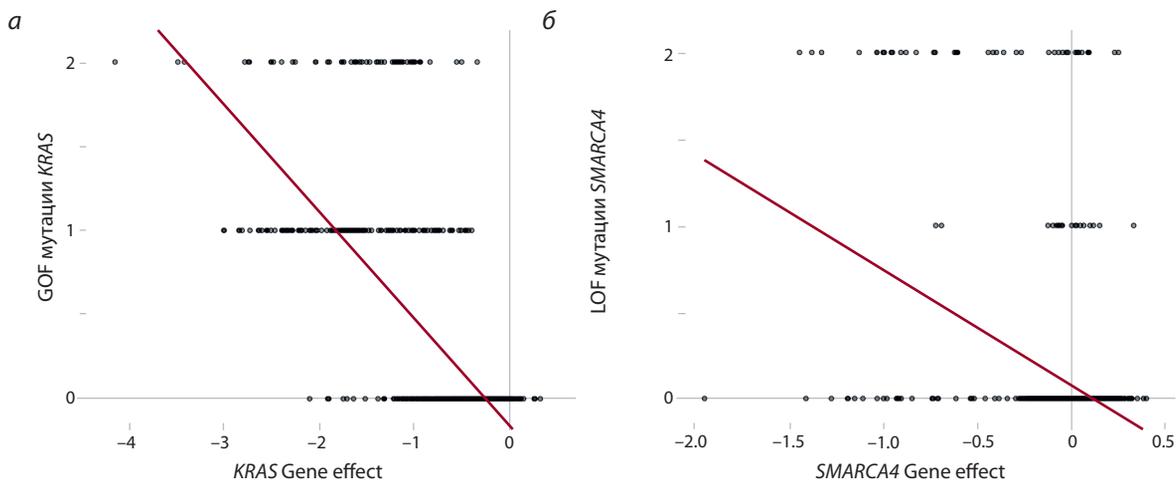
амилоид) образует характерные бляшки в тканях мозга (Chen et al., 2024). Имеются указания на участие *APP* и в канцерогенезе, однако здесь данный фактор изучен недостаточно (Lee et al., 2021). О взаимодействии *SH3GL1* и *APP* в биологии опухолей не сообщалось. Таким образом, анализ корреляций жизнеспособности клеток позволяет не только прогнозировать ответ опухоли на делецию определенных генов, но и обнаруживать ранее неизвестные регуляторные взаимосвязи между генами и их продуктами.

**Генетические мутации.** В 35 случаях наблюдались корреляции между низкой жизнеспособностью опухолевых клеток и нарушениями структуры генов в анализируемых клеточных линиях. Выявленные типы нарушений включают точечные мутации (12 случаев), слияние генов (13) и амплификацию (10 случаев) (табл. 4).

На рис. 4, а показаны корреляции для гена *KRAS* (Kirsten rat sarcoma virus), кодирующего малую ГТФазу – ключевой компонент сигнального пути, активируемого при взаимодействии эпидермального фактора роста с соответствующим рецептором (Seres et al., 2025). В данном случае мутации, активирующие *KRAS* (gain-on-function, GOF), коррелируют с чувствительностью клеток к делеции этого гена.

**Таблица 4.** Изменения структуры генов-супермишеней и мутации других генов, ассоциированные с низкой жизнеспособностью клеток при делеции генов-супермишеней

№ п/п	Супер-мишень	Ген чувствительности	Importance, %	Pearson	<i>p</i> -value	№ п/п	Супер-мишень	Ген чувствительности	Importance, %	Pearson	<i>p</i> -value
GOF мутации						LOF мутации					
1	<i>KRAS</i>	<i>KRAS</i>	69.3	-0.680	3.91E-146	1	<i>STAG1</i>	<i>STAG2</i>	32.8	-0.425	2.65E-48
2	<i>BRAF</i>	<i>BRAF</i>	68.2	-0.689	7.68E-152	2	<i>SMARCA2</i>	<i>SMARCA4</i>	31.9	-0.399	3.33E-42
3	<i>MAPK1</i>	<i>BRAF</i>	15.7	-0.438	2.06E-51	3	<i>CTNNB1</i>	<i>APC</i>	24.2	-0.528	6.34E-78
4	<i>CTNNB1</i>	<i>CTNNB1</i>	13.8	-0.288	6.57E-22	4	<i>WRN</i>	<i>KMT2B</i>	19.4	-0.517	2.50E-74
5	<i>DOCK5</i>	<i>KRAS</i>	13.1	-0.331	9.41E-29	5	<i>WRN</i>	<i>RPL22</i>	12.2	-0.479	2.02E-62
Слияние генов						6	<i>FAM126B</i>	<i>RNF43</i>	14.8	-0.231	1.78E-14
1	<i>BCR</i>	<i>BCR_ABL1</i>	44.8	-0.703	2.50E-150	7	<i>EPAS1</i>	<i>VHL</i>	10.5	-0.292	1.45E-22
2	<i>ABL1</i>	<i>BCR_ABL1</i>	42.6	-0.626	2.73E-110	Амплификации					
3	<i>TRIM8</i>	<i>EWSR1_FLI1</i>	23.4	-0.546	5.04E-79	Отрицательная корреляция					
4	<i>NFE2L2</i>	<i>AKR1C1_AKR1C6P</i>	18.7	-0.328	1.59E-26	1	<i>CPM</i>	<i>CPM</i>	32.0	-0.376	1.30E-37
5	<i>PAX3</i>	<i>PAX3_FOXO1</i>	18.6	-0.662	1.99E-127	2	<i>BEST3</i>	<i>BEST3</i>	20.8	-0.175	6.97E-09
6	<i>MYOG</i>	<i>PAX3_FOXO1</i>	16.5	-0.638	1.00E-115	3	<i>MYCN</i>	<i>NBAS</i>	16.2	-0.42	2.28E-47
7	<i>FLI1</i>	<i>EWSR1_FLI1</i>	13.5	-0.548	1.24E-79	4	<i>NFIA</i>	<i>NFIA</i>	11.9	-0.213	1.74E-12
8	<i>MECOM</i>	<i>AC093821.1_LINC01091</i>	13.3	-0.278	2.86E-19	5	<i>CCNE1</i>	<i>CCNE1</i>	11.0	-0.299	9.61E-24
9	<i>DLL1</i>	<i>PAX3_FOXO1</i>	12.8	-0.41	4.85E-42	6	<i>KLB</i>	<i>UPF1</i>	10.9	-0.093	2.26E-03
10	<i>ZBTB18</i>	<i>PAX3_FOXO1</i>	11.2	-0.482	1.60E-59	7	<i>MEF2B</i>	<i>KIR2DS4</i>	10.8	-0.089	3.60E-03
11	<i>CCNE1</i>	<i>AC093821.1_LINC01091</i>	10.8	-0.298	4.77E-22	8	<i>CAPS2</i>	<i>CAPS2</i>	10.7	-0.176	5.99E-09
12	<i>MYOD1</i>	<i>PAX3_FOXO1</i>	10.5	-0.659	8.23E-126	9	<i>HMGA2</i>	<i>HMGA2</i>	10.0	-0.229	3.01E-14
13	<i>JAK2</i>	<i>HBB_AC104389.5</i>	10.3	-0.365	6.35E-33	Положительная корреляция					
						1	<i>DDX5</i>	<i>DDX17</i>	13.3	0.338	3.28E-30



**Рис. 4.** Генетические мутации, ассоциированные с низкой жизнеспособностью опухолевых клеток при делеции генов-супермишеней.

*а* – корреляция между количеством мутаций гена *KRAS* и чувствительностью клеток к делеции *KRAS*; *б* – корреляция между количеством мутаций гена *SMARCA4* и чувствительностью клеток к делеции *SMARCA2*.

Пример, приведенный на рис. 4, *б*, демонстрирует зависимость показателя Gene effect при делеции *SMARCA2* от количества инактивирующих (loss-of-function, LOF) мутаций в близком гомологе – гене *SMARCA4*. Как указано выше, *SMARCA2* и *SMARCA4* – компоненты хроматин-ремоделирующего комплекса SWI/SNF. Повышенная чувствительность клеток к делеции *SMARCA2* коррелирует как с наличием повреждающих мутаций в гене *SMARCA4* (см. рис. 4, *б*), так и с пониженной экспрессией *SMARCA4* (см. рис. 3, *б*).

## Заключение

Проведен биоинформатический анализ низкой жизнеспособности линий опухолевых клеток при делеции генов-супермишеней. Термин «супермишень» предложен нами для генов, инактивация которых существенно снижает жизнеспособность клеток определенного тканевого происхождения (по данным DepMap). В настоящей работе установлено, что в большинстве случаев низкая жизнеспособность клеток коррелирует с изменениями экспрессии определенных генов – как супермишеней, так и генов, не относящихся к таковым. На жизнеспособность клеток могут влиять и генетические мутации – активирующие, инактивирующие, а также слияние генов и амплификация. Полученные взаимоотношения могут быть использованы для тестирования чувствительности опухолевых клеток различного гистогенеза к инактивации генов-супермишеней с помощью низкомолекулярных соединений, применяемых в практике.

Установленные взаимосвязи имеют отношение к актуальным проблемам медицины, главным образом онкологии – разработке персонализированного лечения, основанного на биологических особенностях опухоли конкретного пациента – ее молекулярном «портрете». Развитие представлений о механизмах чувствительности опухоли к тому или иному лекарственному воздействию

углубляет понимание важной роли генетических и/или эпигенетических факторов.

Положительный результат терапии (т. е. гибель или ингибирование пролиферации опухолевых клеток) определяется не одним признаком, а их сочетаниями. Для эффективности конкретного препарата требуется наличие в опухоли двух (возможно, и более) событий одновременно: гибель клеток при действии препарата особенно выражена, если мутация одного гена сопровождается нарушением второго механизма («синтетическая летальность»). Увеличивается количество примеров персонализированной терапии опухолей с мутационной «нагрузкой» при выявлении пар генов синтетической летальности (Du et al., 2023; Previtali et al., 2024). Для опухолей с транскрипционной «нагрузкой» (в частности, новообразований детского возраста) генетические мутации менее важны, чем эпигенетические нарушения (Comitani et al., 2023). На модели дрозофилы показано, что даже временные транскрипционные нарушения могут приводить к развитию опухоли без генетических мутаций (Pargeno et al., 2024).

Установленные в настоящей работе молекулярные события, обуславливающие летальность клеток при инактивации транскрипционных механизмов, расширяют возможности комбинированных фармакологических воздействий в терапии «мутационных» и «транскрипционных» новообразований.

## Список литературы / References

- Arafah R., Shibue T., Dempster J.M., Hahn W.C., Vazquez F. The present and future of the Cancer Dependency Map. *Nat Rev Cancer*. 2025;25(1):59-73. doi 10.1038/s41568-024-00763-x
- Bahmad H.F., Thiravialingam A., Sriganeshan K., Gonzalez J., Alvarez V., Ocejio S., Abreu A.R., Avellan R., Arzola A.H., Hachem S., Poppi R. Clinical significance of *SOX10* expression in human pathology. *Curr Issues Mol Biol*. 2023;45(12):10131-10158. doi 10.3390/cimb45120633

- Bártů M., Dundr P., Němejcová K., Tichá I., Hojný H., Hájková N. The role of HNF1B in tumorigenesis of solid tumours: a review of current knowledge. *Folia Biol (Praha)*. 2018;64(3):71-83. doi 10.14712/fb2018064030071
- Buchwalter G., Hickey M.M., Cromer A., Selfors L.M., Gunawardane R.N., Frishman J., Jeselsohn R., Lim E., Chi D., Fu X., Schiff R., Brown M., Brugge J.S. PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell*. 2013;23(6):753-767. doi 10.1016/j.ccr.2013.04.026
- Chandra S., Srinivasan S., Batra J. Hepatocyte nuclear factor 1 beta: a perspective in cancer. *Cancer Med*. 2021;10(5):1791-1804. doi 10.1002/cam4.3676
- Chen J., Chen J.-S., Li S., Zhang F., Deng J., Zeng L.-H., Tan J. Amyloid precursor protein: a regulatory hub in Alzheimer's disease. *Aging Dis*. 2024;15(1):201-225. doi 10.14336/AD.2023.0308
- Chetverina D., Vorobyeva N.E., Györffy B., Shtil A.A., Erokhin M. Analyses of genes critical to tumor survival reveal potential 'super-targets': focus on transcription. *Cancers (Basel)*. 2023;15(11):3042. doi 10.3390/cancers15113042
- Comitani F., Nash J.O., Cohen-Gogo S., Chang A.I., Wen T.T., Mareshwari A., Goyal B., ... Behjati S., Malkin D., Villani A., Irwin M.S., Shlien A. Diagnostic classification of childhood cancer using multi-scale transcriptomics. *Nat Med*. 2023;29(3):656-666. doi 10.1038/s41591-023-02221-x
- Du Y., Luo L., Xu X., Yang X., Yang X., Xiong S., Yu J., Liang T., Guo L. Unleashing the power of synthetic lethality: augmenting treatment efficacy through synergistic integration with chemotherapy drugs. *Pharmaceutics*. 2023;15(10):2433. doi 10.3390/pharmaceutics15102433
- Eaton A.F., Merkulova M., Brown D. The H<sup>+</sup>-ATPase (V-ATPase): from proton pump to signaling complex in health and disease. *Am J Physiol Cell Physiol*. 2021;320(3):C392-C414. doi 10.1152/ajpcell.00442.2020
- Jayakumar A., Kang Y., Frederick M.J., Pak S.C., Henderson Y., Holton P.R., Mitsudo K., Silverman G.A., EL-Naggar A.K., Brömme D., Clayman G.L. Inhibition of the cysteine proteinases cathepsins K and L by the serpin headpin (SERPINB13): a kinetic analysis. *Arch Biochem Biophys*. 2003;409(2):367-374. doi 10.1016/s0003-9861(02)00635-5
- Lee H.N., Jeong M.S., Jang S.B. Molecular characteristics of amyloid precursor protein (APP) and its effects in cancer. *Int J Mol Sci*. 2021;22(9):4999. doi 10.3390/ijms22094999
- Lemieux S., Sargeant T., Laperrière D., Ismail H., Boucher G., Rozendaal M., Lavallée V.-P., Ashton-Beaucage D., Wilhelm B., Hébert J., Hilton D.J., Mader S., Sauvageau G. MiSTIC, an integrated platform for the analysis of heterogeneity in large tumour transcriptome datasets. *Nucleic Acids Res*. 2017;45(13):e122. doi 10.1093/nar/gkx338
- Nguyen V.T., Tessema M., Weissman B.E. The SWI/SNF complex: a frequently mutated chromatin remodeling complex in cancer. In: Chen J., Wang G.G., Lu J. (Eds) *Epigenetics in Oncology*. *Cancer Treat Res*. Springer, Cham, 2023;190:211-244. doi 10.1007/978-3-031-45654-1\_7
- Paranjapye A., Mutolo M.J., Ebron J.S., Leir S.-H., Harris A. The FOXA1 transcriptional network coordinates key functions of primary human airway epithelial cells. *Am J Physiol Lung Cell Mol Physiol*. 2020;319(1):L126-L136. doi 10.1152/ajplung.00023.2020
- Parreno V., Loubiere V., Schuettengruber B., Fritsch L., Rawal C.C., Erokhin M., Györffy B., ... Butova N.L., Chiolo I., Chetverina D., Martinez A.-M., Cavalli G. Transient loss of Polycomb components induces an epigenetic cancer fate. *Nature*. 2024;629(8012):688-696. doi 10.1038/s41586-024-07328-w
- Pfohl U., Pflaume A., Regenbrecht M., Finkler S., Graf Adelmann Q., Reinhard C., Regenbrecht C.R.A., Wedeken L. Precision oncology beyond genomics: the future is here – it is just not evenly distributed. *Cells*. 2021;10(4):928. doi 10.3390/cells10040928
- Previtali V., Bagnolini G., Ciamarone A., Ferrandi G., Rinaldi F., Myers S.H., Roberti M., Cavalli A. New horizons of synthetic lethality in cancer: current development and future perspectives. *J Med Chem*. 2024;67(14):11488-11521. doi 10.1021/acs.jmedchem.4c00113
- Reddy D., Bhattacharya S., Workman J.L. (mis)-Targeting of SWI/SNF complex(es) in cancer. *Cancer Metastasis Rev*. 2023;42(2):455-470. doi 10.1007/s10555-023-10102-5
- Sadu Murari L.S., Kunkel S., Shetty A., Bents A., Bhandary A., Rive-ra-Mulia J.C. p63: a master regulator at the crossroads between development, senescence, aging, and cancer. *Cells*. 2025;14(1):43. doi 10.3390/cells14010043
- Seres M., Spacayova K., Sulova Z., Spaldova J., Breier A., Pavlikova L. Dynamic multilevel regulation of EGFR, KRAS, and MYC oncogenes: driving cancer cell proliferation through (epi)genetic and post-transcriptional/translational pathways. *Cancers (Basel)*. 2025;17(2):248. doi 10.3390/cancers17020248
- Suresh S., Shaw A.L., Pemberton J.G., Scott M.K., Harris N.J., Parson M.A.H., Jenkins M.L., Rohilla P., Alvarez-Prats A., Balla T., Yip C.K., Burke J.E. Molecular basis for plasma membrane recruitment of PI4KA by EFR3. *Sci Adv*. 2024;10(51):eadp6660. doi 10.1126/sciadv.adp6660
- Tsherniak A., Vazquez F., Montgomery P.G., Weir B.A., Kryukov G., Cowley G.S., Gill S., ... Garraway L.A., Root D.E., Golub T.R., Boehm J.S., Hahn W.C. Defining a cancer dependency map. *Cell*. 2017;170(3):564-576.e16. doi 10.1016/j.cell.2017.06.010
- Verma M. Personalized medicine and cancer. *J Pers Med*. 2012;2(1):1-14. doi 10.3390/jpm2010001
- Welsh T., Sun J., Irving J.A., Blum R., Smith A.I., Whisstock J.C., Pike R.N., von Mikecz A., Ruzicka T., Bird P.I., Abts H.F. Hurpin is a selective inhibitor of lysosomal cathepsin L and protects keratinocytes from ultraviolet-induced apoptosis. *Biochemistry*. 2003;42(24):7381-7389. doi 10.1021/bi027307q
- Yang L.-Q., Huang A.-F., Xu W.-D. Biology of endophilin and its role in disease. *Front Immunol*. 2023;14:1297506. doi 10.3389/fimmu.2023.1297506
- Zhang X., Zhang P., Ren Q., Li J., Lin H., Huang Y., Wang W. Integrative multi-omic and machine learning approach for prognostic stratification and therapeutic targeting in lung squamous cell carcinoma. *BioFactors*. 2025;51(1):e2128. doi 10.1002/biof.2128

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 27.02.2025. После доработки 08.06.2025. Принята к публикации 10.06.2025.

doi 10.18699/vjgb-26-07

## Особенности полиморфизма гена рецептора тиреоидных гормонов *THRB* у коренного населения Сибири

Б.А. Малярчук , Н.В. Похилиук , Г.А. Денисова , А.Н. Литвинов 

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия

 malbor@mail.ru

**Аннотация.** В процессах адаптации к холоду у человека задействованы гены, входящие в сигнальные пути тиреоидной системы, которыми регулируются термогенез, энергетические затраты и метаболические перестройки. Один из таких генов – *THRB*, кодирующий ядерный рецептор TR $\beta$ , с которым взаимодействует гормон щитовидной железы – трийодтиронин (Т3). От концентрации комплексов TR $\beta$ -Т3 зависит активность термогена UCP1, с помощью которого происходит разобщение окислительного фосфорилирования в митохондриях и усиливается производство тепла. Таким образом, рецепторы тиреоидных гормонов играют важную роль в адаптивном термогенезе. В настоящей работе нами впервые проведен анализ опубликованных данных о полиморфизме гена *THRB* в популяциях коренного населения Сибири с целью поиска вариантов полиморфизма, потенциально связанных с адаптацией к холоду. Анализ полиморфизма экзонов и прилегающих некодирующих участков гена *THRB* показал наличие всего одной нуклеотидной замены в белок-кодирующей области (синонимичная замена в локусе rs3752874), все остальные нуклеотидные замены выявлены преимущественно в 3'-нетранслируемых участках и интронах. Анализ распределения гаплотипов гена *THRB* позволил обнаружить два специфичных для коряков гаплотипа, характеризующиеся заменой rs762175401-A. Популяционный скрининг показал, что эта замена распространена среди коряков с частотой 13.8 %, а также присутствует у сибирских эскимосов, хотя в остальных группах населения мира частота замены rs762175401-A не превышает 0.05 % (у японцев и корейцев) или имеет еще более низкие значения (менее 0.02 %). Анализ нуклеотидной последовательности гена *THRB* показывает, что локус rs762175401 находится в 3'-нетранслируемой области в позиции +2 от терминирующего кодона. Вполне вероятно, что эта замена могла привести к изменениям в эффективности терминации трансляции и в случае повышения эффективности терминации способствовала увеличению скорости синтеза белка и, соответственно, концентрации комплексов TR $\beta$ -Т3. Предполагается, что повышение частоты варианта rs762175401-A у коряков и эскимосов, представляющих древнейшее население Северо-Востока Сибири, обусловлено долговременной адаптацией популяций к холоду.

**Ключевые слова:** ген *THRB*; популяции человека; Сибирь; адаптация к холоду; тиреоидная система; адаптивный термогенез

**Для цитирования:** Малярчук Б.А., Похилиук Н.В., Денисова Г.А., Литвинов А.Н. Особенности полиморфизма гена рецептора тиреоидных гормонов *THRB* у коренного населения Сибири. *Вавиловский журнал генетики и селекции*. 2026;30(1):94-100. doi 10.18699/vjgb-26-07

## The specific features of the thyroid hormone receptor gene *THRB* polymorphism in indigenous populations of Siberia

B.A. Malyarchuk , N.V. Pokhilyuk , G.A. Denisova , A.N. Litvinov 

Institute of Biological Problems of the North of the Far Eastern Branch of the Russian Academy of Sciences, Magadan, Russia

 malbor@mail.ru

**Abstract.** In the process of adaptation to cold in humans, genes belonging to the thyroid system signaling pathways that regulate thermogenesis, energy expenditure, and metabolic rearrangements are implicated. One such gene is the *THRB* gene, which encodes the nuclear receptor TR $\beta$ , with which the thyroid hormone triiodothyronine (T3) interacts. The activity of thermogenin UCP1 is influenced by the concentration of TR $\beta$ -T3 complexes, which serve to uncouple oxidative phosphorylation in mitochondria, thereby enhancing heat production. Consequently, thyroid hormone receptors have been demonstrated to play a significant role in adaptive thermogenesis. In the present study, we conducted a comprehensive analysis of published data on the *THRB* gene polymorphism in Siberian indigenous populations, with the objective of identifying potential associations between polymorphism variants and adaptation to cold. The analysis of exon and adjacent noncoding regions of the *THRB* gene revealed a single nucleotide substitution in the protein-coding region (synonymous substitution in the locus rs3752874). All other nucleotide substitutions were detected primarily in 3'-untranslated regions and introns. Analysis of the *THRB* haplotype distribution revealed two Koryak-specific haplotypes characterized by the rs762175401-A substitution. The results of population screening

demonstrated that this substitution is prevalent among the Koryak population, with a frequency of 13.8 %, and is also present in the Siberian Eskimo population. However, in other global populations, the frequency of the rs762175401-A substitution does not exceed 0.05 % (in the Japanese and Koreans) or has even lower values (less than 0.02 %). The analysis of the nucleotide sequence of the *THRB* gene indicates that the rs762175401 locus is situated in the 3'-untranslated region at position +2 from the terminating codon. It is plausible that this substitution may have led to alterations in translation termination efficiency. In the case of enhanced termination efficiency, it is conceivable that it contributed to an elevated rate of protein synthesis, thereby resulting in an increase in the concentration of TR $\beta$ -T3 complexes. The higher frequency of the rs762175401-A variant in the Koryak and Eskimo populations, representing the oldest populations of Northeastern Siberia, is assumed to be due to long-term adaptation of these populations to cold.

**Key words:** *THRB* gene; human populations; Siberia; cold adaptation; thyroid system; adaptive thermogenesis

**For citation:** Malyarchuk B.A., Pokhilyuk N.V., Denisova G.A., Litvinov A.N. The specific features of the thyroid hormone receptor gene *THRB* polymorphism in indigenous populations of Siberia. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):94-100. doi 10.18699/vjgb-26-07

## Введение

Физиологические исследования показали, что коренное население Сибири характеризуется повышенной скоростью метаболизма (особенно в зимнее время) и синхронными с характером метаболизма изменениями уровней тиреоидных гормонов (Leonard, 2024). У якутов, коренных жителей одного из самых холодных регионов мира, зимой метаболическая теплопродукция увеличивается в среднем на 6 % (Leonard et al., 2014). Предполагается, что в зимнее время ткани организма (прежде всего, бурая жировая ткань, БЖТ) для повышения теплопродукции гораздо активнее поглощают гормоны щитовидной железы, что приводит к резкому снижению уровней трийодтиронина (Т3) и тироксина (Т4) в крови (Levy et al., 2013; Nikanorova et al., 2023). Известно, что в БЖТ экспрессируется белок термогенин UCP1 (uncoupling protein-1), с помощью которого происходит разобщение окислительного фосфорилирования в митохондриях и выделения тепла (Bianco, Silva, 1988). В свою очередь, активация UCP1-опосредованного термогенеза обусловлена взаимодействием тиреоидных гормонов с ядерным рецептором TR $\beta$ : чем больше концентрация комплексов TR $\beta$ -Т3, тем выше активность UCP1 (Martinez de Mena et al., 2010; Lee et al., 2012; Yau, Yen, 2020; Ma et al., 2023). Таким образом, рецепторы тиреоидных гормонов, по всей видимости, играют важную роль в реализации процессов несократительного термогенеза, связанного с адаптацией к холоду.

К настоящему времени достаточно надежно установлено, что распределение вариантов полиморфизма генов разобщающих белков *UCP1*, *UCP2* и *UCP3* в популяциях человека связано с влиянием ряда природно-климатических факторов: географической широты, высоты над уровнем моря и суровости природных условий (Hancock et al., 2011; Nikanorova et al., 2021, 2022; Козлов и др., 2024).

Полиморфизм гена *THRB*, кодирующего ядерный рецептор TR $\beta$ , также охарактеризован в различных базах генетических данных (dbSNP, <https://www.ncbi.nlm.nih.gov/snp/>), однако специальных публикаций, посвященных анализу распределения полиморфных вариантов этого гена в популяциях человека, нами не обнаружено. Основные публикации, связанные с полиморфизмом гена *THRB*, посвящены поиску генетических вариантов, ассоциированных с синдромом резистентности к гормонам щитовидной железы (Dumitrescu, Refetoff, 2013), риском онко-

логических заболеваний (González-Sancho et al., 2003), регуляцией транскрипции и ремоделированием хроматина (Grøntved et al., 2015).

Таким образом, целью настоящей работы являются характеристика полиморфизма гена *THRB* у коренного населения Сибири и поиск вариантов полиморфизма, потенциально связанных с адаптацией к холоду.

## Материалы и методы

Проанализированы опубликованные ранее данные о полноэкзомном полиморфизме в популяциях коренного населения Северо-Восточной Сибири (эскимосы, чукчи, коряки;  $N = 25$ ), Центральной Сибири (эвены, эвенки, якуты;  $N = 29$ ), Южной Сибири (тувинцы, шорцы, алтайцы, буряты;  $N = 28$ ) и Западной Сибири (кеты, ханты, манси, селькупы, ненцы, нганасаны;  $N = 20$ ) – суммарно 102 человека (Pagani et al., 2016). Выполнен анализ полиморфизма всех экзонов и прилегающих к ним некодирующих участков гена *THRB*, расположенного на хромосоме 3 между позициями 24158651 и 24536773. Для выявления гаплотипов из генотипов с неизвестной гаметной фазой применяли алгоритм ELB (Excoffier et al., 2003), реализованный в пакете программ Arlequin 3.5. Для оценки статистической значимости различий в частотах полиморфных вариантов применяли точный тест Фишера. Медианную сеть гаплотипов гена *THRB* конструировали с помощью программы Network 10.2 ([www.fluxus-engineering.com](http://www.fluxus-engineering.com)). Для сравнительного анализа в работе использовали информацию из баз геномных данных в популяциях человека: dbSNP ([www.ncbi.nlm.nih.gov/projects/SNP](http://www.ncbi.nlm.nih.gov/projects/SNP)), 1000 Genomes (<https://www.internationalgenome.org/>) и gnomAD (<https://gnomad.broadinstitute.org/>).

Проведен популяционный скрининг полиморфизма в локусах rs762175401 (нуклеотидная позиция 3:24164373) и rs72619908 (позиция 3:24164268) гена *THRB*. Нумерация нуклеотидов приводится согласно референтной последовательности генома человека GRCh37.p13 (hg19). Материалом для исследования служила тотальная ДНК, выделенная из цельной крови от представителей коренного населения Северо-Эвенского района Магаданской области: коряков ( $N = 98$ ) и эвенов ( $N = 110$ ). На основании данных анкетирования обследованные коряки и эвены относили себя к указанным этническим группам на протяжении не менее двух-трех поколений.

Нуклеотидную последовательность гена *THRB*, включающую в свой состав локусы rs762175401 и rs72619908, амплифицировали с применением олигонуклеотидных праймеров 5'-GCGCCATTTTGCTGACTCAA-3' и 5'-TCTTCTCTCTCCCGCAGA-3'. Праймеры сконструированы на основе нуклеотидной последовательности гена *THRB* (под номером NC\_000003.12 в базе данных GenBank) с помощью программы Primer3 (Untergasser et al., 2012).

Продукты амплификации секвенировали с использованием набора для циклического секвенирования ДНК BrilliantDye™ Terminator Cycle Sequencing kit v3.1 (Нидерланды) и генетического анализатора ABI Prism 3500xL (Applied Biosystems, США). Для выравнивания и анализа нуклеотидных последовательностей использовали пакет программ MEGA5 (Tamura et al., 2011). Частоту аллелей, гетерозиготность и соответствие распределения генотипов равновесию Харди–Вайнберга рассчитывали

с помощью пакета программ Arlequin 3.5 (Excoffier, Lischer, 2010).

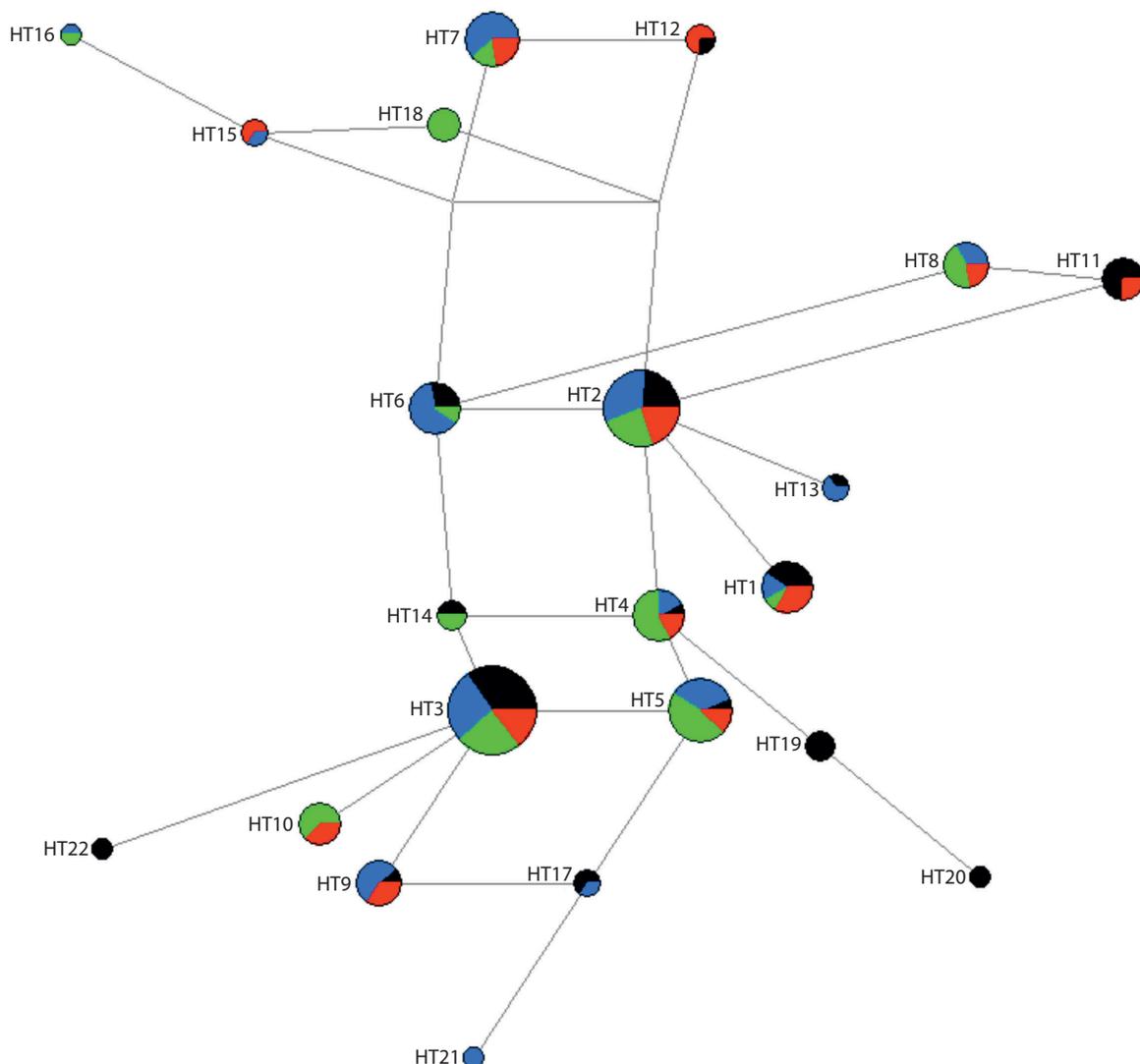
### Результаты и обсуждение

Анализ нуклеотидных последовательностей гена *THRB* у 102 представителей различных этнических групп коренного населения Сибири показал наличие полиморфизма в 22 нуклеотидных позициях этого гена (табл. 1). Однако в экзонах обнаружен лишь один вариант полиморфизма – синонимичная замена в локусе rs3752874 (аминокислотная позиция 245). Все остальные полиморфные варианты найдены в некодирующей области (преимущественно в 3'-нетранслируемых участках и интронах гена). Анализ распределения полиморфных вариантов показывает, что только в двух случаях (для rs56204436-А и rs13326381-Т) выявляется сходное распределение частот аллелей в Сибири, Восточной Азии и Европе ( $p > 0.05$ ).

**Таблица 1.** Частоты вариантов полиморфизма гена *THRB* у коренного населения Сибири, Восточной Азии и Европы

Вариант полиморфизма	Участок	Северо-Восточная Сибирь (N = 25)	Центральная Сибирь (N = 29)	Южная Сибирь (N = 28)	Западная Сибирь (N = 20)	Сибирь (N = 102)	Восточная Азия (N = 1170)	p (Сибирь-Восточная Азия)	Европа (N = 1266)	p (Сибирь-Европа)
rs3752874-A	syn	0.16	0.052	0.107	0.15	0.113	0.056	0.0031	0.149	0.181
rs34833017-G	Интрон	0	0.017	0.161	0.075	0.064	0.006	0	0.191	0.000001
rs13326381-T		0.6	0.397	0.5	0.525	0.5	0.546	<b>0.187</b>	0.472	<b>0.511</b>
rs60502621-C		0	0.017	0.018	0.025	0.015	0.101	0.000002	0.024	0.481
rs58274299-C		0	0.017	0.018	0.025	0.015	0.101	0.000002	0.024	0.481
rs150604595-A		0	0.017	0.018	0.025	0.015	0.101	0.000002	0.024	0.481
rs13090120-C		0.12	0.034	0.054	0.175	0.088	0.012	0	0.113	0.3
rs572372574-C	5'-UTR	0.02	0.069	0	0	0.025	nd	nd	nd	nd
rs75272640-T	3'-UTR	0.12	0.052	0.089	0.15	0.098	0.05	0.0085	0.137	0.134
rs2167116-A		0.06	0.19	0.143	0.225	0.152	0.092	0.0089	0.074	0.0004
rs56204436-A		0.04	0.052	0.107	0.075	0.069	0.046	<b>0.169</b>	0.057	<b>0.53</b>
rs1349265-G		0.54	0.431	0.607	0.375	0.495	0.697	0	0.694	0
rs79270057-T		0.12	0.052	0.089	0.15	0.098	0.05	0.0085	0.139	0.111
rs561918607-T		0.08	0.138	0.018	0.075	0.078	0	0	0.0003	0
rs844107-C		0.52	0.586	0.536	0.65	0.569	0.626	0.114	0.419	0.00005
rs826371-G		0.38	0.397	0.393	0.35	0.382	0.529	0.00007	0.262	0.0004
rs826372-C		0.38	0.397	0.393	0.35	0.382	0.529	0.00007	0.262	0.0004
rs826373-G		0.4	0.534	0.429	0.5	0.466	0.574	0.0032	0.257	0
rs826374-C		0.38	0.397	0.393	0.35	0.382	0.529	0.00007	0.262	0.0004
rs826375-C		0.4	0.534	0.429	0.5	0.466	0.574	0.0032	0.257	0
<b>rs72619908-G</b>		0.02	0.138	0.036	0.15	0.083	0.045	0.024	0	0
<b>rs762175401-A</b>		0.12	0	0	0	0.029	0	0	0.00003	0

Примечание. Данные для населения Восточной Азии и Европы взяты из баз данных 1000 Genomes (<https://www.internationalgenome.org/>) и gnomAD (<https://gnomad.broadinstitute.org/>). Статистическая значимость различий (p) между частотами вариантов полиморфизма в сравниваемых регионах оценивалась с помощью точного теста Фишера. 3'-UTR и 5'-UTR – 3'- и 5'-нетранслируемые области; syn – синонимичная замена в белок-кодирующем участке; nd – нет данных. Полуужирным шрифтом выделены варианты полиморфизма, для которых в настоящей работе проведен дополнительный популяционный скрининг, а также значения  $p > 0.05$  в случаях со сходным распределением частот аллелей в Сибири, Восточной Азии и Европе (для rs56204436-А и rs13326381-Т).



Медианная сеть гаплотипов гена *THRB* у коренного населения Сибири.

Черным цветом отмечены гаплотипы населения Северо-Восточной Сибири, синим – Центральной Сибири, зеленым – Южной Сибири, красным – Западной Сибири.

В большинстве случаев наблюдаются статистически значимые различия по частоте полиморфных вариантов в Сибири в сравнении с Европой и Восточной Азией. По ряду локусов частоты аллелей в сибирских популяциях проявляют большее сходство с европейскими, чем с восточноазиатскими значениями (rs75272640, rs79270057, rs3752874, rs60502621, rs58274299, rs150604595, rs13090120). В трех случаях обнаружены варианты полиморфизма, характерные лишь для сибирских популяций – всех четырех регионов (rs561918607-T) или отдельных популяций (rs572372574-C у чукчей и эвенков и rs762175401-A у эскимосов и коряков) (см. табл. 1).

Для анализа распределения гаплотипов гена *THRB* в популяциях Сибири набор полиморфных локусов нами был редуцирован до 16 (rs75272640, rs2167116, rs56204436, rs1349265, rs79270057, rs561918607, rs844107, rs826371, rs72619908, rs762175401, rs3752874, rs572372574, rs34833017, rs13326381, rs60502621, rs13090120). Это свя-

зано с тем, что ряд локусов, расположенных на небольших расстояниях друг от друга, показывал одинаковые частоты в сибирских, восточноазиатских и европейских выборках и был сцеплен. С помощью алгоритма ELV было выявлено 34 16-локусных гаплотипа (табл. S1 Приложения)<sup>1</sup>. Для дальнейшего анализа филогенетических связей между гаплотипами использовано 22 гаплотипа, зарегистрированных в популяциях более одного раза.

Полученная медианная сеть гаплотипов демонстрирует довольно сложные филогенетические взаимоотношения, что связано, по всей видимости, с повторяющимися мутациями в некодирующих участках гена *THRB* или ошибками секвенирования в локусе rs13326381 (см. рисунок). Исследуемые гаплотипы обнаружены в различных региональных группах Сибири, и поэтому географическая кластеризация групп гаплотипов практически не прослежи-

<sup>1</sup> Табл. S1 Приложения см. по адресу: <https://vavilovj-icg.ru/download/pict-2026-30/appx6.xlsx>

**Таблица 2.** Частоты генотипов и аллелей локуса rs762175401 гена *THRB* у коряков и эвенов

Популяция (N)	Генотипы			Аллели		$H_e$	$p$
	GG	GA	AA	G	A		
Коряки (98)	0.765	0.194	0.041	0.862	0.138	0.239	0.08
Эвены (110)	1.0	0	0	1.0	0	0	1.0

Примечание. Здесь и в табл. 3: N – размер выборки;  $H_e$  – ожидаемая гетерозиготность;  $p$  – статистическая значимость отклонения от равновесия Харди–Вайнберга (достоверно при  $p < 0.05$ ).

**Таблица 3.** Частоты генотипов и аллелей локуса rs72619908 гена *THRB* у коряков и эвенов

Популяция (N)	Генотипы			Аллели		$H_e$	$p$
	CC	CG	GG	C	G		
Коряки (96)	0.948	0.052	0	0.974	0.026	0.051	1.0
Эвены (110)	0.927	0.064	0.009	0.959	0.041	0.079	0.156

ваются. Интерес представляют гаплотипы HT19 и HT20, найденные в выборках из Северо-Восточной Сибири. Эти гаплотипы характеризуются наличием замены G→A в локусе rs762175401, который расположен в 3'-нетранслируемой области гена. Данный вариант полиморфизма обнаружен только у эскимосов (частота 25 %) и коряков (частота 12 %), а по данным dbSNP и gnomAD, аллель rs762175401-A был выявлен с очень низкими частотами у японцев (0.042 %) и корейцев (0.04 %), а также в больших выборках из Восточной Азии (0.013 %), Ближнего Востока (0.016 %), Южной Азии (0.005 %) и Европы (0.002 %).

Поскольку размеры изученных выше сибирских выборок недостаточно велики, нами исследован полиморфизм локуса rs762175401 у коряков и эвенов Магаданской области в более представительных выборках. В исследуемый участок гена *THRB* попал также расположенный на расстоянии 104 п. н. от rs762175401 локус rs72619908 (табл. 2 и 3). Согласно историческим данным, коряки (наряду с чукчами) относятся к северовосточным палеоазиатам – древнейшему населению Северо-Восточной Сибири, а тунгусоязычные эвены начали осваивать места проживания коряков примерно с XVII в. (Хаховская, 2024). Однако, несмотря на длительное соседство и брачные связи, эти этнические группы сохранили особенности своих генофондов, которые, как ожидается, в разной степени адаптированы к экстремальным условиям окружающей их природной среды (Cardona et al., 2014; Derenko et al., 2023; Малярчук, Деренко, 2024).

Анализ полиморфизма локуса rs762175401 показал, что в исследованных выборках только у коряков наблюдается аллель rs762175401-A с частотой 13.8 % (см. табл. 2). Полиморфные варианты локуса rs72619908 распределены в обеих этнических группах примерно одинаковым образом (см. табл. 3). Столь высокая частота варианта rs762175401-A, зарегистрированная у коряков, а также наличие этого аллеля у сибирских эскимосов позволяют предположить, что данный вариант полиморфизма имеет приспособительное значение в отношении адаптивного термогенеза. Локус rs762175401 расположен в 3'-нетранс-

лируемой области гена *THRB* в позиции +2 от терминирующего кодона UAG. Не исключено, что замена в этой нуклеотидной позиции могла привести к изменениям в эффективности терминации трансляции, поскольку ранее было показано, что 3'-контекст стоп-кодонов оказывает влияние на терминацию трансляции у эукариот (Cridge et al., 2018; Соколова и др., 2020). Если замена rs762175401-A способствовала повышению эффективности терминации, то это могло привести к увеличению скорости синтеза белка и тем самым к оптимизации трансляции (повышению стабильности мРНК, рециркуляции рибосом, точности трансляции), что ранее было установлено для *Escherichia coli* и дрожжей (Baggett et al., 2017; Wu et al., 2020).

Улучшение синтеза рецептора тиреоидных гормонов предположительно может способствовать увеличению концентрации комплексов TRβ-T3 и поэтому в условиях холода может иметь адаптивное значение. Таким образом, зарегистрированное нами повышение частоты варианта rs762175401-A у коряков и эскимосов, представляющих древнейшее население Северо-Востока Сибири, можно объяснить долговременным влиянием климатических факторов на функционирование гена *THRB*.

Известно, что в процессах адаптации к холоду у человека задействован ряд генов (*DIO2*, *UCP1*, *UCP3*, *THRB*, *PPARGC1A*, *RXRA*), входящих в сигнальные пути тиреоидной системы, которыми регулируются термогенез, энергозатраты и метаболические перестройки (Laurberg et al., 2005; Bianco et al., 2019; Tsibulnikov et al., 2020). Результаты генетических исследований показали, что в отношении некоторых генов (*UCP1*, *UCP3*) наблюдаются популяционно-генетические эффекты, проявляющиеся в повышении частоты определенных вариантов полиморфизма на Северо-Востоке Азии (Nikanorova et al., 2021, 2022; Козлов и др., 2024). В арктических популяциях под хроническим воздействием холода, вероятнее всего, повышается активность фермента *DIO2*, регулирующего уровень T3 в клетках, что позволяет поддерживать выработку этого гормона и компенсировать высокие метаболические потребности термогенеза (Noahsen et al., 2021).

Вполне возможное действие естественного отбора зарегистрировано также в генах *ANGPTL8* и *PLA2G2A* у коренного населения Сибири (коряков, юкагиров, ненцев и других народов) (Hallmark et al., 2019). Эти гены также задействованы в адаптивном ответе на холод: они активируются гормоном ТЗ и, в свою очередь, участвуют в регуляции метаболизма липидов (Sharma et al., 2014; Tseng et al., 2014). Действие отбора на ген *THRB* (с помощью теста PBS) также зарегистрировано в якутской выборке (Cardona et al., 2014), а анализ действия отбора в региональных группах мира обнаружил статистические связи (тест ЕНН) с функционированием тиреоидной системы, БЖТ и терморегуляцией только в центральносибирской выборке и у африканцев (Pagani et al., 2016).

Вместе с тем недавнее геномное исследование эскимосов Гренландии не выявило отбора, действующего на гены тиреоидной системы (Stæger et al., 2025). Варианты полиморфизма, специфичные для эскимосов Гренландии и Северной Америки и сибирских популяций (*FADS1/2*, *SI*, *CPT1A*, *TBC1D*) или же только для гренландских и американских эскимосов (*LDLR*, *HNFI1A*, *ADCY3*, *ATP8B1*, *PCCA/PCCB*), обнаружены преимущественно в генах липидного и углеводного обмена (Stæger et al., 2025).

Эти данные свидетельствуют о необходимости проведения более детальных исследований полиморфизма генов тиреоидной системы среди коренного населения Крайнего Севера разного этнического происхождения, поскольку адаптивные изменения генофондов могут иметь популяционную специфику.

## Заключение

Результаты проведенного нами исследования позволили впервые оценить распространенность различных вариантов полиморфизма гена *THRB*, кодирующего ядерный рецептор тиреоидных гормонов, в этнических группах коренного населения Сибири. Анализ распределения гаплотипов гена *THRB* показал, что на Северо-Востоке Сибири достаточно широкое распространение получил вариант rs762175401-A, обнаруженный у коряков (с частотой 13.8 %) и также отмеченный у сибирских эскимосов. Вполне вероятно, что повышение частоты этого варианта полиморфизма связано с адаптацией популяций коренных народов Крайнего Севера к холоду и перестройками тиреоидной системы, непосредственно участвующей в процессах несократительного термогенеза. Дальнейшие молекулярно-генетические исследования помогут прояснить механизмы обнаруженного явления.

## Список литературы / References

Козлов А.И., Вершубская Г.Г., Мальярчук Б.А., Нагорная Е.Г., Парфентьева О.И., Балановская Е.В. Варибельность генов разобщающих белков *UCP1* и *UCP3* в связи с климатом в популяциях коренного населения Сибири и Дальнего Востока. *Вестник Московского университета. Серия XXIII. Антропология*. 2024; 3:79-90. doi 10.55959/MSU2074-8132-24-3-7  
[Kozlov A.I., Verhubskaya G.G., Malyarchuk B.A., Nagornaya E.G., Parfenteva O.I., Balanovska E.V. Variability of *UCP1* and *UCP3* uncoupling protein genes in relation to climate in indigenous populations of Siberia and the Far East. *Lomonosov Journal*

*of Anthropology (Moscow University Anthropology Bulletin). Series XXIII. Anthropology*. 2024;3:79-90. doi 10.55959/MSU2074-8132-24-3-7 (in Russian)]  
Мальярчук Б.А., Деренко М.В. Генетическая история коряков и эвенов Магаданской области по данным о полиморфизме Y-хромосомы. *Вавиловский журнал генетики и селекции*. 2024;28(1):90-97. doi 10.18699/vjgb-24-11  
[Malyarchuk B.A., Derenko M.V. Genetic history of the Koryaks and Evens of the Magadan region based on Y-chromosome polymorphism data. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2024;28(1):90-97. doi 10.18699/vjgb-24-11]  
Соколова Е.Е., Власов П.К., Егорова Т.В., Шувалов А.В., Алкалаева Е.З. Влияние A/G-состава 3'-контекстов стоп-кодонов на терминацию трансляции у эукариот. *Молекулярная биология*. 2020; 54(5):837-848. doi 10.31857/S0026898420050080  
[Sokolova E.E., Vlasov P.K., Egorova T.V., Shuvalov A.V., Alkalaeva E.Z. The influence of A/G composition of 3' stop codon contexts on translation termination efficiency in eukaryotes. *Mol Biol*. 2020;54(5):739-748. doi 10.1134/S0026893320050088]  
Хоховская Л.Н. Номенклатура коренных народов Магаданской области в историческом контексте. *Вестник Северо-Восточного научного центра ДВО РАН*. 2024;2:108-115. doi 10.34078/1814-0998-2024-2-108-115  
[Khakhovskaya L.N. Nomenclature of indigenous peoples of Magadan oblast in the historical context. *Bulletin of the North-East Scientific Center of FEB RAS*. 2024;2:108-115. doi 10.34078/1814-0998-2024-2-108-115 (in Russian)]  
Baggett N.E., Zhang Y., Gross C.A. Global analysis of translation termination in *E. coli*. *PLoS Genet*. 2017;13(3):e1006676. doi 10.1371/journal.pgen.1006676  
Bianco A.C., Silva J.E. Cold exposure rapidly induces virtual saturation of brown adipose tissue nuclear T3 receptors. *Am J Physiol*. 1988;255(4):E496-E503. doi 10.1152/ajpendo.1988.255.4.E496  
Bianco A.C., Dumitrescu A., Gereben B., Ribeiro M.O., Fonseca T.L., Fernandes G.W., Bocco B.M.L.C. Paradigms of dynamic control of thyroid hormone signaling. *Endocr. Rev*. 2019;40(4):1000-1047. doi 10.1210/er.2018-00275  
Cardona A., Pagani L., Antao T., Lawson D.J., Eichstaedt C.A., Yngvadottir B., Shwe M.T.T., ... Willerslev E., Tyler-Smith C., Malyarchuk B.A., Derenko M.V., Kivisild T. Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS One*. 2014; 9(5):e98076. doi 10.1371/journal.pone.0098076  
Cridge A.G., Crowe-McAuliffe C., Mathew S.F., Tate W.P. Eukaryotic translational termination efficiency is influenced by the 3' nucleotides within the ribosomal mRNA channel. *Nucleic Acids Res*. 2018; 46(4):1927-1944. doi 10.1093/nar/gkx1315  
Derenko M., Denisova G., Litvinov A., Dambueva I., Malyarchuk B. Mitogenomics of the Koryaks and Evens of the northern coast of the Sea of Okhotsk. *J Hum Genet*. 2023;68(10):705-712. doi 10.1038/s10038-023-01173-x  
Dumitrescu A.M., Refetoff S. The syndromes of reduced sensitivity to thyroid hormone. *Biochim Biophys Acta*. 2013;1830(7):3987-4003. doi 10.1016/j.bbagen.2012.08.005  
Excoffier L., Lischer H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567. doi 10.1111/j.1755-0998.2010.02847.x  
Excoffier L., Laval G., Balding D. Gametic phase estimation over large genomic regions using an adaptive window approach. *Hum Genomics*. 2003;1(1):7-19. doi 10.1186/1479-7364-1-1-7  
González-Sancho J.M., García V., Bonilla F., Muñoz A. Thyroid hormone receptors/THR genes in human cancer. *Cancer Lett*. 2003; 192(2):121-132. doi 10.1016/S0304-3835(02)00614-6  
Grøntved L., Waterfall J.J., Kim D.W., Baek S., Sung M.H., Zhao L., Park J.W., Nielsen R., Walker R.L., Zhu Y.J., Meltzer P.S., Hager G.L., Cheng S.Y. Transcriptional activation by the thyroid hormone receptor through ligand-dependent receptor recruitment and

- chromatin remodelling. *Nat Commun.* 2015;6:7048. doi 10.1038/ncomms8048
- Hallmark B., Karafet T.M., Hsieh P., Osipova L.P., Watkins J.C., Hammer M.F. Genomic evidence of local adaptation to climate and diet in indigenous Siberians. *Mol Biol Evol.* 2019;36(2):315-327. doi 10.1093/molbev/msy211
- Hancock A.M., Clark V.J., Qian Y., Di Rienzo A. Population genetic analysis of the uncoupling proteins supports a role for UCP3 in human cold resistance. *Mol Biol Evol.* 2011;28(1):601-614. doi 10.1093/molbev/msq228
- Laurberg P., Andersen S., Karmisholt J. Cold adaptation and thyroid hormone metabolism. *Horm Metab Res.* 2005;37(9):545-549. doi 10.1055/s-2005-870420
- Lee J.-Y., Takahashi N., Yasubuchi M., Kim Y.-I., Hashizaki H., Kim M.-J., Sakamoto T., Goto T., Kawada T. Triiodothyronine induces UCP-1 expression and mitochondrial biogenesis in human adipocytes. *Am J Physiol Cell Physiol.* 2012;302(2):C463-C472. doi 10.1152/ajpcell.00010.2011
- Leonard W.R. *Pearl Memorial Lecture.* Humans at the extremes: exploring human adaptation to ecological and social stressors. *Am J Hum Biol.* 2024;36(3):e24010. doi 10.1002/ajhb.24010
- Leonard W.R., Levy S.B., Tarskaia L.A., Klimova T.M., Fedorova V.I., Baltakhinova M.E., Krivoschapkin V.G., Snodgrass J.J. Seasonal variation in basal metabolic rates among the Yakut (Sakha) of north-eastern Siberia. *Am J Hum Biol.* 2014;26(4):437-445. doi 10.1002/ajhb.22524
- Levy S.B., Leonard W.R., Tarskaia L.A., Klimova T.M., Fedorova V.I., Baltakhinova M.E., Krivoschapkin V.G., Snodgrass J.J. Seasonal and socioeconomic influences on thyroid function among the Yakut (Sakha) of Eastern Siberia. *Am J Hum Biol.* 2013;25(6):814-820. doi 10.1002/ajhb.22457
- Ma Y., Shen S., Yan Y., Zhang S., Liu S., Tang Z., Yu J., ... Li Y., Hu C., Jiang J., Li Y., Ying H. Adipocyte thyroid hormone receptor-mediated hormone action fine-tunes intracellular glucose and lipid metabolism and systemic homeostasis. *Diabetes.* 2023;72(5):562-574. doi 10.2337/db22-0656
- Martinez de Mena R., Scanlan T.S., Obregon M.J. The T<sub>3</sub> receptor beta1 isoform regulates UCP1 and D2 deiodinase in rat brown adipocytes. *Endocrinology.* 2010;151(10):5074-5083. doi 10.1210/en.2010-0533
- Nikanorova A.A., Barashkov N.A., Pshennikova V.G., Nakhodkin S.S., Gotovtsev N.N., Romanov G.P., Solovyev A.V., Kuzmina S.S., Sazonov N.N., Fedorova S.A. The role of nonshivering thermogenesis genes on leptin levels regulation in residents of the coldest region of Siberia. *Int J Mol Sci.* 2021;22(9):4657. doi 10.3390/ijms22094657
- Nikanorova A.A., Barashkov N.A., Pshennikova V.G., Gotovtsev N.N., Romanov G.P., Solovyev A.V., Kuzmina S.S., Sazonov N.N., Fedorova S.A. Relationships between uncoupling protein genes *UCP1*, *UCP2* and *UCP3* and irisin levels in residents of the coldest region of Siberia. *Genes.* 2022;13(9):1612. doi 10.3390/genes13091612
- Nikanorova A.A., Barashkov N.A., Pshennikova V.G., Teryutin F.M., Nakhodkin S.S., Solovyev A.V., Romanov G.P., Burtseva T.E., Fedorova S.A. A systematic review and meta-analysis of free triiodothyronine (FT3) levels in humans depending on seasonal air temperature changes: is the variation in FT3 levels related to nonshivering thermogenesis? *Int J Mol Sci.* 2023;24(18):14052. doi 10.3390/ijms241814052
- Noahsen P., Rex K.F., Bülow Pedersen I., Mulvad G., Florian-Sørensen H.C., Pedersen M.L., Andersen S. Adaptation to a high iodine intake in Greenland Inuit suggested by thyroid disease pattern. *Thyroid.* 2021;31(12):1850-1857. doi 10.1089/thy.2021.0342
- Pagani L., Lawson D.J., Jagoda E., Mörseburg A., Eriksson A., Mitt M., Clemente F., ... Nielsen R., Villems R., Willerslev E., Kivisild T., Metspalu M. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature.* 2016;538(7624):238-242. doi 10.1038/nature19792
- Sharma P., Levesque T., Boilard E., Park E.A. Thyroid hormone status regulates the expression of secretory phospholipases. *Biochem Biophys Res Commun.* 2014;444(1):56-62. doi 10.1016/j.bbrc.2014.01.003
- Stæger F.F., Andersen M.K., Li Z., Hjerresen J.P., He S., Santander C.G., Jensen R.T., ... Nielsen R., Jørgensen M.E., Hansen T., Moltke I., Albrechtsen A. Genetic architecture in Greenland is shaped by demography, structure and selection. *Nature.* 2025;639(8054):404-410. doi 10.1038/s41586-024-08516-4
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28:2731-2739. doi 10.1093/molbev/msr121
- Tseng Y.H., Ke P.Y., Liao C.J., Wu S.M., Chi H.C., Tsai C.Y., Chen C.Y., Lin Y.H., Lin K.H. Chromosome 19 open reading frame 80 is up-regulated by thyroid hormone and modulates autophagy and lipid metabolism. *Autophagy.* 2014;10(1):20-31. doi 10.4161/auto.26126
- Tsibulnikov S., Maslov L., Voronkov N., Oeltgen P. Thyroid hormones and the mechanisms of adaptation to cold. *Hormones.* 2020;19(3):329-339. doi 10.1007/s42000-020-00200-2
- Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G. Primer3 – new capabilities and interfaces. *Nucleic Acids Res.* 2012;40(15):e115. doi 10.1093/nar/gks596
- Wu C., Roy B., He F., Yan K., Jacobson A. Poly(A)-binding protein regulates the efficiency of translation termination. *Cell Rep.* 2020;33(7):108399. doi 10.1016/j.celrep.2020.108399
- Yau W.W., Yen P.M. Thermogenesis in adipose tissue activated by thyroid hormone. *Int J Mol Sci.* 2020;21(8):3020. doi 10.3390/ijms21083020

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 24.02.2025. После доработки 09.05.2025. Принята к публикации 21.05.2025.

doi 10.18699/vjgb-26-08

## Транскриптомика тяжелой формы COVID-19

А.А. Гусарова , Е.А. Трифонова , А.А. Бабовская , М.М. Гавриленко , В.А. Степанов 

Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия

 anastasia.gusarova@medgenetics.ru

**Аннотация.** В настоящее время выявление биомаркеров, позволяющих эффективно определить пациентов с риском развития тяжелой формы COVID-19, которая может привести к летальному исходу, считается важной задачей. Изучение патогенетических механизмов перехода умеренной формы в тяжелую с помощью анализа транскриптома крови обеспечивает идентификацию дифференциально экспрессирующихся генов (ДЭГ), которые могут стать потенциальными прогностическими биомаркерами тяжести инфекции, а также новыми терапевтическими мишенями в борьбе с осложнениями COVID-19. В данном обзорном исследовании проведены поиск и анализ работ по изучению различий в экспрессии генов при применении подхода секвенирования РНК пула клеток крови (bulk RNA-seq) при сравнении умеренной и тяжелой степени коронавирусной инфекции. По результатам пяти работ были определены пять общих, наиболее значимых дифференциально экспрессирующихся генов (*CD177*, *PPARG*, *PCOLCE2*, *SLC51A* и *ADAMTS2*) и рассмотрена их предполагаемая роль в развитии тяжелой формы COVID-19. Проведен анализ функционального обогащения, который определил общие пути, в которые вовлечены гены, дифференциально экспрессирующиеся при тяжелой форме COVID-19, такие как активация процессов дегрануляции нейтрофилов, путей интерлейкинов, биосинтеза коллагена и подавление путей адаптивного и опосредованного NK-клетками иммунного ответа. Проанализированы также результаты секвенирования РНК единичных клеток (single-cell RNA-seq) в изучении умеренной и тяжелой форм, подтверждающие некоторые результаты bulk RNA-seq. В связи с низкой общностью данных в рассматриваемых работах один из разделов обзора посвящен анализу дизайнов выбранных исследований, включая инструменты анализа, сбор материала и критерии формирования сравниваемых групп. Транскриптомика тяжелой формы COVID-19 раскрывает как клеточные, так и молекулярные механизмы иммунного ответа, дисрегуляция которого может привести к развитию тяжелых проявлений. При этом другие омиксные технологии смогут дополнить пробелы в изучении особенностей тяжелой формы и раскрыть механизмы прогрессирования заболевания для разработки подходов профилактики и терапии COVID-19.

**Ключевые слова:** тяжелая форма COVID-19; секвенирование РНК; транскриптомика; bulk RNA-seq; single-cell RNA-seq; дифференциально экспрессирующиеся гены

**Для цитирования:** Гусарова А.А., Трифонова Е.А., Бабовская А.А., Гавриленко М.М., Степанов В.А. Транскриптомика тяжелой формы COVID-19. *Вавиловский журнал генетики и селекции*. 2026;30(1):101-116. doi 10.18699/vjgb-26-08

**Финансирование.** Исследование выполнено за счет средств государственного задания по теме ФНИ № 122020200083-8.

## Transcriptomics of severe COVID-19

А.А. Gusarova , Е.А. Trifonova , А.А. Babovskaya , М.М. Gavrilenko , V.A. Stepanov 

Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia

 anastasia.gusarova@medgenetics.ru

**Abstract.** Currently, identifying biomarkers that can reliably predict the risk of developing severe COVID-19, potentially leading to fatal outcomes, remains a critical challenge. Studying the pathogenetic mechanisms underlying the progression from moderate to severe disease through blood transcriptome analysis enables the identification of differentially expressed genes (DEGs), which may serve as potential prognostic biomarkers of disease severity and as novel therapeutic targets for managing COVID-19 complications. In this review, we have summarized and analyzed studies that compared gene expression profiles between moderate and severe COVID-19 cases using bulk RNA sequencing of blood cell samples. Based on the results of five studies, five commonly and significantly differentially expressed genes were identified (*CD177*, *PPARG*, *PCOLCE2*, *SLC51A* and *ADAMTS2*), and their potential roles in the progression to severe COVID-19 are discussed. Functional enrichment analysis was performed, and shared pathways associated with severe COVID-19 were identified, including neutrophil degranulation, interleukin signaling, collagen biosynthesis, and suppression of adaptive and NK cell-mediated immune responses. Additionally, single-cell RNA sequencing (scRNA-seq) studies were reviewed, comparing moderate and severe cases, supporting some of the bulk RNA-seq findings. Due to the limited overlap of data in the reviewed articles, one section of this review focuses on the study designs, including analytical tools, sample collection protocols, and criteria used to define comparison groups. Transcriptomic analysis of

the COVID-19 severe form reveals both cellular and molecular mechanisms of the immune response, the dysregulation of which can lead to the development of severe manifestations. RNA-markers seem to be promising predictors of the severity of COVID-19. At the same time, other omics technologies can fill in the gaps in understanding the characteristics of severe COVID-19 and identify mechanisms of disease progression to develop approaches for COVID-19 prevention and treatment.

**Key words:** severe COVID-19; RNA sequencing; transcriptomics; bulk RNA-seq; single-cell RNA-seq; differentially expressed genes

**For citation:** Gusarova A.A., Trifonova E.A., Babovskaya A.A., Gavrilenko M.M., Stepanov V.A. Transcriptomics of severe COVID-19. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):101-116. doi 10.18699/vjgb-26-08

## Введение

COVID-19 (Coronavirus Disease 2019) – острая респираторная инфекция, вызванная вирусом SARS-CoV-2. Клиническая картина COVID-19 имеет широкий спектр проявлений и течения болезни: она варьирует от легкой степени тяжести до крайне тяжелой (Профилактика... COVID-19, 2023). Тяжелая форма может сопровождаться развитием острого респираторного дистресс-синдрома (ОРДС), который характеризуется острым началом, тяжелой гипоксемией, двусторонней инфильтрацией и отеком легких. У большинства пациентов с тяжелой формой наблюдается лимфопения, а у некоторых – тромбоэмболические осложнения. Тяжелая степень COVID-19 может привести к полиорганной недостаточности и смертельному исходу (Berlin et al., 2020; Huang C. et al., 2020).

Для определения методов лечения и профилактики тяжелой формы COVID-19 требуется понимание патогенетической основы заболевания, а также ключевых иммунных и воспалительных процессов, которые отличают тяжелое состояние от нетяжелого (Jovic et al., 2022). В настоящее время для раннего выявления пациентов группы риска тяжелого течения COVID-19 применяют лабораторные показатели крови (Chen et al., 2022; Roessler et al., 2023). Биомаркеры, используемые для прогнозирования тяжести COVID-19, включают показатели общего анализа крови (уровни лейкоцитов, лимфоцитов, нейтрофилов, соотношение нейтрофилов и лимфоцитов, уровень тромбоцитов), маркеры воспаления и провоспалительные цитокины (С-реактивный белок, прокальцитонин, ферритин, ИЛ-6, ИЛ-8, ФНО-альфа и т. д.), маркеры повреждения органов-мишеней (факторы свертывания крови, D-димер, кардио-маркеры и маркеры функции почек), а также маркеры окислительного стресса (уровни активных форм кислорода и антиоксидантов: витамина С, тиоловых белков, сывороточных параметров системы глутатиона, окислительно-восстановительный статус, перекисное окисление липидов) (Polonikov, 2020; Pincemail et al., 2021; Tabassum et al., 2021; Chen et al., 2022; Karu et al., 2022; Roessler et al., 2023; Liu X. et al., 2024). Данные биомаркеры являются важными индикаторами тяжелой формы COVID-19, однако часто они определяются уже на поздних стадиях заболевания (Chen et al., 2022). В связи с этим приобретают актуальность исследования, направленные на обнаружение потенциальных биомаркеров, которые могут быть измерены на более ранних фазах инфицирования и могут прогнозировать особенности иммунной защиты организма от SARS-CoV-2 (Schultze, Aschenbrenner, 2021).

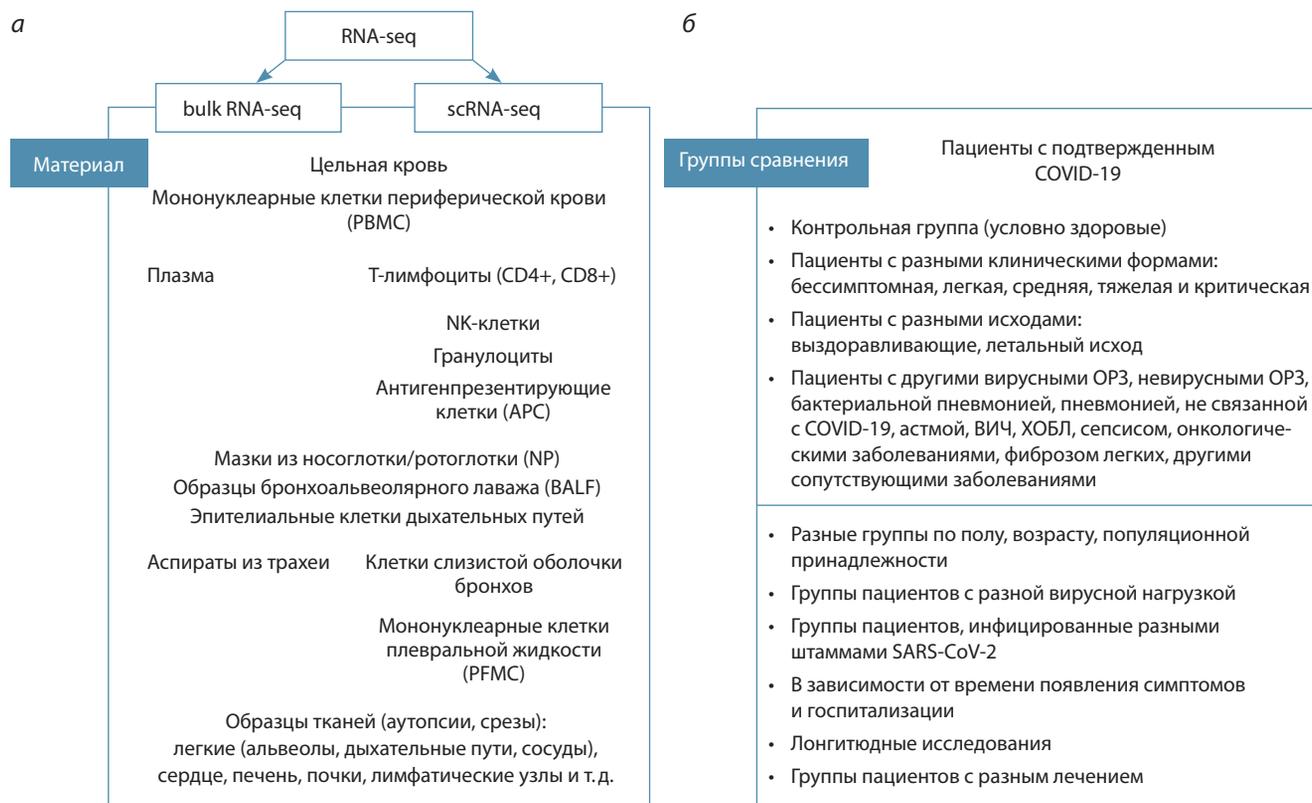
Транскриптомика играет большую роль в определении механизмов, отражающих иммунные реакции организма-хозяина. Изучение транскриптомных профилей пациентов с различной степенью тяжести инфекции может предоставить уникальную информацию о биологических процессах, лежащих в основе тяжести течения, которая может быть использована для выявления прогностических и диагностических биомаркеров и терапевтических мишеней (Lee H.J. et al., 2018; Huang W. et al., 2021; Schultze, Aschenbrenner, 2021). На сегодняшний день наиболее распространенный метод расшифровки транскриптома – высокопроизводительное секвенирование РНК (RNA-seq), которое является количественной системой для профилирования экспрессии в масштабе генома и, следовательно, применимо для характеристики событий, связанных с нарушением регуляции экспрессии генов у пациентов с тяжелой формой COVID-19 (Hegenbarth et al., 2022).

Различия в экспрессии генов, выявляемые при разной клинической тяжести болезни, указывают на то, что существуют потенциальные гены, которые связаны с прогрессированием заболевания, а не с заболеванием в целом. Идентификация и функциональная характеристика дифференциально экспрессирующихся генов при тяжелой и умеренной форме заболевания представляются многообещающей стратегией для выявления биомаркеров тяжести течения COVID-19. Это может обеспечить более глубокое понимание патогенеза COVID-19, помогая в выборе методов лечения благодаря расшифровке сложности иммунной реакции организма и открытию новых терапевтических мишеней (Arunachalam et al., 2020; Bando et al., 2023). Однако в настоящее время в клинической практике не установлено точных РНК-маркеров, позволяющих определить пациентов, у которых может развиваться тяжелое состояние при COVID-19.

Таким образом, в представленной обзорной статье мы сфокусировались на поиске общих генов при анализе дифференциальной экспрессии генов, полученных при исследовании умеренной и тяжелой форм COVID-19 с помощью секвенирования РНК лейкоцитов крови.

## Полнотранскриптомный анализ крови при различной степени тяжести COVID-19

На 2025 г. проведены множественные работы по изучению транскриптома при COVID-19. Количество различных наборов данных в базе данных GEO (<https://www.ncbi.nlm.nih.gov/geo/>) составляет около 300, при этом оригинальных исследований в поисковой системе PubMed насчитыва-



**Рис. 1.** Подходы РНК-секвенирования, материалы (а) и группы сравнения (б) в исследовании средней и тяжелой форм COVID-19.

ется более 60. Их дизайн включает различные подходы секвенирования РНК, группы сравнения и анализируемые материалы (рис. 1).

Так, основным материалом для профилирования экспрессии генов является кровь и ее компоненты. В связи с клиническими проявлениями COVID-19 изучаются транскриптомы мазков из носоглотки/ротоглотки (Chua et al., 2020; Hadzega et al., 2024), образцов бронхоальвеолярного лаважа (BALF) (Xiong et al., 2020; Nassir et al., 2021) и т. д. Кроме того, поскольку рецептор, связывающийся с SARS-CoV-2, обнаружен во многих тканях (<https://www.uniprot.org/uniprotkb/Q9BYF1/entry>), исследуется влияние коронавирусной инфекции и на другие системы органов, включая анализ транскриптомов тканей сердца, печени, почек, лимфатических узлов (Delogeу et al., 2021), а также главной мишени SARS-CoV-2 – ткани легких (см. рис. 1, а). Группы сравнения включают пациентов с подтвержденным COVID-19 и контрольную группу, состоящую из условно здоровых индивидов. В единичных работах сравниваются транскриптомные профили у разных клинических форм: бессимптомной формы, легкой, средней, тяжелой, критической, а также при разных исходах, включая выздоравливающих пациентов и летальные случаи. Помимо того, изучаются различия между коронавирусной инфекцией и другими вирусными и невирусными ОРЗ, бактериальной пневмонией, астмой и ХОБЛ, сепсисом, фиброзом легких, различными сопутствующими заболеваниями (Blanco-Melo et al., 2020; COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium, 2022). Анализ может

включать разные группы пациентов с COVID-19 по полу, возрасту, популяционной принадлежности, с высокой и низкой вирусной нагрузкой, в зависимости от времени появления симптомов и времени взятия материала, группы пациентов без лечения и с лечением определенным препаратом, разные штаммы SARS-CoV-2 (см. рис. 1, б).

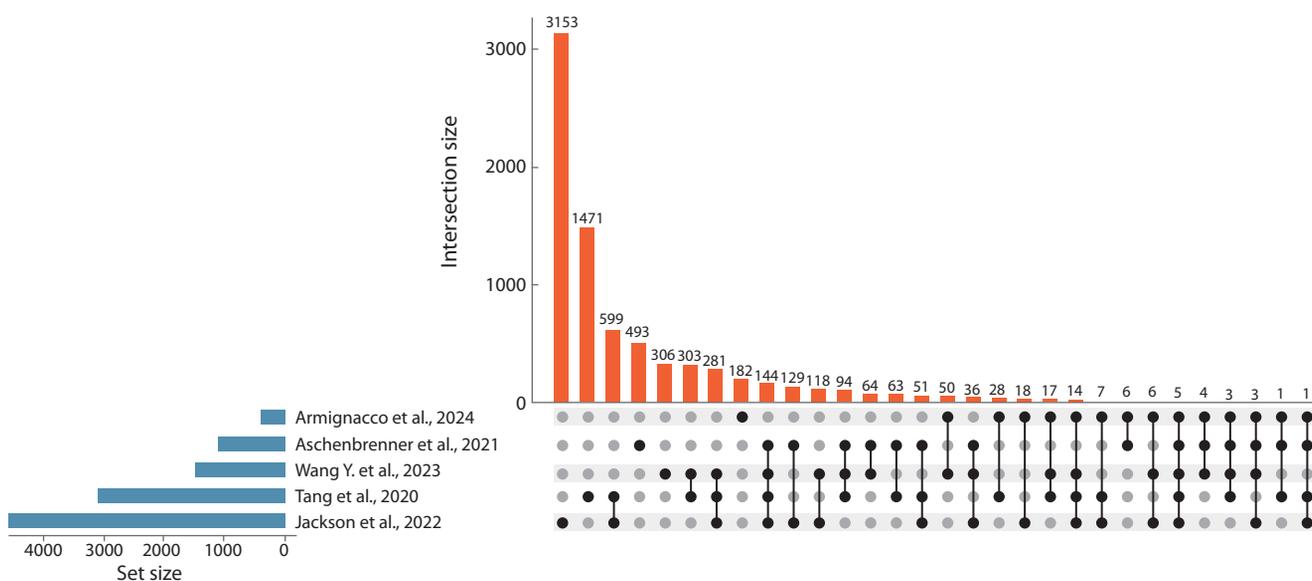
В большинстве работ по транскриптомике COVID-19 периферическая кровь является основным источником обнаружения биомаркеров заболевания. Изменения в транскриптомных профилях крови могут быть вызваны воздействием иммуногенных факторов и/или изменениями в относительном составе иммунных клеток (Chaussabel et al., 2010). Общая дисрегуляция определенных генов может предполагать определенный механизм иммунного ответа. Для выявления наиболее надежных биомаркеров мы провели поиск общих дифференциально экспрессирующихся генов (ДЭГ) в нескольких независимых исследованиях при умеренной и тяжелой формах COVID-19. Это может снижать вероятность того, что наблюдаемый паттерн экспрессии вызван гетерогенностью клеточной популяции (Song et al., 2017).

При поиске в системе PubMed статей, где показаны данные по дифференциальной экспрессии генов при тяжелом и умеренном течении COVID-19 с помощью секвенирования РНК, были использованы такие ключевые слова и выражения, как: “severe COVID-19”, “moderate COVID-19”, “mild COVID-19”, “bulk RNA-sequencing”. Таким образом, критериями отбора являлись анализ транскриптома цельной крови, полученного с помощью секвенирования РНК

**Таблица 1.** Исследования, в которых проведен полнотранскриптомный анализ тяжелой и легкой/средней форм COVID-19

№ п/п	Исследование	Группы сравнения и количество пациентов	Платформа	Результаты
1	Tang et al., 2020	Тяжелая форма (n = 6) и умеренная форма (n = 6)	HiSeq 4000 (Illumina)	3082 ДЭГ (2267↑, 815↓)
2	Aschenbrenner et al., 2021	Тяжелая форма (n = 20) и легкая форма (n = 19)	NovaSeq 6000 (Illumina)	1097 ДЭГ (623↑, 474↓)
3	Jackson et al., 2022	Тяжелая форма (n = 10) и легкая форма (n = 19); тяжелая форма и умеренная форма (n = 26)		7343 ДЭГ (3329↑, 4014↓); 8971 ДЭГ (4380↑, 4591↓)
4	Wang Y. et al., 2023	Тяжелая форма (n = 32) и умеренная форма (n = 25); тяжелая форма и легкая форма (n = 31)		1448 ДЭГ (1013↑, 435↓); 4592 ДЭГ (2617↑, 1975↓)
5	Armignacco et al., 2024	Пневмония тяжелой степени (n = 11) и пневмония легкой степени (n = 53)		345 ДЭГ (237↑, 108↓)

Примечание. ДЭГ – дифференциально экспрессирующиеся гены; ↑ – гиперэкспрессия; ↓ – гипоекспрессия.



**Рис. 2.** График UpSet plot, отражающий количество дифференциально экспрессирующихся генов при сравнении результатов выбранных исследований.

Столбчатая диаграмма, расположенная слева, показывает количество выявленных ДЭГ для каждого исследования. Круги, составляющие матрицу, представляют собой секции диаграммы Венна. Соединенные круги указывают на пересечение генов между определенными исследованиями. Столбчатая диаграмма над матрицей показывает количество уникальных или перекрывающихся ДЭГ. Например, первый и второй столбцы диаграммы отражают количество уникальных генов для исследований (Jackson et al., 2022) и (Tang et al., 2020) – 3153 и 1471 ДЭГ соответственно. Третий столбец показывает количество общих генов только для данной пары (599 ДЭГ). Для работ (Jackson et al., 2022; Wang Y. et al., 2023) были включены общие ДЭГ, полученные при сравнении тяжелой и нетяжелой форм согласно классификации тяжести ВОЗ. Диаграмма построена с помощью <https://intervene.shinyapps.io/intervene/> (Khan, Mathelier, 2017).

пула клеток крови (bulk RNA-seq) у определенных групп сравнения – тяжелой формы и умеренной/легкой формы COVID-19 (severe versus moderate/mild). В анализ были включены работы, точно соответствующие критериям, их характеристика представлена в табл. 1.

Всего при сравнении групп тяжелой формы и умеренной/легкой форм в данных работах было идентифицировано 7650 дифференциально экспрессирующихся генов (рис. 2). Для каждой пары работ количество общих генов составляет от 23 (Aschenbrenner et al., 2021; Armignacco et al., 2024) до 1102 (Tang et al., 2020; Jackson et al., 2022).

При анализе общности ДЭГ отмечалась низкая репликация результатов. Выявлено всего пять генов, которые были идентифицированы при сравнении тяжелой и умеренной форм COVID-19 в каждом из выбранных исследований.

Данные наблюдения могут указывать на значимость изменения экспрессии этих генов в патогенезе тяжелой формы COVID-19.

### Характеристика наиболее значимых генов

Более подробно остановимся на характеристике пяти общих, наиболее значимых дифференциально экспрессирующихся генов: *CD177*, *PPARG*, *PCOLCE2*, *SLC51A* (гиперэкспрессированы) и *ADAMTS2* (гипоекспрессирован в исследовании (Armignacco et al., 2024) и гиперэкспрессирован в исследованиях (Tang et al., 2020; Aschenbrenner et al., 2021; Jackson et al., 2022; Wang Y. et al., 2023)).

Ген *PPARG* кодирует ядерный рецептор PPAR $\gamma$ , активируемый пероксисомным пролифератором гамма. PPAR $\gamma$  регулирует пероксисомальный путь бета-окисле-

ния жирных кислот, а также активацию макрофагов, ингибируя продукцию воспалительных цитокинов моноцитами (<https://www.genecards.org/>).

Согласно базе данных DisGeNET, этот ген значимо ассоциирован с сахарным диабетом 2-го типа у взрослых и с ожирением, для которых показатель ассоциации ген-болезнь (GDA score) равен 1. Данные патологии являются факторами риска тяжелого течения COVID-19 (de Seabra Rodrigues Dias et al., 2022). Кроме того, выявлена ассоциация с COVID-19 при GDA score равном 0.3 (<https://disgenet.com/>).

Предполагается, что SARS-CoV-2 подавляет экспрессию PPAR в легких и нарушает работу противовоспалительного механизма NF-κB, тем самым вызывая гипервоспалительную реакцию у пациентов с тяжелой формой COVID-19 (Desterke et al., 2020; Hasankhani et al., 2024). Наблюдаются различные изменения экспрессии гена, в том числе в разных тканях. Например, в работе (Vlasov et al., 2021) экспрессия *PPARG* была повышена у пациентов с неблагоприятным исходом. Авторы предположили, что повышенный уровень PPARγ может быть признаком неразрешенного воспаления в условиях истощения липидов, характерного для тяжелой формы COVID-19 (Pei et al., 2021). По результатам исследований с использованием сетевых подходов ген *PPARG* был предложен в качестве многообещающей терапевтической мишени контроля воспаления при COVID-19 (Auwul et al., 2021; Oh et al., 2021).

**SLC51A.** Переносчик органических растворенных веществ альфа, кодируемый геном *SLC51A*, является основным компонентом комплекса OSTα/OSTβ, который действует как транспортер, отвечающий за экспорт желчных кислот из энтероцитов.

Согласно базе данных DisGeNET (<https://disgenet.com/>), *SLC51A* значимо ассоциирован с первичным билиарным циррозом печени (GDA score = 0.65), синдромом Байлера (GDA score = 0.5) и с осложнениями сахарного диабета 2-го типа (GDA score = 0.4).

Показано, что *SLC51A* является мишенью для препаратов при лечении COVID-19 (Morselli Gysi et al., 2021). Кроме того, *SLC51A* входит в значимые гены при дифференциации пациентов с сепсисом и определении его эндотипов, клинические характеристики которого часто обсуждаются как общие с тяжелым COVID-19 (Baghela et al., 2023; Fang, Ma, 2023). В инфицированных SARS-CoV-2 органах дыхательных путей и альвеол, полученных из эмбриональных стволовых клеток человека, переносчик *SLC51A* обнаружен в сниженном количестве (Pei et al., 2021). Несмотря на значимость гена *SLC51A* при дифференциации пациентов и определении терапевтических мишеней COVID-19, его роль в механизмах развития тяжести коронавирусной инфекции остается малоизученной.

Ген *CD177* опосредует активацию нейтрофилов, индуцированных TNF-α, включая дегрануляцию и продукцию супероксида, и способствует адгезии нейтрофилов (<https://www.genecards.org/>).

Согласно базе данных DisGeNET, этот ген значимо ассоциирован с миелопролиферативными заболеваниями

(GDA score = 0.35), а также с COVID-19 при GDA score равном 0.25 (<https://disgenet.com/>).

Значимость данного биомаркера при прогнозе тяжелого течения коронавирусной инфекции выявлена в нескольких транскриптомных и протеомных исследованиях (Derakhshani et al., 2021; Lévy et al., 2021; Meizlish et al., 2021; Schimke et al., 2022; Wang Q.S. et al., 2022; Lei, 2024). В работе (Lévy et al., 2021) наблюдались высокая экспрессия *CD177* и более высокий средний уровень данного сывороточного белка в крови у пациентов, находящихся в критическом состоянии. Предполагается, что дегрануляция нейтрофилов вызывает повреждение эндотелия и, следовательно, тромботические осложнения при COVID-19 (Reusch et al., 2021). Таким образом, гиперэкспрессия *CD177* является признаком физиопатологии COVID-19 и может выступать в качестве возможного прогностического фактора прогрессирования заболевания.

**PCOLCE2** кодирует белок-энхансер проколлагеновой C-эндопептидазы 2, который обеспечивает активность связывания коллагена и гепарина.

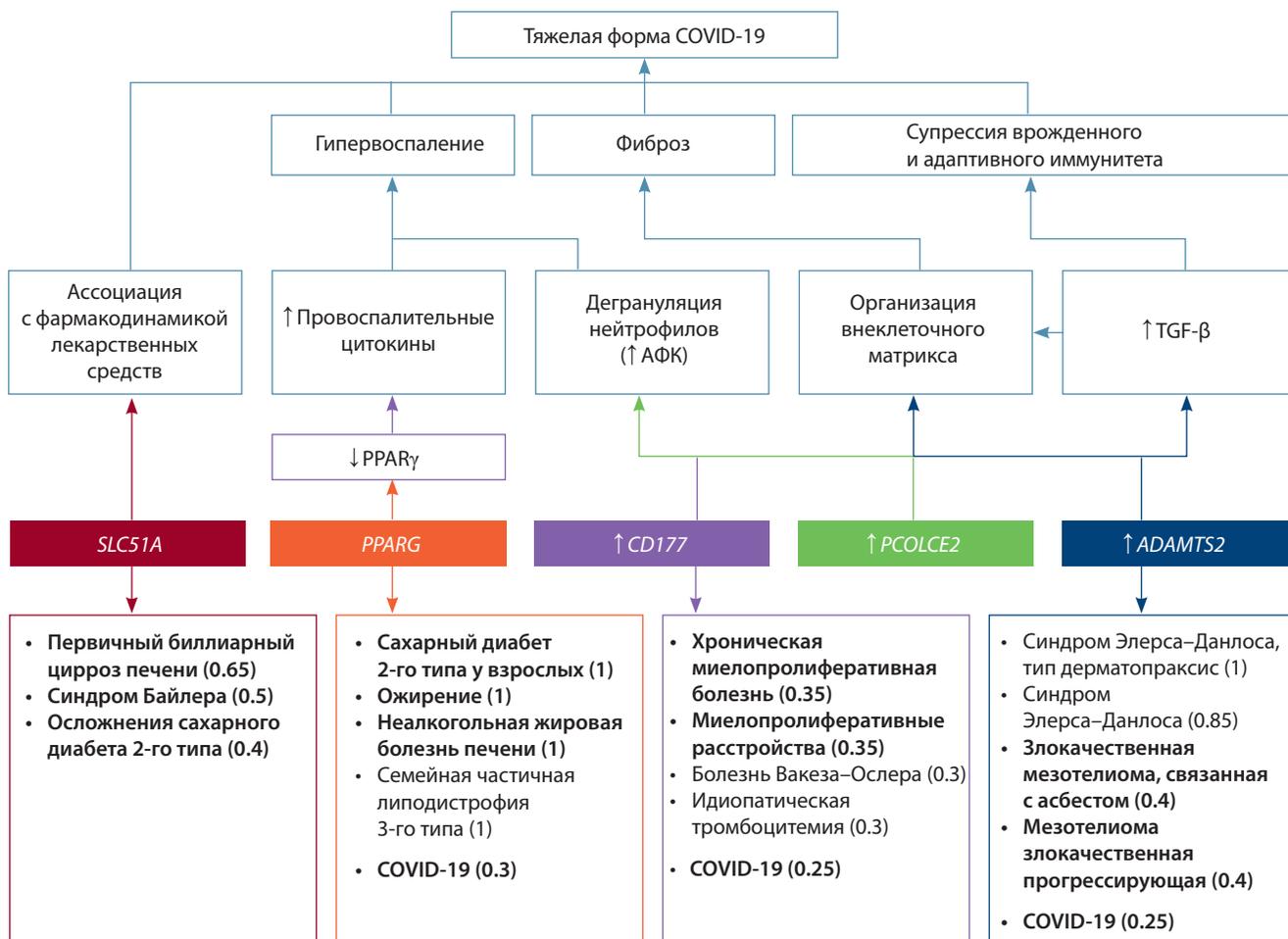
Повышенная экспрессия гена *PCOLCE2* обнаружена у пациентов с тяжелой формой COVID-19 (Alqutami et al., 2021; Che et al., 2022). При анализе обогащения выявлено, что данный ген участвует в организации внеклеточного матрикса и регуляции воспалительного ответа. Отмечается, что *PCOLCE2* может усиливать активность коллагена, а инфекция SARS-CoV-2 вызывает повышение уровня коллагена I в органоидах и способствует активации сигнальных путей фиброза (Jansen et al., 2022). Предполагается также, что *PCOLCE2* стимулирует выработку активных форм кислорода нейтрофилами и формирование внеклеточных ловушек нейтрофилов (NET) (Yoon et al., 2022).

Ген *ADAMTS2* кодирует представителя семейства белков ADAMTS, который играет ключевую роль в превращении фибриллярных предшественников проколлагена в молекулы коллагена (<https://www.genecards.org/>).

Согласно базе данных DisGeNET, *ADAMTS2* значимо ассоциирован с развитием злокачественной мезотелиомы (GDA score = 0.4), а также с COVID-19 при GDA score равном 0.25 (<https://disgenet.com/>).

Белки ADAMTS участвуют в ремоделировании внеклеточного матрикса, что потенциально может иметь важное значение при развитии легочного фиброза, наблюдаемого у пациентов с тяжелым COVID-19. ДЭГ-анализ и анализ генной онтологии (GO) в группах тяжелого, предкритического и критического COVID-19 показали, что в моноцитах в высокой степени экспрессируются гены, связанные с фиброзом (*AREG*, *EREG*, ген цитокина IL-18) и *ADAMTS2* (Zhang Y. et al., 2022).

Высказано предположение, что ADAMTS2 опосредует путь TGF-β, трансформирующего фактора роста-β (de Seabra Rodrigues Dias et al., 2022). Нарушение передачи сигналов TGF-β может способствовать чрезмерному отложению внеклеточного матрикса в тканях, которое может быть вызвано инфекцией и воспалением (Togami et al., 2017; Deng et al., 2024). Кроме того, TGF-β участвует в поддержании иммунного гомеостаза, подавляя активность иммунокомпетентных клеток: данный цитокин препят-



**Рис. 3.** Патогенетические аспекты тяжелой формы COVID-19, включающие общие гены, встречающиеся в пяти выбранных полнотранскриптомных исследованиях средней и тяжелой форм коронавирусной инфекции, и наиболее значимые ассоциации данных генов с заболеваниями согласно базе данных DisGeNET (GDScore  $\geq$  0.3).

Жирным шрифтом выделены заболевания, относящиеся к группам патологии, являющимся факторами риска тяжелой формы COVID-19. АФК – активные формы кислорода.

стует дифференцировке наивных Т-клеток в классические эффекторные Т-клетки, подавляет экспрессию цитотоксических факторов и развитие NK-клеток (Deng et al., 2024). Предполагается также, что TGF-β является ключевым цитокином, регулирующим хроническую иммунную реакцию при тяжелом течении COVID-19 (Ferreira-Gomes et al., 2021).

Благодаря связи *ADAMTS2* с активацией TGF-β пути и формированием внеклеточного матрикса, повышенная экспрессия данного гена при тяжелой форме COVID-19 может отражать иммунопатологические процессы, характерные для прогрессирования заболевания и развития фиброза легких.

Таким образом, были рассмотрены предполагаемые механизмы, посредством которых выявленные гены участвуют в патогенезе тяжелой степени COVID-19. Они отражены в общей схеме (рис. 3), включающей также наиболее значимые ассоциации этих генов с заболеваниями по базе данных DisGeNET. Снижение экспрессии *PPARG* может

привести к низкой экспрессии *PPARG*, что влечет за собой усиление провоспалительных реакций; повышение экспрессии *CD177* приводит к дегрануляции нейтрофилов и высвобождению активных форм кислорода – развивается характерное для тяжелой формы гипервоспаление. Вовлечение в патогенез нейтрофилов возможно и через повышенную экспрессию *PCOLCE2* как фактора, также стимулирующего выработку активных форм кислорода. В развитии фиброза легочной ткани как осложнения тяжелой пневмонии могут быть вовлечены *PCOLCE2* и *ADAMTS2* через различные пути организации внеклеточного матрикса посредством активации синтеза коллагена и TGF-β пути. Изменение в экспрессии *ADAMTS2* тоже может привести к нарушению регуляции TGF-β и, таким образом, вызвать супрессию как врожденного, так и адаптивного иммунитета. Примечательно, что по DisGeNET данные гены связаны с заболеваниями, определенными как факторы риска тяжелой формы COVID-19 (Профилактика... COVID-19, 2023): сахарным диабетом 2-го типа и

его осложнениями (*PPARG*, *SLC51A*), ожирением (*PPARG*), онкологическими заболеваниями (*CD177*, *ADAMTS2*) и хроническими заболеваниями печени (*SLC51A*).

Не все механизмы, посредством которых умеренная форма COVID-19 может перейти в тяжелую, достаточно изучены, и более полную роль данных генов в патогенезе тяжелых проявлений еще предстоит выяснить. Кроме того, различия в направленности их экспрессии, наблюдаемые в разных исследованиях, возможно, предполагают иные закономерности. Сигнальные пути, в которые вовлечены изученные гены, могут раскрыть патогенез тяжелой степени коронавирусной инфекции с другой стороны, учитывая функциональные взаимосвязи генов.

### Анализ функционального обогащения общих дифференциально экспрессирующихся генов, ассоциированных с тяжестью COVID-19

В каждом из рассмотренных исследований для ДЭГ, полученных при сравнении тяжелой и средней форм коронавирусной инфекции, также проведен анализ функционального обогащения. Установлено, что значимые сигнальные пути, ассоциированные с гиперэкспрессией генов, связаны с воспалительным иммунным ответом организма. Например, процессы дегрануляции и активации нейтрофилов. Пути лимфоцитарного иммунного ответа были обогащены как гиперэкспрессирующимися, так и гипоекспрессирующимися генами. Выявлены значительные обогащения в путях, связанных с Т-хелперами 2-го типа, участвующими в гуморальном иммунном ответе (Jackson et al., 2022), и в процессах, связанных с активацией, пролиферацией и регуляцией лимфоцитов (Tang et al., 2020). Наряду с этим, процессы, обогащенные генами с пониженной экспрессией, не пересекались. Таким образом, возникают вопросы о механизмах нарушения взаимодействия врожденного и адаптивного иммунных ответов и перехода от «вирусной» реакции при нетяжелой форме COVID-19 к тяжелому воспалительному процессу (Jackson et al., 2022).

Поскольку при сравнении тяжелой и средней форм коронавирусной инфекции большинство значимых сигнальных путей в анализируемых исследованиях различны, нами был проведен анализ обогащения для общих дифференциально экспрессирующихся генов, чтобы выявить общие закономерности развития тяжелой формы. Были выбраны четыре работы (Tang et al., 2020; Aschenbrenner et al., 2021; Jackson et al., 2022; Wang Y. et al., 2023), в которых найденные дифференциально экспрессирующиеся гены наиболее пересекались. Они включали 149 ДЭГ: 102 гена с повышенной экспрессией и 47 генов с пониженной экспрессией. Анализ функционального обогащения был выполнен с помощью ресурса WebGestalt (<https://www.webgestalt.org/>), а также Reactome (Fabregat et al., 2017).

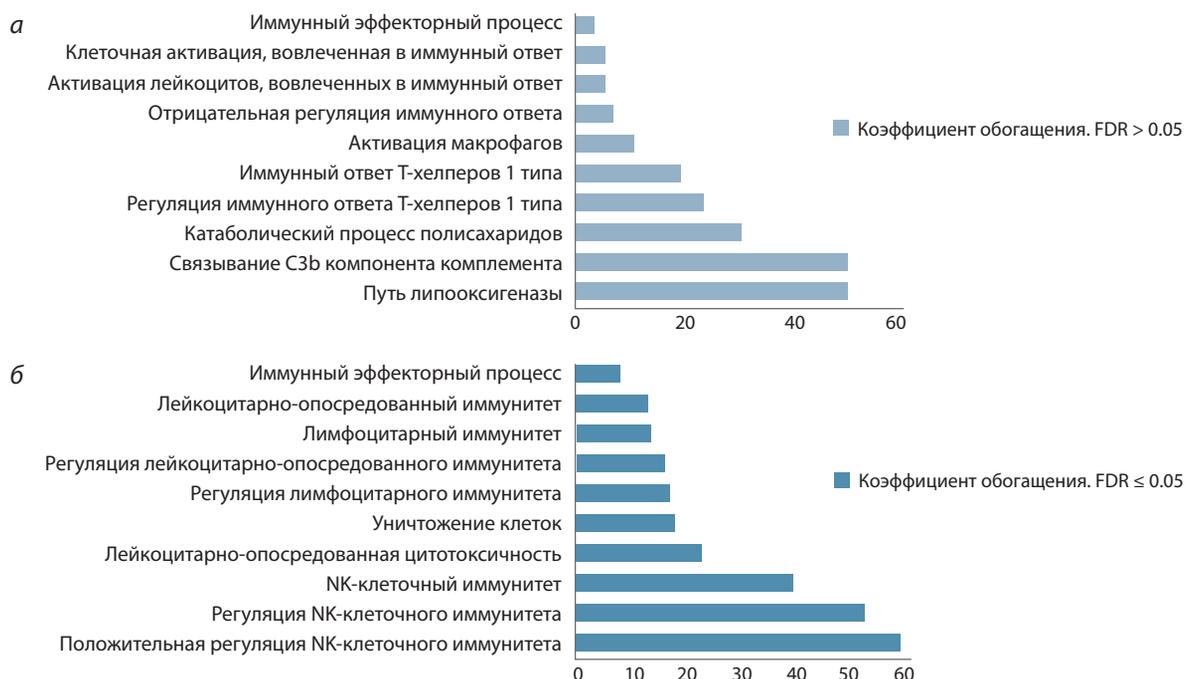
Результаты анализа функционального обогащения с помощью Reactome представлены в табл. 2.

Гены, гиперэкспрессирующиеся в группе с тяжелым течением болезни, связаны с дегрануляцией нейтрофилов (R-HSA-6798695) и с биосинтезом коллагена (R-HSA-1474290). Также была выявлена активация путей передачи сигналов интерлейкинов (R-HSA-449147) – IL-18 (R-HSA-9012546), семейства интерлейкинов-1 (R-HSA-446652), включая IL-33 (R-HSA-9014843), IL-4 и IL-13 (R-HSA-6785807). IL-4, основной цитокин иммунного ответа Th2, как и IL-13, преимущественно связан с фиброзным воспалительным ремоделированием (Vaz de Paula et al., 2020; Wang F. et al., 2020). Кроме того, IL-33, сигнальный белок, предупреждающий иммунную систему о повреждении, может играть важную роль на всех стадиях COVID-19 (Zizzo, Cohen, 2020) и тоже может быть связан с развитием фиброза (Учасова и др., 2018). Сообщалось, что передача сигналов IL-18, участвующего в регуляции Th1-, Th2- и Th17-типов иммунного ответа, в легких активирует инфламмосому, связанные с тяжестью заболевания (Filbin et al., 2021; Насонов, Авдеева, 2022).

Гипоекспрессирующиеся гены при тяжелой форме COVID-19 были связаны с адаптивным иммунитетом

**Таблица 2.** Наиболее значимые пути Reactome при FDR ≤ 0.05

Путь	p-value	FDR
Пути, ассоциированные с гиперэкспрессирующимися генами		
Дегрануляция нейтрофилов (R-HSA-6798695)	7.24e-10	2.83e-07
Передача сигналов интерлейкина-18 (R-HSA-9012546)	1.22e-04	0.024
Образование коллагена (R-HSA-1474290)	2.76e-04	0.035
Передача сигналов с помощью интерлейкинов (R-HSA-449147)	3.86e-04	0.035
Передача сигналов интерлейкина-4 и интерлейкина-13 (R-HSA-6785807)	4.58e-04	0.035
Передача сигналов интерлейкина-33 (R-HSA-9014843)	5.52e-04	0.035
Передача сигналов семейства интерлейкинов-1 (R-HSA-446652)	6.45e-04	0.035
Пути, ассоциированные с гипоекспрессирующимися генами		
Иммунорегуляторные взаимодействия между лимфоидной и нелимфоидной тканью (R-HSA-198933)	1.22e-07	2.53e-05
Экспрессия TGFBR3 (R-HSA-9839394)	1.56e-04	0.016
Адаптивная иммунная система (R-HSA-1280218)	7.70e-04	0.041
Передача сигналов TGFBR3 (R-HSA-9839373)	7.83e-04	0.041



**Рис. 4.** Пути Gene Ontology, обогащенные гиперэкспрессирующимися (а) и гипоекспрессирующимися (б) генами.

(R-HSA-1280218), с экспрессией (R-HSA-9839394) и передачей сигналов TGFBR3 (R-HSA-9839373) – рецептора TGF-β III типа, ингибирующего передачу сигналов TGF-β (Ahn et al., 2010; Chu et al., 2011). Роль данного пути была рассмотрена при характеристике общего гена *ADAMTS2*.

Таким образом, анализ путей, выявленных в отдельных исследованиях, и анализ путей, ассоциированных с общими генами, при переходе в тяжелую форму выявляют дисбалансирующие изменения в иммунном ответе, характеризующиеся активацией путей дегрануляции нейтрофилов, сигналов провоспалительных (IL-1β, IL-18) и противовоспалительных (IL-4/13) цитокинов, а также подавлением адаптивного иммунного ответа, что может быть ключевым фактором, влияющим на тяжесть течения COVID-19.

При этом результаты анализа Gene Ontology (рис. 4) включали значимые сигнальные пути, ассоциированные с гипоекспрессирующимися генами, которые относятся к функциональным категориям иммунитета, опосредованного NK-клетками (GO:0002228), лимфоцитами (GO:0002449), и лейкоцитарно-опосредованной цитотоксичности (GO:0001909). Данные наблюдения отмечают важное участие иммунных клеток в ответе на SARS-CoV-2.

### Дизайн исследований

При установлении причин вариации и низкой репликации генов и сигнальных путей было проведено сравнение дизайнов исследований: характеристик когорт пациентов, платформ секвенирования РНК, а также методов статистической обработки данных. Далее по тексту работам (Tang et al., 2020; Aschenbrenner et al., 2021; Jackson et al., 2022; Wang Y. et al., 2023; Armignacco et al., 2024) будут соответствовать номера 1, 2, 3, 4 и 5.

**Инструменты анализа.** Для получения дифференциально экспрессирующихся генов использовались разные программные пакеты среды R: DESeq2 (Love et al., 2014), edgeR (Robinson et al., 2010) и limma (Ritchie et al., 2015). Значимым считалось скорректированное значение *p* менее 0.05. Порог логарифмического изменения log2FC по модулю составлял ≥1 в анализе (1) и >1.5 в анализах (4, 5).

**Степени тяжести COVID-19.** При включении пациентов в анализируемые выборки соблюдались некоторые общие критерии: возраст 18 лет и старше; положительный результат теста ОТ-ПЦР на SARS-CoV-2 в респираторных образцах (мазок из носа/глотки/мокроты/бронхоальвеолярного лаважа) и/или серологического теста (5) и/или установление диагноза COVID-19 на основании наличия типичных клинических симптомов и результатов КТ (2, 5). Пациенты были классифицированы как имеющие легкую, среднюю или тяжелую форму заболевания. Классификация тяжести COVID-19 в четырех работах была основана на рекомендациях Всемирной организации здравоохранения, однако границы тяжести для формирования групп были различны. Легкая степень соответствовала оценке ВОЗ 1–2 в исследовании (3) и оценке ВОЗ 1–4 в (2, 4). Средняя (умеренная) степень соответствовала уровню ВОЗ 3–4 в (3) и уровню ВОЗ 5 в (4). Тяжелая степень COVID-19 соответствовала уровням тяжести ВОЗ 5–7 в (2), ВОЗ 5–8 в (3) и ВОЗ 6–9 в (4). Шкала ВОЗ по установлению степени тяжести приведена в (3). В исследовании (5) степень тяжести пациентов определялась по течению развивающейся пневмонии. Состояние пациентов классифицировалось как легкая, средняя или тяжелая пневмония в зависимости от начала и развития осложнений COVID-19, включая продолжительность госпитализации, потребность в подаче кислорода, искусственной вентиляции легких или

экстракорпоральной оксигенации. В связи с вариацией формирования групп сравнения анализ ДЭГ проводился с включением групп, которые относились к легкой форме заболевания.

**Сбор материала.** Общий для всех работ период сбора материала составлял год, с февраля 2020 г. по февраль 2021 г. За данный период циркулировали такие варианты SARS-CoV-2, как Beta, Alpha, Delta, Gamma и др. (<https://www.who.int/ru/activities/tracking-SARS-CoV-2-variants>). Предполагается, что вариант вируса может влиять на тяжесть течения заболевания. Были проведены транскриптомные исследования при выявлении различий в реакции организма-хозяина на инфекции с разными вариантами SARS-CoV-2: «пре-VOC» и «VOC» (варианты, вызывающие обеспокоенность). Инфекция при «пре-VOC» характеризовалась умеренными проявлениями и сохранялась намного дольше, чем инфекция во время циркуляции VOC. Инфекция Delta была тяжелой, что приводило к высоким показателям госпитализации и смертности (Hughes et al., 2023; Maurya et al., 2023). Этот вариант вируса способен к усиленной репликации из-за устойчивого подавления врожденного иммунного ответа организма-хозяина, что потенциально способствует тяжелым симптомам и длительному выздоровлению (Laine et al., 2022). Несмотря на выявление дифференцированных транскриптомных ответов, данные по влиянию на тяжесть течения определенного варианта SARS-CoV-2 все же остаются противоречивыми. В некоторых исследованиях на уровне значимости 0.05 не было обнаружено связи между вирусными линиями и тяжестью заболевания на ранней стадии пандемии (Parikh et al., 2022).

Образцы крови пациентов собирались в разные временные периоды: в течение первых 24 часов с момента госпитализации (2, 5) или через 5–7 дней после поступления в стационар (3). Кроме того, кровь собиралась не только в условиях стационара. Например, в работе (3) собирали образцы крови от пациентов с легкой степенью на дому. Сбор крови для дальнейшего выделения РНК проводили в пробирки PAXgene™ и после необходимых манипуляций хранили при температуре –80 °С, при этом в исследовании (1) тип пробирки не раскрывался. Важно отметить, что при сравнении профилей экспрессии генов клеток крови, собранной в разные виды пробирок для выделения РНК, наблюдается неполная сходимость результатов (Menke et al., 2012).

**Критерии формирования выборки пациентов.** Критерием исключения пациентов для всех исследований была выявленная коинфекция (например, вирус иммунодефицита человека). Также в работе (2) были исключены пациенты с нейтропенией, гематологическими злокачественными новообразованиями и/или с активной химиотерапией, трансплантацией солидных органов, аутоиммунными заболеваниями и любым предшествующим приемом иммунодепрессантов (кортикостероидов, антицитокиновых биологических препаратов и модификаторов биологического ответа), в то время как в анализе (3) иммуномодулирующее лечение учитывалось как ковариационный фактор.

При сравнении сформированных групп не было выявлено статистически значимых различий по полу (1, 2, 4, 5) и возрасту (1, 2, 4). Однако в работе (3) показана зависимость между степенью тяжести и полом/возрастом. Различна и популяционная принадлежность пациентов – Восточная Азия, Австралия и Европа.

**Сопутствующая патология.** При анализе учитывался и такой важный фактор для прогрессирования тяжести COVID-19, как сопутствующие заболевания. В работе (2) для каждой из групп сравнения рассчитывался индекс коморбидности Charlson (Charlson et al., 1987). По данному индексу не выявлено статистически значимых различий между группами легкой, средней и тяжелой степени тяжести. В исследовании (3) частота эндокринных сопутствующих заболеваний, курения и ожирения была самой высокой в группе с тяжелой формой заболевания: 70, 50 и 50 % соответственно по сравнению с 53,8, 15,4 и 46,2 % пациентов со средней формой и 21, 5,3 и 5,3 % с легкой формой COVID-19. Значимые различия между пациентами разной степени тяжести среди сопутствующих заболеваний в (5) были выявлены только по сахарному диабету (скорректированное  $p = 0.048$ ).

Таким образом, несмотря на соответствие критериям отбора, обнаружены заметные различия в особенностях групп сравнения, а также инструментах выполнения анализа, что может объяснять наблюдаемую вариацию результатов выбранных исследований. Другим немаловажным фактором, влияющим на вариабельность результатов, является клеточная гетерогенность рассматриваемых образцов. Изучить вклад клеточной гетерогенности и более подробно рассмотреть состав и функциональную характеристику клеток возможно с помощью подхода секвенирования РНК единичных клеток (single-cell RNA-seq).

### Секвенирование РНК отдельных клеток в исследовании тяжелой формы COVID-19

На сегодняшний день проведен ряд исследований с использованием технологии single-cell RNA-seq для изучения иммунного клеточного ландшафта пациентов с COVID-19. В табл. S1 Приложения<sup>1</sup> представлены особенности экспрессии генов и клеточного состава при сравнении тяжелой и умеренной формы COVID-19. По результатам scRNA-seq клеток крови тяжелая форма характеризуется определенными изменениями в составе иммунных клеток: наблюдалось снижение плазматоцитодных дендритных клеток, NK-клеток, неклассических моноцитов и увеличение доли классических моноцитов, зрелых нейтрофилов и незрелых субпопуляций моноцитов и нейтрофилов. При тяжелой форме COVID-19 были обнаружены нейтрофилы низкой плотности (LDN), возникающие в условиях экстренного миелопоэза и связанные с дисфункциональными иммунными реакциями (Schultze et al., 2019), чего не наблюдалось при легкой форме. LDN, как и зрелые нейтрофилы, выделяли высокие уровни аларминов S100A8 и S100A9 (Schulte-Schrepping et al., 2020; Silvin et al., 2020; Ren et al., 2021; Wilk et al., 2021). Алармины высвобож-

<sup>1</sup> Табл. S1 Приложения см. по адресу:  
<https://vavilovj-icg.ru/download/pict-2026-30/appx7.pdf>

даются в условиях воспаления и образуют стабильный гетеродимер, известный как кальпротектин (Wang S. et al., 2018), который участвует в активации и хемотаксисе нейтрофилов (Ruskmann et al., 2003), а также предположительно является причиной синдрома высвобождения цитокинов (Silvin et al., 2020). Однако при тяжелой форме установлена разнонаправленная экспрессия генов некоторых провоспалительных (*IL1B*, *TNF*) и противовоспалительных (*IFNG*) цитокинов при общей повышенной выработке провоспалительных цитокинов иммунными клетками.

Различия в составе и функциональной активности Т-клеток при разной степени тяжести и на разных стадиях COVID-19 могут указывать на сложность Т-клеточных реакций в ответ на инфекцию (Ren et al., 2021). При этом каждый тип клеток, вовлеченных в ответ на коронавирусную инфекцию, отличается специфическим паттерном экспрессии генов и мембранных белков, что отражает сложную патофизиологию тяжелых симптомов COVID-19. Основные различия обнаружены в ответе клеток на интерферон I типа и в экспрессии интерферон-стимулируемых генов (*ISG15* и *IFITM1/2*). Такие наблюдения можно объяснить временными изменениями на разных стадиях прогрессирования заболевания. При тяжелой форме наблюдается временной сдвиг: от ранней, но кратковременной реакции на интерферон I типа к провоспалительной реакции на более поздних стадиях (Arunachalam et al., 2020), поэтому важно учитывать и момент времени взятия материала пациентов при анализе – день от начала заболевания. Однако не во всех работах конкретизируются типы клеток при анализе экспрессии генов, что тоже усложняет выявление общих изменений при тяжелой форме COVID-19.

В нескольких исследованиях наблюдались и общие патофизиологические пути развития тяжелой формы, отражающие явление дисрегуляции иммунного ответа. Например, отмечается нарушение презентации моноцитами антигенов главного комплекса гистосовместимости класса II Т-клеткам (*HLA-DR*), что указывает на развитие иммуносупрессии (Зурочка и др., 2008). Кроме того, установлен активированный путь NF-κB – транскрипционного фактора, способствующего развитию «цитокинового шторма» и окислительного стресса посредством усиления синтеза провоспалительных цитокинов и активных форм кислорода (АФК) активированными макрофагами и нейтрофилами (Болевич С.Б., Болевич С.С., 2020; Kesika et al., 2024). Оксидативный стресс, возникающий в результате повышенного образования АФК и снижения антиоксидантной защиты, способствует патогенезу тяжелой формы COVID-19, приводит к повреждению клеток и тканей через прямое повреждение, перекисное окисление липидов и окисление белков. Окислительный стресс в сочетании с высвобождением цитокинов вызывает дисфункцию эндотелиальных клеток и активирует каскад свертывания крови, провоцируя микрососудистый тромбоз. Провоспалительные цитокины дополнительно стимулируют синтез АФК, усугубляя ОРДС и повреждение легочной ткани и формируя порочный круг между окислительным стрессом и цитокиновым штормом (Polonikov, 2020; Gadotti et al., 2021; Alam, Czajkowsky, 2022; Labarrere, Kassab, 2022).

Анализ транскриптомных данных также подтверждает значение нарушения окислительно-восстановительного баланса в патогенезе тяжелого течения COVID-19. Авторы работы (Saheb Sharif-Askari et al., 2021) провели анализ *in silico* общедоступных данных пациентов с COVID-19 для оценки уровней экспрессии 125 генов, связанных с окислительным стрессом, и обнаружили семь генов (*MPO*, *S100A8*, *S100A9*, *SRXN1*, *GCLM*, *SESN2* и *TXN*), имевших повышенную экспрессию в цельной крови и аутопсиях легких пациентов с тяжелой формой COVID-19 по сравнению с пациентами с нетяжелой формой. В исследовании (Tavassolifar et al., 2023) в мононуклеарных клетках периферической крови пациентов с COVID-19 выявлены повышение уровня внутриклеточных АФК, снижение уровня антиоксиданта глутатиона, а также гиперэкспрессия генов оксидантов и антиоксидантов (*CAT*, *NFE2L2*, *SOD1*, *SOD2* и *CYBB*).

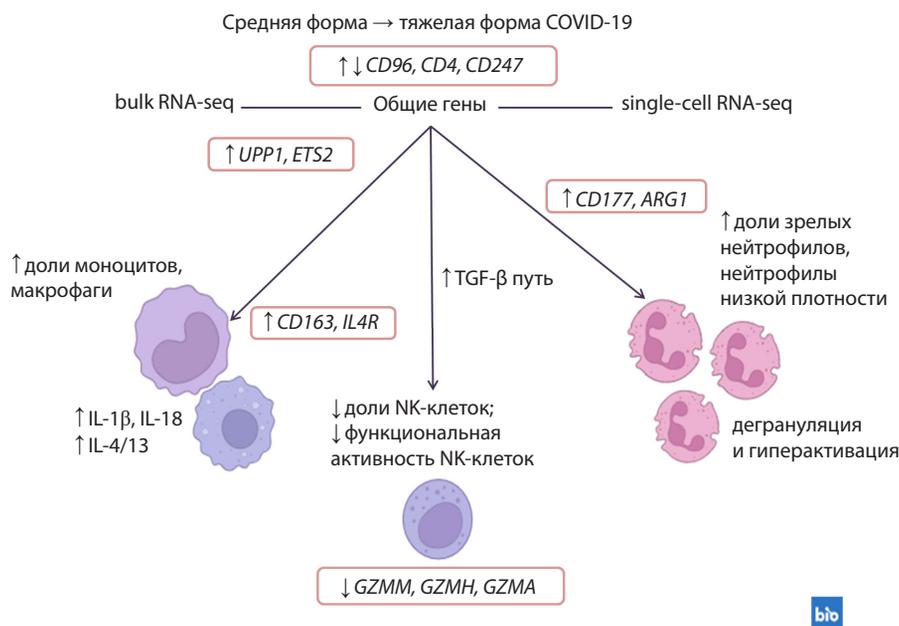
### Общность исследований bulk и single-cell RNA-seq

Некоторые закономерности, выявленные при анализе транскриптомов пула клеток крови пациентов с умеренной и тяжелой формами COVID-19, были установлены и при применении транскриптомики единичных клеток.

Увеличение доли циркулирующих нейтрофилов, их гиперактивация и дисрегуляция, появление незрелых или развивающихся нейтрофилов, связанных с тяжестью COVID-19, обнаружены в нескольких исследованиях с применением scRNA-seq подхода. Анализ выявил гиперэкспрессию генов, кодирующих провоспалительные цитокины, связанные с фагоцитозом и дегрануляцией, а также генов (например, *PADI4*, *MPO*, *ELANE* и *PRTN3*), которые участвуют в образовании нейтрофильных внеклеточных ловушек (Barnes et al., 2020; Schulte-Schrepping et al., 2020; Silvin et al., 2020; Wilk et al., 2021). Избыточное высвобождение NET из нейтрофилов способствует развитию окислительного стресса, гиперкоагуляции, нарушению альвеолярной микроциркуляции и повреждению легочной ткани. Уровни компонентов нейтрофильных внеклеточных ловушек – например, РНК *DEF1* и активность эластазы нейтрофилов (*ELANE*) в крови – рассматриваются как потенциальные биомаркеры тяжелой формы COVID-19 (Wargodsky et al., 2022).

У зрелых активированных нейтрофилов при тяжелой форме COVID-19 обнаружена повышенная экспрессия гена *CD177*, возможная роль которого в патогенезе ранее обсуждалась, и генов *CD274* и *ARG1* (Schulte-Schrepping et al., 2020; Wilk et al., 2021), входящих в состав 149 общих ДЭГ. *CD274* (PD-L1) и *ARG1* связаны с подавлением активации Т-клеток, что позволяет предположить, что нейтрофилы могут выполнять иммуносупрессивные функции при тяжелом течении COVID-19 (Schulte-Schrepping et al., 2020). Эти наблюдения согласуются с подавлением общих путей лимфоцитарного иммунитета.

Обогащение путей сигналов интерлейкинов как с противовоспалительными, так и с провоспалительными функциями обнаружено при анализах bulk и single-cell RNA-seq. Предполагается, что повышение уровня провоспалительных цитокинов играет важную роль в тяжелом прогресс-



**Рис. 5.** Общие дифференциально экспрессирующиеся гены и общие закономерности, выявленные при сравнении средней и тяжелой форм COVID-19 при bulk и single-cell подходах к секвенированию РНК крови.

↑ – увеличение доли или экспрессии; ↓ – снижение доли или экспрессии; ↑ ↓ – разнонаправленная экспрессия. Сделано с помощью <https://BioRender.com>.

сировании COVID-19, вызывая гипервоспалительную реакцию, называемую цитокиновым штормом (Ершов и др., 2020; Guo et al., 2020). Реакция IFN-I может способствовать развитию гипервоспалительного ответа, вызванного IL-1β, при тяжелом прогрессировании COVID-19 (Lee J.S. et al., 2020). Генетическая сигнатура сигнальных путей IL-4/13 и IL-18 в моноцитах значительно увеличивалась при тяжелой форме (Lee J.S. et al., 2020; Liu C. et al., 2021; Jeong et al., 2023).

Естественные киллерные (NK) клетки играют важную роль в реакциях врожденного иммунитета на вирусные инфекции. Снижение функциональной активности NK-клеток связано с острой и хронической вирусной инфекцией (Абакушина и др., 2012). При анализе результатов секвенирования РНК единичных клеток выявлены различия в транскриптоме NK-клеток между группами пациентов с умеренной и тяжелой степенью тяжести COVID-19. Например, обнаружена значительная транскрипционная перестройка, вызванная повышением экспрессии канонических генов активации NK-клеток при тяжелой форме, в том числе повышенная экспрессия генов, кодирующих цитотоксические эффекторные молекулы *GZMB*, *PRF1*, *GZMA*, а также маркер пролиферации *MKI67* (Wilk et al., 2021; Shaymardanov et al., 2022). Однако гены, связанные с цитотоксичностью NK-клеток (*GZMM*, *GZMH*, *GZMA*) были идентифицированы среди общих низкоэкспрессирующихся генов при анализе bulk RNA-seq, и сигнальный путь иммунитета, опосредованного NK-клетками (GO:0002228), был обогащен генами с гипозэкспрессией. Признаки истощения NK-клеток на уровне транскриптома наблюдались у пациентов с тяжелой формой COVID-19 и при применении scRNA-seq (Lee J.S. et al., 2020; Krämer

et al., 2021; Liu C. et al., 2021; Wilk et al., 2021; Witkowski et al., 2021). В работе (Witkowski et al., 2021) NK-клетки демонстрировали дисрегуляцию продукции цитокинов, клеточно-опосредованной цитотоксичности и ответа на вирус, несмотря на высокую экспрессию цитотоксических эффекторных молекул. При этом на транскрипционные сети, отвечающие за активацию NK-клеток, накладывалась доминирующая сигнатура ответа на TGF-β. В исследовании (McClain et al., 2023) дифференциальная экспрессия гена *TGFB1*, кодирующего трансформирующий ростовой фактор бета-1, наблюдалась в моноцитах CD14<sup>+</sup> и была связана с ухудшением течения COVID-19, а в работе (Rep et al., 2021) повышенная экспрессия *TGFB1* наблюдалась в Т-, В-лимфоцитах, NK- и дендритных клетках. Среди основных путей, активирующихся при тяжелой форме в случае Delta штамма коронавируса, был выявлен и путь TGF-β (Shaymardanov et al., 2022). Таким образом, участие TGF-β подтверждается и при анализе транскриптома на уровне единичных клеток. Данные результаты позволяют предположить, что нарушения цитотоксичности NK-клеток, в том числе посредством влияния TGF-β пути, могут быть связаны с механизмами развития тяжелой степени COVID-19 (Su et al., 2020; Lee M.J., Blish, 2023).

Особенности развития тяжелой формы COVID-19 характеризуются многогранной иммунной дисрегуляцией, которая описывается как состояние дисбаланса (Yao et al., 2021). Несмотря на различия в подходах к секвенированию РНК, при сравнении средней и тяжелой форм были идентифицированы общие дифференциально экспрессирующиеся гены и некоторые общие закономерности (рис. 5). Наблюдается нарушение регуляции врожденных (гипервоспалительные реакции, снижение цитотоксично-

сти НК-клеток, активация и дегрануляция нейтрофилов) и адаптивных (угнетение путей лимфоцитарного иммунитета) иммунных реакций. При этом обнаружена также разнонаправленная экспрессия некоторых общих ДЭГ, например *CD96*, *CD4* и *CD247*, экспрессирующихся на разных иммунных клетках и участвующих в иммунном ответе, что согласуется с явлением клеточной гетерогенности. При bulk RNA-seq вклад в общий уровень экспрессии генов может вносить экспрессия генов в клетках, которые содержатся в образцах крови пациентов в разных пропорциях, в то время как при single-cell RNA-seq уровни экспрессии генов получены в отдельных популяциях клеток.

## Заключение

Транскриптомика крови является важным инструментом изучения особенностей иммунитета у пациентов с коронавирусной инфекцией. Динамика иммунного ответа организма при инфекционном процессе отражается в изменениях экспрессии генов и может быть изучена у пациентов с развитием тяжелого состояния для выявления путей, ключевых для прогрессирования заболевания. Предполагается, что обнаруженные гены могут служить биомаркерами критического состояния при COVID-19 и возможными терапевтическими мишенями.

Проведен анализ общих, наиболее значимых дифференциально экспрессирующихся генов и путей, полученных в транскриптомных исследованиях COVID-19. Проанализированы работы, соответствующие критериям отбора, включающие массовое параллельное секвенирование РНК клеток цельной крови и сравнение тяжелой и средней степеней тяжести. Рассмотрены возможные молекулярные изменения, вызванные SARS-CoV-2, которые дифференцируют пациентов с развитием тяжелых осложнений COVID-19 от пациентов с умеренным течением болезни. В исследованиях bulk RNA-seq всего было идентифицировано 7605 дифференциально экспрессирующихся генов, при этом выявлено пять общих генов: *CD177*, *PPARG*, *PCOLCE2*, *SLC51A* и *ADAMTS2*. Рассмотрена их возможная роль в патогенезе тяжелой формы COVID-19. Такие патологические процессы, как гипервоспаление, фиброз и дисрегуляция адаптивного и врожденного иммунитета, могут быть одними из возможных механизмов прогрессирования коронавирусной инфекции. Для четырех наиболее близких по дизайну работ были найдены общие пути обогащения по 149 ДЭГ, которые включали активацию процессов дегрануляции нейтрофилов, сигнальных путей интерлейкинов, биосинтеза коллагена и подавление путей адаптивного и НК-клеточного иммунитета.

В данной работе рассматриваются результаты изучения транскриптомов пациентов с разными клиническими формами COVID-19 с использованием single-cell технологии, которые подтверждают предположение об иммунной дисрегуляции, выявленной по результатам bulk RNA-seq у пациентов при развитии тяжелой формы. Тяжелое течение отличается увеличением доли и активности зрелых и развивающихся нейтрофилов, развитием синдрома вы-

свобождения цитокинов, а также нарушением цитотоксичности НК-клеток с вовлечением TGF- $\beta$  пути.

Исследования в области транскриптомики COVID-19 раскрывают особенности клеточных и молекулярных процессов, которые могут привести к развитию тяжелой формы коронавирусной инфекции. Выявление РНК-биомаркеров у пациентов с COVID-19 может способствовать более эффективному распознаванию пациентов из групп риска, стратификации тяжести заболевания, прогнозированию тяжелого течения и исходов COVID-19. Транскриптомные биомаркеры показывают свою эффективность и в дифференциации пациентов с инфекцией SARS-CoV-2 и пациентов с другими легочными инфекциями. Кроме того, они играют важную роль в определении направлений эффективного лечения инфекции, в том числе длительной COVID-19. РНК-маркеры могут быть использованы в качестве терапевтических мишеней, а также в оценке ответа на лечение. Данные преимущества подчеркивают потенциал РНК-биомаркеров для клинического применения в диагностике, прогнозировании и разработке персонализированных стратегий лечения и реабилитации пациентов.

Однако, несмотря на высокую чувствительность РНК-биомаркеров, их интеграция в клиническую практику сталкивается со значительными трудностями. Особенности дизайнов исследований нередко затрудняют обобщение результатов и ограничивают возможность идентификации согласованных РНК-биомаркеров тяжести течения COVID-19. Вариабельность данных РНК секвенирования может быть обусловлена и влиянием индивидуальных физиологических различий пациентов и факторов окружающей среды. Результаты могут быть нерепрезентативными для пациентов разных популяций, в том числе из-за популяционной специфичности иммунного ответа (Nédélec et al., 2016; Randolph et al., 2024). Важно учитывать и особенности получения материала в клинических условиях: точное измерение РНК затруднено нестабильностью РНК в крови и сложностью очистки. В перспективе интеграция новых технологий, таких как РНК-секвенирование единичных клеток, проверка эффективности выявленных предикторов в более крупных многоцентровых испытаниях и разработка стандартизированных протоколов исследований помогут преодолеть некоторые ограничения клинического использования РНК-маркеров (Schultze, Aschenbrenner, 2021; Chen et al., 2022; Wargodsky et al., 2022; Eldien et al., 2025; Shimansky et al., 2025). Кроме того, для полного раскрытия сложного патогенеза тяжелой формы COVID-19 важно применить комплексный, «омиксный» подход, который помимо транскриптомики будет включать иные методы изучения особенностей иммунного ответа на SARS-CoV-2: протеомику, метаболомику, эпигенетику, мультиплексные измерения цитокинов/хемокинов и др. На основании результатов применения мультиомиксных технологий может быть обеспечена разработка новых подходов в прогнозировании исходов, профилактике и лечении тяжелого COVID-19.

## Список литературы / References

- Абакушина Е.В., Кузьмина Е.Г., Коваленко Е.И. Основные свойства и функции НК-клеток человека. *Иммунология*. 2012;33(4): 220-224  
[Abakushina E.V., Kuzmina E.G., Kovalenko E.I. The main characteristics of human natural killer cells. *Immunologiya = Immunology*. 2012;33(4):220-224 (in Russian)]
- Болевич С.Б., Болевич С.С. Комплексный механизм развития COVID-19. *Сеченовский вестник*. 2020;11(2):50-61. doi 10.47093/2218-7332.2020.11.2.50-61  
[Bolevich S.B., Bolevich S.S. Complex mechanism of COVID-19 development. *Sechenov Med J*. 2020;11(2):50-61. doi 10.47093/2218-7332.2020.11.2.50-61 (in Russian)]
- Ершов А.В., Суrowa В.Д., Долгих В.Т., Долгих Т.И. Цитокиновый шторм при новой коронавирусной инфекции и способы его коррекции. *Антибиотики и химиотерапия*. 2020;65(11-12): 27-37. doi 10.37489/0235-2990-2020-65-11-12-27-37  
[Ershov A.V., Surova V.D., Dolgikh V.T., Dolgikh T.I. Cytokine storm in the novel coronavirus infection and methods of its correction. *Antibiotiki i Khimioterapiya = Antibiotics and Chemotherapy*. 2020; 65(11-12):27-37. doi 10.37489/0235-2990-2020-65-11-12-27-37 (in Russian)]
- Зурочка А.В., Котляров А.Н., Кувайцев М.В., Квятковская С.В., Зурочка В.А., Рябова Л.В., Хайдуков С.В. Изменения экспрессии HLA-DR-антигенов на моноцитах у детей и ее клиническая значимость при сепсисе. *Медицинская иммунология*. 2008; 10(4-5):379-388. doi 10.15789/1563-0625-2008-4-5-379-388  
[Zurochka A.V., Kotlyarov A.N., Kuvaytsev M.V., Kvyatkovskaya S.V., Zurochka V.A., Ryabova L.V., Khaidukov S.V. Changes of HLA-DR antigen expression on monocytes in children and their clinical significance in sepsis. *Meditsinskaya Immunologiya = Medical Immunology*. 2008;10(4-5):379-388. doi 10.15789/1563-0625-2008-4-5-379-388 (in Russian)]
- Насонов Е.Л., Авдеева А.С. Интерлейкин 18 при иммуновоспалительных ревматических заболеваниях и COVID-19. *Научно-практическая ревматология*. 2022;60(2):195-204. doi 10.47360/1995-4484-2022-195-204  
[Nasonov E.L., Avdeeva A.S. Interleukin 18 in immune-mediated rheumatic diseases and COVID-19. *Rheumatology Science and Practice*. 2022;60(2):195-204. doi 10.47360/1995-4484-2022-195-204 (in Russian)]
- Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID-19). Временные методические рекомендации. Версия 18. Министерство здравоохранения Российской Федерации, 2023  
[Prevention, Diagnosis and Treatment of Novel Coronavirus Infection (COVID-19). Temporary guidelines. Version 18. The Ministry of Health of the Russian Federation, 2023 (in Russian)]
- Учасова Е.Г., Груздева О.В., Дылева Ю.А., Каретникова В.Н. Интерлейкин-33 и фиброз: современный взгляд на патогенез. *Медицинская иммунология*. 2018;20(4):477-484. doi 10.15789/1563-0625-2018-4-477-484  
[Uchasova E.G., Gruzdeva O.V., Dileva Yu.A., Karetnikova V.N. Interleukin 33 and fibrosis: pathogenesis updated. *Meditsinskaya Immunologiya = Medical Immunology*. 2018;20(4):477-484. doi 10.15789/1563-0625-2018-4-477-484 (in Russian)]
- Ahn J.Y., Park S., Yun Y.S., Song J.Y. Inhibition of type III TGF- $\beta$  receptor aggravates lung fibrotic process. *Biomed Pharmacother*. 2010;64(7):472-476. doi 10.1016/j.biopha.2010.01.006
- Alam M.S., Czajkowsky D.M. SARS-CoV-2 infection and oxidative stress: pathophysiological insight into thrombosis and therapeutic opportunities. *Cytokine Growth Factor Rev*. 2022;63:44-57. doi 10.1016/j.cytogfr.2021.11.001
- Alqutami F., Senok A., Hachim M. COVID-19 transcriptomic atlas: a comprehensive analysis of COVID-19 related transcriptomics datasets. *Front Genet*. 2021;12:755222. doi 10.3389/fgene.2021.755222
- Armignacco R., Carlier N., Jouinot A., Birtolo M.F., de Murat D., Tubach F., Hausfater P., ... Beurton A., Goulet H., Manivet P., Bertherat J., Assié G.; COVIDeF group. Whole blood transcriptome signature predicts severe forms of COVID-19: results from the COVIDeF cohort study. *Funct Integr Genomics*. 2024;24(3):107. doi 10.1007/s10142-024-01359-2
- Arunachalam P.S., Wimmers F., Mok C.K.P., Perera R.A.P.M., Scott M., Hagan T., Sigal N., ... Maecker H.T., Khatri P., Rouphael N., Peiris M., Pulendran B. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science*. 2020; 369(6508):1210-1220. doi 10.1126/science.abc6261
- Aschenbrenner A.C., Mouktaroudi M., Krämer B., Oestreich M., Antonakos N., Nuesch-Germano M., Gkizeli K., ... Nattermann J., Koutsoukou A., Giamarellos-Bourboulis E.J., Ulas T. German COVID-19 Omics Initiative (DeCOI). Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med*. 2021;13(1):7. doi 10.1186/s13073-020-00823-5
- Auwul M.R., Rahman M.R., Gov E., Shahjaman M., Moni M.A. Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. *Brief Bioinform*. 2021;22(5): bbab120. doi 10.1093/bib/bbab120
- Baghela A., An A., Zhang P., Acton E., Gauthier J., Brunet-Ratnasingham E., Blimkie T., Freue G.C., Kaufmann D., Lee A.H.Y., Levesque R.C., Hancock R.E.W. Predicting severity in COVID-19 disease using sepsis blood gene expression signatures. *Sci Rep*. 2023;13(1):1247. doi 10.1038/s41598-023-28259-y
- Bando S.Y., Bertonha F.B., Vieira S.E., de Oliveira D.B.L., Chalup V.N., Durigon E.L., Palmeira P., ... Antonangelo L., Lauterbach G.D.P., Regalio F.A., Cesar R.M., Jr, Moreira-Filho C.A. Blood leukocyte transcriptional modules and differentially expressed genes associated with disease severity and age in COVID-19 patients. *Sci Rep*. 2023; 13(1):898. doi 10.1038/s41598-023-28227-6
- Barnes B.J., Adrover J.M., Baxter-Stoltzfus A., Borczuk A., Cools-Lartigue J., Crawford J.M., Daßler-Plenker J., Guerci P., Huynh C., Knight J.S. Targeting potential drivers of COVID-19: neutrophil extracellular traps. *J Exp Med*. 2020;217:e20200652. doi 10.1084/jem.20200652
- Berlin D.A., Gulick R.M., Martinez F.J. Severe Covid-19. *N Engl J Med*. 2020;383(25):2451-2460. doi 10.1056/NEJMcp2009575
- Blanco-Melo D., Nilsson-Payant B.E., Liu W.C., Uhl S., Hoagland D., Möller R., Jordan T.X., ... Wang T.T., Schwartz R.E., Lim J.K., Albrecht R.A., ten Oever B.R. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*. 2020;181(5):1036-1045.e9. doi 10.1016/j.cell.2020.04.026
- Charlson M.E., Pompei P., Ales K.L., MacKenzie C.R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis*. 1987;40(5):373-383. doi 10.1016/0021-9681(87)90171-8
- Chaussabel D., Pascual V., Banchereau J. Assessing the human immune system through blood transcriptomics. *BMC Biol*. 2010;8:84. doi 10.1186/1741-7007-8-84
- Che Y., Jiang D., Zhang Y., Zhang J., Xu T., Sun Y., Fan J., ... Ding J., Hu C., Huang Y., Zhang J., Yang K. Elevated ubiquitination contributes to protective immunity against severe SARS-CoV-2 infection. *Clin Transl Med*. 2022;12(12):e1103. doi 10.1002/ctm2.1103
- Chen C.H., Lin S.W., Shen C.F., Hsieh K.S., Cheng C.M. Biomarkers during COVID-19: mechanisms of change and implications for patient outcomes. *Diagnostics (Basel)*. 2022;12(2):509. doi 10.3390/diagnostics12020509
- Chu W., Li X., Li C., Wan L., Shi H., Song X., Liu X., Chen X., Zhang C., Shan H., Lu Y., Yang B. TGF $\beta$ 3, a potential negative regulator of TGF- $\beta$  signaling, protects cardiac fibroblasts from hypoxia-induced apoptosis. *J Cell Physiol*. 2011;226(10):2586-2594. doi 10.1002/jcp.22604
- Chua R.L., Lukassen S., Trump S., Hennig B.P., Wendisch D., Pott F., Debnath O., ... Laudi S., Lehmann I., Conrad C., Sander L.E., Eils R. COVID-19 severity correlates with airway epithelium-immune cell

- interactions identified by single-cell analysis. *Nat Biotechnol.* 2020; 38(8):970-979. doi 10.1038/s41587-020-0602-4
- COVID-19 Multi-omics Blood Atlas (COMBAT) Consortium. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell.* 2022;185(5):916-938.e58. doi 10.1016/j.cell.2022.01.012
- Delorey T.M., Ziegler C.G.K., Heimberg G., Normand R., Yang Y., Segerstolpe Å., Abbondanza D., ... Li B., Shalek A.K., Villani A.C., Rozenblatt-Rosen O., Regev A. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature.* 2021;595(7865): 107-113. doi 10.1038/s41586-021-03570-8
- Deng Z., Fan T., Xiao C., Tian H., Zheng Y., Li C., He J. TGF- $\beta$  signaling in health, disease, and therapeutics. *Signal Transduct Target Ther.* 2024;9(1):61. doi 10.1038/s41392-024-01764-w
- Derakhshani A., Hemmat N., Asadzadeh Z., Ghaseminiya M., Shadbad M.A., Jadideslam G., Silvestris N., Racanelli V., Baradaran B. Arginase 1 (*Arg1*) as an up-regulated gene in COVID-19 patients: a promising marker in COVID-19 immunopathy. *J Clin Med.* 2021; 10(5):1051. doi 10.3390/jcm10051051
- de Seabra Rodrigues Dias I.R., Cao Z., Kwok H.F. Adamalysins in COVID-19 – potential mechanisms behind exacerbating the disease. *Biomed Pharmacother.* 2022;150:112970. doi 10.1016/j.biopha.2022.112970
- Desterke C., Turhan A.G., Bennaceur-Griscelli A., Griscelli F. PPAR $\gamma$  cistrome repression during activation of lung monocyte-macrophages in severe COVID-19. *iScience.* 2020;23(10):101611. doi 10.1016/j.isci.2020.101611
- Eldien H.M.S., Almaeen A.H., El Fath A.A., Taha A.E., Ahmed R., Elfadil H., Hetta H.F. Unlocking the potential of RNA sequencing in COVID-19: toward accurate diagnosis and personalized medicine. *Diagnostics (Basel).* 2025;15(2):229. doi 10.3390/diagnostics15020229
- Fabregat A., Sidiropoulos K., Viteri G., Forner O., Marin-Garcia P., Arnau V., D'Eustachio P., Stein L., Hermjakob H. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics.* 2017;18(1):142. doi 10.1186/s12859-017-1559-2
- Fang C., Ma Y. Peripheral blood genes crosstalk between COVID-19 and sepsis. *Int J Mol Sci.* 2023;24(3):2591. doi 10.3390/ijms24032591
- Ferreira-Gomes M., Kruglov A., Durek P., Heinrich F., Tizian C., Heinz G.A., Pascual-Reguant A., ... Radbruch H., Witkowski M., Melchers F., Radbruch A., Mashreghi M.F. SARS-CoV-2 in severe COVID-19 induces a TGF- $\beta$ -dominated chronic immune response that does not target itself. *Nat Commun.* 2021;12(1):1961. doi 10.1038/s41467-021-22210-3
- Filbin M.R., Mehta A., Schneider A.M., Kays K.R., Guess J.R., Gentili M., Fenyves B.G., ... Parry B.A., Villani A.C., Sade-Feldman M., Hacohen N., Goldberg M.B. Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Rep Med.* 2021;2(5): 100287. doi 10.1016/j.xcrn.2021.100287
- Gadotti A.C., Lipinski A.L., Vasconcellos F.T., Marqueze L.F., Cunha E.B., Campos A.C., Oliveira C.F., Amaral A.N., Baena C.P., Telles J.P., Tuon F.F., Pinho R.A. Susceptibility of the patients infected with Sars-Cov2 to oxidative stress and possible interplay with severity of the disease. *Free Radic Biol Med.* 2021;165:184-190. doi 10.1016/j.freeradbiomed.2021.01.044
- Guo C., Li B., Ma H., Wang X., Cai P., Yu Q., Zhu L., ... Weng J., Wei H., Jin T., Lin J., Qu K. Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm. *Nat Commun.* 2020;11(1):3924. doi 10.1038/s41467-020-17834-w
- Hadjadj J., Yatim N., Barnabei L., Corneau A., Boussier J., Smith N., Péré H., ... Fischer A., Duffy D., Rieux-Laucat F., Kernéis S., Terrier B. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science.* 2020;369(6504):718-724. doi 10.1126/science.abc6027
- Hadzega D., Babisova K., Hyblova M., Janostiakova N., Sabaka P., Janega P., Minarik G. Analysis of transcriptomics data from COVID-19 patients: a pilot research. *Folia Microbiol (Praha).* 2024; 69(1):155-164. doi 10.1007/s12223-024-01130-x
- Hasankhani A., Bahrami A., Tavakoli-Far B., Iranshahi S., Ghaemi F., Akbarzadeh M.R., Amin A.H., Abedi Kiasari B., Mohammadzadeh Shabestari A. Corrigendum: the role of peroxisome proliferator-activated receptors in the modulation of hyperinflammation induced by SARS-CoV-2 infection: a perspective for COVID-19 therapy. *Front Immunol.* 2024;15:1422261. doi 10.3389/fimmu.2024.1422261
- Hegenbarth J.C., Lezocco G., De Windt L.J., Stoll M. Perspectives on bulk-tissue RNA sequencing and single-cell RNA sequencing for cardiac transcriptomics. *Front Mol Med.* 2022;2:839338. doi 10.3389/fmmed.2022.839338
- Huang C., Wang Y., Li X., Ren L., Zhao J., Hu Y., Zhang L., ... Jiang R., Gao Z., Jin Q., Wang J., Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395(10223):497-506. doi 10.1016/S0140-6736(20)30183-5
- Huang W., Wang D., Yao Y.F. Understanding the pathogenesis of infectious diseases by single-cell RNA sequencing. *Microb Cell.* 2021; 8(9):208-222. doi 10.15698/mic2021.09.759
- Hughes T.D., Subramanian A., Chakraborty R., Cotton S.A., Del Pilar Giraldo Herrera M., Huang Y., Lambert N., Pinto M.D., Rahmani A.M., Sierra C.J., Downs C.A. The effect of SARS-CoV-2 variant on respiratory features and mortality. *Sci Rep.* 2023;13(1): 4503. doi 10.1038/s41598-023-31761-y
- Jackson H., Rivero Calle I., Broderick C., Habgood-Coote D., D'Souza G., Nichols S., Vito O., ... Levin M., Martinon-Torres F., Kaforou M.; PERFORM consortium; GEN-COVID (www.gen.covid.eu) study group. Characterisation of the blood RNA host response underpinning severity in COVID-19 patients. *Sci Rep.* 2022; 12(1):12216. doi 10.1038/s41598-022-15547-2
- Jansen J., Reimer K.C., Nagai J.S., Varghese F.S., Overheul G.J., de Beer M., Roverts R., ... COVID Moonshot consortium; van Rij R.P., Costa I.G., Schneider R.K., Smeets B., Kramann R. SARS-CoV-2 infects the human kidney and drives fibrosis in kidney organoids. *Cell Stem Cell.* 2022;29(2):217-231.e8. doi 10.1016/j.stem.2021.12.010
- Jeong K., Kim Y., Jeon J., Kim K. Subtyping of COVID-19 samples based on cell-cell interaction in single cell transcriptomes. *Sci Rep.* 2023;13(1):19629. doi 10.1038/s41598-023-46350-2
- Jovic D., Liang X., Zeng H., Lin L., Xu F., Luo Y. Single-cell RNA sequencing technologies and applications: a brief overview. *Clin Transl Med.* 2022;12(3):e694. doi 10.1002/ctm2.694
- Karu N., Kindt A., van Gammeren A.J., Ermens A.A.M., Harms A.C., Portengen L., Vermeulen R.C.H., Dik W.A., Langerak A.W., van der Velden V.H.J., Hankemeier T. Severe COVID-19 is characterised by perturbations in plasma amines correlated with immune response markers, and linked to inflammation and oxidative stress. *Metabolites.* 2022;12(7):618. doi 10.3390/metabo12070618
- Kesika P., Thangaleela S., Sisubalan N., Radha A., Sivamaruthi B.S., Chaiyasut C. The role of the nuclear factor-kappa B (NF- $\kappa$ B) pathway in SARS-CoV-2 infection. *Pathogens.* 2024;13(2):164. doi 10.3390/pathogens13020164
- Khan A., Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics.* 2017;18(1):287. doi 10.1186/s12859-017-1708-7
- Krämer B., Knoll R., Bonaguro L., ToVinh M., Raabe J., Astaburua-ga-García R., Schulte-Schrepping J., ... Sawitzki B.; Deutsche COVID-19 OMICS Initiative (DeCOI); Aschenbrenner A.C., Schultze J.L., Nattermann J. Early IFN- $\alpha$  signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19. *Immunity.* 2021;54(11):2650-2669.e14. doi 10.1016/j.immuni.2021.09.002
- Labarrere C.A., Kassab G.S. Glutathione deficiency in the pathogenesis of SARS-CoV-2 infection and its effects upon the host immune response in severe COVID-19 disease. *Front Microbiol.* 2022;13: 979719. doi 10.3389/fmicb.2022.979719
- Laine L., Skön M., Väisänen E., Julkunen I., Österlund P. SARS-CoV-2 variants Alpha, Beta, Delta and Omicron show a slower host cell

- interferon response compared to an early pandemic variant. *Front Immunol.* 2022;13:1016108. doi 10.3389/fimmu.2022.1016108
- Lee H.J., Georgiadou A., Walther M., Nwakanma D., Stewart L.B., Levin M., Otto T.D., Conway D.J., Coin L.J., Cunningham A.J. Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci Transl Med.* 2018; 10(447):eaar3619. doi 10.1126/scitranslmed.aar3619
- Lee J.S., Park S., Jeong H.W., Ahn J.Y., Choi S.J., Lee H., Choi B., ... Park S.H., Choi J.Y., Kim S.H., Jung I., Shin E.C. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci Immunol.* 2020; 5(49):eabd1554. doi 10.1126/sciimmunol.abd1554
- Lee M.J., Blish C.A. Defining the role of natural killer cells in COVID-19. *Nat Immunol.* 2023;24(10):1628-1638. doi 10.1038/s41590-023-01560-8
- Lei H. Hypoxia and activation of neutrophil degranulation-related genes in the peripheral blood of COVID-19 patients. *Viruses.* 2024;16(2): 201. doi 10.3390/v16020201
- Lévy Y., Wiedemann A., Hejblum B.P., Durand M., Lefebvre C., Surénaud M., Lacabaratz C., ... Yazdanpanah Y., Pantaleo G., Hocini H., Thiébaud R.; French COVID cohort study group. CD177, a specific marker of neutrophil activation, is associated with coronavirus disease 2019 severity and death. *iScience.* 2021;24(7):102711. doi 10.1016/j.isci.2021.102711
- Liu C., Martins A.J., Lau W.W., Rachmaninoff N., Chen J., Imberti L., Mostaghimi D., ... COVID Clinicians; Rossi C., Su H.C., Kuhns D.B., Cohen J.I., Notarangelo L.D., Tsang J.S. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell.* 2021;184(7):1836-1857.e22. doi 10.1016/j.cell.2021.02.018
- Liu X., Chen R., Li B., Zhang J., Liu P., Li B., Li F., Zhang W., Lyu X., Hu M. Oxidative stress indexes as biomarkers of the severity in COVID-19 patients. *Int J Med Sci.* 2024;21(15):3034-3045. doi 10.7150/ijms.102879
- 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
- Maurya R., Swaminathan A., Shamim U., Arora S., Mishra P., Raina A., Ravi V., Tarai B., Budhiraja S., Pandey R. Co-evolution of SARS-CoV-2 variants and host immune response trajectories underlie COVID-19 pandemic to epidemic transition. *iScience.* 2023;26(12): 108336. doi 10.1016/j.isci.2023.108336
- McClain M.T., Zhbannikov I., Satterwhite L.L., Henao R., Giroux N.S., Ding S., Burke T.W., Tsalik E.L., Nix C., Balcazar J.P., Petzold E.A., Shen X., Woods C.W. Epigenetic and transcriptional responses in circulating leukocytes are associated with future decompensation during SARS-CoV-2 infection. *iScience.* 2023;27(1):108288. doi 10.1016/j.isci.2023.108288
- Meizlish M.L., Pine A.B., Bishai J.D., Goshua G., Nadelmann E.R., Simonov M., Chang C.H., ... Halene S., Damsky W., van Dijk D., Lee A.I., Chun H.J. A neutrophil activation signature predicts critical illness and mortality in COVID-19. *Blood Adv.* 2021;5(5):1164-1177. doi 10.1182/bloodadvances.2020003568
- Menke A., Rex-Haffner M., Klengel T., Binder E.B., Mehta D. Peripheral blood gene expression: it all boils down to the RNA collection tubes. *BMC Res Notes.* 2012;5:1. doi 10.1186/1756-0500-5-1
- Morselli Gysi D., do Valle Í., Zitnik M., Ameli A., Gan X., Varol O., Ghiassian S.D., Patten J.J., Davey R.A., Loscalzo J., Barabási A.L. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc Natl Acad Sci USA.* 2021;118(19): e2025581118. doi 10.1073/pnas.2025581118
- Nassir N., Tambi R., Bankapur A., Al Heialy S., Karuvantevida N., Khansaheb H.H., Zehra B., ... Yaseen Hachim M., Casanova J.L., Berdiev B.K., Alsheikh-Ali A., Uddin M. Single-cell transcriptome identifies *FCGR3B* upregulated subtype of alveolar macrophages in patients with critical COVID-19. *iScience.* 2021;24(9):103030. doi 10.1016/j.isci.2021.103030
- Nédélec Y., Sanz J., Baharian G., Szpiech Z.A., Pacis A., Dumaine A., Grenier J.C., ... Hernandez R.D., Pique-Regi R., Tung J., Yotova V., Barreiro L.B. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell.* 2016; 167(3):657-669.e21. doi 10.1016/j.cell.2016.09.025
- Oh K.K., Adnan M., Cho D.H. Network pharmacology approach to decipher signaling pathways associated with target proteins of NSAIDs against COVID-19. *Sci Rep.* 2021;11(1):9606. doi 10.1038/s41598-021-88313-5
- Parikh V.N., Ioannidis A.G., Jimenez-Morales D., Gorzynski J.E., De Jong H.N., Liu X., Roque J., ... Palacios J.A., Pinsky B.A., Bustamante C.D., Rivas M.A., Ashley E.A. Deconvoluting complex correlates of COVID-19 severity with a multi-omic pandemic tracking strategy. *Nat Commun.* 2022;13(1):5107. doi 10.1038/s41467-022-32397-8
- Pei R., Feng J., Zhang Y., Sun H., Li L., Yang X., He J., ... Wen K., Zhou H., Chen J., Rong Z., Chen X. Host metabolism dysregulation and cell tropism identification in human airway and alveolar organoids upon SARS-CoV-2 infection. *Protein Cell.* 2021;12(9):717-733. doi 10.1007/s13238-020-00811-w
- Pincemail J., Cavalier E., Charlier C., Cheramy-Bien J.P., Brevers E., Courtois A., Fadeur M., Meziane S., Goff C.L., Misset B., Albert A., Defraigne J.O., Rousseau A.F. Oxidative stress status in COVID-19 patients hospitalized in intensive care unit for severe pneumonia. A pilot study. *Antioxidants (Basel).* 2021;10(2):257. doi 10.3390/antiox10020257
- Polonikov A. Endogenous deficiency of glutathione as the most likely cause of serious manifestations and death in COVID-19 patients. *ACS Infect Dis.* 2020;6(7):1558-1562. doi 10.1021/acscinfecdis.0c00288
- Randolph H.E., Aracena K.A., Lin Y.L., Mu Z., Barreiro L.B. Shaping immunity: the influence of natural selection on population immune diversity. *Immunol Rev.* 2024;323(1):227-240. doi 10.1111/imr.13329
- Ren X., Wen W., Fan X., Hou W., Su B., Cai P., Li J., ... Qu K., Wang X., Chen J., Jin R., Zhang Z. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell.* 2021;184(7):1895-1913.e19. doi 10.1016/j.cell.2021.01.053
- Reusch N., De Domenico E., Bonaguro L., Schulte-Schrepping J., Baßler K., Schultze J.L., Aschenbrenner A.C. Neutrophils in COVID-19. *Front Immunol.* 2021;12:652470. doi 10.3389/fimmu.2021.652470
- Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47. doi 10.1093/nar/gkv007
- 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
- Roessler C., de Oliveira K.C.S., de Oliveira Portella A.X., Fortes P.C.N., Machado F.R., Araujo S.M., Prigol M., Lucio L.C., Benvegnú D.M., Ferreto L.E.D. Evaluation of oxidative stress level: reactive oxygen species, reduced glutathione, and D-dimer in patients hospitalized due to COVID-19. *Redox Rep.* 2023;28(1):1-6. doi 10.1080/13510002.2023.2272384
- Ryckman C., Vandal K., Rouleau P., Talbot M., Tessier P.A. Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. *J Immunol.* 2003; 170(6):3233-3242. doi 10.4049/jimmunol.170.6.3233
- Saheb Sharif-Askari N., Saheb Sharif-Askari F., Mdkhana B., Hussain Alsayed H.A., Alsafar H., Alrais Z.F., Hamid Q., Halwani R. Upregulation of oxidative stress gene markers during SARS-COV-2 viral infection. *Free Radic Biol Med.* 2021;172:688-698. doi 10.1016/j.freeradbiomed.2021.06.018
- Saichi M., Ladjemi M.Z., Korniotis S., Rousseau C., Ait Hamou Z., Massenet-Regad L., Amblard E., Noel F., Marie Y., Bouteiller D., Medvedovic J., Pène F., Soumelis V. Single-cell RNA sequencing of blood antigen-presenting cells in severe COVID-19 reveals multi-process defects in antiviral immunity. *Nat Cell Biol.* 2021;23(5): 538-551. doi 10.1038/s41556-021-00681-2

- Schimke L.F., Marques A.H.C., Baiocchi G.C., de Souza Prado C.A., Fonseca D.L.M., Freire P.P., Rodrigues Praça D., ... Ulas T., Schultze J.L., Nakaya H.I., Jurisica I., Cabral-Marques O. Severe COVID-19 shares a common neutrophil activation signature with other acute inflammatory states. *Cells*. 2022;11(5):847. doi 10.3390/cells11050847
- Schulte-Schrepping J., Reusch N., Paclik D., Baßler K., Schlickeiser S., Zhang B., Krämer B., ... Li Y., Nattermann J., Sawitzki B., Saliba A.E., Sander L.E.; Deutsche COVID-19 OMICS Initiative (DeCOI). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell*. 2020;182(6):1419-1440.e23. doi 10.1016/j.cell.2020.08.001
- Schultze J.L., Aschenbrenner A.C. COVID-19 and the human innate immune system. *Cell*. 2021;184(7):1671-1692. doi 10.1016/j.cell.2021.02.029
- Schultze J.L., Mass E., Schlitzer A. Emerging principles in myelopoiesis at homeostasis and during infection and inflammation. *Immunity*. 2019;50(2):288-301. doi 10.1016/j.immuni.2019.01.019
- Shaymardanov A.M., Antonova O.A., Sokol A.D., Deinichenko K.A., Kazakova P.G., Milovanov M.M., Zakubansky A.V., ... Keskinov A.A., Kraevoy S.A., Snigir E.A., Svetlichnyy D.V., Skvortsova V.I. Single-cell gene expression analysis revealed immune cell signatures of Delta COVID-19. *Cells*. 2022;11(19):2950. doi 10.3390/cells11192950
- Shimansky V., Popov O., Kel A., Goryanin I., Klochkova T., Apalko S., Shushentseva N., Anisenkova A., Mosenko S., Shcherbak S. Analysis of the expression profile in COVID-19 patients in the Russian population considering disease severity, mortality, and cytokine storm. *Biomedicines*. 2025;13(4):863. doi 10.3390/biomedicines13040863
- Silvin A., Chapuis N., Dunsmore G., Goubet A.G., Dubuisson A., Derosa L., Almiere C., ... André F., Zitvogel L., Ginhoux F., Fontenay M., Solary E. Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell*. 2020;182(6):1401-1418.e18. doi 10.1016/j.cell.2020.08.002
- Song F., Qian Y., Peng X., Li X., Xing P., Ye D., Lei H. The frontline of immune response in peripheral blood. *PLoS One*. 2017;12(8):e0182294. doi 10.1371/journal.pone.0182294
- Su Y., Chen D., Yuan D., Lausted C., Choi J., Dai C.L., Voillet V., ... Greenberg P.D., Gottardo R., Davis M.M., Goldman J.D., Heath J.R. Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell*. 2020;183(6):1479-1495.e20. doi 10.1016/j.cell.2020.10.037
- Tabassum T., Rahman A., Araf Y., Ullah M.A., Hosen M.J. Prospective selected biomarkers in COVID-19 diagnosis and treatment. *Biomark Med*. 2021;15(15):1435-1449. doi 10.2217/bmm-2021-0038
- Tang H., Gao Y., Li Z., Miao Y., Huang Z., Liu X., Xie L., Li H., Wen W., Zheng Y., Su W. The noncoding and coding transcriptional landscape of the peripheral immune response in patients with COVID-19. *Clin Transl Med*. 2020;10(6):e200. doi 10.1002/ctm2.200
- Tavassolifar M.J., Aghdaei H.A., Sadatpour O., Maleknia S., Fayazadeh S., Mohebbi S.R., Montazer F., Rabbani A., Zali M.R., Izad M., Meyfour A. New insights into extracellular and intracellular redox status in COVID-19 patients. *Redox Biol*. 2023;59:102563. doi 10.1016/j.redox.2022.102563
- Togami K., Yamaguchi K., Chono S., Tada H. Evaluation of permeability alteration and epithelial-mesenchymal transition induced by transforming growth factor- $\beta_1$  in A549, NCI-H441, and Calu-3 cells: development of an *in vitro* model of respiratory epithelial cells in idiopathic pulmonary fibrosis. *J Pharmacol Toxicol Methods*. 2017; 86:19-27. doi 10.1016/j.vascn.2017.02.023
- Vaz de Paula C.B., de Azevedo M.L.V., Nagashima S., Martins A.P.C., Malaquias M.A.S., Miggiolaro A.F.R.D.S., da Silva Motta Júnior J., Avelino G., do Carmo L.A.P., Carstens L.B., de Noronha L. IL-4/IL-13 remodeling pathway of COVID-19 lung injury. *Sci Rep*. 2020; 10(1):18689. doi 10.1038/s41598-020-75659-5
- Vlasov I., Panteleeva A., Usenko T., Nikolaev M., Izumchenko A., Gavrilova E., Shlyk I., Miroshnikova V., Shadrina M., Polushin Y., Pchelina S., Slonimsky P. Transcriptomic profiles reveal downregulation of low-density lipoprotein particle receptor pathway activity in patients surviving severe COVID-19. *Cells*. 2021;10(12):3495. doi 10.3390/cells10123495
- Wang F., Xia H., Yao S. Regulatory T cells are a double-edged sword in pulmonary fibrosis. *Int Immunopharmacol*. 2020;84:106443. doi 10.1016/j.intimp.2020.106443
- Wang Q.S., Edahiro R., Namkoong H., Hasegawa T., Shirai Y., Sonehara K., Tanaka H., ... Miyano S., Ogawa S., Kanai T., Fukunaga K., Okada Y. The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 task force. *Nat Commun*. 2022;13(1):4830. doi 10.1038/s41467-022-32276-2
- Wang S., Song R., Wang Z., Jing Z., Wang S., Ma J. S100A8/A9 in inflammation. *Front Immunol*. 2018;9:1298. doi 10.3389/fimmu.2018.01298
- Wang Y., Schughart K., Pelala T.M., Chew T., Kim K., Karvunidis T., Knippenberg B., ... Burcham J., McLean A.; PREDICT-19 consortium; Tang B., Shojaei M. Blood transcriptome responses in patients correlate with severity of COVID-19 disease. *Front Immunol*. 2023;13:1043219. doi 10.3389/fimmu.2022.1043219
- Wargodsky R., Dela Cruz P., LaFleur J., Yamane D., Kim J.S., Benjenk I., Heinz E., ... Farrar K., Toma I., Jordan T., Goldman J., McCaffrey T.A. RNA sequencing in COVID-19 patients identifies neutrophil activation biomarkers as a promising diagnostic platform for infections. *PLoS One*. 2022;17(1):e0261679. doi 10.1371/journal.pone.0261679
- Wilk A.J., Lee M.J., Wei B., Parks B., Pi R., Martínez-Colón G.J., Ranganath T., ... Holmes S., Rabinovitch M., Rogers A.J., Greenleaf W.J., Blish C.A. Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. *J Exp Med*. 2021;218(8):e20210582. doi 10.1084/jem.20210582
- Witkowski M., Tizian C., Ferreira-Gomes M., Niemeyer D., Jones T.C., Heinrich F., Frischbutter S., ... Durek P., Kruglov A., Radbruch A., Mashreghi M.F., Diefenbach A. Untimely TGF $\beta$  responses in COVID-19 limit antiviral functions of NK cells. *Nature*. 2021; 600(7888):295-301. doi 10.1038/s41586-021-04142-6
- Xiong Y., Liu Y., Cao L., Wang D., Guo M., Jiang A., Guo D., ... Shi M., Liu Y., Zhou Y., Lan K., Chen Y. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect*. 2020;9(1):761-770. doi 10.1080/22221751.2020.1747363
- Yao C., Bora S.A., Parimon T., Zaman T., Friedman O.A., Palatinus J.A., Surapaneni N.S., ... Martins G.A., Stripp B.R., Gharib S.A., Goodridge H.S., Chen P. Cell-type-specific immune dysregulation in severely ill COVID-19 patients. *Cell Rep*. 2021;34(1):108590. doi 10.1016/j.celrep.2020.108590
- Yoon H.S., Kim H.Y., Cho K.A., Kim Y.H., Woo S.Y., Kim H.S., Kang J.L., Ryu K.H., Park J.W. Procollagen C-endopeptidase enhancer 2 secreted by tonsil-derived mesenchymal stem cells increases the oxidative burst of promyelocytic HL-60 cells. *Biology (Basel)*. 2022;11(2):255. doi 10.3390/biology11020255
- Zhang J.Y., Wang X.M., Xing X., Xu Z., Zhang C., Song J.W., Fan X., ... Zhang Y., Zumla A., Maeurer M., Bai F., Wang F.S. Single-cell landscape of immunological responses in patients with COVID-19. *Nat Immunol*. 2020;21(9):1107-1118. doi 10.1038/s41590-020-0762-x
- Zhang Y., Wang S., Xia H., Guo J., He K., Huang C., Luo R., Chen Y., Xu K., Gao H., Sheng J., Li L. Identification of monocytes associated with severe COVID-19 in the PBMCs of severely infected patients through single-cell transcriptome sequencing. *Engineering (Beijing)*. 2022;17:161-169. doi 10.1016/j.eng.2021.05.009
- Zizzo G., Cohen P.L. Imperfect storm: is interleukin-33 the Achilles heel of COVID-19? *Lancet Rheumatol*. 2020;2(12):e779-e790. doi 10.1016/S2665-9913(20)30340-4

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 16.07.2025. После доработки 05.12.2025. Принята к публикации 05.12.2025.

doi 10.18699/vjgb-26-12

# Однонуклеотидные полиморфизмы в геномах вирусов клещевого энцефалита и Западного Нила при тройной природной инфекции у садовой камышовки (*Acrocephalus dumetorum*)

Е.П. Пономарева <sup>1</sup> , В.А. Терновой <sup>1</sup>, Е.В. Протопопова <sup>1</sup>, Н.А. Тупота <sup>1</sup>, В.Б. Локтев <sup>1, 2</sup><sup>1</sup> Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р. п. Кольцово, Новосибирская область, Россия<sup>2</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия ponomareva\_ep@vector.nsc.ru

**Аннотация.** Вовлечение различных видов диких птиц в формирование природных очагов вирусов клещевого энцефалита (ВКЭ) и Западного Нила (ВЗН) обеспечивает быстрое распространение этих ортофлавирусов в различных географических районах и формирование новых природных очагов данных инфекций. Однако роль популяционной variability ВКЭ и ВЗН в формировании природных очагов, возможность появления (селекции) новых вирусных вариантов, патогенных для человека, в этих природных очагах остаются еще недостаточно изученными. Цель настоящего исследования – оценить популяционную гетерогенность геномов ВКЭ сибирского и дальневосточного генотипов и ВЗН, которые были одновременно выделены из тканей одной особи садовой камышовки (*Acrocephalus dumetorum*), отловленной в природных ландшафтах пригорода Томска. С этой целью были использованы методы выделения вирусных штаммов на различных культурах клеток в сочетании с анализом полных вирусных геномов полученных изолятов, определенных методами традиционного и высокопроизводительного секвенирования (NGS). Консенсусные полногеномные нуклеотидные последовательности вирусов получали секвенированием по Сэнгеру и сравнивали с последовательностями, полученными методом NGS, с однонуклеотидными заменами (single nucleotide variant, SNV) 2 % и выше, в пределах изучаемой популяции. Обнаруженный однонуклеотидный полиморфизм (SNP) ассоциировался как с синонимичными, так и несинонимичными нуклеотидными заменами преимущественно локализующихся в генах неструктурных белков ВКЭ и ВЗН. При этом возможных рекомбинационных событий в геномах изолированных ортофлавирусов обнаружить не удалось. Результаты показали, что изоляты ВЗН Tomsk/bird/2006/A4, ВКЭ RT12 и RT122, выделенные из *A. dumetorum*, представлены гетерогенными вирусными популяциями, в которых обнаруживаются множественные SNV, возникающие с частотой от 1.75 до 19.88 % для ВЗН и от 2.08 до 23.73 % для ВКЭ. Для большинства выявленных SNP найдены аналогичные нуклеотидные замены в геномах уже известных штаммов ВКЭ и ВЗН, что может свидетельствовать о важной роли этих SNV-спектров в вирусной адаптации и обеспечении генетического и фенотипического разнообразия этих вирусов в природе.

**Ключевые слова:** вирус клещевого энцефалита; вирус Западного Нила; однонуклеотидный полиморфизм; вирусный геном; высокопроизводительное секвенирование; садовая камышовка; ортофлавирусы

**Для цитирования:** Пономарева Е.П., Терновой В.А., Протопопова Е.В., Тупота Н.А., Локтев В.Б. Однонуклеотидные полиморфизмы в геномах вирусов клещевого энцефалита и Западного Нила при тройной природной инфекции у садовой камышовки (*Acrocephalus dumetorum*). *Вавиловский журнал генетики и селекции*. 2026;30(1):117-125. doi 10.18699/vjgb-26-12

**Финансирование.** Исследование выполнено в рамках государственного задания 7/21 и 9/21 ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора.

**Благодарности.** Авторы выражают глубокую благодарность  А.Н. Швалову и  Т.П. Микрюковой за помощь в проведении экспериментов, секвенировании и выполнении анализа вирусных геномов.

## Single nucleotide polymorphisms are typical for tick-borne encephalitis and West Nile viruses during triple natural mixed infections in Blyth's reed warbler (*Acrocephalus dumetorum*)

Е.П. Ponomareva <sup>1</sup> , V.A. Ternovoi <sup>1</sup>, E.V. Protopopova <sup>1</sup>, N.L. Tupota <sup>1</sup>, V.B. Loktev <sup>1, 2</sup><sup>1</sup> State Research Center of Virology and Biotechnology "Vector", Koltsovo, Novosibirsk Region, Russia<sup>2</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia ponomareva\_ep@vector.nsc.ru

**Abstract.** Wild bird species contribute significantly to the rapid geographic dissemination of tick-borne encephalitis viruses (TBEV) and West Nile virus (WNV), facilitating the establishment of new natural foci of these *orthoflaviviruses*. However, the impact of TBEV and WNV population variability on shaping these foci, as well as the potential emergence of new human-pathogenic viral variants, remain underexplored. This study aimed to assess the genetic heterogeneity of TBEV (Siberian and Far Eastern genotypes) and WNV, isolated simultaneously from the tissues of a single garden reed warbler (*Acrocephalus dumetorum*) collected in the suburbs of Tomsk. The methods of viral strain isolation on various cell cultures were used in combination with a whole-genome analysis of isolates through traditional and high-throughput sequencing (NGS) methods. Consensus full-genome nucleotide sequences of the viruses were obtained by Sanger sequencing and compared with those obtained by NGS, with single nucleotide substitutions (single nucleotide variants, SNVs) accounting for 2 % or higher within the population under study. Our findings revealed single nucleotide polymorphisms (SNPs) associated with both synonymous and non-synonymous nucleotide substitutions, primarily located within the non-structural protein genes of TBEV and WNV. Notably, recombination events were not detected in the genomes of isolated *orthoflaviviruses*. The WNV isolate, Tomsk/bird/2006/A4, and the TBEV isolates, PT12 and PT122, obtained from *A. dumetorum*, exhibited heterogeneous viral populations, with SNVs ranging in frequency from 1.75 to 19.88 % for WNV and from 2.08 to 23.73 % for TBEV. Most identified SNPs shared similar nucleotide substitutions in the genomes of already known strains of TBEV and WNV, suggesting that these SNVs could play a crucial role in viral adaptation and underscore the genetic and phenotypic diversity of these viruses in nature.

**Key words:** tick-borne encephalitis virus; West Nile virus; single nucleotide polymorphism; viral genome; high-throughput sequencing; garden warbler; *orthoflaviviruses*

**For citation:** Ponomareva E.P., Ternovoi V.A., Protopopova E.V., Tupota N.L., Loktev V.B. Single nucleotide polymorphisms are typical for tick-borne encephalitis and West Nile viruses during triple natural mixed infections in Blyth's reed warbler (*Acrocephalus dumetorum*). *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):117-125. doi 10.18699/vjgb-26-12

## Введение

Вирус клещевого энцефалита (*Orthoflavivirus encephaliti-dis*) и вирус Западного Нила (*Orthoflavivirus nilense*) относятся к семейству *Flaviviridae*, роду *Orthoflavivirus* (Current ICTV Taxonomy Release; Postler et al., 2023). Вирусы клещевого энцефалита (ВКЭ) и Западного Нила (ВЗН) способны вызывать заболевания человека, которые могут привести к развитию тяжелых поражений центральной нервной системы (Woraku, 2023; Singh et al., 2024). Вирусы формируют природные очаги, а человек, как правило, заражается через укус комара (ВЗН) или клеща (ВКЭ). Природные очаги ВКЭ характерны для северной Евразии, а ВЗН распространен фактически глобально (Pustijanac et al., 2023; Simonin, 2024). На территории юга Западной Сибири отмечена одновременная циркуляция этих двух ортофлавивирусов (Терновой и др., 2004; Кононова и др., 2006).

Для ВКЭ и ВЗН характерно генетическое разнообразие, так, для ВКЭ описано существование не менее пяти основных генотипов, а для ВЗН – по меньшей мере девяти генотипов (Dai et al., 2018; Kozlova et al., 2018; Simonin, 2024). Уровень нуклеотидных различий между разными генотипами может достигать 18–20 %. Для сибирского региона характерна циркуляция сибирского и дальневосточного генотипов ВКЭ и генотипа 1a ВЗН.

Современные методы высокопроизводительного секвенирования дают возможность выявить наличие SNP (однонуклеотидный полиморфизм) в относительно небольших вирусных популяциях. Так, исследование наличия SNV (single nucleotide variant) в популяциях адаптированного к различным условиям культивирования штамма ЕК-328 ВКЭ и его клонированных вариантов показало, что SNV возникают с частотой около 1 % в популяциях лабораторного штамма ВКЭ (Litov et al., 2018). Это позволило предположить, что существование минорных SNV ВКЭ обеспечивает микроэволюцию и возможность быстрой

адаптации вирусной популяции к изменению условий даже при культивировании в лабораторных условиях. Данную гипотезу подтвердили исследования по изменчивости генома штамма С11-13 ВКЭ сибирского генотипа при культивировании в лабораторных условиях (Ternovoi et al., 2024). Наличие однонуклеотидного полиморфизма также описано в геномах вирусов Зика, денге, японского энцефалита и ВЗН (Kaiser et al., 2019; Zaráte et al., 2019; Borda et al., 2021). Однако попытка использования SNP, характерных для гена белка Е вакцинного штамма вируса японского энцефалита с целью аттенуации ВЗН не принесла успеха. Было высказано предположение, что профиль SNP характерен для каждого вида ортофлавивирусов.

Ранее участие диких птиц в циркуляции ВКЭ и ВЗН в Томской и Новосибирской областях было показано для 42 из 60 исследованных видов птиц (Mikryukova et al., 2014; Moskvitina et al., 2014; Korobitsyn et al., 2021). Причем у птиц и клещей, снятых с них, отмечено наличие микстинфекций в 2.13 % случаев для клещей и до 39 % для птиц. Такое видовое разнообразие птиц, обеспечивающее циркуляцию ВКЭ и ВЗН в природных условиях, закономерно ставит вопрос о механизмах адаптации этих ортофлавивирусов к новому хозяину. Было высказано предположение, что вовлечение различных видов диких птиц в формирование природных очагов ВКЭ и ВЗН обеспечивает быстрое распространение этих ортофлавивирусов в разных географических районах и формирование новых природных очагов данных инфекций. Однако роль популяционной изменчивости ВКЭ и ВЗН в формировании природных очагов, возможность появления (селекции) новых вирусных вариантов, патогенных для человека в этих природных очагах, остаются еще недостаточно изученными.

В настоящем исследовании мы попытались оценить популяционную гетерогенность геномов ВЗН и ВКЭ в тканях одной садовой камышовки, которая была ранее отлов-

лена в пригороде Томска. С этой целью мы применили методы выделения вирусных штаммов на различных культурах клеток в сочетании с анализом полных вирусных геномов полученных изолятов, определенных методами традиционного и высокопроизводительного секвенирования.

## Материалы и методы

**Исследованные образцы** – объединенный 10 % гомогенат селезенки и печени садовой камышовки (*Acrocephalus dumetorum*), отловленной в Томской области в 2006 г. (Mikryukova et al., 2014). Для изоляции ортофлавивирусов использовали клеточные культуры почки эмбриона свиньи (СПЭВ) и комара *Aedes albopictus* (С6/36), полученные из коллекции клеточных культур ФБУН ГНЦ ВБ «Вектор». Клеточные культуры культивировали с помощью среды DMEM/F12 (ГНЦ ВБ «Вектор», РФ), содержащей 10 % сыворотки плода коровы (Gibco, США) и 80 мкг/мл сульфата гентамицина, при 37 °С для клеток СПЭВ и 28 °С для клеток С6/36.

Определение инфекционной активности вирусных изолятов проводили по развитию цитопатогенного действия. С этой целью клетки СПЭВ рассеивали в 96-луночные культуральные микропланшеты в количестве 10<sup>5</sup> клеток в объеме 50 мкл на лунку и осуществляли заражение вирусосодержащим материалом. Вирусы титровали в среде DMEM/F12 с содержанием 2 % сыворотки плода коровы, в объеме 100 мкл на лунку. Результат учитывали через пять суток с вычислением инфекционного титра вирусов, как описано в статье (Svatchenko et al., 2021).

**Иммуноферментный анализ.** Выявление антигенов ВКЭ и ВЗН в культуральной среде проводили с использованием в качестве подложки мышинных моноклональных антител 13F6 против ВКЭ и 9E2 против ВЗН (Razumov et al., 2005; Shanshin et al., 2024). Связавшийся антиген находили при помощи моноклональных антител 10Н10 (ВКЭ) и 5Н6 (ВЗН), меченных биотином и стрептавидин-пероксидазным конъюгатом (ICN, США), как описано ранее (Korobitsyn et al., 2021).

**Пробоподготовка.** После обработки бензоназой (Law et al., 2013) тотальную РНК экстрагировали с помощью набора «Реагент Extract RNA» («Евроген», Россия) согласно протоколу производителя. Построение первой цепи ДНК выполняли с использованием набора MMLV RT kit («Евроген») согласно инструкции производителя. Постановку ПЦР осуществляли с помощью «БиоМастер LR HS-ПЦР» («BioLabMix», Россия) с праймерами для выявления РНК-вируса клещевого энцефалита и вируса Западного Нила соответственно (Приложения 1 и 2)<sup>1</sup>. Режим проведения ПЦР (амплификатор С1000, Bio-Rad, США): 94 °С – 10 с, 58 °С – 20 с, 72 °С – 30 с (40 циклов), 72 °С – 7 минут.

**Электрофоретический анализ и выделение фрагментов вирусной ДНК из геля.** Продукты амплификации анализировали в 2 % агарозном геле в буфере TAEх1 (40 мМ Трис, 1 мМ Na<sub>2</sub>ЭДТА). Для выделения продуктов

амплификации из агарозного геля применяли набор *diaGene* («Диаэм», Россия).

**Определение нуклеотидных последовательностей вирусной кДНК.** Определение нуклеотидных последовательностей продуктов амплификации осуществляли при использовании автоматического генетического анализатора ABI 3130xl (Applied Biosystems (США), BigDye Terminator v3.1 Cycle Sequencing Kit) и набора реактивов Applied Biosystems (США) согласно инструкции производителя. Выравнивание нуклеотидных последовательностей выполняли при помощи приложения *Lasergene 7* (DNASTAR).

Синтез первой цепи кДНК для NGS проводили с применением модуля NEBNext® Ultra Directional (NEB). Синтез второй цепи осуществляли при использовании UMI Second Strand Synthesis Module for QuantSeq FWD (Illumina, Lexogen, Австрия). *Cutadapt* (версия 1.18) и *SAMtools* (версия 0.1.18) использовали для удаления адаптеров Illumina и повторного чтения. Контиги были собраны *de novo* с применением ассемблера MIRA (версия 4.9.6).

Нуклеотидные последовательности сравнения были взяты из базы данных GenBank. Множественное выравнивание нуклеотидных последовательностей проводили при помощи приложения *AlignX*, программного пакета *Vector NTI 11* (InforMax, США). Анализ полученных нуклеотидных последовательностей осуществляли с помощью программ *Unipro UGENE v.1.30* и *MEGA 7/10* (Kumar et al., 2018). Филогенетические деревья рассчитаны по методу максимального правдоподобия с использованием 1000 реплик бутстрапов.

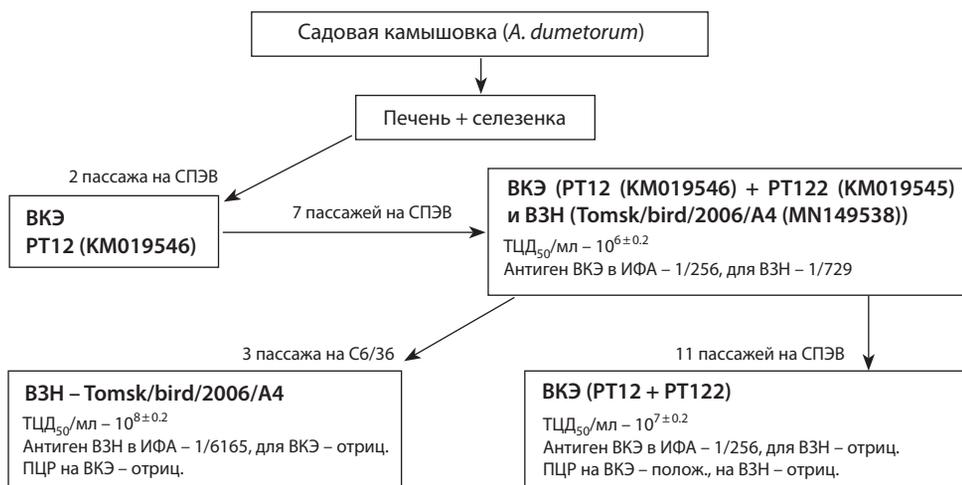
## Результаты

В нашей работе задействован объединенный образец из печени и селезенки садовой камышовки (*A. dumetorum*), отловленной в Томской области в 2006 г. (Mikryukova et al., 2014). На рис. 1 представлена схема выделения вирусных изолятов из данного образца на культурах клеток С6/36 и СПЭВ.

Уже после проведения двух пассажей было зарегистрировано первое появление цитопатогенного эффекта, которое, предположительно, связано с репликацией ВКЭ. Однако после семи дополнительных пассажей в пробе обнаружены антиген и генетический материал ВЗН и ВКЭ. Пассирование инфекционного материала в клетках комара С6/36 позволило получить чистую культуру ВЗН. Дополнительные пассажи этого же материала на клетках СПЭВ фактически элиминировали популяцию ВЗН и сформировали стабильную популяцию вируса клещевого энцефалита.

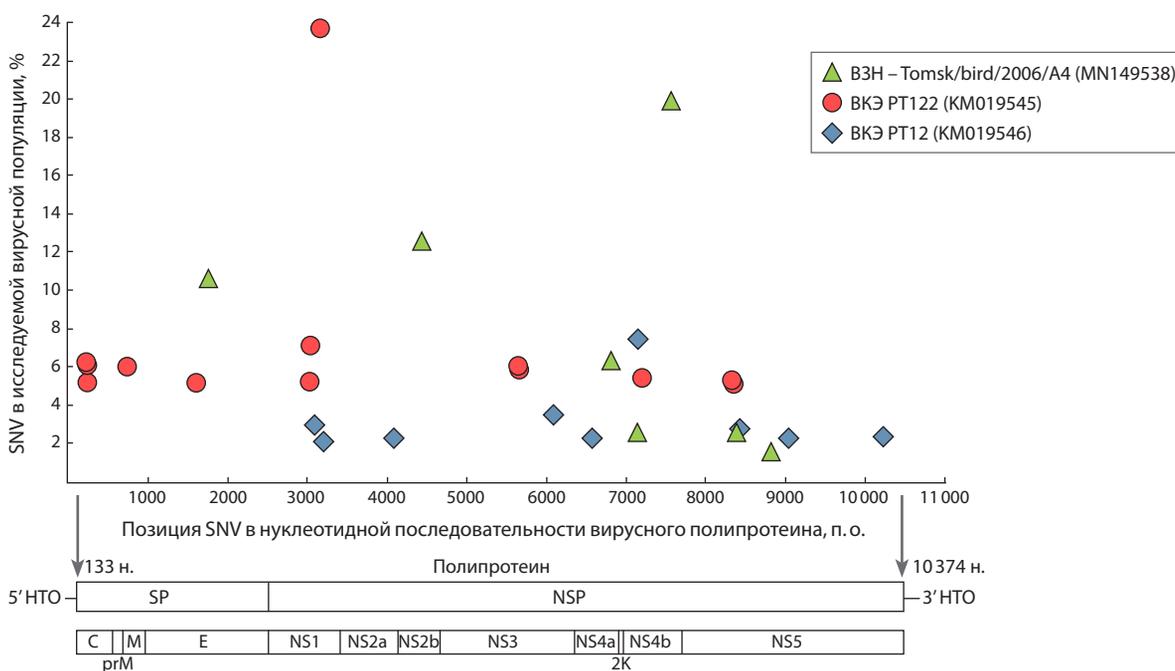
Секвенирование образцов дало возможность определить нуклеотидные последовательности геномов для двух изолятов ВКЭ (РТ12 и РТ122) и одного изолята ВЗН (Tomsk/bird/2006/A4). Филогенетический анализ полногеномных последовательностей показал, что изолят РТ12 относился к дальневосточному генотипу ВКЭ, а изолят РТ122 представлен сибирским генотипом ВКЭ. Изолят Tomsk/bird/2006/A4 был генотипирован как вирус, относящийся к генотипу Ia ВЗН (Приложение 3).

<sup>1</sup> Приложения 1–4 см. по адресу:  
<https://vavilovj-icg.ru/download/pict-2026-30/appx8.pdf>



**Рис. 1.** Схема выделения изолятов ВКЭ и ВЗН из объединенного гомогената селезенки и печени садовой камышовки на культурах клеток.

По мере проведения пассажей на клеточных культурах в каждом отмеченном на схеме блоке выполнялись определение инфекционной активности, наличия вирусного антигена в ИФА с помощью моноклональных антител, проведение ПЦР и секвенирование образца методами Сэнгера и высокопроизводительного секвенирования, как описано в разделе «Материалы и методы».



**Рис. 2.** Картирование однонуклеотидных замещений в геномах ВКЭ и ВЗН методом высокопроизводительного секвенирования.

Анализ набора нуклеотидных последовательностей в результате высокопроизводительного секвенирования позволил выявить наличие SNV в вирусных популяциях. Обнаруженные однонуклеотидные замены, возникающие с частотой 1.75 % и выше, для трех исследованных вирусных популяций представлены на рис. 2.

Полученные результаты показали, что изоляты ВЗН Tomsk/bird/2006/A4 и ВКЭ – PT12 и PT122, выделенные из *A. dumetorum*, представлены гетерогенными вирусными

популяциями, в которых обнаруживаются множественные SNV, возникающие с частотой от 1.75 до 19.88 % для ВЗН и от 2.08 до 23.73 % для ВКЭ (см. таблицу). Выявленные SNP картируются на протяжении всего вирусного генома как в генах структурных вирусных белков, так и в генах неструктурных белков. При этом изолят PT12 дальневосточного генотипа ВКЭ был полностью лишен SNV в области генома, кодирующей структурные белки вириона. Наибольшее количество SNP в этом районе генома обна-

Однонуклеотидные замещения в геномной РНК ВЗН и ВКЭ при ассоциированной инфекции у *A. dumetorum*

№ п/п	Позиция в нуклеотидной последовательности вирусного полипротеина, п.о.	Превалирующий нуклеотид/SNV в исследуемой вирусной популяции	Частота возникновения SNP, %	Прототипные штаммы вирусов клещевого энцефалита и Западного Нила (номера доступа GenBank) с выявленными аналогичными нуклеотидными замещениями в геномах	Ген	Аминокислотная замена
ВКЭ РТ12 (KM019546), дальневосточный генотип						
1	3093	C/T	2.95	JF819648 JN229223 JN003205 и др.	NS1	Y <sub>1031</sub> Y
2	3201	T/C	2.08	KF880803 KU761572 EU816451 и др.		D <sub>1067</sub> D
3	4083	T/C	2.23	JQ825154 EF469661 KY069125 и др.	NS2a	S <sub>1361</sub> S
4	6087	C/T	3.48	JN003205 AF069066 KP716978 и др.	NS3	T <sub>2029</sub> T
5	6572	T/C	2.23	JF819648 JN003205 JN003206	NS4a	M <sub>2191</sub> T
6	7150	G/T	7.41	Нет	NS4b	V <sub>2383</sub> L
7	8430	C/T	2.76	JQ825147 JQ650523 JX534167 и др.	NS5	H <sub>2810</sub> H
8	9038	G/A	2.27	KJ633033 DQ862460 KF880803 и др.		G <sub>3013</sub> E
9	10217	A/C	2.33	Нет		E <sub>3406</sub> A
ВКЭ РТ122 (KM019545), сибирский генотип						
1	218	T/C	6.32	EU816451 KT321430 JN003208 и др.	C	V <sub>73</sub> A
2	231	A/G	6.10	JQ825155 DQ862460 OP902895 и др.		K <sub>77</sub> K
3	234	C/T	5.18	KJ626343 JX498939 JN003205 и др.		I <sub>78</sub> I
4	732	C/T	6.01	JQ825147 FJ402886 JF819648 и др.	preM	N <sub>244</sub> N
5	1605	T/C	5.17	JN003207 EF469661 AF069066 JQ650523 и др.	E	Y <sub>534</sub> Y
6	3027	C/T	5.22	KJ000002 KJ739729 KC414090 FJ572210 и др.	NS1	D <sub>1009</sub> D
7	3031	G/C	7.09	FJ968751		A <sub>1011</sub> P
8	3159	A/C	23.73	U27495 KP716978 KJ922516 KY069125 и др.		T <sub>1053</sub> I
9	5643	C/T	6.05	KJ633033 EU816451 JQ825147 и др.	N NS3	N <sub>1881</sub> N
10	5658	G/A	5.86	KJ922516 KC414090 MF774565 и др.		E <sub>1886</sub> E

**Окончание таблицы**

№ п/п	Позиция в нуклеотидной последовательности вирусного полипротеина, п.о.	Превалирующий нуклеотид/SNV в исследуемой вирусной популяции	Частота возникновения SNP, %	Прототипные штаммы вирусов клещевого энцефалита и Западного Нила (номера доступа GenBank) с выявленными аналогичными нуклеотидными замещениями в геномах	Ген	Аминокислотная замена
<b>БКЭ РТ122 (KM019545), сибирский генотип</b>						
11	7194	C/T	5.42	MF774565 KJ633033	NS4b	V <sub>2498</sub> V
12	8220	C/T	5.20	JN003208 KF826914 JQ654701 и др.	NS5	N <sub>2740</sub> N
13	8238	C/T	5.15	KU761572 JQ825155 KP716978 и др.		N <sub>2746</sub> N
<b>ВЗН Tomsk/bird/2006/A4 (MN149538), генотип 1a</b>						
1	1748	A/G	10.55	KX547363 KX547219 MH507756 и др.	E	E <sub>583</sub> G
2	4434	C/T	12.58	AF196835 EF657887 AF404754 и др.	NS2b	C <sub>1478</sub> C
3	6807	C/T	6.29	AY262283 DQ786572	2k	S <sub>2269</sub> S
	7149	G/T	2.52	AF196835 EF657887 AF404754 и др.	NS4b	V <sub>2383</sub> V
5	7566	C/A	19.88	KJ958922 GQ851607 GQ851608 и др.		D <sub>2522</sub> E
6	8388	G/A	2.55	AY701413 JN858069 GQ851607 и др.	NS5	G <sub>2796</sub> G
7	8907	C/T	1.75	FJ766332		R <sub>2969</sub> R

Примечание. Темно-серым фоном выделены несинонимичные нуклеотидные замены; светло-серым – отсутствие прототипных штаммов с выявленными аналогичными нуклеотидными замещениями в геномах. Обозначения аминокислот указаны в соответствии с общепринятым международным однобуквенным кодом. SNV/SNP – однонуклеотидные замены (single nucleotide variant)/однонуклеотидный полиморфизм (single nucleotide polymorphism). Нуклеотиды: A, T, C, G – аденин, тимин, цитозин и гуанин соответственно.

ружено для сибирского генотипа вируса клещевого энцефалита.

Интересно отметить, что все 29 выявленных SNP в трех изолятах ортофлавивирусов были оригинальными. Это показывает, что обнаруженный спектр SNV характерен и специфичен для исследованных вирусных популяций БКЭ и ВЗН в одном инфицированном хозяине, что косвенно подтверждается ранее полученными данными по несовпадению спектров SNV для вакцинного штамма вируса японского энцефалита и ВЗН (Kaiser et al., 2019). Однако при сборке полных геномов этих вирусов нам не удалось выявить наличие признаков рекомбинации протяженными фрагментами между выделенными изолятами БКЭ и ВЗН, которые были обнаружены в тканях одной дикой птицы. Уникальный паттерн SNV для всех трех изолятов подтверждает отсутствие рекомбинации при одновременной циркуляции трех вирусов в одном хозяине.

Наибольшее количество SNP (13) обнаружено для изолята РТ122 сибирского генотипа БКЭ, наименьшее – для ВЗН (7). Более 60 % SNP картировалось в генах, кодирующих неструктурные вирусные белки. Средняя пред-

ставленность частоты SNP в отношении консенсусной геномной последовательности составила 3.1 % для изолята РТ12 БКЭ (дальневосточный генотип), 7.1 % – для РТ122 БКЭ (сибирский генотип) и 8.0 % – для изолята ВЗН генотипа 1a (см. таблицу).

Примечательно, что для выявленных нами SNP были найдены аналогичные подтвержденные замены в геномах уже известных штаммов вирусов БКЭ и ВЗН. Только SNP<sub>7150</sub> в белке NS4b и SNP<sub>10217</sub>NS5, картированные для дальневосточного генотипа БКЭ, не имели аналогов среди известных геномов БКЭ. Это позволяет предположить, что наличие идентичных нуклеотидных замен в геномах описанных ранее вирусных изолятов свидетельствует о неслучайном характере возникновения SNV и говорит об их потенциальной роли в обеспечении генетической изменчивости этих флавивирусов.

**Обсуждение**

Известно, что изоляты различных РНК-содержащих вирусов представляют собой, как правило, гетерогенную популяцию близкородственных вариантов, которую часто трак-

туют как квазивид (Eigen et al., 1988; Holland et al., 1992; Domingo, Holland, 1997; Domingo et al., 2012; Karbowiak et al., 2016). Вирусная популяция обычно представлена совокупностью большого количества вирусных частиц, ее размер может достигать  $10^{10}$ – $10^{12}$  и более вирионов на инфицированный макроорганизм (Marí Saéz et al., 2015; Diallo et al., 2016; Thorson et al., 2016; Domingo et al., 2021). При этом для инфицирования чувствительного организма бывает достаточно всего нескольких вирусных частиц. Дальнейшая репликация РНК-содержащих вирусов в организме хозяина обеспечивает формирование гетерогенной вирусной популяции.

Гетерогенность квазивида может формироваться в силу целого ряда причин и приводить к появлению новых вирусных вариантов. Наиболее часто изменчивость вирусного РНК генома ассоциируется со значительным уровнем ошибок РНК полимеразы, что приводит к высокому уровню возникновения нуклеотидных замен; отсутствием механизмов коррекции (восстановления) нуклеотидной последовательности вирусной РНК; действием различных селективных факторов на вирусную популяцию при размножении в организме хозяина.

Ранее в городских и пригородных биотопах Томска была обнаружена одновременная циркуляция ВЗН и ВКЭ (Москвитина и др., 2008; Чаусов и др., 2009). Характерной особенностью природных очагов данных инфекций в этом регионе было участие более 42 видов диких птиц (Mikryukova et al., 2014; Moskvitina et al., 2014; Korobitsyn et al., 2021). При этом совместное присутствие в пробе генетических маркеров ВКЭ и ВЗН было зарегистрировано в 1.7 % исследованных проб. Известны также и другие случаи микст-инфекций, вызванных различными генотипами ВКЭ (Безрукова и др., 2008; Ковалев и др., 2008; Козлова и др., 2010; Погодина и др., 2012). Причем частота обнаружения разных подтипов ВКЭ в иксодовом клеще может достигать 4.4–15 %.

Уровень генетических различий между европейским, сибирским и дальневосточными генотипами ВКЭ может быть 18–20 % (Ternovoi et al., 2007). ВКЭ и ВЗН относятся к разным видам ортофлавивирусов и имеют существенно более выраженные отличия геномов, уровень которых может достигать 28–32 % и более. Такие генетические отличия, характерные для этих двух видов ортофлавивирусов, позволили выделить не менее пяти основных генотипов для ВКЭ и девяти генотипов ВЗН (Dai et al., 2018; Kozlova et al., 2018; Simonin, 2024). Как правило, различные генотипы этих вирусов ассоциируются с природными очагами, расположенными в разных географических районах мира.

В последнее время накапливаются многочисленные данные, говорящие о возможности совместной циркуляции и широком распространении различных ортофлавивирусов в новых регионах. Циркуляция ВКЭ и ВЗН в городских и пригородных биотопах Томска подтверждает эту тенденцию, а обнаружение совместной циркуляции нескольких ортофлавивирусов в одной особи садовой камышовки наглядно иллюстрирует особенности распространения этих вирусов на юге Западной Сибири.

Выявление целого набора SNV в популяциях изолятов Tomsk/bird/2006/A4 ВЗН, PT12 ВКЭ и PT122 ВКЭ в тканях одной дикой птицы свидетельствует о выраженной гетерогенности этих вирусных популяций. Возможность возникновения случайных нуклеотидных замен в геномной РНК исключать нельзя. Однако большинство обнаруженных SNP имеет аналогичные нуклеотидные замены в геномах уже известных ВЗН и ВКЭ. Вероятно, существование целого ряда SNP определяет пределы для генетической вариабельности этих флавивирусов и предопределяет возможности генетической изменчивости данных ортофлавивирусов в природных очагах. Картирование оригинальных паттернов SNV, характерных для каждого из трех различных изолятов ортофлавивирусов и циркулирующих одновременно в одной инфицированной особи, говорит о независимости механизмов поддержания гетерогенности вирусной популяции в пределах сибирского и дальневосточного генотипа ВКЭ и генотипа Ia ВЗН. Важно отметить, что нам не удалось выявить наличие признаков новых рекомбинационных событий при сборке и анализе консенсусных геномов этих вирусов (Приложение 4).

Таким образом, полученные результаты показывают существование выраженной гетерогенности популяции геномной РНК флавивирусов, которая сохраняется при коинфекции тремя флавивирусами одной чувствительной особи. По всей вероятности, существование множественных SNV внутри ограниченной вирусной популяции ортофлавивирусов может быть важным механизмом формирования новых геновариантов ВКЭ и ВЗН даже при репликации этих вирусов в тканях инфицированного хозяина.

## Заключение

Оценка возможной популяционной гетерогенности вирусов клещевого энцефалита и Западного Нила методами метагеномного анализа показала наличие множественных SNV в вирусных популяциях всех трех исследованных изолятов ортофлавивирусов, выделенных из тканей одной особи садовой камышовки. Частота обнаружения SNV у исследованных изолятов ВКЭ и ВЗН колебалась от 1.75 до 23.73 %.

Выявленные SNP были ассоциированы как с синонимичными, так и с несинонимичными одиночными нуклеотидными заменами, которые преимущественно локализовались в генах вирусных неструктурных белков. Наибольшее количество (13) SNP обнаружено у PT122 ВКЭ (сибирский генотип). Они характеризовались высокой частотой встречаемости в вирусной популяции, которая достигала 23.73 % для SNP<sub>3159</sub> в белке NS1. У штамма PT12 ВКЭ дальневосточного генотипа наиболее часто встречался SNP<sub>7150</sub>NS4b с аминокислотной заменой V<sub>2383</sub>L, возникающей с частотой 7.41 % и не описанной ранее для других известных штаммов ВКЭ, так же, как и SNP<sub>10271</sub>NS5. Наименьшее количество (7) SNV выявлено у ВЗН, однако SNP<sub>1748</sub>E, SNP<sub>4434</sub>NS3b и SVP<sub>7566</sub>NS4b имели высокую частоту встречаемости, от 10.55 до 19.88 %, в исследованной вирусной популяции.

Всего в исследованных штаммах ВКЭ и ВЗН обнаружено 29 SNP, приводящих к существованию девяти вирусных вариантов с аминокислотными заменами. Совокупность полученных результатов свидетельствует о значимости SNP для обеспечения генетического разнообразия ортофлавириусов.

## Список литературы / References

- Безрукова (Гамова) Е.Г., Погодина В.В., Левина Л.С., Карань Л.С., Маленко Г.В. Изучение штаммов вируса клещевого энцефалита различных генотипов, выделенных от больных с хроническим течением болезни. *Медицина в Кузбассе*. 2008;55:21-28 [Bezrukova (Gamova) E.G., Pogodina V.V., Levina L.S., Karan L.S., Malenko G.V. The research of different genotypes of TBE strain virus isolated from patients with chronic course of the disease. *Medicina v Kuzbasse = Medicine in Kuzbass*. 2008;55:21-28 (in Russian)]
- Ковалев С.Ю., Умпелова Т.В., Снитковская Т.Э., Килячина А.С., Романенко В.В., Кокорев В.С., Глинских Н.П. Молекулярно-эпидемиологическая характеристика вируса клещевого энцефалита на территории Свердловской области на основе генотип-специфической ОТ-ПЦР. *Вопросы вирусологии*. 2008;53(2):27-31 [Kovalev S.Yu., Umpelova T.V., Snitkovskaya T.E., Kilyachina A.S., Romanenko V.V., Kokorev V.S., Glinskikh N.P. Molecular and epidemiological characteristics of tick-borne encephalitis virus in the Sverdlovsk region on the basis of genotype-specific RT-PCR. *Voprosy Virusologii = Problems of Virology*. 2008;53(2):27-31 (in Russian)]
- Козлова И.В., Верхожина М.М., Демина Т.В., Джиоев Ю.П., Дорошенко Е.К., Лисак О.В., Карань Л.С., Колясникова Н.М., Рар В.А., Фоменко Н.В., Ткачев С.Е., Богомазова О.Л., Борисов В.А., Туваков М.К., Злобин В.И. Сочетанные очаги трансмиссивных клещевых инфекций на территории Прибайкалья. *Эпидемиология и вакцинопрофилактика*. 2010;4(53):40-46 [Kozlova I.V., Verkhovina M.M., Demina T.V., Dzhioev Yu.P., Doroshchenko E.K., Lisak O.V., Karan L.S., Koliashnikova N.M., Rar V.A., Fomenko N.V., Tkachev S.E., Bogomazova O.L., Borisov V.A., Tuvakov M.K., Zlobin V.I. Combined natural foci of tick-borne infections in Baikal region. *Epidemiologiya i Vakcinoprofilaktika = Epidemiology and Vaccinal Prevention*. 2010;4(53):40-46 (in Russian)]
- Кононова Ю.В., Терновой В.А., Щелканов М.Ю., Протопопова Е.В., Золотых С.И., Юрлов А.К., Друзяка А.В., Славский А.А., Шестопалов А.М., Львов Д.К., Локтев В.Б. Генотипирование вируса Западного Нила в популяциях диких птиц наземного и древесно-кустарникового комплексов на территориях Барабинской вирус лесостепи и Кулундинской степи (2003-2004 гг.). *Вопросы вирусологии*. 2006;51(4):19-23 [Kononova Yu.V., Ternovoi V.A., Shchelkanov M.Yu., Protopopova E.V., Zolotykh S.I., Yurlov A.K., Druzyaka A.V., Slavskii A.A., Shestopalov A.M., Lvov D.K., Loktev V.B. West Nile virus genotyping among wild birds belonging to ground and tree-brush bird populations on the territories of the Baraba forest-steppe and Kulunda steppe (2003-2004). *Voprosy Virusologii = Problems of Virology*. 2006;51(4):19-23 (in Russian)]
- Москвитина Н.С., Романенко В.Н., Терновой В.А., Иванова Н.В., Протопопова Е.В., Кравченко Л.Б., Кононова Ю.В., Куранова В.Н., Чаусов Е.В., Москвитин С.С., Першикова Н.Л., Гашков С.И., Коновалова С.Н., Большакова Н.П., Локтев В.Б. Выявление вируса Западного Нила и его генотипирование в иксодовых клещах (Acari: Ixodidae) в Томске и его пригородах. *Паразитология*. 2008;42(3):210-225 [Moskvitina N.S., Romanenko V.N., Ternovoi V.A., Ivanova N.V., Protopopova E.V., Kravchenko L.B., Kononova Yu.V., Kuranova V.N., Chausov E.V., Moskvitin S.S., Pershikova N.L., Gashkov S.I., Konovalova S.N., Bolshakova N.P., Loktev V.B. Detection of the West Nile virus and its genetic typing in ixodid ticks (Parasitiformes: Ixodidae) in Tomsk and its suburbs. *Parazitologiya*. 2008;42(3):210-225 (in Russian)]
- Погодина В.В., Карань Л.С., Колясникова Н.М., Герасимов С.Г., Левина Л.С., Бочкова Н.Г., Андаев Е.И., Трухина А.Г., Борисова Т.И., Сидорова Е.А., Нагибина О.А., Маленко Г.В., Безрукова Е.Г. Политиповые штаммы в генофонде вируса клещевого энцефалита. *Вопросы вирусологии*. 2012;57(3):30-37 [Pogodina V.V., Karan L.S., Kolyasnikova N.M., Gerasimov S.G., Levina L.S., Bochkova N.G., Andayev E.I., Trukhina A.G., Borisova T.I., Sidorova E.A., Nagibina O.A., Malenko G.V., Bezrukova E.G. Polytypic strains in the genofund of tick-borne encephalitis virus. *Voprosy Virusologii = Problems of Virology*. 2012;57(3):30-37 (in Russian)]
- Терновой В.А., Щелканов М.Ю., Шестопалов А.М., Аристова В.А., Протопопова Е.В., Громашевский В.Л., Друзяка А.В., Славский А.А., Золотых С.И., Локтев В.Б., Львов Д.К. Выявление вируса Западного Нила у птиц территории Барабинской и Кулундинской низменностей (западносибирский пролетный путь) в лето-осенний период 2002 г. *Вопросы вирусологии*. 2004;49(3):52-56 [Ternovoi V.A., Shchelkanov M.Yu., Shestopalov A.M., Aristova V.A., Protopopova E.V., Gromashevsky V.L., Druzyaka A.V., Slavsky A.A., Zolotykh S.I., Loktev V.B., Lvov D.K. Detection of West Nile virus in birds in the territories of Baraba and Kulunda lowlands (West Siberian migration way) during summer-autumn of 2002. *Voprosy Virusologii = Problems of Virology*. 2004;49(3):52-56 (in Russian)]
- Чаусов Е.В., Терновой В.А., Протопопова Е.В., Коновалова С.Н., Кононова Ю.В., Першикова Н.Л., Москвитина Н.С., Романенко В.Н., Иванова Н.В., Большакова Н.П., Москвитин С.С., Коробицын И.Г., Гашков С.И., Тютенков О.И., Куранова В.Н., Кравченко Л.Б., Сучкова Н.Г., Агулова Л.П., Локтев В.Б. Генетическое разнообразие инфекционных агентов, переносимых иксодовыми клещами в г. Томске и его пригородах. *Паразитология*. 2009;43(5):374-388 [Chausov E.V., Ternovoi V.A., Protopopova E.V., Konovalova S.N., Kononova J.V., Pershikova N.L., Moskvitina N.S., Romanenko V.N., Ivanova N.V., Bolshakova N.P., Moskvitin S.S., Korobitsyn I.G., Gashkov S.I., Tutenkov O.I., Kuranova V.N., Kravchenko L.B., Suchkova N.G., Agulova L.P., Loktev V.B. Genetic diversity of ixodid tick-borne pathogens in Tomsk City and suburbs. *Parazitologiya*. 2009;43(5):374-388 (in Russian)]
- Borda V., da Silva Francisco Junior R., Carvalho J.B., Morais G.L., Duque Rossi A., Pezzuto P., Azevedo G.S., ... Tanuri A., Stratakis C.A., Aguiar R.S., Cardoso C.C., Vasconcelos A.T.R. Whole-exome sequencing reveals insights into genetic susceptibility to Congenital Zika Syndrome. *PLoS Negl Trop Dis*. 2021;15(6):e0009507. doi 10.1371/journal.pntd.0009507
- Current ICTV Taxonomy Release. Taxonomy Browser. Ch. Family: Flaviviridae. Available at: <https://ictv.global/report/chapter/flaviviridae/flaviviridae/orthoflavivirus>
- Dai X., Shang G., Lu S., Yang J., Xu J. A new subtype of eastern tick-borne encephalitis virus discovered in Qinghai-Tibet Plateau, China. *Emerg Microbes Infect*. 2018;7:74. doi 10.1038/s41426-018-0081-6
- Diallo B., Sissoko D., Loman N.J., Bah H.A., Bah H., Worrell M.C., Conde L.S., ... Formenty P., Keita S., Günther S., Rambaut A., Duraffour S. Resurgence of Ebola virus disease in Guinea linked to a survivor with virus persistence in seminal fluid for more than 500 days. *Clin Infect Dis*. 2016;63(10):1353-1356. doi 10.1093/cid/ciw601
- Domingo E., Holland J.J. RNA virus mutations and fitness for survival. *Annu Rev Microbiol*. 1997;51:151-178. doi 10.1146/annurev.micro.51.1.151
- Domingo E., Sheldon J., Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev*. 2012;76(2):159-216. doi 10.1128/mmr.05023-11
- Domingo E., García-Crespo C., Perales C. Historical perspective on the discovery of the quasispecies concept. *Annu Rev Virol*. 2021;8(1):51-72. doi 10.1146/annurev-virology-091919-105900
- Eigen M., McCaskill J., Schuster P. Molecular quasi-species. *J Phys Chem*. 1988;92(24):6881-6891. doi 10.1021/j100335a010

- Holland J.J., De La Torre J.C., Steinhauer D.A. RNA-virus populations as quasispecies. In: Holland J.J. (Ed.) Genetic Diversity of RNA Viruses. Current Topics in Microbiology and Immunology. Vol. 176. Berlin; Heidelberg: Springer, 1992;176:1-20. doi 10.1007/978-3-642-77011-1\_1
- Kaiser J.A., Luo H., Widen S.G., Wood T.G., Huang C.Y., Wang T., Barrett A.D.T. Japanese encephalitis vaccine-specific envelope protein E138K mutation does not attenuate virulence of West Nile virus. *NPJ Vaccines*. 2019;4:50. doi 10.1038/s41541-019-0146-0
- Karbowiak G., Biernat B., Werszko J., Rychlik L. The transstadial persistence of tick-borne encephalitis virus in *Dermacentor reticulatus* ticks in natural conditions. *Acta Parasitol*. 2016;61(1):201-203. doi 10.1515/ap-2016-0028
- Korobitsyn I.G., Moskvitina N.S., Tyutenkov O.Y., Gashkov S.I., Kononova Y.V., Moskvitin S.S., Romanenko V.N., ... Kononova S.N., Tupota N.L., Sementsova A.O., Ternovoi V.A., Loktev V.B. Detection of tick-borne pathogens in wild birds and their ticks in Western Siberia and high level of their mismatch. *Folia Parasitol (Praha)*. 2021;68:024. doi 10.14411/fp.2021.024
- Kozlova I.V., Demina T.V., Tkachev S.E., Doroshchenko E.K., Lisak O.V., Verkhozina M.M., Karan L.S., ... Savinova Y.S., Chernovanova O.O., Ruzek D., Tikunova N.V., Zlobin V.I. Characteristics of the Baikal subtype of tick-borne encephalitis virus circulating in Eastern Siberia. *Acta Biomed Scientifica*. 2018;3(4):53-60. doi 10.29413/ABS.2018-3.4.9
- Kumar S., Stecher G., Li M., Nknyaz C., Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547-1549. doi 10.1093/molbev/msy096
- Law J., Jovel J., Patterson J., Ford G., O'Keefe S., Wang W., Meng B., ... Mitchell T., Jordan T., Carpenter E., Mason A.L., Wong G.K. Identification of hepatotropic viruses from plasma using deep sequencing: a next generation diagnostic tool. *PLoS One*. 2013;8(4):e60595. doi 10.1371/journal.pone.0060595
- Litov A.G., Deviatkin A.A., Goptar I.A., Dedkov V.G., Gmyl A.P., Markelov M.L., Shipulin G.A., Karganova G.G. Evaluation of the population heterogeneity of TBEV laboratory variants using high-throughput sequencing. *J Gen Virol*. 2018;99(2):240-245. doi 10.1099/jgv.0.001003
- Marí Saéz A., Weiss S., Nowak K., Lapeyre V., Zimmermann F., Düx A., Kühl H.S., ... Fahr J., Borchner M., Gogarten J.F., Calvignac-Spencer S., Leendertz F.H. Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Mol Med*. 2015;7(1):17-23. doi 10.15252/emmm.201404792
- Mikryukova T.P., Moskvitina N.S., Kononova Y.V., Korobitsyn I.G., Kartashov M.Y., Tyutenkov O.Y., Protopopova E.V., ... Moskvitin S.S., Tupota N.L., Sementsova A.O., Ternovoi V.A., Loktev V.B. Surveillance of tick-borne encephalitis virus in wild birds and ticks in Tomsk city and its suburbs (Western Siberia). *Ticks Tick Borne Dis*. 2014;5(2):145-151. doi 10.1016/j.ttbdis.2013.10.004
- Moskvitina N.S., Korobitsyn I.G., Tyutenkov O.Y., Gashkov S.I., Kononova Y.V., Moskvitin S.S., Romanenko V.N., ... Kononova S.N., Tupota N.L., Sementsova A.O., Ternovoi V.A., Loktev V.B. The role of birds in the maintenance of tick borne infections in the Tomsk anthropurgic foci. *Biol Bull Russ Acad Sci*. 2014;41(4):387-393. doi 10.1134/S1062359014040086
- Postler T.S., Beer M., Blitvich B.J., Bukh J., de Lamballerie X., Drexler J.F., Imrie A., Kapoor A., Karganova G.G., Lemey P., Lohmann V., Simmonds P., Smith D.B., Stapleton J.T., Kuhn J.H. Renaming of the genus *Flavivirus* to *Orthoflavivirus* and extension of binomial species names within the family *Flaviviridae*. *Arch Virol*. 2023;168(9):224. doi 10.1007/s00705-023-05835-1
- Pustijanac E., Buršić M., Talapko J., Škrlec I., Meštrović T., Lišnjić D. Tick-borne encephalitis virus: a comprehensive review of transmission, pathogenesis, epidemiology, clinical manifestations, diagnosis, and prevention. *Microorganisms*. 2023;11(7):1634. doi 10.3390/microorganisms11071634
- Razumov I.A., Kazachinskaja E.I., Ternovoi V.A., Protopopova E.V., Galkina I.V., Gromashevskii V.L., Prilipov A.G., Kachko A.V., Ivanova A.V., L'vov D.K., Loktev V.B. Neutralizing monoclonal antibodies against Russian strain of the West Nile virus. *Viral Immunol*. 2005;18(3):558-568. doi 10.1089/vim.2005.18.558
- Shanshin D.V., Borisevich S.S., Shaprova O.N., Nesmeyanova V.S., Bondar A.A., Porozov Y.B., Khamitov E.M., Kolosova E.A., Shelemba A.A., Ushkalenko N.D., Protopopova E.V., Sergeev A.A., Loktev V.B., Shcherbakov D.N. Phage display revealed the complex structure of the epitope of the monoclonal antibody 10H10. *Int J Mol Sci*. 2024;25(19):10311. doi 10.3390/ijms251910311
- Simonin Y. Circulation of West Nile virus and Usutu virus in Europe: overview and challenges. *Viruses*. 2024;16(4):599. doi 10.3390/v16040599
- Singh P., Khatib M.N., Ballal S., Kaur M., Nathiya D., Sharma S., Siva Prasad G.V., ... Lakhanpal S., Shabil M., Bushi G., Sah S., Serhan H.A. West Nile virus in a changing climate: epidemiology, pathology, advances in diagnosis and treatment, vaccine designing and control strategies, emerging public health challenges – a comprehensive review. *Emerg Microbes Infect*. 2024;30:2437244. doi 10.1080/22221751.2024.2437244
- Svyatchenko V.A., Nikonov S.D., Mayorov A.P., Gelfond M.L., Loktev V.B. Antiviral photodynamic therapy: inactivation and inhibition of SARS-CoV-2 *in vitro* using methylene blue and Radachlorin. *Photodiagnosis Photodyn Ther*. 2021;33:102-112. doi 10.1016/j.pdpdt.2020.102112
- Ternovoi V.A., Protopopova E.V., Chausov E.V., Novikov D.V., Leonova G.N., Netesov S.V., Loktev V.B. Novel variant of tickborne encephalitis virus, Russia. *Emerg Infect Dis*. 2007;13(10):1574-1578. doi 10.3201/eid1310.070158
- Ternovoi V.A., Ponomareva E.P., Protopopova E.V., Tupota N.L., Mikryukova T.P., Loktev V.B. Changes in the genome of the Tick-Borne encephalitis virus during cultivation. *Mol Biol*. 2024;58(2):266-278. doi 10.1134/S0026893324020146
- Thorson A., Formenty P., Lofthouse C., Broutet N. Systematic review of the literature on viral persistence and sexual transmission from recovered Ebola survivors: evidence and recommendations. *BMJ Open*. 2016;6:e008859. doi 10.1136/bmjopen-2015-008859
- Worku D.A. Tick-borne encephalitis (TBE): from tick to pathology. *J Clin Med*. 2023;12(21):6859. doi 10.3390/jcm12216859
- Zárate S., Hernández-Pérez F., Taboada B., Martínez N.E., Alcaráz-Estrada S.L., Del Moral O., Yocupicio-Monroy M. Complete genome of DENV2 isolated from mosquitoes in Mexico. *Infect Genet Evol*. 2019;71:98-107. doi 10.1016/j.meegid.2019.03.018

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 28.01.2025. После доработки 23.05.2025. Принята к публикации 30.05.2025.

doi 10.18699/vjgb-26-13

## Новые аргументы в дискуссии о природе пикобирнавирусов

А.Ю. Кашников  , Н.В. Епифанова , Н.А. Новикова 

Нижегородский научно-исследовательский институт эпидемиологии и микробиологии им. академика И.Н. Блохиной Роспотребнадзора, Нижний Новгород, Россия

 a.kashn@yandex.ru

**Аннотация.** Пикобирнавирусы (ПБВ) из семейства *Picobirnaviridae* находят у самых разных хозяев – высших и низших эукариот, а также в грибах и бактериях. Однако среди ученых в отношении ПБВ на сегодняшний день нет однозначного понимания, кто является их истинным хозяином. Принадлежность ПБВ к вирусам высших эукариот не доказана, поскольку не подобраны ни культура животных клеток для их размножения, ни животное-гнотобионт. В связи с обнаружением прокариотических участков (мотивов) в сегментах генома ПБВ было высказано предположение о прокариотической природе их хозяев. Однако и это открытие не закрепило одного конкретного хозяина за ПБВ, так как затем были обнаружены ПБВ-подобные геномы, не характерные для изученных штаммов ПБВ, – с митохондриальным генетическим кодом, свойственным низшим эукариотам (плесени и беспозвоночным). А недавно появилась новая версия происхождения ПБВ от вирусов позвоночных и грибов, отрицающая их фаговую природу. Для понимания природы генетически разнородных штаммов ПБВ, обнаруженных у разных организмов, исследователи руководствовались информацией о присутствии специфических для вирусного семейства мотивов в геноме, используемом генетическом коде и способе распространения. Существующая в настоящее время информация вселяет уверенность в скором завершении продолжающейся дискуссии о возможных хозяевах ПБВ. В частности, недавно появилась гипотеза, демонстрирующая вероятный механизм замены генетического кода у РНК-вирусов, которая позволяет объяснить происхождение форм ПБВ с митохондриальным генетическим кодом, способных к репродукции в клетках низших эукариот на примере фагов. Однако еще не представлена эволюционно детерминированная модель, демонстрирующая путь формирования ПБВ с генетическим кодом клеток плесени и беспозвоночных. Этот эволюционный путь в представлении авторов данного обзора обусловлен эндосимбиотическими отношениями между предполагаемыми хозяевами ПБВ, способствующими горизонтальному распространению вируса. Цель нашей статьи – попытка описания возможного пути формирования из предковой формы ПБВ ее производных эволюционных форм, одни из которых унаследовали геном с прокариотическим мотивом и стандартным генетическим кодом, а другие обрели нестандартную форму генома с кодом низших эукариот. В статье делается акцент на ведущей роли горизонтальной передачи в формировании нестандартных промежуточных форм пикобирнавирусов.

**Ключевые слова:** пикобирнавирус; сегмент генома; клетка-хозяин; митохондриальный генетический код; реассортация

**Для цитирования:** Кашников А.Ю., Епифанова Н.В., Новикова Н.А. Новые аргументы в дискуссии о природе пикобирнавирусов. *Вавиловский журнал генетики и селекции*. 2026;30(1):126-135. doi 10.18699/vjgb-26-13

## New arguments in the discussion about the nature of picobirnaviruses

A.Yu. Kashnikov  , N.V. Epifanova , N.A. Novikova 

I.N. Blokhina Nizhny Novgorod Research Institute of Epidemiology and Microbiology, Nizhny Novgorod, Russia

 a.kashn@yandex.ru

**Abstract.** Picobirnaviruses (PBVs), members of the *Picobirnaviridae* family, are found in a wide range of hosts, including eukaryotes (both higher and lower), fungi, and bacteria. However, scientists are unsure about their “true master” or primary host. While often found in animals, including cases of gastroenteritis, they are also detected in environmental samples and have shown genetic links to bacterial and fungal viruses. The lack of a reliable cell culture or animal model for PBV propagation further complicates determining their host specificity. Due to the discovery of prokaryotic regions (motifs) in segments of the PBV genome, it was suggested that their hosts are prokaryotic. However, even this discovery did not pin one specific host to PBVs; since then PBV-like genomes not characteristic of the studied PBV strains, with a mitochondrial genetic code characteristic of lower eukaryotes (molds and invertebrates), were discovered. And recently, a new version of the origin of PBVs from vertebrate viruses and fungi has appeared, denying their phage nature. To understand the nature of genetically diverse PBV strains detected in different organisms, researchers were guided by information about the presence of motifs specific to the viral family in the genome, the genetic code used, and the

method of distribution. Recent research suggests that PBVs, previously thought to have a vertebrate origin, may have also evolved from fungal sources denying their phage nature. Some PBV-like sequences have been found to utilize the fungal mitochondrial genetic code, indicating a possible fungal origin or a close relationship with fungal viruses like mitoviruses. This discovery challenges the previously held view of PBVs as exclusively vertebrate viruses and suggests a more complex evolutionary history. The information available today inspires confidence in the imminent conclusion of the ongoing discussion about the possible PBV hosts. In particular, a hypothesis has recently emerged demonstrating a possible mechanism for the replacement of the genetic code in RNA viruses, which makes it possible to explain the origin of PBV forms with the mitochondrial genetic code capable of reproduction in cells of lower eukaryotes using the example of phages. However, an evolutionarily deterministic model demonstrating the path of PBV formation with the genetic code of mold and invertebrate cells has not yet been presented. According to the authors of this review, this evolutionary path is due to the endosymbiotic relationships between the putative PBV hosts, contributing to the horizontal virus spread. The purpose of this review article is to attempt to describe a possible path of formation from the ancestral PBV form and its derived evolutionary forms, some of which inherited a genome with a prokaryotic motif and a standard genetic code, while others acquired a non-standard form of the genome with the code of lower eukaryotes. This review article focuses on the leading role of horizontal transmission in the formation of non-standard intermediate PBV forms.

**Key words:** picobirnavirus; genome segment; host cell; mitochondrial genetic code; reassortment

**For citation:** Kashnikov A.Yu., Epifanova N.V., Novikova N.A. New arguments in the discussion about the nature of picobirnaviruses. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(1):126-135. doi 10.18699/vjgb-26-13

## Дискуссия о природе пикобирнавирусов

*Picobirnavirus* (пикобирнавирус, ПБВ) – единственный род в семействе *Picobirnaviridae*, относящийся к классу *Duplopiviricetes*, типу *Pisuviricota* (Reddy et al., 2023). Частицы ПБВ диаметром 33–37 нм с однослойной белковой оболочкой имеют икосаэдрическую симметрию. Липопротеиновая оболочка отсутствует. Геном состоит из двух сегментов двуцепочечной дцРНК (Delmas et al., 2019). Сегмент 1 (более крупный, размером 2.4–2.6 т. п. н.) кодирует два белка – капсидный белок и белок, функция которого до сих пор неизвестна. Сегмент 2 (меньший, размером 1.5–1.9 т. п. н.) кодирует фермент РНК-зависимую РНК-полимеразу (RdRp семейства Pfam RdRp\_1), по первичной структуре которого ПБВ делят на геногруппы. Основными, наиболее распространенными геногруппами являются I (GI) и II (GII) (Malik et al., 2014; Reddy et al., 2023).

Первоначально ПБВ были обнаружены в содержимом кишечника и в дыхательных путях высших эукариот (Pereira et al., 1988; Новикова и др., 2003; Delmas et al., 2019; Kumar et al., 2020; Ghosh, Malik, 2021). При симптоматических инфекциях ПБВ часто обнаруживали в кишечнике животных и человека в ассоциации с вирусами, патогенность которых установлена. Однако их также находили в стуле животных без симптомов инфекции и у здоровых животных. По этой причине ПБВ традиционно считались условно-патогенными кишечными вирусами млекопитающих и птиц (Shi et al., 2016; Delmas et al., 2019; Kumar et al., 2020). Но поскольку успешного размножения ПБВ в культурах клеток млекопитающих или животных-гнотобионтов добиться не удалось, у исследователей появилось сомнение о связи ПБВ, присутствующих в кишечнике, с заболеванием животных (Ghosh, Malik, 2021; Sadiq et al., 2024). ПБВ находили у беспозвоночных (моллюсков, членистоногих, насекомых) (Shi et al., 2016), а более поздние исследования показали, что эти вирусы, вероятно, могут инфицировать прокариотические или грибковые клетки-хозяева (Adriaenssens et al., 2018; Boros

et al., 2018; Krishnamurthy, Wang, 2018; Yinda et al., 2018; Kleymann et al., 2020).

В настоящее время обсуждаются три альтернативные версии о природе хозяев ПБВ. Согласно первой версии, хозяева ПБВ – клетки высших эукариот (животных). По второй версии, ПБВ могут быть вирусами прокариот. Такое предположение связано с обнаружением в геноме ПБВ участков (мотивов), характерных для вирусов прокариот. Согласно этой версии, поскольку ПБВ-подобные нуклеотидные последовательности часто обнаруживают в фекалиях разных животных, они могут иметь отношение не к клеткам этих животных, а к присутствию в микробиоме кишечника соответствующих бактерий. Однако в дальнейшем из микробиома кишечника животных были выделены атипичные ПБВ-подобные геномы с прокариотическими мотивами, но с не свойственным прокариотам митохондриальным генетическим кодом, характерным для низших эукариот – плесневых грибов и беспозвоночных (Yinda et al., 2018; Kleymann et al., 2020). Так возникла третья версия: кроме клеток прокариот, некоторые ПБВ, выделенные из содержимого кишечника животных, могут инфицировать митохондрии грибов или беспозвоночных (Shi et al., 2016; Ghosh, Malik, 2021).

По дискутируемому вопросу до сих пор нет однозначного ответа – какой организм (организмы) является хозяином ПБВ. Хотя на протяжении ряда лет альтернативные точки зрения о природе ПБВ пополнялись новыми аргументами в свою поддержку. В частности, в последнее время в пользу гипотезы, отрицающей фаговую природу ПБВ, возникла новая версия о параллельной эволюции ПБВ от двух разных предков, которые были паразитами грибов и позвоночных (Perez et al., 2023).

## Версия, согласно которой предками ПБВ являются вирусы позвоночных и грибов

Эта версия строится на наличии иммунного ответа у позвоночного хозяина при обнаружении у него ПБВ со стандартным генетическим кодом и отсутствии иммун-

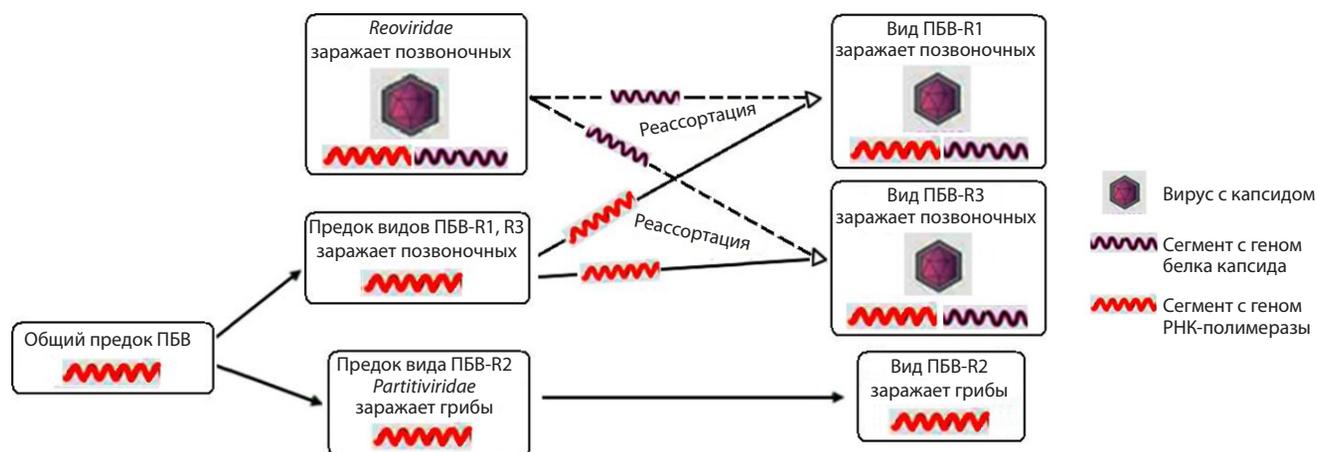


Рис. 1. Первая модель эволюционного развития семейства ПБВ по версии L.J. Perez с коллегами (2023).

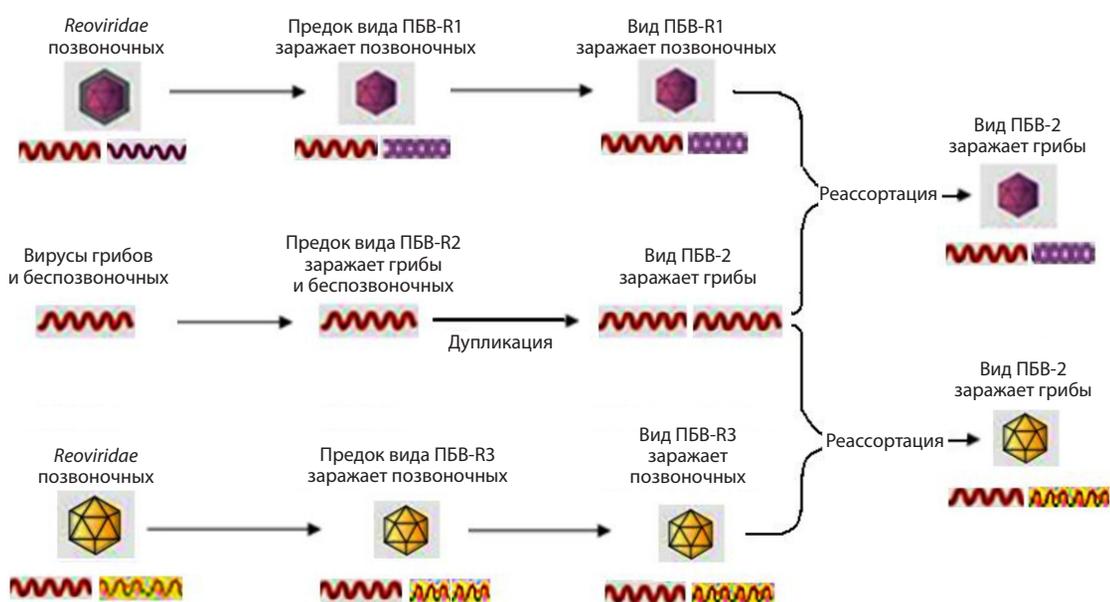


Рис. 2. Вторая модель эволюционного развития семейства ПБВ по версии L.J. Perez с коллегами (2023).

Бордовым и желтым цветом обозначены сегменты с генами белка капсида; красным – сегменты с генами РНК-полимеразы.

ного ответа при выявлении ПБВ с альтернативным генетическим кодом гриба. Версия предусматривает две разные модели эволюционного развития семейства ПБВ, которое сложилось из трех разных предковых линий. Согласно версии L.J. Perez с коллегами (2023), предковые линии ПБВ (ПБВ-R1 и ПБВ-R3) происходят от реовирусов позвоночных (поскольку создают иммунитет у позвоночного, в кишечнике которого их находят), а предковая линия ПБВ-R2 происходит от грибковых партитивирусов (иммунитет у животного, у которого они были найдены, не наблюдается).

В соответствии с первой моделью эволюции сначала общий предок ПБВ, содержащий только ген RdRp, разделился на два вида. Один из них в дальнейшем разделился еще на два вида с одновременным эволюционным приобретением обоими видами капсида в результате события

реассортации сегментов их геномов с сегментами генома их общего предка, относящегося к семейству *Reoviridae*. В результате два капсидированных предковых вида ПБВ расположились на одной ветви филогенетического дерева. Вместе с отделившимся ранее видом ПБВ-R2 при дальнейшей диверсификации образовалось три параллельно эволюционирующих предковых вида (линии) ПБВ (рис. 1).

Согласно второй модели эволюции, три ПБВ-подобных предковых вида, содержащих ген RdRp, произошли от разных предков, при этом два вида обладали капсидом с момента своего появления и, как в первой модели, сформировали общую ветвь, предположительно, эволюционируя от реовирусов позвоночных (рис. 2).

Третий грибковый вид (ПБВ-R2), изначально лишенный капсида, приобрел сегмент с геном капсида позже в результате дупликации и реассортации сегментов генома

с одним из двух капсидированных видов ПБВ, инфицирующих позвоночных, по механизму, описанному в работе (Luo et al., 2018). Этот предковый вид сначала напоминал митовирусы с митохондриальным кодом, аналогично ПБВ-подобным штаммам, обнаруженным С.К. Yinda с коллегами (2018) и А. Kleumann с коллегами (2020), и в дальнейшем расположился на одной ветви с семейством *Partitiviridae*, заражающим нехордовых эукариот (грибов и беспозвоночных).

Параллельная эволюция этих трех предков ПБВ в сходных условиях существования могла привести к формированию похожих и в то же время генетически диверсифицированных представителей семейства *Picobirnaviridae* (Perez et al., 2023). Как и гипотеза о фаговой природе ПБВ, данное предположение имеет право на существование. Но и сторонники гипотезы, отстаивающей фаговую природу ПБВ, также приводят новые аргументы в пользу своей точки зрения.

### Точка зрения, согласно которой ПБВ могут инфицировать клетки прокариот

Сторонники гипотезы, предполагающей, что хозяевами ПБВ могут быть клетки прокариот, в поддержку фаговой природы ПБВ приводят следующие аргументы.

1. Геном ПБВ обогащен (по сравнению с вирусами эукариот) прокариотическими мотивами – последовательностями Шайна–Дальгарно (Shine–Dalgarno sequence или SD – гексамер AGGAGG), подобно геномам бактериофагов семейства *Cystoviridae* (Ghosh, Malik, 2021), у которых этот мотив встречается даже с меньшей частотой (Krishnamurthy, Wang, 2018). Например, в недавнем исследовании S. Sadiq с коллегами (2024) показано, что геном ПБВ был обогащен мотивом SD в 83–85 % ORF (open reading frame – открытые рамки считывания) в геномах *Picobirnavirus*, что превышает обогащение этим мотивом геномов РНК-бактериофагов *Leviviricetes* и *Cystoviridae*.
2. ПБВ обнаруживаются почти исключительно в фекалиях животных и до сих пор не могут быть культивированы в каких-либо линиях эукариотических клеток. Это также может означать, что ПБВ на самом деле являются вирусами бактерий, входящих в состав микрофлоры кишечника животных или компонентов их пищи (Adriaenssens et al., 2018; Delmas et al., 2019; Bell et al., 2020; Guajardo-Leiva et al., 2020; Ghosh, Malik, 2021; Knox et al., 2023; Sadiq et al., 2024). Такое предположение согласуется с тем, что их ближайших родственников – представителей семейства *Partitiviridae*, инфицирующих растения, грибы и простейших (Vainio et al., 2018), также обнаруживают в микрофлоре кишечника животных (Chen et al., 2021) и в микробиоме беспозвоночных (Shi et al., 2016; Le Lay et al., 2020).
3. Расположение сегментов генома у типичных ПБВ наблюдается в одной частице, как у вирусов бактерий, в то время как у большинства вирусов грибов с сегментированным геномом сегменты инкапсулированы отдельно (Luque et al., 2018).

4. Предположение противников фаговой природы ПБВ (Perez et al., 2023), что вирусы, формирующие иммунитет у инфицированных животных (или человека), должны относиться к эукариотическим вирусам, не может однозначно свидетельствовать в пользу версии эукариотической природы хозяев ПБВ, поскольку установлено, что иммунные ответы хозяина могут возникать и против бактериальных вирусов (Dabrowska et al., 2005; Górski et al., 2006). Следовательно, ПБВ могут вызывать иммунный ответ на заражение не клеток самого животного, а бактериальных клеток, составляющих его микробиом (Ghosh, Malik, 2021).
5. Идентификация уникального для вируса белка – ключевой шаг по определению его принадлежности к конкретному хозяину, поскольку присутствие бактериолизитических свойств у его владельца может убедительно свидетельствовать, что данный вирус является бактериофагом (Кашников и др., 2023; Gan, Wang, 2023). Как выяснилось, в капсиде у ПБВ присутствует белок с лизирующей функцией, который лизирует *Escherichia coli*, а такие белки кодируются только генами двух известных семейств РНК-фагов, *Leviviridae* и *Cystoviridae* (Cai et al., 2021). Однако, возможно, не все, а только некоторые ПБВ кодируют бактериолизины (Gan, Wang, 2023).
6. В подтверждение предположения о фаговой природе ПБВ можно привести и недавнее исследование S. Sadiq с коллегами (2024), в котором на основе проведенной кластеризации выявлено семь кластеров ПБВ в микробиоме, хозяевами которых могли быть только клетки разных микробов, поскольку филогенетической группировки по принадлежности ПБВ к животным-хозяевам при высокой генетической вариабельности ПБВ не наблюдалось. Отсутствие группировки между вирусом и хозяином свидетельствует о филогенетическом несоответствии между ними (Duraisamy et al., 2018; Woo et al., 2019; Mahar et al., 2020). На основании этого S. Sadiq с коллегами (2024) выдвинуто предположение, что ПБВ инфицируют различные микробные организмы, имеющие связь с позвоночными и беспозвоночными через их рацион питания и среду обитания.

Руководствуясь этими наблюдениями, сторонники фаговой природы ПБВ считают, что широкое разнообразие ПБВ отсутствовало бы, если их хозяевами являлись бы млекопитающие, птицы и беспозвоночные (Sadiq et al., 2024). Они предполагают, что кластеризация, связанная с бактериями, подтверждается высокой скоростью распространения ПБВ среди животных (даже более высокой, чем в любом другом семействе РНК-вирусов). Такая скорость распространения более соответствует принадлежности ПБВ к прокариотическим хозяевам с их обширной межвидовой передачей. Кроме того, сторонники фаговой природы ПБВ считают, что сохранение прокариотических SD-последовательностей невозможно у вирусов, инфицирующих грибы (Ullah et al., 2022; Wang, 2022). Данный аргумент подкрепляется вероятной способностью ПБВ подобно некоторым фагам изменять свой генетический код.

### Смена генетического кода как тенденция в эволюции фагов

В геномах некоторых микроорганизмов возможна замена стандартного генетического кода альтернативным (митохондриальным). Эта замена, по мнению Y. Shulgina и S.R. Eddy (2021), связана со снижением в ходе эволюции содержания GC-оснований за счет редукции генома у микроорганизмов, которая может быть обусловлена их паразитическим способом существования – эндосимбиозом (McCutcheon, Moran, 2011). В частности, уменьшение количества GC-оснований в геноме у организмов ведет к снижению доли стоп-кодона TGA (с повышением вероятности переназначения этого кодона) (Korkmaz et al., 2014).

Открытие альтернативных генетических кодов в геноме митохондрий клеток плесневых грибов и беспозвоночных у бактерий и архей демонстрирует способность генетического кода эволюционировать (Shackelton, Holmes, 2008; Kollmar, Mühlhausen, 2017; Shulgina, Eddy, 2021). Смена генетического кода альтернативным у митохондрий эукариотических клеток у бактерий и архей, предположительно, могла быть связана с их стремлением спастись от бактериофагов, которые изначально пользовались тем же кодом, что и бактерии (Bender et al., 2008).

Бактериофаги также способны к смене своего генетического кода, когда возникает необходимость обмануть защитные противовирусные системы бактерии-хозяина. Например, установлено, что генетический код в геноме некоторых фагов, выделенных из фекалий павианов, соответствует «альтернативному» генетическому коду геномов бактерий *Bacilli* (Al-Shayeb et al., 2020). В исследовании N. Yutin с коллегами (2021) сообщается о замене (перекодировании) стоп-кодонов (TAG) у ДНК-бактериофагов *stAssphage* при трансляции на кодон, кодирующий аминокислоту глутамин. Переназначение стандартного стоп-кодона TGA на альтернативный, кодирующий аминокислоту триптофан, соответствующий коду митохондрий грибов и беспозвоночных, наблюдали в некоторых ПБВ-подобных геномах С.К. Yinda с коллегами (2018), впервые обнаружившие это явление у пикобирнавирусов.

Согласно D. Wang (2022), тенденция к смене генетического кода ПБВ, как и у бактериофагов, вероятно, связана со способностью считывать или перекодировать стоп-кодон TGA при трансляции, независимо от таксономической принадлежности клеток, которые они заражают (поскольку фаги, вероятно, нацелены на митохондрии в эукариотических клетках, учитывая их бактериальное происхождение). Возможно, некоторые штаммы ПБВ, подобно бактериофагам, оказавшись в клетке гриба, способны захватывать кодируемую хозяином супрессорную транспортную РНК (тРНК) или каким-то образом нарушать механизм трансляции хозяина (Wang, 2022).

По информации S.L. Peters с коллегами (2022), некоторые фаги для трансляции своих белков, вероятно, используют бактериальные рибосомы с применением как стандартного, так и альтернативного кода. При этом переназначение нормальных стоп-кодонов TAG и TGA для трансляции на глутамин и триптофан особенно распро-

странено в фагах, инфицирующих грамположительные бактерии типов *Firmicutes* и *Bacteroidetes* (Peters et al., 2022). У *Firmicutes*, известных своим низким содержанием GC в геноме (меньше 50 %), связанным, вероятно, с их паразитическим способом существования – «эндосимбиозом», часто наблюдается смена стандартного кода альтернативным. Ранее отмечалось, что *Firmicutes* гипотетически более всего подходят в качестве хозяев для ПБВ, которые, предположительно, могут быть фагами (Krishnamurthy, Wang, 2018). Возможно, у атипичных ПБВ-штаммов с «альтернативным» генетическим кодом гриба (Yinda et al., 2018; Kleymann et al., 2020), подобно фагам, произошло перекодирование при трансляции (Peters et al., 2022) вслед за своим хозяином *Clostridia Firmicutes*, паразитирующим в клетке гриба.

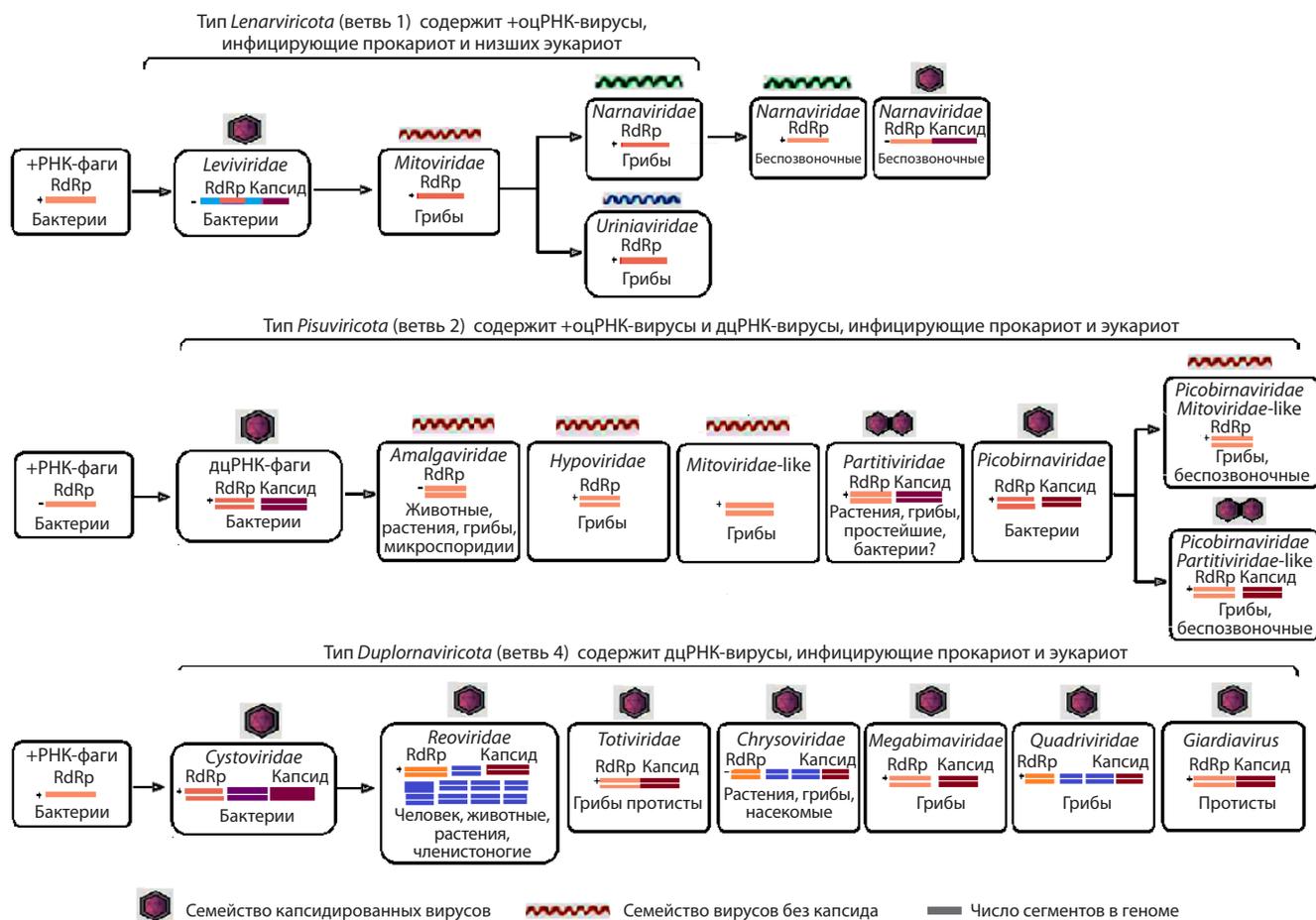
Таким образом, остается вероятность того, что ПБВ, использующие грибной митохондриальный код, являются бактериофагами, у которых смена кода обусловлена эволюцией, связанной с хозяином-прокариотом. Более того, E.V. Koonin с коллегами (2020) предполагают, что ПБВ, считавшиеся ранее вирусами животных, – исключительно бактериальные вирусы, в отличие от филогенетически близкого к ним семейства *Partitiviridae*, включающего эукариотические и, по последним данным, бактериальные вирусы, поскольку у них, как у ПБВ, были обнаружены бактериолизины. По мнению U. Neri с коллегами (2022), семейство *Picobirnaviridae* может представлять собой третью кладу РНК-бактериофагов вместе с семействами *Leviviridae* и *Cystoviridae*.

### Могут ли клетки плесневых грибов и беспозвоночных быть хозяевами ПБВ?

Некоторые аргументы сторонников фаговой природы ПБВ ставят под сомнение существование форм ПБВ, способных инфицировать клетки низших эукариот. Однако факт выявления нетипичных форм ПБВ, генетический код которых соответствует вирусам плесневых грибов и беспозвоночных, позволяет предположить, что не все ПБВ кодируют бактериолизины и, вероятно, у некоторых форм ПБВ, подобно *Mitoviridae* или *Partitiviridae*, вместо клеток бактерий хозяевами могут быть клетки низших эукариот (Luo et al., 2018; Shi et al., 2018; Yinda et al., 2018; Ghosh, Malik, 2021; Ullah et al., 2022; Reddy et al., 2023). Предпосылками, указывающими на такую возможность, являются: происхождение РНК-вирусов эукариот от +РНК-фагов (Wolf et al., 2018) и их предрасположенность к горизонтальной передаче (Son et al., 2015; Dolja, Koonin, 2018).

### Происхождение РНК-вирусов эукариот от +РНК-фагов как свидетельство способности вирусов к изменению своей таксономической природы

Имеющиеся на сегодняшний день результаты филогенетических исследований свидетельствуют о существовании родственной связи между семействами РНК-вирусов эукариот и прокариот и их общем происхождении (рис. 3) (Dolja, Koonin, 2018; Wolf et al., 2018). В частности, согласно E.V. Koonin с коллегами (2015), бактериофаги семейства *Cystoviridae* эволюционно связаны с семейством



**Рис. 3.** Модель эволюционного развития вирусных семейств, ведущих начало от РНК-вирусов прокариот, согласно версии, поддерживаемой исследователями: E.V. Koonin с коллегами (2015); Y.I. Wolf с коллегами (2018); S. Ghosh и Y.S. Malik (2021); S. Sadiq с коллегами (2022).

*Reoviridae* (вирусов эукариот) и могут быть прямыми предками семейства *Picobirnaviridae*.

На основании анализа РНК-зависимой РНК-полимеразы доказано родство между одноцепочечными РНК-бактериофагами семейства *Leviviridae* из класса *Leviricetes* и митовирусами (Wang, 2022). По предположению исследователей, семейства РНК-вирусов типа *Lenarviricota* (*Mitoviridae*, *Narnaviridae*, *Urniaviridae*), эволюционно представляют собой переходные формы от РНК-бактериофагов к ранним РНК-вирусам, инфицирующим низших эукариот (Wolf et al., 2018; Sadiq et al., 2022). Как считают исследователи, эволюционный переход вирусов прокариот к формам РНК-вирусов, заражающих эукариот, произошел примерно 1.45 млрд лет назад вследствие перехода  $\alpha$ -протеобактерий к эндосимбиотическому способу существования. При этом формы, утратившие капсид вследствие данного эволюционного перехода, такие, как *Mitoviridae*, перешли к репродукции в митохондриях, а в дальнейшем, подобно *Narnaviridae*, вышли в цитозоль клетки (Dolja, Koonin, 2018; Wolf et al., 2018).

Y.I. Wolf с коллегами (2018) считают, что разные группы РНК-вирусов могли эволюционировать от прокариотических +РНК-вирусов независимо, поскольку гены RdR

у них обладают большим сходством с генами RdR предковых +РНК-вирусов, чем с генами других РНК-вирусов из параллельных ветвей филогенетического дерева.

### Роль способа распространения РНК-вирусов в формировании их таксономической принадлежности

Данные филогенетических исследований показали, что РНК-вирусы, принадлежащие к одному семейству, могут поражать хозяев из разных таксонов, включая грибы, растения, животных и простейших (Son et al., 2015). Причем подавляющее большинство семейств РНК-вирусов заражают одноклеточных эукариот, используя соответствующие этим организмам альтернативные генетические коды (Neri et al., 2022). Эти исследования согласуются с «гипотезой древней коэволюции» (Pearson et al., 2009), которая утверждает, что вирусы могут переходить от хозяина, относящегося к одной таксономической категории, к хозяину из другой таксономической категории. Подобная связь с организмами из разных таксонов наблюдается, например, у РНК-вирусов семейств *Partitiviridae*, *Reoviridae*, *Amalgaviridae*, *Totiviridae*, *Chrysoviridae*. Члены РНК-вирусов типов *Lenarviricota*, *Duplornaviricota* и *Pisuviricota* (к *Pisuviricota* относят ПБВ) очень разнообразны, распростра-

нены практически в любой среде и связаны с широким спектром хозяев, включая бактерий, простейших, грибы и растения (Dolja, Koonin, 2018; Sadiq et al., 2022).

Согласно «гипотезе древней коэволюции», бескапсидные митовирусоподобные формы ПБВ, подобно митовирусам, могут представлять собой переходную эволюционную форму от предкового +РНК-вируса прокариот к простейшим вирусам эукариот (Wolf et al., 2018). Унаследовав от своего предка, прокариотического вируса, ген RdRp с прокариотическим мотивом и специфическими мотивами (DFXKFD, SGSGGT и GDD), эта форма при переходе к новому хозяину – грибу могла приобрести его генетический код, который ее первоначальный хозяин – прокариот мог заимствовать, находясь в эндосимбиотических отношениях с грибом.

Эндосимбиоз (симбиогенез) – разновидность симбиоза – когда один из партнеров живет внутри клетки другого. При эндосимбиозе более крупного из партнеров обычно называют хозяином, а партнера, организма, живущего внутри своего хозяина, – паразитом в том случае, если он конкурентно влияет на своего хозяина, подавляя его репродукцию. Во многих ассоциациях микроорганизмы-паразиты переходят к постоянному внутриклеточному существованию и передаются по наследству.

Потеря капсида переходной формой ПБВ (подобной *Mitoviridae*) требовала репродукции внутри митохондрий или вакуолей (как у *Narnaviridae*), чтобы избежать избирательного уничтожения системой защиты клетки ее двуцепочечного генома, соответствующего репликативной форме генома *Mitoviridae*, формируемой на промежуточной стадии репликации. В данном случае капсидированную ПБВ-подобную форму P16-366 (Yinda et al., 2018) можно рассматривать как эволюционно более продвинутую переходную форму, поскольку наличие капсида у этой формы ПБВ не требует укрытия ее двуцепочечного генома в митохондриях гриба (Wolf et al., 2018).

Подтверждением возможности такого эволюционного сценария может быть недавнее масштабное метагеномное исследование эукариотических +РНК-вирусов семейства *Narnaviridae* – одного из двух известных семейств-потомков +РНК-фагов, хозяевами которого являются грибы. Данное исследование обнаружило среди представителей *Narnaviridae* многочисленные нарна-подобные геноварианты, как без гена белка капсида, так и с геном капсида, хозяевами которых были разнообразные беспозвоночные (Shi et al., 2016). В современной таксономии происходящие от бактериофагов мито- и нарна-подобные РНК-вирусы с митохондриальным генетическим кодом гриба получили статус семейства и определяются как РНК-вирусы эукариот. Причем, если на уровне семейства спектр хозяев РНК-вирусов может быть широк, то на уровне рода этот спектр обычно ограничен, и существуют четкие филогенетические различия между родами, заражающими хозяев из разных таксонов (Sadiq et al., 2024).

Согласно Y.I. Wolf с коллегами (2018), происхождение ПБВ или партитивирусов объясняется реассортацией между сегментом генома RdRp вируса с +оцРНК из клады партитивирус-пикобирнавирус (с митовирусоподобным

способом репродукции) и сегментом с геном капсида вируса с дцРНК-геномом из клады цистовирусов, тотивирусов и реовирусов.

Реассортация – форма обмена генетической информацией между вирусами с сегментированным геномом. У вирусов с сегментированным геномом обмен сегментами возможен, когда два (или более) генетически различных вируса одновременно инфицируют одну и ту же клетку. Вирусы, между которыми происходит реассортация, образуют новый вид или штамм вируса, обладающий иными качествами и в большинстве случаев повышенной патогенностью. Реассортация в природе чаще всего происходит внутри одного вида, но также может происходить в пределах одного рода.

Возможность реассортации между сегментами вирусных геномов, образованных дцРНК и +оцРНК, можно объяснить участием в реассортации со стороны +РНК-вирусов сегментов, находящихся на стадии репликативных промежуточных форм РНК (дцРНК). Такой вариант происхождения ПБВ объясняет факт выявления С.К. Yinda с коллегами (2018) бескапсидных и капсидированных форм ПБВ с альтернативным генетическим кодом, напоминающих вирусы грибов по способу репродукции.

Вместе с тем в соответствии с «теорией эндосимбиоза» бескапсидные ПБВ-подобные репликоны с митохондриальным генетическим кодом гриба (P11-300, P11-378, P14-90, P15-218, WGML128211, M17A), обнаруженные С.К. Yinda с коллегами (2018) и А. Kleymann с коллегами (2020), можно рассматривать не только как продукты обмена гомологичными сегментами между семействами родственных РНК-вирусов, но и как результат симбиотических отношений. По имеющимся на сегодняшний день данным, геномы ПБВ с прокариотическими мотивами и мотивами плесневых грибов и беспозвоночных обнаружены в кишечниках животных в составе одной микробиоты, состоящей из клеток бактерий и простейших эукариот. Выявление в одном микробиоме ПБВ с генетическим кодом клеток бактерий и плесневых грибов предполагает наличие эндосимбиотических отношений между их хозяевами (Bruto et al., 2014). Недавние исследования состава кишечной микробиоты, в том числе бактерий и грибов, свидетельствуют о значительной взаимозависимости между грибами и бактериями (Li et al., 2022).

Существование симбиотических отношений между таксономически разными хозяевами РНК-вирусов допускает преодоление вирусом барьера между клетками-симбионтами путем горизонтального переноса с возможностью дальнейшей репродукции в клетках нового хозяина. Расширенный филогенетический анализ доказывает повсеместное распространение горизонтального переноса РНК-вирусов между различными хозяевами и его ведущую роль в эволюции этих вирусов (Dolja, Koonin, 2018). Установлено, что бактерии способны проникать в цитоплазму грибов и находиться там длительное время (<https://www.pravda.ru/news/science/2024301-simbioz/>). Причем горизонтальный перенос генов или вирусов, как правило, осуществляется от бактерий к грибам. В обратном направлении перенос генов (вирусов) вряд ли возможен (микровирусы передают-

ся клетками грибов исключительно по вертикали – от родителей к потомкам) (Bruto et al., 2014).

Фирмикуты (*Firmicutes*), предполагаемые хозяева ПБВ (Krishnamurthy, Wang, 2018), являясь наиболее распространенным микроорганизмом в кишечнике животных и человека, могли оказаться там вместе с клетками гриба. Имеются данные, свидетельствующие о том, что некоторые фирмикуты, относящиеся к классу *Clostridia*, конкурентно взаимодействуют в кишечнике с клетками грибка *Candida albicans*, препятствуя колонизации этим грибом клеток высших эукариот (Shulgina, Eddy, 2021). Паразитический способ существования (эндосимбиоз), связывающий *Firmicutes* с одноклеточными грибами *C. albicans*, допускает горизонтальный перенос между ними вирусов и генов (Taggart et al., 2023).

Обнаружение геномов ПБВ с генетическим кодом, характерным для разных в таксономическом отношении хозяев прокариот, низших эукариот, объясняется и некоторой общей закономерностью в эволюции дцРНК-вирусов, замеченной исследователями. Эта закономерность связана со способностью небольших дцРНК-вирусов с минимальными геномами, образованными минимальным количеством сегментов, кодирующих один белок RdRp (*Narnaviridae*) или два, RdRp и белок капсида (*Totiviridae*, *Partitiviridae*, *Picobirnaviridae*), к неспецифическому горизонтальному распространению среди разных в таксономическом отношении хозяев (Dolja, Koonin, 2018). По этой причине ПБВ, вероятно, могли проникнуть и в клетки беспозвоночных, так как беспозвоночные являются особенно неразборчивыми хозяевами для вирусов.

Часто отдаленно родственные беспозвоночные могут быть хозяевами одной и той же группы вирусов (Wolf et al., 2018). С другой стороны, дцРНК-вирусы с максимально возможным для них размером генома, такие как семейство *Reoviridae*, характеризуются гораздо большей степенью специфичности к хозяину, вероятно, из-за большей адаптации к нему посредством приобретения генов, участвующих во взаимодействиях вирус–хозяин (Dolja, Koonin, 2018).

Считается, что вирусы бактерий не могут напрямую инфицировать клетки органов высших эукариот и в эти органы они могут попасть только путем неспецифической транслокации с помощью бактерий, в которых они размножаются (Dabrowska et al., 2005). И лишь недавно появилась информация, что клетки млекопитающих могут непосредственно интернализировать бактериофагов (Bichet et al., 2023) через макропиноцитоз (неспецифическая интернализация) и в редких случаях с помощью рецептор-опосредованного эндоцитоза (специфической интернализации) (Bichet et al., 2023). Однако до настоящего времени способность проникновения ПБВ в клетки животных не доказана.

## Заключение

Ссылаясь на имеющиеся в нашем распоряжении источники, в которых дана молекулярно-генетическая характеристика обнаруженных до настоящего времени штаммов ПБВ, можно предположить, что вероятнее всего хозяевами существующих в природе штаммов ПБВ являются прока-

риоты, а в одноклеточных эукариотах (плесневых грибах и беспозвоночных) могут присутствовать их переходные эволюционные формы.

В пользу такого предположения мы приводим следующие аргументы, которые, как нам представляется, наиболее весомы.

### Аргументы, подтверждающие возможность репродукции ПБВ в клетках прокариот

- Геном у ПБВ обогащен последовательностями Шайна–Дальгарно, которые присущи прокариотам (Krishnamurthy, Wang, 2018).
- Присутствие бактериолитического белка у ПБВ может служить убедительной аргументацией, что ПБВ являются бактериофагами (Gan, Wang, 2023).
- Склонность ПБВ к генетическим изменениям (характерная для вирусов с сегментированным геномом) в большей степени свойственна вирусам прокариот.
- Повсеместное обнаружение ПБВ – в кишечнике животных разного уровня организации (позвоночных и беспозвоночных) и в сточных водах (Ghosh, Malik, 2021).
- Невозможность культивировать ПБВ в лабораторных условиях или выделить их из образцов тканей животных (Sadiq et al., 2024).
- Более частая межвидовая передача ПБВ, чем РНК-вирусов, животных из любого другого семейства (Sadiq et al., 2024).
- Явление переназначения при трансляции стоп-кодонов TAG и TGA на альтернативные (кодирующие аминокислоты глутамин и триптофан) особенно распространено в фагах, инфицирующих *Firmicutes* и *Bacteroidetes* (Peters et al., 2022).

### Аргументы, указывающие на возможность существования некоторых переходных эволюционных форм ПБВ в плесневых грибах и беспозвоночных

- Общее происхождение семейств прокариотических и эукариотических РНК-вирусов допускает на уровне семейства существование широкого спектра хозяев РНК-вируса, имеющих отношение к разным таксонам (Wolf et al., 2018).
- Эндосимбиотические отношения между *Firmicutes* (предполагаемыми хозяевами ПБВ) и грибами *C. albicans* (Peters et al., 2022), занимающими одну экологическую нишу (кишечник животного или человека), предполагают возможность горизонтального переноса генов и вирусов между ними (Shulgina, Eddy, 2021).
- Эндосимбиотические отношения между *Firmicutes* и *C. albicans* являются причиной смены кода у ПБВ вслед за своим хозяином *Firmicutes* (Bender et al., 2008).
- Механизм изменения ПБВ своего генетического кода (Wang, 2022) может демонстрировать не только возможность фагов к репродукции в клетках независимо от их таксономической принадлежности, но и вероятный путь формирования переходных эволюционных форм ПБВ, обнаруженных в клетках некоторых низших эукариот.
- Только у некоторых ПБВ в капсиде обнаружены белки с бактериолитической функцией (Wang, 2022), как и у близкородственного им семейства *Partitiviridae*, представитель которого относят к вирусам грибов (Neri et al., 2022).

Эти наблюдения подтверждают возможность существования отдельных форм ПБВ, таксономически связанных как с бактериями, так и с одноклеточными эукариотами – плесневыми грибами и беспозвоночными. Поэтому, как справедливо заметили D. Wang с коллегами (2022), окончательные выводы относительно истинного хозяина (хозяев) выявленных ПБВ не могут быть обобщены на уровне семейства, а требуют исследований всего разнообразия ПБВ, чтобы определить таксономическую принадлежность всего спектра их хозяев.

## Список литературы / References

- Кашников А.Ю., Епифанова Н.В., Новикова Н.А. О природе пикобирнавирусов. *Вавиловский журнал генетики и селекции*. 2023; 27(3):264-275. doi 10.18699/VJGB-23-32  
[Kashnikov A.Yu., Epifanova N.V., Novikova N.A. On the nature of picobirnaviruses. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2023;27(3):264-275. doi 10.18699/VJGB-23-32]
- Новикова Н.А., Епифанова Н.В., Федорова О.Ф., Голицына Л.Н., Куприянова Н.В. Обнаружение пикобирнавирусов методом электрофореза РНК в полиакриламидном геле. *Вопросы вирусологии*. 2003;48(6):41-43  
[Novikova N.A., Epifanova N.V., Fedorova O.F., Golitsyna L.N., Kupriianova N.V. The detection of picobirnavirus by RNA electrophoresis in polyacrylamide gel. *Voprosy Virusologii = Problems of Virology*. 2003;48(6):41-43 (in Russian)]
- Adriaenssens E.M., Farkas K., Harrison C., Jones D.L., Allison H.E., McCarthy A.J. Viromic analysis of wastewater input to a river catchment reveals a diverse assemblage of RNA viruses. *mSystems*. 2018;3(3):e00025-18. doi 10.1128/mSystems.00025-18
- Al-Shayeb B., Sachdeva R., Chen L.X., Ward F., Munk P., Devoto A., Castelle C.J., ... Lehours A.C., Warren L., Cate J.H.D., Santini J.M., Banfield J.F. Clades of huge phages from across Earth's ecosystems. *Nature*. 2020;578(7795):425-431. doi 10.1038/s41586-020-2007-4
- Bell N., Khamrin P., Kumthip K., Rojjanadumrongkul K., Nantachit N., Maneekarn N. Molecular detection and characterization of picobirnavirus in environmental water in Thailand. *Clin Lab*. 2020;66(5): 855-858. doi 10.7754/Clin.Lab.2019.191013
- Bender A., Hajieva P., Moosmann B. Adaptive antioxidant methionine accumulation in respiratory chain complexes explains the use of a deviant genetic code in mitochondria. *Proc Natl Acad Sci USA*. 2008;105(43):16496-16501. doi 10.1073/pnas.0802779105
- Bichet M.C., Adderley J., Avellaneda-Franco L., Magnin-Bougma I., Torriero-Smith N., Gearing L.J., Deffrasnes C., ... Lin R.C., Patwa R., Moseley G.W., Doerig C., Barr J.J. Mammalian cells internalize bacteriophages and use them as a resource to enhance cellular growth and survival. *PLoS Biol*. 2023;21(10):e3002341. doi 10.1371/journal.pbio.3002341
- Boros Á., Polgár B., Pankovics P., Fenyvesi H., Engelmann P., Phan T.G., Delwart E., Reuter G. Multiple divergent picobirnaviruses with functional prokaryotic Shine–Dalgarno ribosome binding sites present in cloacal sample of a diarrheic chicken. *Virology*. 2018; 525:62-72. doi 10.1016/j.virol.2018.09.008
- Bruto M., Prigent-Combaret C., Luis P., Moëhne-Loccoz Y., Muller D. Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes. *Proc Biol Sci*. 2014;281(1789): 20140848. doi 10.1098/rspb.2014.0848
- Cai X., Tian F., Teng L., Liu H., Tong Y., Le S., Zhang T. Cultivation of a lytic double-stranded RNA bacteriophage infecting *Microvirgula aerodenitrificans* reveals a mutualistic parasitic lifestyle. *J Virol*. 2021;95(17):e0039921. doi 10.1128/jvi.00399-21
- Chen Y.-M., Sadiq S., Tian J.-H., Chen X., Lin X.-D., Shen J.-J., Chen H., ... Xu Q.-Y., Wang W., Gao W.-H., Holmes E.C., Zhang Y.-Z. RNA virome composition is shaped by sampling ecotype. *SSRN Electron J*. 2021. doi 10.2139/ssrn.3934022
- Dabrowska K., Switała-Jelen K., Opolski A., Weber-Dabrowska B., Gorski A. Bacteriophage penetration in vertebrates. *J Appl Microbiol*. 2005;98(1):7-13. doi 10.1111/j.1365-2672.2004.02422.x
- Delmas B., Attoui H., Ghosh S., Malik Y.S., Mundt E., Vakharia V.N. ICTV virus taxonomy profile: *Picobirnaviridae*. *J Gen Virol*. 2019; 100(2):133-134. doi 10.1099/jgv.0.001186
- Dolja V.V., Koonin E.V. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res*. 2018;244:36-52. doi 10.1016/j.virusres.2017.10.020
- Duraisamy R., Akiana J., Davoust B., Mediannikov O., Michelle C., Robert C., Parra H.-J., Raoult D., Biagini P., Desnues C. Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses. *Virus Genes*. 2018;54(2):256-271. doi 10.1007/s11262-018-1543-6
- Gan T., Wang D. Picobirnaviruses encode proteins that are functional bacterial lysins. *Proc Natl Acad Sci USA*. 2023;120(37): e2309647120. doi 10.1073/pnas.2309647120
- Ghosh S., Malik Y.S. The true host/s of picobirnaviruses. *Front Vet Sci*. 2021;7:615293. doi 10.3389/fvets.2020.615293
- Górski A., Kniotek M., Perkowska-Ptasińska A., Mróz A., Przerwa A., Górczyca W., Dabrowska K., Weber-Dabrowska B., Nowaczyk M. Bacteriophages and transplantation tolerance. *Transplant Proc*. 2006;38(1):331-333. doi 10.1016/j.transproceed.2005.12.073
- Guajardo-Leiva S., Chnaiderman J., Gaggero A., Diez B. Metagenomic insights into the sewage RNA virosphere of a large city. *Viruses*. 2020;12(9):1050. doi 10.3390/v12091050
- Kleymann A., Becker A.A.M.J., Malik Y.S., Kobayashi N., Ghosh S. Detection and molecular characterization of picobirnaviruses (PBVs) in the mongoose: identification of a novel PBV using an alternative genetic code. *Viruses*. 2020;12(1):99. doi 10.3390/v12010099
- Knox M.A., Wierenga J., Biggs P.J., Gedyke K., Almeida V., Hall R., Kalem-Zikusoka G., Rubanga S., Ngabirano A., Valdivia-Granda W., Hayman D.T.S. Abundant dsRNA picobirnaviruses show little geographic or host association in terrestrial systems. *Infect Genet Evol*. 2023;112:105456. doi 10.1016/j.meegid.2023.105456
- Kollmar M., Mühlhausen S. Nuclear codon reassignments in the genomics era and mechanisms behind their evolution. *BioEssays*. 2017;39(5):1600221. doi 10.1002/bies.201600221
- Koonin E.V., Dolja V.V., Krupovic M. Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology*. 2015;479-480: 2-25. doi 10.1016/j.virol.2015.02.039
- Koonin E.V., Dolja V.V., Krupovic M., Varsani A., Wolf Y.I., Yutin N., Zerbini F.M., Kuhn J.H. Global organization and proposed mega-taxonomy of the virus world. *Microbiol Mol Biol Rev*. 2020;84(2): e00061-19. doi 10.1128/MMBR.00061-19
- Korkmaz G., Holm M., Wiens T., Sanyal S. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J Biol Chem*. 2014;289(44):30334-30342. doi 10.1074/jbc.M114.606632
- Krishnamurthy S.R., Wang D. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. *Virology*. 2018;516:108-114. doi 10.1016/j.virol.2018.01.006
- Kumar N., Mascarenhas J.D.A.P., Ghosh S., Masachessi G., da Silva Bandeira R., Nates S.V., Dhama K., Singh R.K., Malik Y.S. Picobirnavirus. In: Malik Y.S., Singh R.K., Dhama K. (Eds) *Animal-Origin Viral Zoonoses*. Singapore: Springer, 2020;291-312. doi 10.1007/978-981-15-2651-0\_13
- Le Lay C., Shi M., Buček A., Bourguignon T., Lo N., Holmes E.C. Unmapped RNA virus diversity in termites and their symbionts. *Viruses*. 2020;12(10):1145. doi 10.3390/v12101145
- Li H., Miao M.X., Jia C.L., Cao Y.B., Yan T.H., Jiang Y.Y., Yang F. Interactions between *Candida albicans* and the resident microbiota. *Front Microbiol*. 2022;13:930495. doi 10.3389/fmicb.2022.930495
- Luo X.L., Lu S., Jin D., Yang J., Wu S.S., Xu J. *Marmota himalayana* in the Qinghai-Tibetan plateau as a special host for bi-segmented and unsegmented picobirnaviruses. *Emerg Microbes Infect*. 2018; 7(1):20. doi 10.1038/s41426-018-0020-6

- Luque D., Mata C.P., Suzuki N., Ghabrial S.A., Castón J.R. Capsid structure of dsRNA fungal viruses. *Viruses*. 2018;10(9):481. doi 10.3390/v10090481
- Mahar J.E., Shi M., Hall R.N., Strive T., Holmes E.C. Comparative analysis of RNA virome composition in rabbits and associated ectoparasites. *J Virol*. 2020;94(11):e02119-19. doi 10.1128/JVI.02119-19
- Malik Y.S., Kumar N., Sharma K., Dhama K., Shabbir M.Z., Ganesh B., Kobayashi N., Banyai K. Epidemiology, phylogeny and evolution of emerging enteric Picobirnaviruses of animal origin and their relationship to human strains. *Biomed Res Int*. 2014;2014:780752. doi 10.1155/2014/780752
- McCutcheon J.P., Moran N.A. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 2011;10(1):13-26. doi 10.1038/nrmicro2670
- Neri U., Wolf Y.I., Roux S., Camargo A.P., Lee B., Kazlauskas D., Chen I.M., ... Krupovic M., Dolja V.V., Kyrpides N.C., Koonin E.V., Gophna U. Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell*. 2022;185(21):4023-4037.e18. doi 10.1016/j.cell.2022.08.023
- Pearson M.N., Beaver R.E., Boine B., Arthur K. Mycoviruses of filamentous fungi and their relevance to plant pathology. *Mol Plant Pathol*. 2009;10(1):115-128. doi 10.1111/j.1364-3703.2008.00503.x
- Pereira H.G., Fialho A.M., Flewett T.H., Teixeira J.M.S., Andrade Z.P. Novel viruses in human faeces. *Lancet*. 1988;332(8602):103-104. doi 10.1016/S0140-6736(88)90032-3
- Perez L.J., Cloherty G.A., Berg M.G. Parallel evolution of picobirnaviruses from distinct ancestral origins. *Microbiol Spectr*. 2023;11(6):e02693-23. doi 10.1128/spectrum.02693-23
- Peters S.L., Borges A.L., Giannone R.J., Morowitz M.J., Banfield J.F., Hettich R.L. Experimental validation that human microbiome phages use alternative genetic coding. *Nat Commun*. 2022;13(1):5710. doi 10.1038/s41467-022-32979-6
- Reddy M.V., Gupta V., Nayak A., Tiwari S.P. Picobirnaviruses in animals: a review. *Mol Biol Rep*. 2023;50(2):1785-1797. doi 10.1007/s11033-022-08133-2
- Sadiq S., Chen Y.M., Zhang Y.Z., Holmes E.C. Resolving deep evolutionary relationships within the RNA virus phylum *Lenarviricota*. *Virus Evol*. 2022;8(1):veac055. doi 10.1093/ve/veac055
- Sadiq S., Holmes E.C., Mahar J.E. Genomic and phylogenetic features of the *Picobirnaviridae* suggest microbial rather than animal hosts. *Virus Evol*. 2024;10(1):veae033. doi 10.1093/ve/veae033
- Shackelton L.A., Holmes E.C. The role of alternative genetic codes in viral evolution and emergence. *J Theor Biol*. 2008;254(1):128-134. doi 10.1016/j.jtbi.2008.05.024
- Shi M., Lin X.D., Tian J.H., Chen L.J., Chen X., Li C.X., Qin X.C., ... Buchmann J., Wang W., Xu J., Holmes E.C., Zhang Y.-Z. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540(7634):539-543. doi 10.1038/nature 20167
- Shi M., Lin X.D., Chen X., Tian J.H., Chen L.J., Li K., Wang W., Eden J.S., Shen J.J., Liu L., Holmes E.C., Zhang Y.Z. The evolutionary history of vertebrate RNA viruses. *Nature*. 2018;556(7700):197-202. doi 10.1038/s41586-018-0012-7
- Shulgina Y., Eddy S.R. A computational screen for alternative genetic codes in over 250,000 genomes. *eLife*. 2021;10:e71402. doi 10.7554/eLife.71402
- Son M., Yu J., Kim K.-H. Five questions about Mycoviruses. *PLoS Pathog*. 2015;11(11):e1005172. doi 10.1371/journal.ppat.1005172
- Taggart N.T., Crabtree A.M., Creagh J.W., Bizarría R. Jr., Li S., de la Higuera I., Barnes J.E., Shipley M.A., Boyer J.M., Stedman K.M., Ytreberg F.M., Rowley P.A. Novel viruses of the family *Partitiviridae* discovered in *Saccharomyces cerevisiae*. *PLoS Pathog*. 2023;19(6):e1011418. doi 10.1371/journal.ppat.1011418
- Ullah K., Mehmood A., Chen X., Dar M.A., Yang S., Zhang W. Detection and molecular characterization of picobirnaviruses in the wild birds: identification of a novel picobirnavirus possessing yeast mitochondrial genetic code. *Virus Res*. 2022;308:198624. doi 10.1016/j.virusres.2021.198624
- Vainio E.J., Chiba S., Ghabrial S.A., Maiss E., Roossinck M., Sabana-dzovic S., Suzuki N., Xie J., Nibert M. ICTV virus taxonomy profile: *Partitiviridae*. *J Gen Virol*. 2018;99(1):17-18. doi 10.1099/jgv.0.000985
- Wang D. The enigma of picobirnaviruses: viruses of animals, fungi, or bacteria? *Curr Opin Virol*. 2022;54:101232. doi 10.1016/j.coviro.2022.101232
- Wolf Y.I., Kazlauskas D., Iranzo J., Lucia-Sanz A., Kuhn J.H., Krupovic M., Dolja V.V., Koonin E.V. Origins and evolution of the global RNA virome. *MBio*. 2018;9(6):e02329-18. doi 10.1128/mBio.02329-18
- Woo P.C.Y., Teng J.L.L., Bai R., Tang Y., Wong A.Y.P., Li K.S.M., Lam C.S.F., Fan R.Y.Y., Lau S.K.P., Yuen K.-Y. Novel picobirnaviruses in respiratory and alimentary tracts of cattle and monkeys with large intra- and inter-host diversity. *Viruses*. 2019;11(6):574. doi 10.3390/v11060574
- Yinda C.K., Ghogomu S.M., Conceição-Neto N., Beller L., Deboutte W., Vanhulle E., Maes P., Van Ranst M., Matthijnsens J. Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. *Virus Evol*. 2018;4(1):vey008. doi 10.1093/ve/vey008
- Yutin N., Benler S., Shmakov S.A., Wolf Y.I., Tolstoy I., Rayko M., Antipov D., Pevzner P.A., Koonin E.V. Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features. *Nat Commun*. 2021;12(1):1044. doi 10.1038/s41467-021-21350-w

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 17.02.2025. После доработки 16.04.2025. Принята к публикации 25.06.2025.

doi 10.18699/vjgb-26-15

## Эволюция порядка генов в мтДНК байкальских эндемичных амфипод и ее возможные механизмы

Е.А. Сиротина , Д.Ю. Щербаков , Е.В. Романова 

Лимнологический институт Сибирского отделения Российской академии наук, Иркутск, Россия  
 haleo.inc@gmail.com

**Аннотация.** Значительное разнообразие в порядке генов митохондриальных (мт) геномов беспозвоночных отмечается в подтипе Crustacea, и, в частности, в отряде Amphipoda. Амфиподы озера Байкал также известны как группа с уникальными порядками генов в мт геномах. Чтобы оценить разнообразие порядков белок-кодирующих генов (ПГ) амфипод, был проведен сравнительный анализ генных перестроек в мт геномах байкальских и небайкальских видов. В некоторых случаях данные о перестройках генов и истории их перемещений у различных таксономических групп также могут информативно дополнять данные филогенетического анализа. Для 13 ранее полученных нуклеотидных последовательностей мт геномов байкальских видов мы определили 4 варианта ПГ, для 114 мт геномов небайкальских видов – 14 вариантов ПГ. Были также рассчитаны типы и число шагов перестроек (от 1 до 3), требующихся для перехода от одного порядка генов к другому, и число мт генов, подвергшихся перестройкам в каждом ПГ (от 1 до 5). Байкальские амфиподы принадлежат к двум линиям (I и II) в соответствии с молекулярными данными, указывающими на их происхождение от двух независимых вселений предковых видов в озеро. Все случаи перестроек порядка мт генов обнаружены у видов из линии I, тогда как порядок мт генов в линии II консервативен у всех изученных видов и соответствует модели Pancrustacean pattern (ПанПГ). ПанПГ был определен как предковый порядок генов для обеих линий байкальских амфипод. В исследовании обсуждаются возможные механизмы перестроек порядка мт генов, такие, как полная или частичная дупликация мт генома и последующие случайные делеции. Высказывается предположение, что повышенная скорость мутаций, ослабление стабилизирующего отбора и другие особые факторы могут влиять на вероятность появления и фиксации различных ПГ в мт геномах байкальских амфипод.

**Ключевые слова:** амфиподы; озеро Байкал; митохондриальный геном; перестройки генов

**Для цитирования:** Сиротина Е.А., Щербаков Д.Ю., Романова Е.В. Эволюция порядка генов в мтДНК байкальских эндемичных амфипод и ее возможные механизмы. *Вавиловский журнал генетики и селекции*. 2026;30(1):136-145. doi 10.18699/vjgb-26-15

## Evolution of gene order in mtDNA of Baikal endemic amphipods and its possible mechanisms

E.A. Sirotnina , D.Yu. Sherbakov , E.V. Romanova 

Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia  
 haleo.inc@gmail.com

**Abstract.** Significant gene order diversity of mitochondrial (mt) genomes of invertebrates is peculiar to subphylum Crustacea, and to order Amphipoda in particular. Amphipods from Lake Baikal are also known as a group with unique gene orders in their mt genomes. To estimate the diversity of protein-coding gene orders (GOs) in amphipods, a comparative analysis of gene rearrangements in the mt genomes of Baikal and non-Baikal species was performed. In some cases, gene rearrangement data and the history of gene relocation in different taxonomic groups can also supplement the results of phylogenetic inferences. Among the thirteen mt genomes of Baikal species sequenced in previous studies, four gene order patterns were identified, and fourteen gene order patterns for 114 mt genomes of non-Baikal species were observed. The type and number of rearrangement steps (from 1 to 3) required to transition from one order to another and the number of mt genes rearranged in each GO (from 1 to 5) were also defined. Baikalian amphipods belong to two lineages (I and II) according to molecular data which reveal their origin from two independent introductions of ancestral species into the lake. All cases of mt gene order rearrangements have been detected in species from the first lineage, whereas the mt gene order in the second lineage is conserved in all species studied and corresponds to the Pancrustacean pattern (PanGO). PanGO has been determined as the ancestral gene order for both Baikalian amphipod lineages. The possible mechanisms of mt gene order rearrangements such as a complete or partial duplication of mt genome and subsequent random deletions are discussed in our study. It is supposed that increased mutation rate, weakening of stabilizing selection and other specific factors may influence the probability of emergence and fixation of different GOs in mt genomes of Baikalian amphipods.

**Key words:** amphipods; Lake Baikal; mitochondrial genome; gene rearrangement

**For citation:** Sirotnina E.A., Sherbakov D.Yu., Romanova E.V. Evolution of gene order in mtDNA of Baikal endemic amphipods and its possible mechanisms. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2026;30(1):136-145. doi 10.18699/vjgb-26-15

## Введение

Разнообразие порядков генов митохондриального (мт) генома эукариот очень велико. В зависимости от путей эволюции в разных линиях, типы организаций мт генома (линейные, кольцевые, оформленные в несколько хромосом), состав генов (потеря и приобретение генов) и их расположение различаются (Sterling-Montealegre, Prada, 2024). Широкий разброс в размерах мт геномов (от 11–50 т. п. н. у животных до 66 т. п. н.–11.3 млн п. н. у растений) наблюдается во многом из-за вариаций в числе генов, а также длины некодирующей мтДНК. Разнообразие порядков генов (ПГ) определяется как изменением их положений относительно друг друга, так и наличием делеций и дупликаций. У некоторых крупных таксонов ПГ консервативен (например, у позвоночных), у других ПГ широко варьирует (в некоторых группах беспозвоночных) (Zardoya, 2020). Порядки генов в мт геномах ракообразных являются одними из самых разнообразных среди беспозвоночных (подтип Crustacea), что проявляется в изменении положений генов относительно базового порядка, характерного также для насекомых (Pancrustacean pattern, ПанПГ) (Boore, 1999; Kilpert, Podsiadlowski, 2006; Sterling-Montealegre, Prada, 2024).

Типы изменений ПГ могут быть разными. В недавнем исследовании R.A. Sterling-Montealegre и C.F. Prada (2024) среди 464 мт геномов 47 отрядов членистоногих проанализировано 299 геномов подтипа Crustacea и выделено 87 ПГ, в том числе четыре ПГ, наиболее часто встречающихся у ракообразных.

Согласно этим данным, в мт геномах Crustacea на тРНК гены пришлось 70.1 % от числа всех перестроек, 29.9 % составили изменения положений других генов. Изменения положений белок-кодирующих и рРНК генов встречаются реже (Castellucci et al., 2022), чем у тРНК генов, положения которых различаются порой даже у видов одного семейства (Jühling et al., 2012). Ранее была показана положительная ассоциация значительных показателей генных перестроек с высокой скоростью нуклеотидных замен в белок-кодирующих генах подкласса Caenogastropoda Cox, 1959 (брюхоногие моллюски) (Fourdrilis et al., 2018). У байкальских амфипод, группы с разнообразным ПГ в мт геномах, также обнаружено значимое ускорение эволюции в нескольких белок-кодирующих генах по сравнению с родственными видами пресноводных амфипод рода *Gammarus* (Romanova, Sherbakov, 2019). Тем не менее такая связь прослеживается не во всех группах организмов (Shao et al., 2003; Xu et al., 2006). Примечательно, что в некоторых линиях ракообразных порядок генов в мт геномах остается неизменным в течение длительных периодов времени, в то время как в других группах этот признак сильно варьирует (Kilpert et al., 2012; Tan et al., 2019; Zardoya, 2020).

Филогенетический и статистический анализы позволяют реконструировать историю изменения ПГ у разных таксонов. В некоторых случаях ПГ могут являться информативными синапоморфиями для отдельных групп, дополнительно помогая уточнить родственные связи на уровне инфраотрядов, надсемейств и семейств (Tan et al., 2019).

Мт геномы с измененным порядком генов часто встречаются у амфипод из озера Байкал (Romanova et al., 2016, 2020; Drozdova et al., 2024). Это группа беспозвоночных животных, возникшая в озере в результате адаптивной радиации и существующая на протяжении длительного времени (Sherbakov, 1999; Mats et al., 2011). Причины, приводящие к перестройкам ПГ в данной группе эндемичных беспозвоночных, не известны, как и их предполагаемые механизмы (Mueller, Boore, 2005).

Статья посвящена оценке разнообразия ПГ в мт геномах амфипод с помощью сравнительного анализа генных перестроек у этой группы и предположений о механизмах их появления в контексте эволюционной истории. Результаты позволили определить ПанПГ как предковый для байкальских видов, а также некоторые закономерности возникновения перестроек в этой группе. Также были выделены группы с неодинаковыми ПГ в выборке видов.

## Материалы и методы

**Набор данных ПГ амфипод.** Последовательности мт геномов амфипод для этого исследования взяты из базы данных GenBank (Приложение 1)<sup>1</sup>. Мт геномы байкальских видов были опубликованы в более ранних исследованиях (Rivarola-Duarte et al., 2014; Romanova et al., 2016, 2021; Mamos et al., 2021). Места сбора образцов, а также детали выделения ДНК, проведения ПЦР, секвенирования, сборки и аннотации мт геномов описаны в публикациях E.V. Romanova с коллегами (2016, 2021), а также в Приложении 2.

В набор данных для сравнения порядков генов включили только те мт геномы (127 видов), для которых была полностью определена нуклеотидная последовательность кодирующей части (13 белок-кодирующих генов). Из набора были исключены рибосомные гены, гены тРНК и КР (контрольный регион), так как у ряда видов количество генов было неполным или имелись дублицированные копии, что затруднило бы их анализ. Мт геном байкальского вида *Linevichella vortex* (Dybowski, 1874) был также исключен из анализа ПГ из-за неполного генома, но включен в филогенетический анализ. Схема ПанПГ была использована для попарных сравнений. Следует отметить, что способ выделения ПГ на основании положений 13 мт генов без учета тРНК, рРНК и положения КР, в отличие от анализа 37 генов, чаще определяет порядки генов как одинаковые – относительно ПанПГ или друг друга. Полные ПГ этих же видов не будут идентичными из-за разных положений остальных генов.

**Анализ порядка генов в программе CREx.** Тип перестроек мт генов у амфипод оценивали с помощью программы CREx (Common Interval Rearrangement Explorer), входящей в состав CREx2 (Bernt et al., 2007; Hartmann et al., 2019), на сервере Galaxy (The Galaxy Community..., 2024) (<https://usegalaxy.eu/>). В алгоритме CREx применяется понятие общих интервалов – групп генов, последовательно расположенных в сравниваемых ПГ. Количество общих интервалов (NSCI) используется для оценки сходства ПГ;

<sup>1</sup> Приложения 1–4 см. по адресу:  
<https://vavilovj-icg.ru/download/pict-2026-30/appx9.zip>

чем выше этот показатель, тем более идентичными они являются. При помощи моделей событий CREx определялись возможные сценарии перестроек, включающие тандемную дупликацию сегмента соседних генов с последующим случайным удалением части копий генов (TDRL), транспозицию (Т) – перемещение гена в пределах той же кодирующей цепи мтДНК, инверсию гена (I) на другую кодирующую цепь мтДНК и инверсную транспозицию (iT) (Bernt et al., 2007; Basso et al., 2017).

В рамках алгоритма CREx названные типы перестроек оцениваются только при помощи математических моделей. TDRL чаще предполагается в сценарии более масштабных перестроек мт генома и перемещения большего числа генов за один шаг, чем при транспозиции, а также является несимметричной перестройкой, которая может быть произведена только в одном направлении для двух сравниваемых ПГ, позволяя определить предковое состояние (Bernt et al., 2007). При предсказании двух и более альтернативных сценариев выбирался тот, который предполагал наименьшее количество шагов.

**Ручной подсчет изменений положений отдельных генов.** При попарном сравнении ПанПГ с ПГ каждого вида рассматривали порядок генов внутри генных блоков на двух цепях мтДНК, а также проверяли, сохраняется ли порядок блоков в геноме. Двигаясь по направлению от *cox1* к *nad1*, определяли визуально, где находятся границы каждого консервативного блока, относительно которых произошла перестройка генов. Гены, находящиеся в той же позиции, что и у ПанПГ, обозначались как «0», отдельные гены, положение которых изменено относительно консервативных генных блоков в ПанПГ, – как «1» (Приложения 3 и 4). Результаты подсчета были визуализированы в виде гистограмм, схемы соответствующих ПГ были совмещены с филогенетическими деревьями.

**Филогенетические реконструкции.** Для построения филогенетических деревьев использовали аминокислотные последовательности 13 белок-кодирующих генов мт геномов 128 видов амфипод (доступные в базе данных GenBank к 10.12.2023), в том числе 14 байкальских видов (включая вид *L. vortex* с неполным мт геномом) и 3 вида из отряда Isopoda (*Ligia oceanica* (Linnaeus, 1767), *Eophreatoicus karrkkanj* Wilson & Humphrey, 2020, *Neomysis japonica* Nakazawa, 1910), взятых в качестве внешней группы.

Аминокислотные последовательности каждого белок-кодирующего гена выравнивались по кодонам в TranslatorX (Abascal et al., 2010) с помощью алгоритма ClustalW и затем были объединены в программе SeaView 4.5.4 (Gouy et al., 2010). В ModelFinder (Kalyaanamoorthy et al., 2017) была выбрана оптимальная модель аминокислотных замен JTT+F+I+R9. С помощью IQ-TREE 2 (Minh et al., 2020) методом максимального правдоподобия построено филогенетическое дерево, которое затем было визуализировано в FigTree 1.4.3 (Rambaut, 2018).

Филогенетическое дерево амфипод совместно с данными о порядках генов было использовано для выявления закономерностей генных перестроек, возникших в ходе эволюции исследуемых таксонов. Отмечены ветви и узлы дерева, в которых произошли изменения положе-

ний отдельных генов, варианты ПГ разных видов. Также оценивалось соответствие между изменениями в ПГ и филогенией видов.

## Результаты

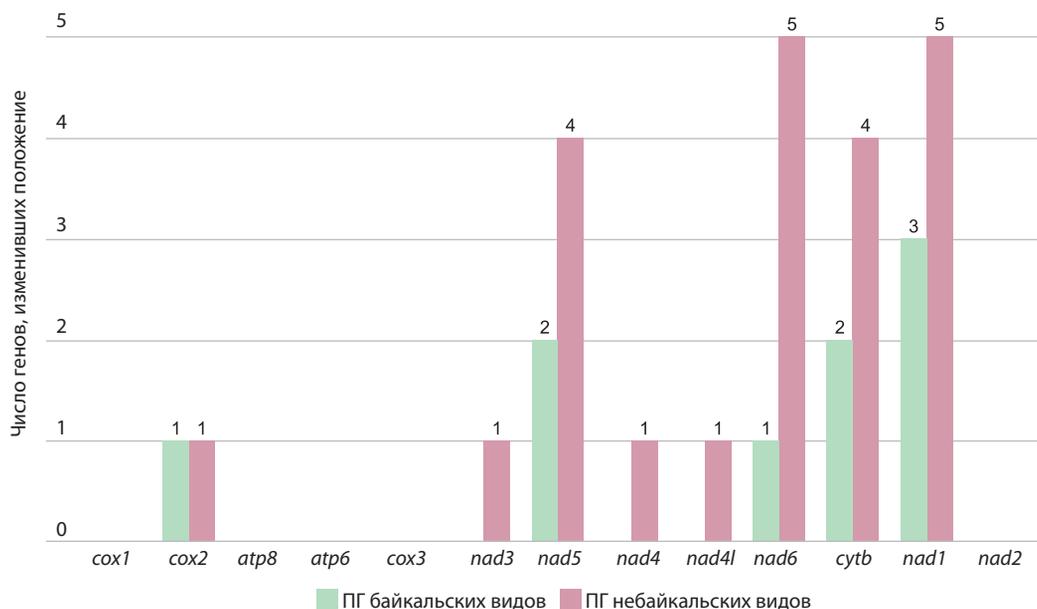
### Анализ перестроек порядка генов

Среди 127 проанализированных полных мт геномов амфипод выделено 18 различных порядков белок-кодирующих генов (ПГ1–ПГ18). С помощью CREx рассчитано от 1 до 3 шагов перестроек, минимально необходимых для перехода от одного порядка к другому. Виды амфипод, порядки генов мт геномов и типы перестроек приведены в Приложении 3. Дальнейший анализ генных перестроек проводили, учитывая только уникальные ПГ. Было подсчитано вручную и отмечено общее число мт генов, изменивших положение по сравнению с ПанПГ: в двух группах амфипод – байкальских и небайкальских видов (рис. 1), а также отдельно в каждом ПГ (рис. 2, см. Приложение 2). Такие перестройки встречаются чаще в более крупной и экологически разнообразной выборке небайкальских видов.

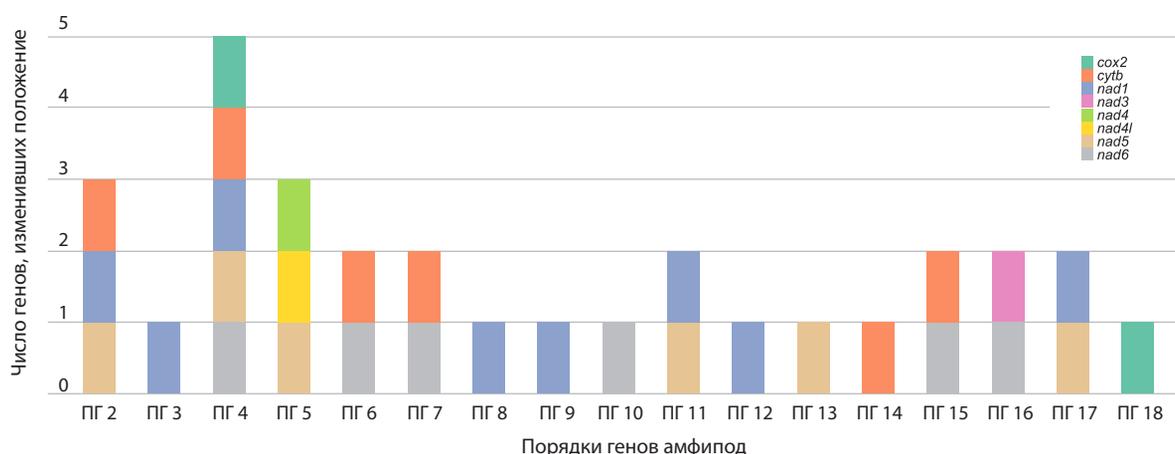
Ген *nad1* является наиболее часто меняющим положение по сравнению с ПанПГ (в восьми ПГ) как в группе байкальских, так и небайкальских видов (см. рис. 1, см. Приложение 4). Показано, что 8 из 13 белок-кодирующих генов (*nad1*, *cytb*, *nad6*, *nad5*, *cox2*, *nad3*, *nad4*, *nad4l*) чаще других подвергались перестройкам (в частности, инверсиям) (см. рис. 1). ПГ с наибольшим количеством генов, изменивших положение среди всех амфипод, имеют виды *Macrohectopus branickii* (Dybowsky, 1874) (ПГ 4, пять генов), два вида рода *Caprella* Lamarck, 1801, *Cyamus boopis* Lütken, 1870 (ПГ 5, три гена) и *Crypturopsis tuberculatus*, *C. inflatus* (Dybowsky, 1874) (ПГ 2, три гена) (см. рис. 2, см. Приложение 3).

У 20 видов отмечены сложные перестройки, соответствующие девяти ПГ (ПГ 2, ПГ 4, ПГ 5, ПГ 6, ПГ 7, ПГ 11, ПГ 15, ПГ 16, ПГ 17), в которых изменено положение от 2 до 5 отдельных генов (см. Приложение 4). У остальных ПГ изменено положение 1 гена. Инверсии (I) и инверсные транспозиции (iT) генов встречаются часто – у 7 из 17 измененных ПГ (см. Приложение 3). Самый распространенный тип перестройки генов в мт геномах амфипод – транспозиция (у 11 ПГ).

К сложному типу мы относим перестройки, требующие число шагов, равное 2 или более (согласно расчету при помощи CREx), сочетание перестроек разных типов (инверсий, транспозиций) либо не менее одной перестройки типа TDRL (Bernt et al., 2007; Castellucci et al., 2022). Сложный тип перестроек TDRL отмечен только у байкальских видов *M. branickii*, *C. tuberculatus*, *C. inflatus*. У небайкальских видов найдены сочетания разных типов перестроек, а TDRL отсутствовали. Только у байкальского вида линии I (*M. branickii*) отмечены как большое количество шагов перестроек, так и высокая сложность (инверсия, тандемная дупликация со случайной потерей генов, транспозиция), в то время как у видов линии II мт геномы по порядку белок-кодирующих генов соответствуют ПанПГ (см. Приложение 3).



**Рис. 1.** Число случаев перестроек каждого белок-кодирующего гена мт генома по сравнению с их положениями в ПанПГ, подсчитанное для группы байкальских и небайкальских амфипод.



**Рис. 2.** Порядки генов амфипод (ПГ 2–ПГ 18) с соответствующим числом белок-кодирующих генов мт генома, изменивших положение по сравнению с ПанПГ. ПГ 1, соответствующий ПанПГ, не показан.

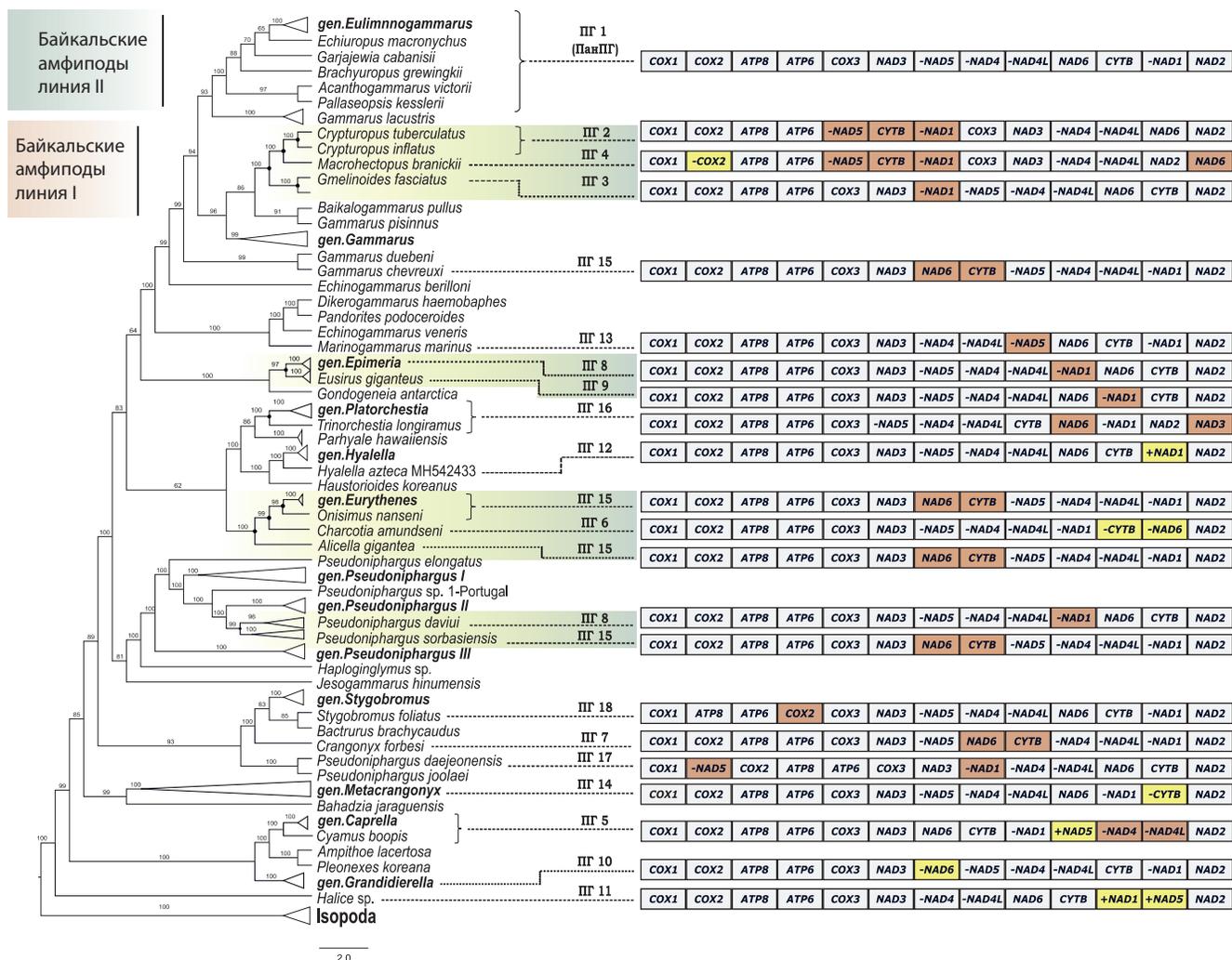
### Филогенетический анализ

Филогенетическое дерево было построено по аминокислотным последовательностям 13 белок-кодирующих генов 128 видов амфипод. Одна клада представлена аутгруппой – изоподами (рис. 3). Группа байкальских амфипод представлена 13 видами. Байкальские виды разделились на две клады, соответствующие ранее выделенным линиям I и II.

Каждая байкальская линия кластеризуется с разными видами из рода *Gammarus* Fabricius, 1775, I – с *G. pisinus* Hou, Li & Li, 2014, *G. fossarum* Koch, 1836, *Baikalogammarus pullus* (Dybowski, 1874), *G. roeselii* Gervais, 1835; II – с *G. lacustris* G.O. Sars, 1863. Таким образом, род *Gammarus* является парафилетичным. Кластеризация филогенетического дерева показывает, что обе линии байкальских

видов имеют общего предка с видами рода *Gammarus*, так как *G. duebeni* Lilljeborg, 1852 и *G. chevreuxi* Sexton, 1913 занимают базальное положение по отношению к этим линиям. Линия I включает виды семейств Micrurropodidae Kamal'tynov, 1999, Macrohectopodidae Sowinsky, 1915 и Crypturropodidae Kamal'tynov, 2002; линию II составляют виды семейств Eulimnogammaridae Kamal'tynov, 1999, Acanthogammaridae Garajeff, 1901 и Pallaseidae Tachteew, 2001.

Виды родов *Metacrangonyx* Chevreux, 1909 (обитают в морской солоноватой воде, редко в пресных водах) и *Pseudoniphargus* Chevreux, 1901 (стигобионты, обитают в среде от солоноватых колодцев до горных рек) объединены в монофилетические группы. Вид *Echinogammarus berilloni* (Catta, 1878), встречающийся в реках и ручьях За-



**Рис. 3.** Филогенетическое дерево, построенное на основе аминокислотных последовательностей 13 белок-кодирующих мт генов 128 видов амфипод, со схемой порядков белок-кодирующих генов соответствующих мт геномов.

Дерево было визуально трансформировано по пропорциям; длины ветвей масштабированы. Выделенные зеленым цветом группы видов с неодинаковыми ПГ. Вид *L. vortex* не подписан, так как не был включен в анализ ПГ. Желтым цветом на схеме выделены гены, изменившие кодирующую цепь (I, IT), красным – гены, подвергшиеся перестройкам типов T и TDRL.

падной Европы, занимает базальное положение у группы, включающей байкальские виды и виды рода *Gammarus*. Порядок генов мт генома вида *E. berilloni* идентичен ПГ последних (кроме *G. chevreuxi*).

Данные филогенетического анализа вкупе с данными ручного анализа перестроек позволили обнаружить таксоны амфипод, порядок генов мт генома которых подвергся наибольшим изменениям. Из рис. 2, 3 и Приложения 3 видно, что у ряда далеких друг от друга видов одни и те же белок-кодирующие гены изменили положение в мт геноме, и в некоторых случаях (например, *Eurythenes maldoror* d’Udekem d’Acoz & Navermans, 2015 и *G. chevreuxi* в ПГ 15) эти перестройки случайным образом привели к образованию одинаковых ПГ (гомоплазия). Анализ мт геномов видов рода *Gammarus* позволил нам показать, что ПанПГ является синплезиоморфией для обеих линий байкальских видов.

## Обсуждение

Полученные результаты анализа ПГ в мт геномах амфипод характеризуют их как группу с часто встречающимися и разнообразными перестройками генов в мт геномах (18 ПГ по белок-кодирующим генам для 127 видов). Четыре из 13 байкальских видов соответствуют трем ПГ с изменениями, а 53 из 114 небайкальских видов соответствуют 14 ПГ с изменениями (см. рис. 1 и 2). Таким образом, изменения ПГ наблюдаются у 57 видов из выборки, что составляет 17 ПГ. Ранее описанный ПанПГ является самым часто встречающимся ПГ по белок-кодирующим генам среди амфипод, хотя не доказано, что он – самый древний для этой группы видов. ПГ 11 у базального для группы амфипод вида *Halice* sp. (Li et al., 2019) заметно отличается от ПанПГ.

Вероятно, из-за того, что амфиподы – древняя группа беспозвоночных животных, в ходе эволюции у них об-

разовалось много вариантов порядков генов (Hou et al., 2014; Mamos et al., 2021). М.Н. Тан с коллегами, исследуя разнообразие мт геномов отряда Decapoda внутри инфраотрядов и между ними, проанализировали филогенетическое дерево с отмеченными на нем ПГ и выделили так называемые «горячие точки» (Tan et al., 2019). Это группы близких видов, в которых встречается одновременно два или более уникальных ПГ, отличающихся друг от друга и от предка. Всего на дереве выделено четыре группы амфипод, содержащих неодинаковые ПГ. Два ПГ в группе надсемейства Lysianassoidea Dana, 1849 (ПГ 6 и ПГ 15, в последнем одни и те же гены изменены у шести видов), два – в группе рода *Epimeria* A. Costa in Hope, 1851 (ПГ 8 и ПГ 9), два – в группе рода *Pseudoniphargus* (ПГ 8 и ПГ 15) и три в байкальской линии I (ПГ 2, ПГ 3, ПГ 4, есть гены, измененные у 4 из 5 видов) (см. рис. 3).

По количеству шагов перестроек генов, согласно расчету при помощи CREx, линия I байкальских видов (1–3 шага) схожа с группой видов рода *Caprella*, *C. boopis* (2 шага) и базальным видом *Halice* sp. (2 шага). В этих группах сосредоточены ПГ с более сложными перестройками (рис. 4). Образование ПГ у видов рода *Platorchestia* Bousfield, 1982 (2Т), вида *P. daejeonensis* (2Т) предполагает более простые сценарии перестроек, но также требующие 2 шагов. Таким образом, линия I байкальских амфипод является группой с наибольшей интенсивностью эволюционных процессов, так как в ней обнаружено максимальное количество как сложных перестроек ПГ, так и уникальных ПГ.

Виды амфипод с измененными ПГ, имеющие различные биологические и экологические особенности, широко встречаются у разных таксонов на дереве (см. рис. 4), что указывает на неустойчивость ПГ как признака для таксономии. Новые ПГ и ПанПГ выявляются как у давно отделившихся видов, так и в более молодых радиациях. Ожидаемо чаще можно наблюдать примеры того, как виды из одного семейства, формирующие одну кладу, имеют одинаковые ПГ, например *Caprella mutica* Schurin, 1935, *C. boopis* (ПГ 5).

Также существуют неродственные виды амфипод, имеющие одинаковый ПГ при анализе выборки белок-кодирующих генов. Гомоплазия отмечается в трех отдаленных группах видов, где встречается одинаковое измененное положение генов *nad6* и *cytb*, соответствующее ПГ 15 (см. Приложение 3, рис. 3 и 4). При включении в анализ тРНК, рРНК и КР порядки генов таких видов в основном не будут идентичными.

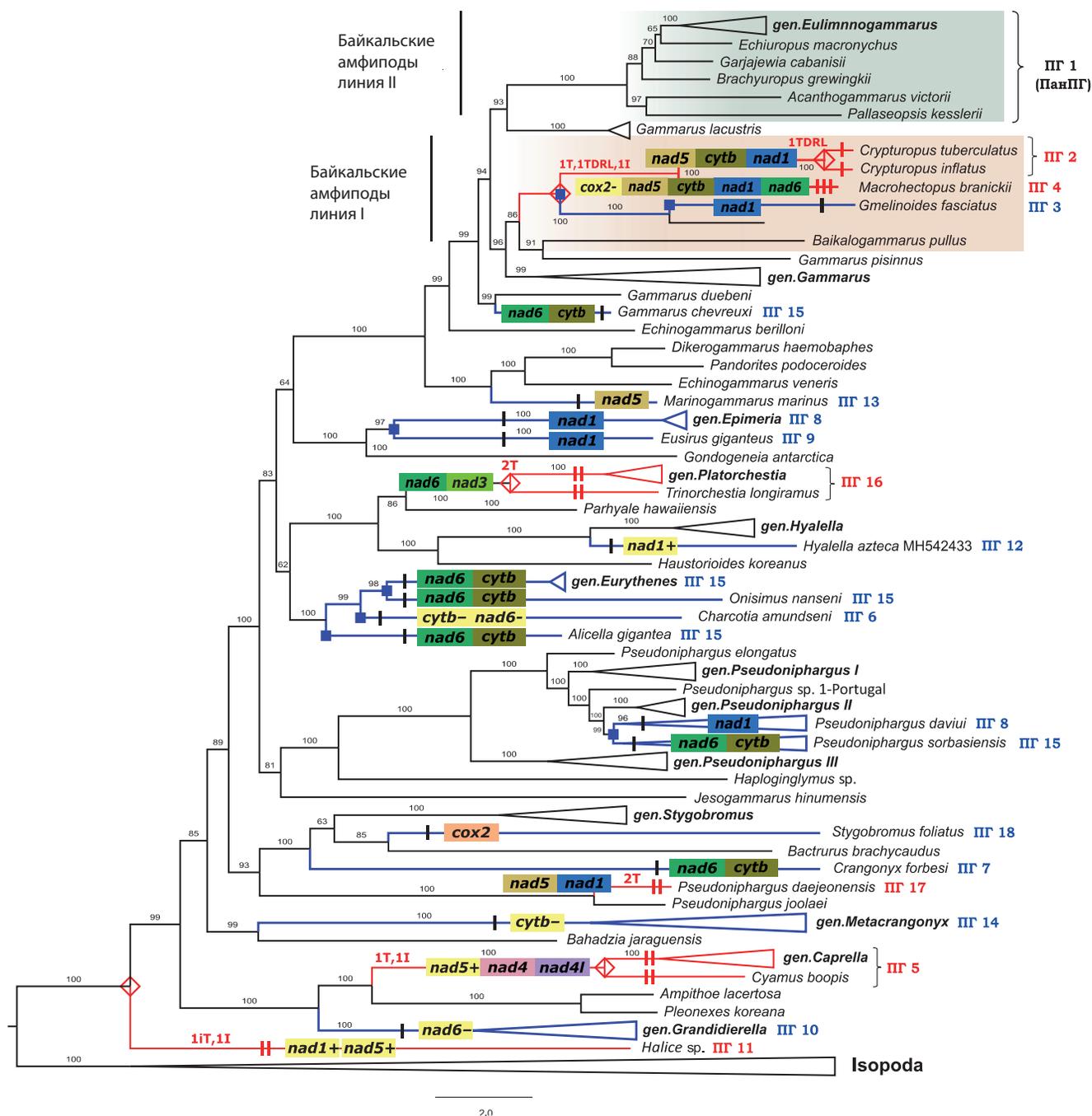
Гомоплазии ПГ мт генома нередко можно встретить у разных видов беспозвоночных (Kilpert et al., 2012; Tan et al., 2019; Castellucci et al., 2022). Мы отметили это явление с учетом более редкой встречаемости изменения положений белок-кодирующих генов по сравнению с тРНК (Jühling et al., 2012), а также как возможный конвергентный признак у отдаленных групп амфипод. У многих пресноводных и морских амфипод наблюдаются схожие для разных видов перестройки в одном и том же кластере генов *nad5-nad4---nad6---nad1*. Перестройки у четырех байкальских видов также затрагивают данный кластер генов.

В мт геномах беспозвоночных ПГ не всегда совпадает с их принадлежностью к виду, роду и другим таксонам, выявленным на основании морфологических признаков. Тем не менее, используя филогенетический анализ совместно с анализом ПГ, можно обнаружить клады видов с наиболее разнообразными ПГ и узлы на дереве, внутри которых данный признак был подвержен наибольшим изменениям (группы с неодинаковыми ПГ). Например, среди видов одного рода *Pseudoniphargus* и у видов байкальской линии I, где ряд ПГ не совпадает с ПГ, характерными для этой группы.

В литературе отмечается, что неравномерность распределения перестроек генов на филогенетическом дереве характерна для подтипа Нехарода и ракообразных, что, возможно, объясняется параллельной эволюцией данного признака (Tan et al., 2017; Moreno-Carmona et al., 2021). Примером могут служить перестройки генов у ряда таксонов кладоцер (дафний, босмин и др.), где мт геном соответствует ПанПГ, но не у всех видов (17 из 32 видов соответствуют ПанПГ, 15 видов – 9 другим ПГ) (Castellucci et al., 2022) и др. Среди видов рода *Gammarus*, большинство из которых соответствует ПанПГ, есть один вид, имеющий изменения положения генов *cytb* и *nad6*, – *G. chevreuxi*. Еще более заметное различие в порядке генов встречается у мт геномов вида *Hyalella azteca* (Saussure, 1858): в геноме MH542433 указана инверсия гена *nad1*, отсутствующая в геноме MT672041. F. Zapelloni с коллегами (2021) также отмечают инверсии и инверсные транспозиции внутри рода *Hyalella* S.I. Smith, 1874, предполагая, что они могут быть связаны с принадлежностью к южным и северным популяциям.

Так как ПанПГ является предковым для обеих линий байкальских видов, все изменения в ПГ, вероятно, возникли у разных видов байкальских амфипод в ходе адаптивной радиации в озере. Перестройки мт геномов, по-видимому, происходили в несколько шагов в направлении более разнообразных ПГ байкальских видов из линии I. Наименее измененные порядки генов (ПГ 3, ПГ 8–10, ПГ 12–14) являются плезиоморфными. Байкальские амфиподы из линии I, относящиеся к семействам Crypturopodidae, Micruropodidae и Macrohectopodidae (Kamaltynov, 1999), имеют сильно измененные порядки генов, которые не встречаются у других амфипод. Следует отметить также, что неполная нуклеотидная последовательность мт генома вида *L. vortex* (Micruropodidae) имеет тот же порядок генов, что и ПанПГ. Виды этой группы (за исключением вида *M. branickii*) – мелководные и устойчивые к повышенным температурам (*Gmelinoides fasciatus* (Stebbing, 1899), *C. tuberculatus*, *L. vortex*) (Камалтынов, 2001).

Один из возможных факторов, обеспечивающих повышенную вероятность изменений порядка генов в части видов, – ослабление стабилизирующего отбора, вызванное низким эффективным размером и небольшим генетическим разнообразием популяции предка (Charlesworth, 2009; Lavrov, Pett, 2016) в ходе эволюционной истории видов (Shao et al., 2003). Мелководные виды амфипод озера Байкал могут быть, вероятно, более склонны к резкому



**Рис. 4.** Филогенетическое дерево, построенное на основе аминокислотных последовательностей 13 белок-кодирующих мт генов 128 видов амфипод.

На ветвях отмечены гены, изменившие положение по сравнению с ПанПГ. Вертикальные черты на ветвях соответствуют числу шагов перестроек (CREX). Гены, изменившие кодирующую цепь (i, iT), выделены желтым цветом. Ветви и ПГ, отмеченные красным цветом, показывают сложные типы перестроек в данной группе видов, синим выделены простые типы изменений. Вид *L. vortex* не подписан, так как не был включен в анализ ПГ.

уменьшению численности своей популяции из-за изменений окружающей среды. В частности, периоды ледниковья в ходе геологической истории озера Байкал приводили к периодам вымирания видов (Goldberg et al., 2010; Mats et al., 2011), что было показано для юго-западной популяции *G. fasciatus* (Bukin et al., 2018) и *M. branickii* (Петунина и др., 2023). Кроме того, повышенный уровень зараженности паразитами (микроспоридиями), наблюдающийся

у *G. fasciatus* и *M. branickii* (Петунина и др., 2023), может вести к уменьшению числа самцов в популяции. Поскольку паразиты не имеют митохондрий и используют аденозинтрифосфат (АТФ) хозяина, вполне вероятно, что для адаптации к их негативному воздействию у ряда видов амфипод возникла необходимость в перестройках мтДНК для повышения эффективности продукции АТФ (Bukin et al., 2018).

Виды амфипод с полностью определенными нуклеотидными последовательностями мт геномов в настоящее время принадлежат к разным таксонам и обитают в разнообразных биотопах озера Байкал. Многие исследователи предпринимают попытки определить, существует ли ассоциация изменений в ПГ с экологическими особенностями видов, временем их дивергенции или особенностями жизненного цикла, местообитаний и т. д., приходя к выводу, что изменения ПГ многофакторны, могут быть связаны с эволюцией ряда биохимических, метаболических признаков (Romanova, Sherbakov, 2019; Tan et al., 2019; Castellucci et al., 2022; Benito et al., 2024). В исследовании изопод вида *Janira maculosa* Leach, 1814 F. Kilpert с коллегами предположили, что часто наблюдаемые перемещения генов в мт геноме частично могут быть объяснены низкой степенью сложности мт генома по сравнению с ядерным (Kilpert et al., 2012). Перемещения генов, особенно на противоположную кодирующую цепь, затрагивают процесс транскрипции мт генома и могут влиять на ее эффективность.

Байкальские амфиподы из линии II – более таксономически и экологически разнообразная группа видов, порядок белок-кодирующих генов которых остался идентичным по сравнению с предковым состоянием, но в то же время у разных видов есть различия в количестве и порядке тРНК генов (Romanova et al., 2016; Romanova, Sherbakov, 2019). В некоторых работах высказывается мнение, что изменения в положениях и количестве генов тРНК в мт геномах происходят чаще, чем в положениях рРНК и белок-кодирующих генов (Pääbo et al., 1991; Jühling et al., 2012).

Один из предполагаемых механизмов изменения порядка генов мт генома – частичная или полная дупликация мт генома с последующими делециями дублированных участков (Jühling et al., 2012). Существующие модели расчета сценариев преобразования одного ПГ в другой основаны на предположении, что происходит равное количество событий дубликаций и делеций на каждом этапе TDRL. Известно, что оценка расстояния между двумя порядками генов в случае TDRL несимметрична и требует разного числа шагов при расчете сценариев в прямом и обратном направлениях (Bernt et al., 2007), тем не менее она не учитывает возможность того, что соотношение преобразований в процессе TDRL (1 дупликация на несколько делеций или наоборот) может быть более сложным, а рассчитанное минимальное число шагов для возникновения перестройки может отличаться в большую сторону от того, которое минимально необходимо согласно реальному сценарию событий.

Во время удвоения полного генома до димерной молекулы при репликации с помощью механизма «катящегося кольца» в мтДНК (Fučíková et al., 2016; Xia et al., 2016; Wang et al., 2022) и делеции любой из его копий могут быть дублированы как тРНК, так и группа белок-кодирующих генов. Известно о различных механизмах образования делеций у разных групп, реализуемых через рекомбинацию и ошибки репликации (Nissanka et al., 2019; Oliveira et al., 2020). Если в димере мтДНК есть мутация

промотора в одной из дублированных копий, это может привести к образованию псевдогенов (Lavrov et al., 2002).

Косвенным свидетельством существования такого механизма у байкальских видов могут служить копия гена *cox2* и три фрагмента копий *atp8* у *M. branickii*, а также присутствие укороченных копий генов в протяженных некодирующих последовательностях у *G. fasciatus* (*atp6*, *nad4l*), *Brachyuropus grewingkii* (Dybowski, 1874) (*atp8*, *cox2*, *nad2*), *Garjajewia cabanisii* (Dybowski, 1874) (*cytb*), включая копии тРНК генов (Romanova et al., 2020). Такие области, обычно не характерные для мт геномов животных, могут свидетельствовать о прошлых событиях дубликации и последующей дегенерации этих участков (Boore, 1999), а также служить фактором, влияющим на появление перестроек генов.

В некоторых случаях порядок генов в мт геномах может стать дополнительным диагностическим признаком в таксономии (Lavrov, Lang, 2005; Tan et al., 2019). Чтобы объяснить причины повышенной скорости эволюции в мтДНК и высокую вариабельность порядков генов у байкальских и других видов амфипод, потребуются дальнейшие исследования.

## Заключение

Анализ порядка белок-кодирующих генов в мт геномах байкальских амфипод позволил определить, что ПанПГ является предковым для обеих линий байкальских видов. Изменения ПГ в мт геномах нескольких байкальских видов произошли в ходе эволюционной истории данных видов в озере. Рассчитанные при помощи CREx сценарии, скорее всего, предполагают упрощенные механизмы генных перестроек, поскольку события дубликаций и делеций, приводящие к ним, могут быть неравновероятны.

Анализ перестроек ПГ и филогенетический анализ белок-кодирующих последовательностей мт геномов амфипод, доступных в базе данных, позволили выявить четыре группы неодинаковых ПГ со сложным типом перестроек генов: это клады, объединяющие виды надсемейства *Lysianassoidea*, рода *Epimeria*, рода *Pseudoniphargus*, а также виды байкальских амфипод линии I. Мы предполагаем, что более древнее происхождение видов линии I байкальских видов по сравнению с линией II может частично объяснять более разнообразный ПГ. Также предполагается, что низкий эффективный размер популяции у байкальских амфипод линии I мог быть одним из факторов, ослабляющим действие стабилизирующего отбора и позволяющим закрепиться мутациям в виде перестроек генов мт генома.

## Список литературы / References

- Камалтынов Р.М. Амфиподы (Amphipoda: Gammaroidea). В: Аннотированный список фауны озера Байкал и его водосборного бассейна. Новосибирск: Наука, 2001;572-831  
[Kamaltynov R.M. Amphipods (Amphipoda: Gammaroidea). In: Index of Animal Species Inhabiting Lake Baikal and its Catchment Area. Novosibirsk: Nauka Publ., 2001;572-831 (in Russian)]  
Петунина Ж.В., Ваврищук Н.В., Букин Ю.С., Романова Е.В. Вариабельность морфологических и генетических признаков *Macrohectopus branickii* (Dyb., 1874). Известия Иркутского государ-

- ственного университета. Серия: Биология. Экология. 2023;46: 18-28. doi 10.26516/2073-3372.2023.46.18
- [Petunina J.V., Vavrishchuk N.V., Bukin Yu.S., Romanova E.V. Variability of morphological and genetic characteristics of *Macrohectopus branickii* (Dyb., 1874) (Amphipoda, Macrohectopidae). *Izvestiya Irkutskogo Gosudarstvennogo Universiteta. Seriya: Biologiya. Ecologia = The Bulletin of Irkutsk State University. Series: Biology. Ecology*. 2023;46:18-28. doi 10.26516/2073-3372.2023.46.18 (in Russian)]
- Abascal F., Zardoya R., Telford M.J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*. 2010;38(Suppl. 2):W7-W13. doi 10.1093/nar/gkq291
- Basso A., Babbucci M., Pauletto M., Riginella E., Patarnello T., Negrisolo E. The highly rearranged mitochondrial genomes of the crabs *Maja crispata* and *Maja squinado* (Majidae) and gene order evolution in Brachyura. *Sci Rep*. 2017;7(1):4096. doi 10.1038/s41598-017-04168-9
- Benito J.B., Porter M.L., Niemiller M.L. Comparative mitogenomic analysis of subterranean and surface amphipods (Crustacea, Amphipoda) with special reference to the family Crangonyctidae. *BMC Genomics*. 2024;25(1):298. doi 10.1186/s12864-024-10111-w
- Bernt M., Merkle D., Ramsch K., Fritzsche G., Perseke M., Bernhard D., Schlegel M., Stadler P.F., Middendorf M. CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics*. 2007; 23(21):2957-2958. doi 10.1093/bioinformatics/btm468
- Boore J.L. Animal mitochondrial genomes. *Nucleic Acids Res*. 1999; 27(8):1767-1780. doi 10.1093/nar/27.8.1767
- Bukin Yu.S., Petunina J.V., Sherbakov D.Yu. The mechanisms for genetic diversity of Baikal endemic amphipod *Gmelinoides fasciatus*: relationships between the population processes and paleoclimatic history of the lake. *Russ J Genet*. 2018;54(9):1059-1068. doi 10.1134/S1022795418090053
- Castellucci F., Luchetti A., Mantovani B. Exploring mitogenome evolution in Branchiopoda (Crustacea) lineages reveals gene order rearrangements in Cladocera. *Sci Rep*. 2022;12(1):4931. doi 10.1038/s41598-022-08873-y
- Charlesworth B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 2009;10(3):195-205.
- Drozdova P.B., Madyarova E.V., Gurkov A.N., Saranchina A.E., Romanova E.V., Petunina J.V., Peretolchina T.E., Sherbakov D.Y., Timofeyev M.A. Lake Baikal amphipods and their genomes, great and small. *Vavilov J Genet Breed*. 2024;28(3):317-325. doi 10.18699/vjgb-24-36
- Fourdril S., de Frias Martins A.M., Bäckeljau T. Relation between mitochondrial DNA hyperdiversity, mutation rate and mitochondrial genome evolution in *Melarhaphe neritoides* (Gastropoda: Littorinidae) and other Caenogastropoda. *Sci Rep*. 2018;8(1):17964. doi 10.1038/s41598-018-36428-7
- Fučíková K., Lewis P.O., González-Halphen D., Lewis L.A. Gene arrangement convergence, diverse intron content, and genetic code modifications in mitochondrial genomes of Sphaeropleales (Chlorophyta). *Genome Biol Evol*. 2016;6(8):2170-2180. doi 10.1093/gbe/evu172
- Goldberg E.L., Chebykin E.P., Zhuchenko N.A., Vorobyeva S.S., Stepanova O.G., Khlystov O.M., Ivanov E.V., Weinberg E., Gvozdkov A.N. Uranium isotopes as proxies of the environmental history of the Lake Baikal watershed (East Siberia) during the past 150ka. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2010;294(1-2):16-29. doi 10.1016/j.palaeo.2009.08.030
- Gouy M., Guindon S., Gascuel O. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010;27(2):221-224. doi 10.1093/molbev/msp259
- Hartmann T., Bernt M., Middendorf M. An exact algorithm for sorting by weighted preserving genome rearrangements. *IEEE/ACM Trans Comput Biol Bioinform*. 2019;16(1):52-62. doi 10.1109/TCBB.2018.2831661
- Hou Z., Sket B., Li S. Phylogenetic analyses of Gammaridae crustacean reveal different diversification patterns among sister lineages in the Tethyan region. *Cladistics*. 2014;30(4):352-365. doi 10.1111/cla.12055
- Jühling F., Pütz J., Bernt M., Donath A., Middendorf M., Florentz C., Stadler P.F. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Res*. 2012;40(7):2833-2845. doi 10.1093/nar/gkr1131
- Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., Von Haeseler A., Jermini L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587-589. doi 10.1038/nmeth.4285
- Kamaltynov R.M. On the higher classification of Lake Baikal amphipods. *Crustaceana*. 1999;72(8):933-944. doi 10.1163/156854099503834
- Kilpert F., Podsiadlowski L. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics*. 2006;7:241. doi 10.1186/1471-2164-7-241
- Kilpert F., Held C., Podsiadlowski L. Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. *Mol Phylogenet Evol*. 2012;64(1):106-117. doi 10.1016/j.ympev.2012.03.013
- Lavrov D.V., Lang B.F. Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. *Syst Biol*. 2005;54(4):651-659. doi 10.1080/10635150500221044
- Lavrov D.V., Pett W. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol Evol*. 2016;8(9):2896-2913. doi 10.1093/gbe/evw195
- Lavrov D.V., Boore J.L., Brown W.M. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. *Mol Biol Evol*. 2002; 19(2):163-169. doi 10.1093/oxfordjournals.molbev.a004068
- Li J.Y., Zeng C., Yan G.Y., He L.S. Characterization of the mitochondrial genome of an ancient amphipod *Halice* sp. MT-2017 (Pardaliscidae) from 10,908 m in the Mariana Trench. *Sci Rep*. 2019;9(1): 2610. doi 10.1038/s41598-019-38735-z
- Mamos T., Grabowski M., Rewicz T., Bojko J., Strapagiel D., Burzyński A. Mitochondrial genomes, phylogenetic associations, and SNP recovery for the key invasive Ponto-Caspian amphipods in Europe. *Int J Mol Sci*. 2021;22(19):10300. doi 10.3390/ijms221910300
- Mats V.D., Shcherbakov D.Y., Efimova I.M. Late Cretaceous-Cenozoic history of the Lake Baikal depression and formation of its unique biodiversity. *Stratigr Geol Correl*. 2011;19(4):404-423. doi 10.1134/S0869593811040058
- Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R., Teeling E. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530-1534. doi 10.1093/molbev/msaa015
- Moreno-Carmona M., Cameron S.L., Prada Quiroga C.F. How are the mitochondrial genomes reorganized in Hexapoda? Differential evolution and the first report of convergences within Hexapoda. *Gene*. 2021;791:145719. doi 10.1016/j.gene.2021.145719
- Mueller R.L., Boore J.L. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Mol Biol Evol*. 2005;22(10):2104-2112. doi 10.1093/molbev/msi204
- Nissanka N., Minczuk M., Moraes C.T. Mechanisms of mitochondrial DNA deletion formation. *Trends Genet*. 2019;35(3):235-244. doi 10.1016/j.tig.2019.01.001
- Oliveira M.T., Pontes C.B., Ciesielski G.L. Roles of the mitochondrial replisome in mitochondrial DNA deletion formation. *Genet Mol Biol*. 2020;43(Suppl. 1):e20190069. doi 10.1590/1678-4685-GMB-2019-0069

- Pääbo S., Thomas W.K., Whitfield K.M., Kumazawa Y., Wilson A.C. Rearrangements of mitochondrial transfer RNA genes in marsupials. *J Mol Evol.* 1991;33(5):426-430. doi 10.1007/BF02103134
- Rambaut A. FigTree. Latest Version – v1.4.4. 2018. Available at: <http://tree.bio.ed.ac.uk/software/figtree/>
- Rivarola-Duarte L., Otto C., Jühling F., Schreiber S., Bedulina D., Jakob L., Gurkov A., ... Sartoris F., Pörtlner H.O., Timofeyev M., Luckenbach T., Stadler P.F. A first Glimpse at the genome of the Baikalian amphipod *Eulimnogammarus verrucosus*. *J Exp Zool B Mol Dev Evol.* 2014;322(3):177-189. doi 10.1002/jez.b.22560
- Romanova E.V., Sherbakov D.Y. Different rates of molecular evolution of mitochondrial genes in Baikalian and non-Baikalian amphipods. *Limnol Freshwater Biol.* 2019;(6):339-344. doi 10.31951/2658-3518-2019-A-6-339
- Romanova E.V., Aleoshin V.V., Kamaltynov R.M., Mikhailov K.V., Logacheva M.D., Sirotnina E.A., Gornov A.Y., Anikin A.S., Sherbakov D.Y. Evolution of mitochondrial genomes in Baikalian amphipods. *BMC Genomics.* 2016;17(Suppl. 14):1016. doi 10.1186/s12864-016-3357-z
- Romanova E.V., Bukin Y.S., Mikhailov K.V., Logacheva M.D., Aleoshin V.V., Sherbakov D.Y. Hidden cases of tRNA gene duplication and remolding in mitochondrial genomes of amphipods. *Mol Phylogenet Evol.* 2020;144:106710. doi 10.1016/j.ympev.2019.106710
- Romanova E.V., Bukin Y.S., Mikhailov K.V., Logacheva M.D., Aleoshin V.V., Sherbakov D.Y. The mitochondrial genome of a freshwater pelagic amphipod *Macrohectopus branickii* is among the longest in Metazoa. *Genes (Basel).* 2021;12(12):2030. doi 10.3390/genes12122030
- Shao R., Downton M., Murrell A., Barker S.C. Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. *Mol Biol Evol.* 2003;20(10):1612-1619. doi 10.1093/molbev/msg176
- Sherbakov D.Y. Molecular phylogenetic studies on the origin of biodiversity in Lake Baikal. *Trends Ecol Evol.* 1999;14(3):92-95. doi 10.1016/s0169-5347(98)01543-2
- Sterling-Montealegre R.A., Prada C.F. Variability and evolution of gene order rearrangement in mitochondrial genomes of arthropods (except Hexapoda). *Gene.* 2024;892:147906. doi 10.1016/j.gene.2023.147906
- Tan M.H., Gan H.M., Lee Y.P., Poore G.C.B., Austin C.M. Digging deeper: new gene order rearrangements and distinct patterns of codons usage in mitochondrial genomes among shrimps from the Axiidea, Gebiidea and Caridea (Crustacea: Decapoda). *PeerJ.* 2017;5:e2982. doi 10.7717/peerj.2982
- Tan M.H., Gan H.M., Lee Y.P., Bracken-Grissom H., Chan T.Y., Miller A.D., Austin C.M. Comparative mitogenomics of the Decapoda reveals evolutionary heterogeneity in architecture and composition. *Sci Rep.* 2019;9(1):10756. doi 10.1038/s41598-019-47145-0
- The Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* 2024;52(W1):83-94. <https://doi.org/10.1093/nar/gkae410>
- Wang R., Li X., Qi J. The complete paternally inherited mitochondrial genomes of three clam species in genus *Macridiscus* (Bivalvia: Veneridae): a TDRL model of dimer-mitogenome rearrangement of doubly uniparental inheritance. *Front Mar Sci.* 2022;9:1016779. doi 10.3389/fmars.2022.1016779
- Xia Y., Zheng Y., Murphy R.W., Zeng X. Intraspecific rearrangement of mitochondrial genome suggests the prevalence of the tandem duplication-random loss (TDLR) mechanism in *Quasipaa bou-lengeri*. *BMC Genomics.* 2016;17(1):965. doi 10.1186/s12864-016-3309-7
- Xu W., Jameson D., Tang B., Higgs P.G. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. *J Mol Evol.* 2006;63(3):375-392. doi 10.1007/s00239-005-0246-5
- Zapelloni F., Jurado-Rivera J.A., Jaume D., Juan C., Pons J. Comparative mitogenomics in *Hyaella* (Amphipoda: Crustacea). *Genes (Basel).* 2021;12(2):292. doi 10.3390/genes12020292
- Zardoya R. Recent advances in understanding mitochondrial genome diversity. *F1000Res.* 2020;9:270. doi 10.12688/f1000research.21490.1

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 23.12.2024. После доработки 09.07.2025. Принята к публикации 14.08.2025.

doi 10.18699/vjgb-26-16

## Морфологический и молекулярно-генетический анализ полиморфизма рода *Iris* L. в Республике Башкортостан и Оренбургской области

Э.А. Аухадиева<sup>1</sup>, А.Г. Блинов<sup>2</sup>, Е.И. Вшивцева<sup>2, 4</sup> , З.Х. Шигапов<sup>3</sup>, А.Р. Ишбирдин<sup>5</sup>, Я.М. Голованов<sup>3</sup>, А.В. Крюкова<sup>3</sup>

<sup>1</sup> Уфимский научно-исследовательский институт медицины труда и экологии человека, Уфа, Россия

<sup>2</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>3</sup> Южно-Уральский ботанический сад-институт – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия

<sup>4</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>5</sup> Уфимский университет науки и технологий, Уфа, Россия

 [evsivceva4@gmail.com](mailto:evsivceva4@gmail.com)

**Аннотация.** *Iris* L. (Iridaceae Juss.) – это космополитный род, включающий от 200 до 340 видов, распространенных по всему Северному полушарию. Хотя *Iris* является самой разнообразной группой в семействе Iridaceae, существует множество неопределенностей относительно его таксономического состава и систематики. Цель настоящего исследования – поиск молекулярных маркеров и таксономически значимых морфологических признаков генеративной и вегетативной сферы и оценка их информативности для выявления филогенетического родства представителей рода *Iris*. Образцы ирисов взяты из природных популяций Республики Башкортостан и Оренбургской области, а также получены из ботанических садов Санкт-Петербурга, Бонна, Байройта и Брно. В результате построения структуры изменчивости морфометрических показателей 11 видов найдено 10 таксономических индикаторов, общих для анализируемых таксонов и характеризующихся относительно низкими общей и согласованной изменчивостью: длина и ширина наружных долей околоцветника, длина и ширина внутренних долей околоцветника, длина тычиночной нити, пыльника и пестика, ширина плода, а также длина и ширина семени. Установлены нуклеотидные последовательности *trnL-trnF* фрагментов хлоропластной ДНК для 13 образцов четырех видов дикорастущей флоры Республики Башкортостан и Оренбургской области: *Iris pumila* L., *I. scariosa* Willd. ex Link., *I. pseudacorus* L., *I. sibirica* L. Полученные последовательности были использованы для построения филогенетического дерева совместно с *trnL-trnF* последовательностями еще семи видов ирисов, извлеченных из базы данных. На филогенетическом древе формируется пять групп (кластеров): 1) *I. pumila* L., *I. scariosa* Willd. ex Link.; 2) *I. pseudacorus* L., *I. setosa* Pall. ex Link.; 3) *I. lactea* Pall.; 4) *I. sibirica* L., *I. sanguinea* Hornem.; 5) *I. spuria* L., *I. xanthospuria* Mathew & Baytop., *I. foetidissima* L., *I. sintenisii* Janka. Выявленные кластеры по составу входящих в них видов практически полностью совпадают с кластерами, обнаруженными при морфологическом анализе. Для подтверждения полученных результатов проведен филогенетический анализ интересующих видов еще на двух хлоропластных последовательностях, доступных в базе данных: *matK* и *trnS-trnG*. Кластеризация исследованных видов на *trnS-trnG* и *matK* полностью совпадает с кластеризацией на *trnL-trnF*. Таким образом, можно констатировать, что выявленные для родового комплекса *Iris* морфологические признаки демонстрируют значимую диагностическую ценность при проведении таксономических исследований.  
**Ключевые слова:** *Iris*; таксономия; таксономические индикаторы; нуклеотидные последовательности; *trnL-trnF* последовательность хлоропластной ДНК; филогенетическое древо

**Для цитирования:** Аухадиева Э.А., Блинов А.Г., Вшивцева Е.И., Шигапов З.Х., Ишбирдин А.Р., Голованов Я.М., Крюкова А.В. Морфологический и молекулярно-генетический анализ полиморфизма рода *Iris* L. в Республике Башкортостан и Оренбургской области. *Вавиловский журнал генетики и селекции*. 2026;30(1):146-156. doi 10.18699/vjgb-26-16

**Финансирование.** Работа выполнена при финансовой поддержке государственного бюджетного проекта FWNR-2022-0016.

## Morphological and molecular genetic analysis of the genus *Iris* L. polymorphism in the Republic of Bashkortostan and the Orenburg Region

E.A. Aukhadieva<sup>1</sup>, A.G. Blinov<sup>2</sup>, E.I. Vshivtseva<sup>2, 4</sup> , Z.Kh. Shigapov<sup>3</sup>, A.R. Ishbirdin<sup>5</sup>, Ya.M. Golovanov<sup>3</sup>, A.V. Kryukova<sup>3</sup>

<sup>1</sup> Ufa Research Institute of Occupational Health and Human Ecology, Ufa, Russia

<sup>2</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>3</sup> South-Ural Botanical Garden-Institute – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia

<sup>4</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>5</sup> Ufa University of Science and Technology, Ufa, Russia

 [evsivceva4@gmail.com](mailto:evsivceva4@gmail.com)

**Abstract.** *Iris* is a cosmopolitan genus comprising 200 to 340 species distributed throughout the Northern Hemisphere. Although *Iris* is the most diverse group in the family Iridaceae, there are many uncertainties regarding its taxonomic composition and systematics. The aim of this study was to search for taxonomically significant morphological characters of the generative and vegetative spheres and molecular markers with subsequent assessment of their informativeness in identifying phylogenetic relationships and compliance with the most relevant modern classification systems of the genus *Iris*. As a result of constructing the structure of variability of morphometric parameters of 11 species, 10 taxonomic indicators were identified that were common to the analyzed taxa and were characterized by relatively low total and coordinated variability: length and width of the outer perianth lobes, length and width of the inner perianth lobes, length of the filament, anther and pistil, fruit width, as well as seed length and width. Nucleotide sequences of *trnL-trnF* fragments of chloroplast DNA were established for 13 samples of four species of wild flora of the Republic of Bashkortostan and the Orenburg Region: *Iris pumila* L., *I. scariosa* Willd. ex Link., *I. pseudacorus* L., *I. sibirica* L. The obtained sequences were used to construct a phylogenetic tree together with *trnL-trnF* sequences of seven more iris species extracted from the database. The tree contained five clusters: (1) *I. pumila*, *I. scariosa*; (2) *I. pseudacorus*, *I. setosa* Pall. ex Link; (3) *I. lactea* Pall.; (4) *I. sibirica*, *I. sanguinea* Hornem.; (5) *I. spuria* L., *I. xanthospuria* Mathew & Baytop., *I. foetidissima* L., *I. sintenisii* Janka. By the composition of their species, the identified clusters almost completely coincided with the clusters found during the morphological analysis. To confirm the obtained results, a phylogenetic analysis of the species of interest was performed on two more chloroplast sequences available in the database: *matK* and *trnS-trnG*. Clustering of the studied species on *trnS-trnG* and *matK* completely coincided with clustering on *trnL-trnF*. Thus, we can state that the morphological features identified for the *Iris* generic complex work in the taxonomic direction. The analysis also showed that *I. scariosa* from natural populations of the Republic of Bashkortostan and the Orenburg Region were identified correctly.

**Key words:** *Iris*; taxonomy; taxonomic indicators; nucleotide sequences; *trnL-trnF* chloroplast DNA sequence; phylogenetic tree

**For citation:** Aukhadieva E.A., Blinov A.G., Vshivtseva E.I., Shigapov Z.Kh., Ishbirdin A.R., Golovanov Ya.M., Kryukova A.V. Morphological and molecular genetic analysis of the genus *Iris* L. polymorphism in the Republic of Bashkortostan and the Orenburg Region. *Vavilovskii Zhurnal Genetiki i Selektcii = Vavilov J Genet Breed.* 2026;30(1):146-156. doi 10.18699/vjgb-26-16

## Введение

Род *Iris* L. – самый крупный и космополитный в семействе Iridaceae, распространен в основном в умеренных зонах Северного полушария, включает от 200 до 340 видов (Dorofeeva, Zhurbenko, 2020). Несмотря на огромный прогресс в изучении рода *Iris*, все еще существует множество неопределенностей относительно его таксономического состава и систематики. Родовые границы ирисов спорны, и последние данные, по-видимому, говорят в пользу гораздо более узких границ (Crespo et al., 2015; Boltentkov et al., 2020). К тому же состав исследуемого рода периодически дополняется новыми описанными видами (Zhao, 1992; Mitic, 2002; Alexeeva, 2013), что чаще связано с морфологической изменчивостью и, как следствие, неоднократным описанием видов с широким ареалом (Болтенков и др., 2018).

Классификационные схемы рода *Iris* традиционно опираются на сравнительный анализ анатомо-морфологических и цитогенетических признаков (Доронькин, Красников, 1984; Mathew, 1989; Макаревич и др., 2001), дополняемый в последнее время данными молекулярно-биологических и биохимических исследований (Дорогина и др., 2012; Weber et al., 2020). Сложность заключается в том, что среди систематиков одинаково распространено как широкое, так и узкое понимание рода *Iris*. Единой, общепринятой классификации рода не существует, а наиболее популярные классификационные схемы (Родионенко, 1961; Mathew, 1989; Доронькин, 2006) различаются положением не только отдельных видов, но и более крупных систематических единиц – подродов и секций. Попытки разрешить противоречия с помощью современных методов зачастую дают неоднозначные результаты. Так, молекулярный RAPD-анализ И.Ф. Макаревич с коллегами (2001) показал большее соответствие системе Г.И. Родионенко (1961) в установлении филогенетических связей.

Исследования сибирских видов выявили неожиданные группировки: представители *Xyridion* и *I. sibirica* (*Limniris*) объединились в одну группу, тогда как *I. lactea* и *I. setosa* (*Limniris*) сформировали группу с представителями подрода *Iris*. Эти данные вступают в противоречие с существующими схемами, особенно в отношении полиморфного подрода *Limniris*. Отсутствие консенсуса по составу рода *Iris* среди современных исследователей требует проведения дальнейших комплексных исследований с целью уточнения филогенетических взаимоотношений видов, входящих в состав этого рода.

В литературе имеются публикации по использованию в биологии оценки структуры морфологической изменчивости для классификации морфологических признаков некоторых видов растений и животных по соотношению общей и согласованной изменчивости. Предложена классификация, объединяющая индикаторные признаки в четыре группы: эколого-биологические системные, биологические, генетические (таксономические) и экологические индикаторы (Ростова, 2002; Ишбирдин и др., 2005).

Эколого-биологические системные индикаторы характеризуются высокой общей и согласованной изменчивостью, сильной зависимостью от условий среды и способностью вызывать согласованные изменения всей морфосистемы организма (число бутонов, длина побега у *Cephalanthera rubra*, длина листовой пластинки у *Triticum aestivum* L.).

Биологические индикаторы обладают умеренной зависимостью от среды, низкой общей и высокой согласованной изменчивостью, определяют морфоструктуру растения (например, высоту побега, параметры листа у *C. rubra*, число метамеров вегетативной части годичного побега у *Helianthus annuus*).

Генетические (таксономические) индикаторы отличаются низкой общей и согласованной изменчивостью, высокой автономностью и слабой зависимостью от внешних усло-

вий (например, число листьев у *C. rubra*), являются наиболее информативными для таксономических исследований.

Экологические индикаторы проявляют сильное, относительно независимое варьирование, и их изменения слабо согласованы с общей системой организма, но они чувствительны даже к незначительным внешним воздействиям (например, число неразвитых цветков, признаки ветвления у *Rhinanthus* L.). Для таких видов, как *H. annuus* L., *C. rubra* L., *Panicum miliaceum* L. и ряда других цветковых растений, данные признаки уже обнаружены, для родового комплекса *Iris* корреляционный анализ морфологических признаков с целью выявления индикаторов проведен впервые.

Для установления филогенетических взаимоотношений видов рода *Iris* успешно применяются полиморфизмы ДНК в некодирующих регионах хлоропластного генома (Pleines et al., 2009). Среди наиболее часто используемых участков хлДНК – интрон *trnL* и межгенный спейсер *trnL-trnF*, а также другие вариабельные участки хлоропластного генома: межгенный спейсер *atpB-rbcL*, *trnS-trnG* и *trnH-psbA* (Козыренко и др., 2009; Болтенков и др., 2016).

Цель настоящей работы – поиск таксономически значимых морфологических признаков генеративной и вегетативной сферы и молекулярных маркеров с последующей оценкой их информативности при выявлении филогенетического родства и соответствия наиболее актуальным современным классификационным системам рода *Iris*.

## Материалы и методы

**Растительный материал.** Объектами морфологических исследований были дикорастущие представители видов рода *Iris* (*I. pseudacorus* L., *I. sibirica* L., *I. pumila* L., *I. scariosa* Willd. ex Link), собранные в природных популяциях Республики Башкортостан и Оренбургской области и интродуцированные на коллекционном участке Южно-Уральского ботанического сада-института Уфимского федерального исследовательского центра Российской академии наук в период 2019–2021 гг., а также виды, выращенные из семенного материала, полученного по делектусам из ботанических садов Санкт-Петербурга (*I. lactea* Pall., *I. setosa* Pall., *I. halophila* Pall.), Бонна (*I. sanguinea* Hornem., *I. spuria* L.), Байройта (*I. cartholiniae* Fomin) и Брно (*I. graminea* L.). Список исследованных видов, места сбора и географические координаты указаны в табл. 1.

В связи с большим количеством дискуссионных вопросов относительно таксономии видов рода *Iris* L. в настоящей работе мы ориентировались на классификации трех авторов: В. Mathew (1989), Г.И. Родионенко (1961, 2013) и В.М. Доронькина (2006).

**Морфологический анализ.** Морфометрические показатели были учтены у 25 растений каждого вида. Для оценки изменчивости проводили анализ следующих параметров: высота генеративного побега, длина и ширина листа, длина и ширина долей околоцветника, длина репродуктивных элементов (пестика и тычинки) – в фазе массового цветения, а также длина и ширина плодов и семян – после полного их созревания. Замеры морфометрических показателей побегов, цветков и плодов выполнены с помощью

линейки, семян – с применением микроскопа Levenhuk DTX 90 (Голубев, 1962). Стандартную статистическую обработку полученных данных проводили с использованием программ MS Excel 2010 и IBM SPSS Statistics 21 (Зайцев, 1984; Доспехов, 1985). Рассчитывали средние арифметические значения, ошибку среднего арифметического значения, среднеквадратичное отклонение. Для оценки внутривидовой изменчивости морфологических признаков вычисляли коэффициент вариации (CV, %) (Зайцев, 1990).

Выполнено сравнение биометрических показателей растений на предмет статистической значимости их различий за 2019–2021 гг. вегетации. Величины исследуемых показателей были проверены на нормальность распределения по критерию Колмогорова–Смирнова. Для сравнения независимых выборок с нормальным распределением использован однофакторный дисперсионный анализ, а для выборок, не подчиняющихся закону нормального распределения, – критерий Краскела–Уоллиса. Анализ показал, что для всех исследуемых параметров различие незначимо при уровне значимости  $W = 5\%$ , что позволяет провести оценку структуры изменчивости морфологических признаков и классифицировать их по группам.

Структура изменчивости признаков построена по методу Н.С. Ростовской (2002) с применением программ MS Excel 2010 и IBM SPSS Statistics 21. Морфологические признаки скоординированы в пространстве общей (коэффициент вариации признака) и согласованной (усредненный по признаку «квадрат коэффициента корреляции») изменчивости. По индикаторной роли признаки распределены в три группы: таксономические (генетические), биологические и эколого-биологические системные.

**Выделение суммарной ДНК, ПЦР амплификация и установление нуклеотидных последовательностей и филогенетический анализ.** Выделение суммарной ДНК производили с помощью набора DNeasy Plant Mini Kit (QIAGEN, Германия) согласно протоколу производителя. Для выделения ДНК использовали 50–100 мг высушенных листьев, полученных с растений каждого образца. Количество и качество выделенной ДНК определяли с помощью спектрофотометра NanoDrop2000 (Thermo Scientific, США) и электрофоретического разделения в 1 % агарозном геле, содержащем бромистый этидий (0.5 мг/мл) в 1хТАЕ.

ПЦР-амплификацию осуществляли, как описано в работе (Makarevich et al., 2003), с праймерами, специфичными для *trnL/trnF* последовательностям хлоропластных генов: *trnL* (5'-CGAAATCGGTAGACGCTACG-3'); *trnF* (5'-ATTTGAACTGGTGACACGAG-3').

Реакцию проводили в объеме 20 мкл с использованием BioMaster LR HS-PCR-Color (“Biolabmix”, Россия) в присутствии 10 пмоль каждого праймера и 30 нг суммарной ДНК. Полученные ПЦР-фрагменты разделяли в 1 % агарозном геле и выделяли из геля с помощью QIAquick Gel Extraction Kit (QIAGEN), а затем секвенировали. Реакции секвенирования выполняли с использованием 200 нг ДНК фрагмента и набора BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Scientific) на генетическом анали-

**Таблица 1.** Образцы ирисов, собранных в природных популяциях, места и географические координаты сборов

Вид	Место сбора материала	Географические координаты
<i>I. pseudacorus</i>	Россия, Республика Башкортостан, Кушнаренковский район, д. Ильмурзино	54.952323 с. ш., 55.811026 в. д.
	Россия, Республика Башкортостан, Кушнаренковский район, д. Тарабердино	55.092243 с. ш., 55.418299 в. д.
	Россия, Республика Башкортостан, Демский район г. Уфы, д. Романовка	54.736672 с. ш., 55.834274 в. д.
<i>I. sibirica</i>	Россия, Республика Башкортостан, г. Бирск	55.454523 с. ш., 55.537686 в. д.
	Россия, Республика Башкортостан, г. Агидель	55.844276 с. ш., 53.935868 в. д.
<i>I. scariosa</i>	Россия, Республика Башкортостан, Хайбуллинский район, Таштугайские горы	51.910798 с. ш., 58.494056 в. д.
	Россия, Оренбургская область, Кувандыкский район, д. Рамазаново	51.526943 с. ш., 57.446162 в. д.
	Россия, Оренбургская область, Светлинский район, пос. Тобольский	51.430173 с. ш., 61.164358 в. д.
<i>I. pumila</i> (желтые цветки)	Россия, Республика Башкортостан, Куюргазинский район, д. Лена	52.802508 с. ш., 55.603231 в. д.
<i>I. pumila</i> (синие цветки)	Россия, Республика Башкортостан, Куюргазинский район, д. Лена	52.802508 с. ш., 55.603231 в. д.
<i>I. pumila</i> (желтые цветки)	Россия, Республика Башкортостан, Куюргазинский район, с. Якшимбетово	52.578021 с. ш., 55.654848 в. д.
<i>I. pumila</i> (синие цветки)	Россия, Республика Башкортостан, Куюргазинский район, с. Якшимбетово	52.578021 с. ш., 55.654848 в. д.
<i>I. pumila</i> (желтые цветки)	Россия, Оренбургская область, Переволоцкий район, с. Рычковка	51.721765 с. ш., 54.523130 в. д.

затаре ABI 3130XL (Applied Biosystems, США) в Центре геномики СО РАН (<http://www.niboch.nsc.ru/doku.php/corefacility>). Нуклеотидные последовательности фрагментов *trnL-trnF* хлоропластной ДНК представлены в GenBank (№ PV335670–PV335682).

Для выполнения филогенетического анализа были использованы *trnL-trnF* последовательности хлоропластной ДНК ирисов, извлеченные из базы данных GenBank и полученные в настоящей работе. Выравнивание нуклеотидных последовательностей проводили с использованием программы MAFFT v7.312 (Katoh, Standley, 2013). Выбор модели нуклеотидных замен осуществлялся с помощью Байесовского информационного критерия. Построение филогенетического дерева выполняли по методу максимального правдоподобия с использованием модели Джукса–Кантора в программе IQ-tree (Trifunopoulos et al., 2016). Достоверность построенного дерева проверяли методом Bootstrap с количеством повторений, равным 1000.

## Результаты

### Анализ морфометрических параметров

Морфологические данные проанализированы за три года исследований. Проведено сравнение биометрических показателей на предмет статистической значимости их различий по годам. Анализ показал, что для всех исследуемых

параметров различие незначимо при уровне значимости  $W = 5 \%$ , что позволяет провести оценку структуры изменчивости морфологических признаков по соотношению общей и согласованной изменчивости и классифицировать их по группам. Обобщенные морфометрические параметры исследуемых видов приведены в табл. 2.

Анализ варибельности морфометрических параметров видов рода *Iris* выявил следующие закономерности:

1. Наибольший уровень изменчивости характерен для длины листовой пластинки (CV достигает 20.7 %) и высоты генеративных побегов (CV до 16.7 %).
2. Наиболее стабильные показатели – ширина внутренних долей околоцветника (CV = 2.2–5.3 %) и длина пестика (CV = 0.8–3.4 %), демонстрирующие наименьший разброс значений коэффициента вариации.
3. В характеристиках плодов и семян наблюдаются более высокая варибельность длины плода (CV до 9.4 %), по сравнению с его шириной (CV до 5.8 %), и умеренная изменчивость размеров семян (CV в диапазоне 0.9–5.9 %).

Морфологические признаки скоординированы в пространстве общей (коэффициент вариации признака) и согласованной (усредненный по признаку «квадрат коэффициента корреляции») изменчивости. Составлены графики по структуре изменчивости морфологических признаков. На рис. 1 в качестве примера показана структура изменчивости для *I. pseudacorus*. Структуры изменчивости для

**Таблица 2.** Обобщенные морфометрические параметры видов рода *Iris* (2019–2021 гг.)

Параметр	Диапазоны значений (min–max)	Коэффициенты вариации (CV, %)
Надземные побеги		
Высота генеративного побега, см	9.0–97.5	8.5–16.7
Длина листа, см	7.1–97.2	6.8–20.7
Ширина листа, см	0.5–1.9	4.1–10.9
Цветок		
Длина наружных долей, см	4.0–6.1	1.4–4.2
Ширина наружных долей, см	1.1–3.3	0.8–6.0
Длина внутренних долей, см	1.0–5.1	1.6–6.7
Ширина внутренних долей, см	0.2–2.2	2.2–5.3
Длина тычиночной нити, см	1.1–1.6	1.4–3.8
Длина пыльника, см	1.1–2.0	1.2–5.0
Длина пестика, см	2.8–4.4	0.8–3.4
Плоды и семена		
Длина плода, см	3.3–6.3	3.1–9.4
Ширина плода, см	1.0–2.1	0.7–5.8
Длина семени, см	0.4–0.8	0.9–4.5
Ширина семени, см	0.2–0.8	1.2–5.9

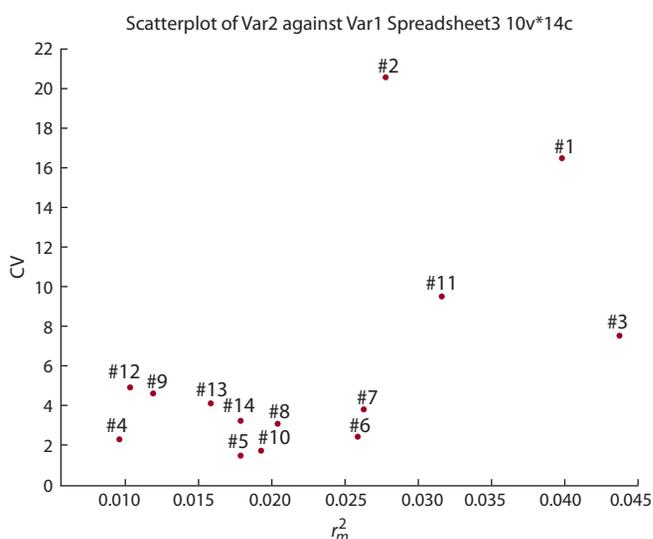
*I. sibirica*, *I. sanguinea*, *I. setosa*, *I. halophila*, *I. graminea*, *I. carthaliniae*, *I. spuria*, *I. pumila*, *I. scariosa*, *I. lactea* приведены в Приложении<sup>1</sup>.

В настоящем исследовании по индикаторной роли рассматриваемые признаки распределены в три группы: таксономические (генетические), биологические и эколого-биологические (см. рис. 1). В результате построения структуры изменчивости выявлено 10 таксономических индикаторов, общих для анализируемых таксонов и характеризующихся относительно низкими общей (CV = 0.8–6.0 %) и согласованной ( $r_m^2 = 0.005–0.078$ ) изменчивостью: длина и ширина наружных долей околоцветника, длина и ширина внутренних долей околоцветника, длина тычиночной нити, пыльника и пестика, ширина плода, а также длина и ширина семени.

Для установления родства исследуемых представителей рода *Iris* проведен кластерный анализ (иерархическая классификация, метод Варда) и построена дендрограмма различия-сходства видов по выявленным диагностическим маркерам (рис. 2). Виды на дендрограмме формируют шесть кластеров:

1. *I. carthaliniae*, *I. halophila*, *I. spuria*, представители подрода *Xyridion* (Родионенко, 1961) или *Limniris* (Mathew, 1989);
2. *I. lactea*, подрод *Limniris* (Родионенко, 1961; Mathew, 1989) или *Eremiris* (Доронькин, 2006);
3. *I. sibirica*, *I. sanguinea*, подрод *Limniris*;
4. *I. pumila*, *I. scariosa*, подрод *Iris*;
5. *I. graminea*, подрод *Xyridion* или *Limniris*;
6. *I. pseudacorus*, *I. setosa*, подрод *Limniris*.

<sup>1</sup> Приложение см. по адресу:  
<https://vavilovj-icg.ru/download/pict-2026-30/appx10.pdf>

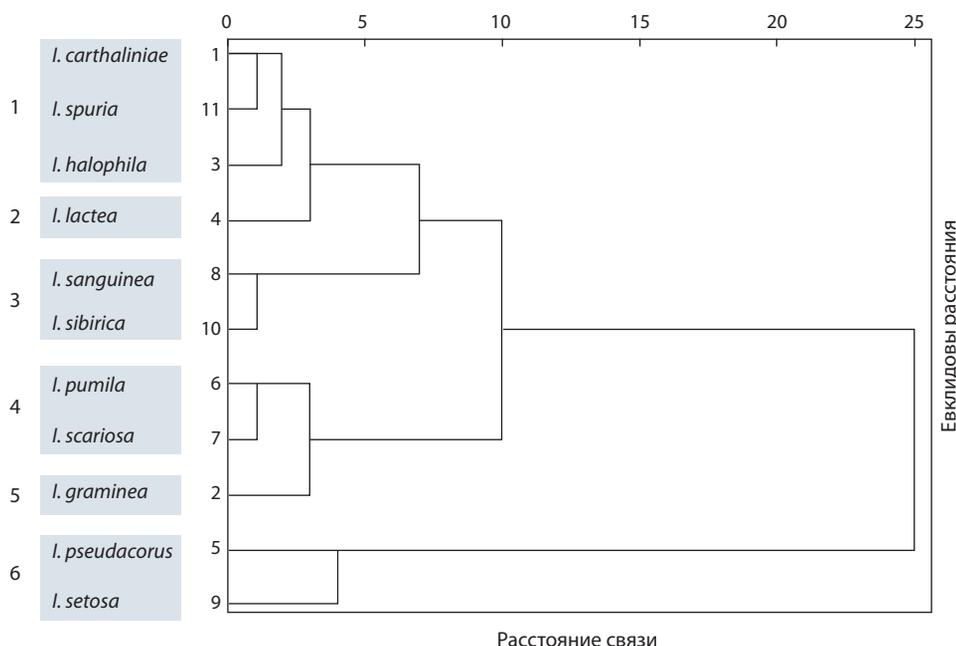


**Рис. 1.** Структура изменчивости морфологических признаков *I. pseudacorus*.

По оси абсцисс – согласованная изменчивость; по оси ординат – общая изменчивость. 1 – высота генеративного побега; 2 – длина листа; 3 – ширина листа; 4 – длина наружных долей околоцветника; 5 – ширина наружных долей околоцветника; 6 – длина внутренних долей околоцветника; 7 – ширина внутренних долей околоцветника; 8 – длина тычиночной нити; 9 – длина пыльника; 10 – длина пестика; 11 – длина плода; 12 – ширина плода; 13 – длина семени; 14 – ширина семени.

### Молекулярно-филогенетический анализ видов рода *Iris*

В ходе проведенной работы установлены нуклеотидные последовательности *trnL-trnF* фрагментов хлоропластной ДНК для 13 образцов четырех видов ирисов: *I. pumila*, *I. pseudacorus*, *I. sibirica*, *I. scariosa*. Эти последовательности были использованы для построения филогенетиче-



**Рис. 2.** Дендрограмма различия-сходства видов рода *Iris* по таксономическим индикаторам.

ского дерева совместно с *trnL-trnF* последовательностями еще семи видов ирисов, извлеченных из базы данных: *I. setosa*, *I. lactea*, *I. sanguinea*, *I. spuria*, *I. xanthospuria* Mathew & Baytop., *I. foetidissima* L., *I. sintenisii* Janka. Были проанализированы *trnL-trnF* последовательности из 63 образцов 11 видов. Полученное филогенетическое дерево представлено на рис. 3. Места сбора исследованных образцов ирисов отмечены на древе.

На древе присутствует пять кластеров:

1. *I. spuria*, *I. xanthospuria*, *I. sintenisii*, *I. foetidissima*;
2. *I. lactea*;
3. *I. sibirica* и *I. sanguinea*;
4. *I. pumila* и *I. scariosa*;
5. *I. pseudacorus* и *I. setosa*.

Данные кластеры по составу входящих в них видов практически полностью совпадают с кластерами, обнаруженными при морфологическом анализе (см. рис. 2).

Для подтверждения полученных результатов мы провели филогенетический анализ интересующих нас видов ирисов с использованием еще двух хлоропластных последовательностей, *matK* и *trnS-trnG*, доступных в базе данных NCBI. Дендрограмма, полученная с использованием последовательности *matK*, показала разделение образцов на пять групп, сходных с группировкой для последовательности *trnL-trnF* (рис. 4):

1. *I. spuria* и *I. halophila*;
2. *I. lactea*;
3. *I. sibirica* и *I. sanguinea*;
4. *I. pumila* и *I. scariosa*;
5. *I. pseudacorus* и *I. setosa*.

Однако есть небольшие различия между деревьями, полученными с помощью *trnL-trnF* и *matK*. При использовании последовательностей *matK* и *trnL-trnF* не подтверждены видовые различия в парах *I. pumila*/*I. scariosa*,

*I. sibirica*/*I. sanguinea* и *I. halophila*/*I. spuria*. А виды *I. pseudacorus* и *I. setosa* достоверно отличались друг от друга. Если межпопуляционные различия *trnL-trnF* последовательностей обнаружены только у двух видов, *I. scariosa* и *I. setosa*, то в случае *matK* последовательностей такие различия присутствуют у большего количества видов: *I. setosa*, *I. pseudacorus*, *I. sibirica*, *I. sanguinea*, *I. lactea*. Возможно, они присутствуют и у видов *I. pumila* и *I. scariosa*, но в базе данных мы обнаружили только три последовательности гена *matK* этих видов.

Информации о *trnS-trnG* последовательностях в базе данных меньше, чем о двух предыдущих. На рис. 5 представлено дерево, построенное на *trnS-trnG* последовательностях. На нем присутствуют 4 из 5 кластеров, полученных на предыдущих деревьях: 1) *I. lactea*; 2) *I. sibirica* и *I. sanguinea*; 3) *I. pumila* и *I. scariosa*; 4) *I. pseudacorus* и *I. setosa*. Пятый кластер отсутствует, потому что *trnS-trnG* последовательностей *I. spuria* и *I. halophila* в базе данных нет. Детально останавливаться на анализе *trnS-trnG* дерева мы не будем по причине малочисленности образцов некоторых видов. Однако основной вывод очевиден: кластеризация исследованных видов на *trnS-trnG* полностью совпадает с таковой на *trnL-trnF*- и *matK*-деревьях.

## Обсуждение

### Морфометрические параметры

Сравнительный анализ морфометрических параметров исследуемых видов с литературными данными (Волкова, 2010) выявил высокую степень соответствия показателей надземной части у большинства видов, с незначительными отклонениями у некоторых видов. Обнаруженные отличия, предположительно, обусловлены спецификой условий культивирования, что подтверждается, в част-



**Рис. 3.** Филогенетическое древо, построенное на основе *trnL-trnF* последовательностей хлДНК ирисов с использованием метода ML.

Здесь и на рис. 4 и 5: места сборов и номера нуклеотидных последовательностей из базы данных GenBank отмечены на древе.

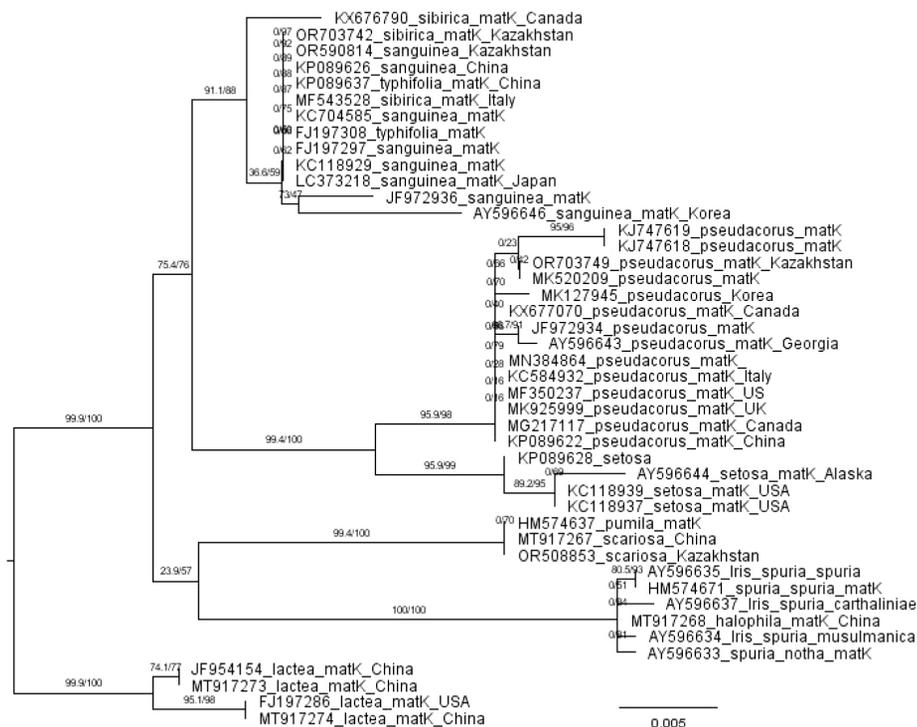
ности, влиянием амплитуды температур на изменчивость морфометрических параметров генеративного побега и листа у *I. pumila* из популяций Нижнего Поволжья и Южного Урала (Кашин и др., 2022).

По строению и морфометрии околоцветника виды разделены на условные группы: 1) с длинными узкими долями (*I. carthaliniae*, *I. halophila*, *I. spuria*, *I. graminea*, *I. lactea*); 2) с крупными наружными и редуцированными внутренними долями (*I. pseudacorus*, *I. setosa*); 3) с крупными широкими долями (*I. sibirica*, *I. sanguinea*); 4) с долями средних размеров (*I. pumila*, *I. scariosa*), что свидетельствует о высокой степени достоверности и обоснованности современных классификаций рода.

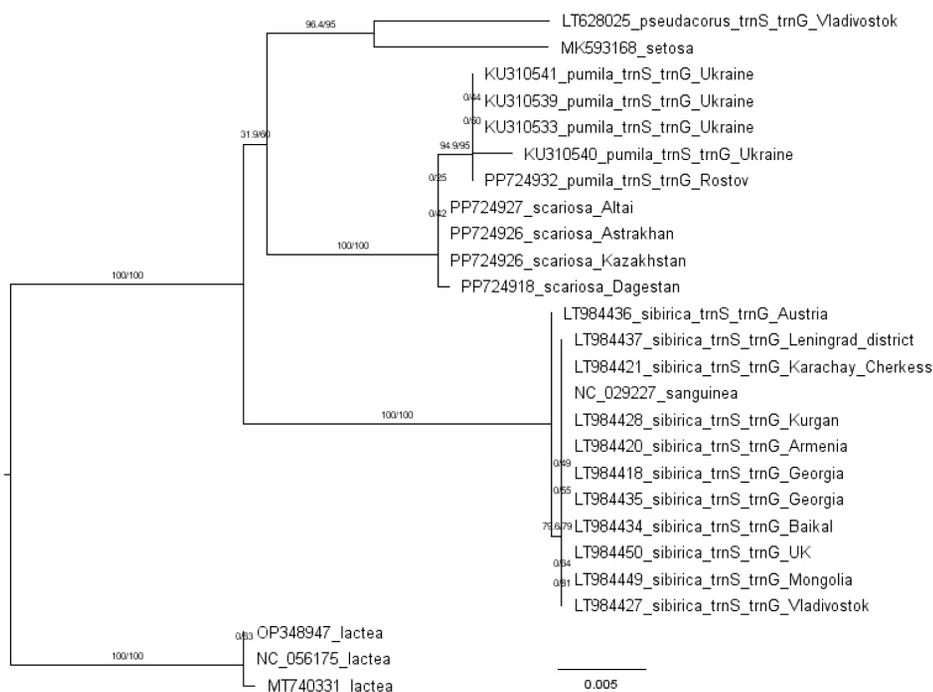
Сравнительный анализ морфометрических характеристик плодов и семян *Iris* из природных популяций Республики Башкортостан и Оренбургской области с литературными данными иных регионов (Алексеева, 2020) выявил умеренный географический полиморфизм. По размерам семян наиболее мелкие семена характерны для популяций *I. sibirica* (Саратовская область), *I. halophila* и *I. pumila*

(Волгоградская область); крупносемянная форма отмечена у *I. pseudacorus* (Приморский край). По параметрам плодов незначительное уменьшение наблюдается у *I. setosa*, *I. graminea*, *I. spuria* и *I. pumila* (Белгородская область). Обнаруженная незначительная вариабельность размеров, а также пигментации репродуктивных структур отражает микропопуляционную дифференциацию видов и локальные адаптивные реакции на условия произрастания, обеспечивающие их успешное воспроизводство в разнообразных условиях среды.

Коэффициент вариации морфологических признаков служит полезным инструментом для предварительной оценки стабильности морфологических характеристик, однако судить о таксономической значимости признаков исключительно по коэффициенту вариации неверно. Но таксономические решения должны основываться на синтезе морфологических, генетических и экологических данных. Для рода *Iris* это особенно актуально из-за его полиморфизма и склонности к гибридизации. Метод выявления таксономических признаков, предложенный



**Рис. 4.** Филогенетическое древо, построенное на основе *matK* последовательностей хлДНК ирисов с использованием метода ML.



**Рис. 5.** Филогенетическое древо, построенное на основе *trnS-trnG* последовательностей хлДНК ирисов с использованием метода ML.

Н.С. Ростовской (2002), основан на анализе корреляционных плеяд – групп взаимосвязанных морфологических признаков, который находит устойчивые связи между признаками. Метод помогает отличить консервативные таксономические признаки от экологически зависимых.

Следует отметить, что *I. scariosa* характеризуется высокой изменчивостью морфологических признаков, особенно, как и близкородственный *I. pumila*, в окраске цветка. Вид описан с самой западной окраины своего основного ареала по экземпляру Палласа, собранному им восточнее

Волги. Встречается преимущественно в северо-западной части Прикаспийской низменности и Восточном Предкавказье; эндемичный прикаспийский европейско-кавказский вид (Красная книга..., 2024).

Таксономическое положение близких видов *I. scariosa* и *I. glaucescens* Bunge вызывает определенные трудности. Оба вида либо синонимизируются, либо рассматриваются в качестве отдельных видов (Цвелев, 1979; Алексеева, 2020), при этом приоритетным видом является *I. scariosa*. *I. glaucescens* описан в Средней Азии. Вид представлен немногочисленными популяциями, встречающимися в России на границе ареала на юге Западной Сибири. За пределами России он известен в Северо-Восточном Казахстане, в Северо-Западных районах Монголии и Китая (Алексеева, 2020). На Южном Урале наблюдается схожая тенденция: в Республике Башкортостан *I. scariosa* включен в региональную Красную книгу (Красная книга..., 2021–2024), а в соседней Челябинской области приводится *I. glaucescens* (Куликов, 2005). Таким образом, таксономическое разграничение этих близких видов вызывает трудности. В литературе имеются подтвержденные молекулярным, морфологическим и палиноморфологическим анализами данные, что *I. glaucescens*, а также *I. timofejewii* и *I. curvifolia* – синонимы *I. scariosa* (Boltenkov, Artyukova, 2024).

Как следует из дендрограммы различия-сходства видов по выявленным диагностическим маркерам наиболее тесная филогенетическая связь обнаруживается между парами видов: *I. carthaliniae* и *I. spuria*; *I. sibirica* и *I. sanguinea*; *I. pumila* и *I. scariosa*. Родство их очевидно и по габитусу растений. Так, пара *I. carthaliniae* и *I. spuria* – это высокорослые растения, имеют длинные корневища с утолщенными звеньями, мощные, ветвистые стебли, широкие мечевидные листья, многоцветковые соцветия (3–5 цветков). Пара *I. sibirica* и *I. sanguinea* – среднерослые ирисы, имеют короткие корневища с узкими звеньями, полые, слабоветвящиеся стебли, узкие линейные листья в прикорневых куртинах, малоцветковые соцветия (2–3 цветка). Пара *I. pumila* и *I. scariosa* – низкорослые растения с толстыми ползучими корневищами, несут очень короткие стебли, широколинейные или ланцетные листья в прикорневом пучке, крупные (относительно размера растения), немногочисленные цветки (1–2), расположенные близко к земле.

#### Молекулярно-филогенетический анализ видов рода *Iris*

Результаты морфологического анализа показали, что *I. sibirica* и *I. sanguinea* очень близки друг другу, этот факт был подтвержден результатами молекулярно-филогенетического анализа. Более того, согласно полученным данным, мы можем предположить, что *I. sibirica* и *I. sanguinea* являются одним видом, поскольку *trnL-trnF* последовательности всех проанализированных образцов этих видов идентичны, вне зависимости от места сбора. К такому выводу, проанализировав ряд морфологических признаков и используя молекулярно-филогенетические данные, ранее приходили и другие исследователи (Boltenkov et al., 2020).

Очень схожая ситуация во второй паре близкородственных видов – *I. pumila* и *I. scariosa*. У всех проанализированных образцов *I. pumila* как с желтыми, так и с фиолетовыми цветками *trnL-trnF* последовательности идентичны. Вид полихромен и характеризуется цветовым полиморфизмом, что определяется нормой реакции особей на условия внешней среды (Кашин и др., 2022), и вариации окраски долей околоцветника не связаны с молекулярно-генетическим полиморфизмом популяций.

У *I. scariosa* выявлено три варианта последовательностей. Первый вариант (из Республики Дагестан) абсолютно идентичен *trnL-trnF* последовательности *I. pumila*, что, скорее, указывает на то, что это *I. pumila*. Два других варианта отличаются от *I. pumila* заменой нескольких нуклеотидов. Вторым вариантом *I. scariosa* обнаружен у образцов из Республики Казахстан и Республики Алтай, а третий вариант – у образцов из Республики Башкортостан и Оренбургской области; последние два варианта отличаются друг от друга лишь одной нуклеотидной заменой. Таким образом, сравнительный анализ некодирующих последовательностей *trnL-trnF* района хлоропластной ДНК *I. scariosa* из природных популяций, а также анализ литературных данных демонстрируют, что в Республике Башкортостан и Оренбургской области данный вид надежно идентифицирован.

Совершенно другая ситуация в третьей паре родственных видов – *I. setosa* и *I. pseudacorus*. Прежде всего, это явно два разных, хотя и близкородственных, вида, что хорошо видно на филогенетическом древе и ранее было отмечено другими исследователями (Choi, Lee, 2024). А вот относительно внутривидовой вариативности эти виды различаются. Все исследованные образцы *I. pseudacorus* показывают идентичные *trnL-trnF* последовательности, а пять проанализированных образцов *I. setosa* продемонстрировали наличие в них двух типов *trnL-trnF* последовательностей, причем все образцы были взяты либо в Китае, либо в Корею. *I. setosa* гетерогенен и таксономически достаточно четко отграничен от других видов, в связи с чем из его состава выделены новые виды и внутривидовые таксоны (Илюшко и др., 2001). Н.Б. Алексеева (2013) пришла к выводу, что полиморфный комплекс *I. setosa* состоит из отличающихся по морфологии и экологии пяти видов. Конечно, в данном случае необходимо проанализировать больше фактического материала.

Все пять *trnL-trnF* последовательностей *I. lactea* идентичны друг другу, независимо от их места произрастания. *Iris* Linnaeus ser. Lacteae Doronkin (Доронкин, 1990) включает виды, распространенные в умеренных азиатских регионах Северного полушария (Родионенко, 2006; Boltenkov, 2018). Первоначально *I. lactea* был отнесен в состав подрода *Limniris*, секция *Limniris* (Родионенко, 1961; Mathew, 1989), затем в состав подрода *Eremiris*, секция *Haloiris* (Доронкин, 2006). Впоследствии он стал рассматриваться в составе отдельного рода *Eremiris* (Родионенко, 2006), но это не согласуется с мнением С.А. Уилсона (2011), основанным на его молекулярно-генетических исследованиях. Долгое время считалось, что серия Lacteae

Doronkin представлена только одним полиморфным видом, однако молекулярные исследования подтвердили филогенетическое разветвление как минимум в трех линиях, соответствующих таксонам *I. lactea*, *I. oxypetala* и *I. tibetica* (Boltenkov et al., 2018).

## Заключение

На основе проведенной оценки структуры морфологической изменчивости представителей рода *Iris* выявлено 10 таксономических индикаторов: длина и ширина наружных долей околоцветника, длина и ширина внутренних долей околоцветника, длина тычиночной нити, пыльника и пестика, ширина плода, а также длина и ширина семени.

Установление филогенетических взаимоотношений исследуемых видов по обнаруженным таксономическим индикаторам и проведенному молекулярно-филогенетическому анализу на основе *trnL-trnF*, *matK* и *trnS-trnG* хлоропластной ДНК позволило объединить исследованные виды ирисов в несколько филогенетических групп.

Таким образом, можно констатировать, что морфологический и молекулярно-филогенетический анализы показывают одинаковые результаты относительно анализа филогении исследованных видов рода *Iris*, из чего следует, что выявленные морфологические индикаторы вполне определенно демонстрируют таксономическую достоверность.

## Список литературы / References

Алексеева Н.Б. О видах рода *Iris* L. (Iridaceae) из родства *Iris setosa* Pall. ex Link. *Turczaninowia*. 2013;16(2):13-29  
[Alexeeva N.B. On the taxa of *Iris setosa* Pall. ex Link affinity (Iridaceae). *Turczaninowia*. 2013;16(2):13-29 (in Russian)]

Алексеева Н.Б. Ирисы – *Iris* L. (Iridaceae Juss.) России. СПб.: Пер- вый ИПХ, 2020  
[Alexeeva N.B. *Iris* L. (Iridaceae Juss.) in the Russia. St. Petersburg, 2020 (in Russian)]

Болтенков Е.В., Артюкова Е.В., Козыренко М.М. Дивергенция видов серии *Lacteae* рода *Iris* (Iridaceae) в России и сопредельных странах на основе анализа хлоропластной ДНК. *Генетика*. 2016; 52(5):579-589. doi 10.7868/S0016675816040032  
[Boltenkov E.V., Artyukova E.V., Kozyrenko M.M. Species divergence in *Iris* series *Lacteae* (Iridaceae) in Russia and adjacent countries based on chloroplast DNA sequence data. *Russ J Genet*. 2016; 52(5):507-516. doi 10.1134/S1022795416040037]

Болтенков Е.В., Артюкова Е.В., Козыренко М.М. Таксономическое исследование рода *Iris* (Iridaceae): морфологические и молекулярные данные. В: Ботаника в современном мире. Т. 1. Махачкала, 2018;13-16. Доступно: [https://www.binran.ru/files/publications/Proceedings/Proceedings\\_RBO/XIV\\_RBO\\_Proceedings\\_T1.pdf](https://www.binran.ru/files/publications/Proceedings/Proceedings_RBO/XIV_RBO_Proceedings_T1.pdf)  
[Boltenkov E.V., Artyukova E.V., Kozyrenko M.M. A taxonomic study of the genus *Iris* (Iridaceae): morphological and molecular data. In: Botany in the Contemporary World. V. 1. Mahachkala, 2018;13-16. Available: [https://www.binran.ru/files/publications/Proceedings/Proceedings\\_RBO/XIV\\_RBO\\_Proceedings\\_T1.pdf](https://www.binran.ru/files/publications/Proceedings/Proceedings_RBO/XIV_RBO_Proceedings_T1.pdf) (in Russian)]

Волкова Г.А. Итоги интродукции природных видов рода *Iris* L. на европейском Севере. В: Ботанические сады. Проблемы интродукции. Томск, 2010;274:127-129  
[Volkova G.A. The results of introduction of natural species of the *Iris* L. genus in northern Europe. In: Botanical Gardens: Issues of Introduction. Tomsk, 2010;274:127-129 (in Russian)]

Голубев В.Н. Основы биоморфологии травянистых растений центральной лесостепи. Воронеж, 1962

[Golubev V.N. Fundamentals of Biomorphology of Herbaceous Plants in the Central Forest-steppe. Voronezh, 1962 (in Russian)]

Дорогина О.В., Доронькин В.М., Селюткина И.Ю., Конищенко Е.С. Структура рода *Iris* L. (Iridaceae Juss.) азиатской России, выявленная SDS-электрофорезом запасных белков семян. *Turczaninowia*. 2012;15(4):76-81  
[Dorogina O.V., Doronkin V.M., Selyutina I.Yu., Konichenko E.S. Structure of the genus *Iris* L. (Iridaceae Juss.) in Asiatic Russia, revealed by SDS-electrophoresis of seed storage proteins. *Turczaninowia*. 2012;15(4):76-81 (in Russian)]

Доронькин В.М. Обзор сибирских видов рода *Iris* (Iridaceae). *Ботанический журнал*. 1990;75(3):409-416  
[Doronkin V.M. The synopsis of Siberian species of the genus *Iris* (Iridaceae). *Botanicheskii Zhurnal*. 1990;75(3):409-416 (in Russian)]

Доронькин В.М. Система рода *Iris* L. (Iridaceae Juss.) Азиатской России. В: Роль ботанических садов в сохранении биоразнообразия растительного мира Азиатской России: настоящее и будущее. Новосибирск, 2006;101-103  
[Doronkin V.M. Taxonomy of the genus *Iris* L. (Iridaceae Juss.) in Asian Russia. In: Role of Botanical Gardens in the Preservation of Vegetation Diversity in Asian Russia: the Present and the Future. Novosibirsk, 2006;101-103 (in Russian)]

Доронькин В.М., Красников А.А. Цитотаксономическое исследование сибирских видов рода *Iris* (Iridaceae). *Ботанический журнал*. 1984;69(5):683-685  
[Doronkin V.M., Krasnikov A.A. Cytotaxonomic studies in some Siberian species of the genus *Iris* (Iridaceae). *Botanicheskii Zhurnal*. 1984;69(5):683-685 (in Russian)]

Доронькин В.М., Байков К.С. Моделирование филогенетических связей сибирских видов рода *Iris* (Iridaceae) по методу SYNAP. *Turczaninowia*. 2004;7(2):45-57  
[Doronkin V.M., Baikov K.S. Modeling phylogenetic relationships of Siberian species of the genus *Iris* (Iridaceae) using SYNAP method. *Turczaninowia*. 2004;7(2):45-57 (in Russian)]

Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). М., 1985  
[Dospikhov B.A. Methodology of Field Experiments (with the Basics of Statistical Processing of Research Results). Moscow, 1985 (in Russian)]

Зайцев Г.Н. Математическая статистика в экспериментальной ботанике. М., 1984  
[Zaitsev G.N. Mathematical Statistics in Experimental Botany. Moscow, 1984 (in Russian)]

Зайцев Г.Н. Математика в экспериментальной ботанике. М., 1990  
[Zaitsev G.N. Mathematics in Experimental Botany. Moscow, 1990 (in Russian)]

Илюшко М.В., Каратавцев Ю.Ф., Журавлев Ю.Н. Изменчивость морфологических признаков *I. setosa* (Iridaceae) на российском Дальнем Востоке. *Ботанический журнал*. 2001;86(3):60-72  
[Ilyushko M.V., Karatavcev Yu.F., Zhuravlev Yu.N. Variability of morphological characters of *I. setosa* (Iridaceae) in the Russian Far East. *Botanicheskii Zhurnal*. 2001;86(3):60-72 (in Russian)]

Ишбирдин А.Р., Ишмуратова М.М., Жирнова Т.В. Стратегии жизни ценопопуляции *Cephalanthera rubra* L. (Rich) на территории Башкирского государственного заповедника. *Вестник Нижнегородского университета им. Н.И. Лобачевского*. 2005;1:85-98  
[Ishbirdin A.R., Ishmuratova M.M., Zhirnova T.V. Life strategies of the cenopopulation *Cephalanthera rubra* L. (Rich) on the territory of the Bashkir State Nature Reserve. *Vestnik of Lobachevsky University of Nizhni Novgorod*. 2005;1:85-98 (in Russian)]

Кашин А.С., Пархоменко А.С., Абрамова Л.М., Кондратьева А.О., Богослов А.В., Шилова И.В., Крюкова А.В. Морфологическая изменчивость *Iris pumila* (Iridaceae) в Нижнем Поволжье и на Южном Урале. *Ботанический журнал*. 2022;107(2):180-197. doi 10.31857/S0006813622020065  
[Kashin A.S., Parkhomenko A.S., Abramova L.M., Kondratyeva A.O., Bogoslov A.V., Shilova I.V., Kryukova A.V. Morphological vari-

- ability of *Iris pumila* (Iridaceae) in the Lower Volga region and the South Urals. *Botanicheskii Zhurnal*. 2022;107(2):180-197. doi 10.31857/S0006813622020065 (in Russian)]
- Козыренко М.М., Артюкова Е.В., Журавлев Ю.Н. Самостоятельность видов *Iris vorobievii* N.S. Pavlova, *Iris mandshurica* Maxim. и *Iris humilis* Georgi (Iridaceae): доказательство из ядерного и хлоропластного геномов. *Генетика*. 2009;45(11):1575-1584 [Kozyrenko M.M., Artyukova E.V., Zhuravlev Y.N. Independent species status of *Iris vorobievii* N.S. Pavlova, *Iris mandshurica* Maxim., and *Iris humilis* Georgi (Iridaceae): evidence from the nuclear and chloroplast genomes. *Russ J Genet*. 2009;45(11):1394-1402. doi 10.1134/S1022795409110143]
- Красная книга Республики Башкортостан. В 2 т. Уфа, 2021-2024 [Red Book of the Republic of Bashkortostan. Vols. 1 and 2. Ufa, 2021-2024 (in Russian)]
- Красная книга Российской Федерации. Растения и грибы. М., 2024 [Red Book of the Russian Federation. Plants and Mushrooms. Moscow, 2024 (in Russian)]
- Куликов П.В. Конспект флоры Челябинской области (сосудистые растения). Екатеринбург-Миасс, 2005 [Kulikov P.V. Synopsis of the Flora of the Chelyabinsk Region (Vascular Plants). Ekaterinburg-Miass, 2005 (in Russian)]
- Макаревич И.Ф., Доронкин В.М., Щербик С.В., Блинов А.Г. Филогенетические взаимоотношения сибирских видов рода *Iris* L. (Iridaceae), выявленные с помощью RAPD-PCR метода. *Turczaninowia*. 2001;4(4):80-92 [Makarevich I.F., Doronkin V.M., Shcherbik S.V., Blinov A.G. Application of RAPD-PCR method in the study of phylogenetic relationships between Siberian species of the genus *Iris* L. (Iridaceae). *Turczaninowia*. 2001;4(4):80-92 (in Russian)]
- Родионенко Г.И. Род Ирис – *Iris* L. М.; Л., 1961 [Rodionenko G.I. Genus *Iris*. Moscow; Leningrad, 1961 (in Russian)]
- Родионенко Г.И. *Eremiris* – новый род семейства Iridaceae. *Ботанический журнал*. 2006;91(11):1707-1712 [Rodionenko G.I. *Eremiris*, a new genus of the family Iridaceae. *Botanicheskii Zhurnal*. 2006;91(11):1707-1712 (in Russian)]
- Родионенко Г.И. Постигая тайны природы (Судьба моя – ирисы). СПб., 2013 [Rodionenko G.I. Understanding the Secrets of Nature (My Destiny is Irises). St. Petersburg, 2013 (in Russian)]
- Ростова Н.С. Корреляции: структура и изменчивость. СПб., 2002 [Rostova N.S. Correlations: Structure and Variability. St. Petersburg, 2002 (in Russian)]
- Цвелев Н.Н. Род Касатик – *Iris* L. В: Флора Европейской части СССР. Т. 4. 1979;299-307 [Tsvelev N.N. Genus *Iris* L. In: Flora of the European Part of the USSR. Vol. 4. 1979;299-307 (in Russian)]
- Alexeeva N.B. *Iris lokiae*, a new species of genus *Iris* (Iridaceae). *Botanicheskii Zhurnal*. 2013;98(11):1415-1420 (in Russian)
- Boltenkov E.V. Taxonomic notes on *Iris* ser. *Lacteae* (Iridaceae) with typifications of fifteen names and one new combination. *Phytotaxa*. 2018;383(3):283-292. doi 10.11646/phytotaxa.383.3.5
- Boltenkov E.V., Artyukova E.V. Updated taxonomy of *Iris scariosa* (Iridaceae) inferred from morphological and chloroplast DNA sequence data with remarks on classification of *Iris* subg. *Iris*. *Plants*. 2024;13(17):2349. doi 10.3390/plants13172349
- Boltenkov E.V., Artyukova E.V., Kozyrenko M.M., Trias-Blasi A. *Iris tibetica*, a new combination in *I*. ser. *Lacteae* (Iridaceae) from China: evidence from morphological and chloroplast DNA analyses. *Phytotaxa*. 2018;338(3):223-240. doi 10.11646/phytotaxa.338.3.1
- Boltenkov E., Artyukova E., Kozyrenko M., Erst A., Trias-Blasi A. *Iris sanguinea* is conspecific with *I. sibirica* (Iridaceae) according to morphology and plastid DNA sequence data. *Peer J*. 2020;8:e10088. doi 10.7717/peerj.10088
- Choi T.-Y., Lee S.-R. Complete plastid genome of *Iris orchioides* and comparative analysis with 19 *Iris* plastomes. *PLoS One*. 2024;19(4):e0301346. doi 10.1371/journal.pone.0301346
- Crespo M.B., Martínez-Azorín M., Mavrodiev E.V. Can a rainbow consist of a single colour? A new comprehensive generic arrangement of the '*Iris sensu latissimo*' clade (Iridaceae), congruent with morphology and molecular data. *Phytotaxa*. 2015;232(1):1-78. doi 10.11646/phytotaxa.232.1.1
- Dorofeeva M.M., Zhurbenko P.M. Comparative study of ovule structure and development in some species of *Iris* subgenus *Limniris* (Iridaceae). *Botanicheskii Zhurnal*. 2020;105(1):15-31. doi 10.31857/S0006813620010068
- Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780. doi 10.1093/molbev/mst010
- Makarevich I., Golovnina K., Scherbik S., Blinov A. Phylogenetic relationships of the Siberian *Iris* species inferred from noncoding chloroplast DNA sequences. *Int J Plant Sci*. 2003;164(2):229-237. doi 10.1086/346160
- Mathew B. The Iris. Portland, 1989
- Mitic B. *Iris adriatica* (Iridaceae), a new species from Dalmatia (Croatia). *Phyton*. 2002;42(2):305-314
- Plaines T., Jakob S.S., Blattner F.R. Application of non-coding DNA regions in intraspecific analyses. *Plant Syst Evol*. 2009;282:281-294. doi 10.1007/s00606-008-0036-9
- Trifinopoulos J., Nguyen L.-T., Haeseler A., Minh B.Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016;44(W1):W232-W235. doi 10.1093/nar/gkw256
- Weber T., Jakše J., Sladonja B., Hruševar D., Landeka N., Brana S., Bohanec B., Milović M., Vladović D., Mitić B., Poljuha D. Molecular study of selected taxonomically critical taxa of the genus *Iris* L. from the broader Alpine-Dinaric area. *Plants*. 2020;9(9):1229. doi 10.3390/plants9091229
- Wilson C.A. Subgeneric classification in *Iris* re-examined using chloroplast sequence data. *Taxon*. 2011;60(1):27-35. doi 10.1002/tax.601004
- Zhao Y.T. A new species of *Iris* from China. *J Syst Evol*. 1992;30(2):181-182

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 30.04.2025. После доработки 10.08.2025. Принята к публикации 09.09.2025.

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](https://elibrary.ru/title_about.asp?id=32440)  
англоязычная версия – на сайте [vavilov.elpub.ru/index.php/jour](http://vavilov.elpub.ru/index.php/jour)  
и платформе PubMed Central, <https://www.ncbi.nlm.nih.gov/pmc/journals/3805/>

При перепечатке материалов ссылка обязательна.

✉ email: [vavilov\\_journal@bionet.nsc.ru](mailto: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 27.02.2026. Формат 60 × 84 1/8. Уч.-изд. л. 20.9.

