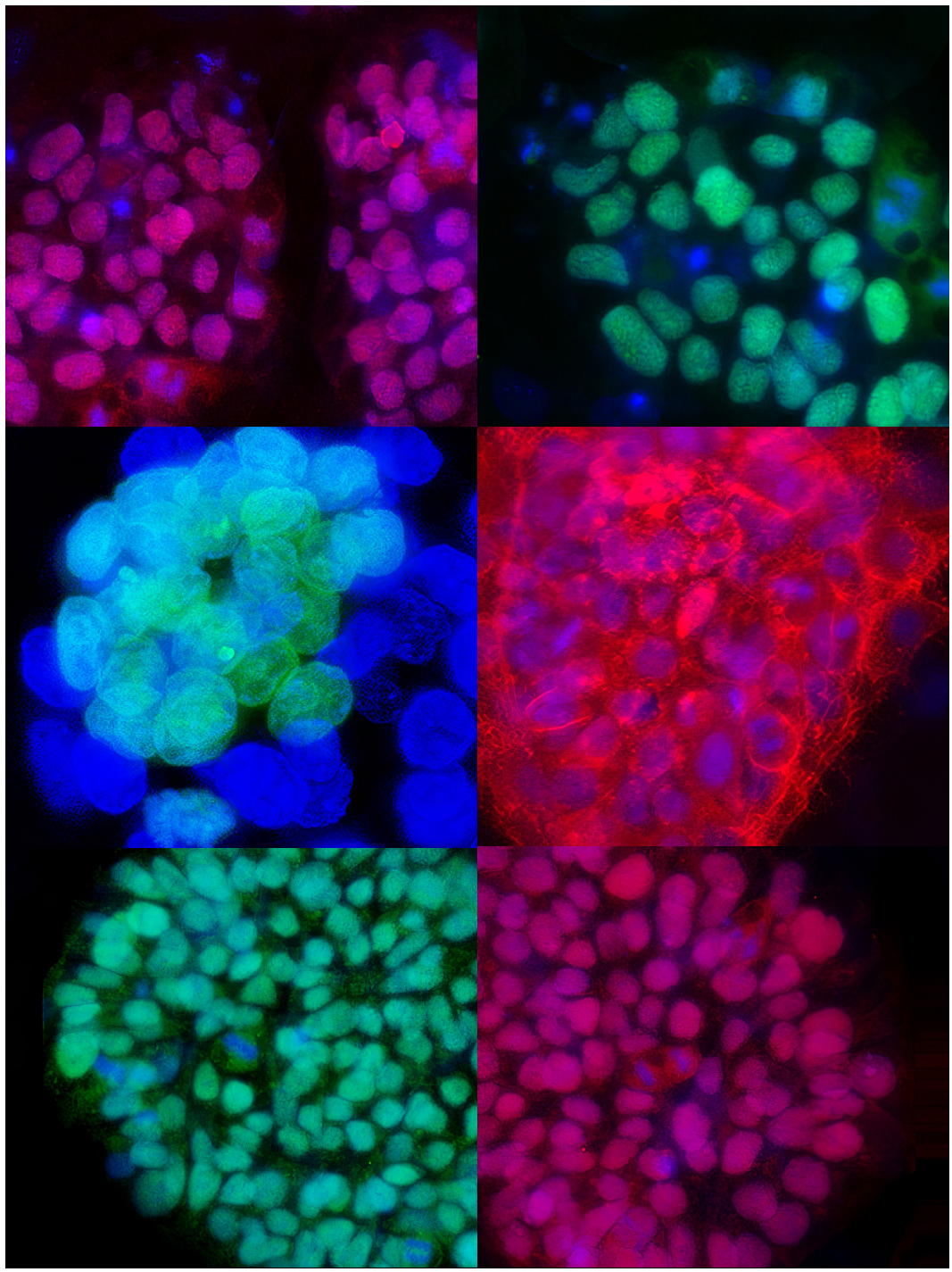
Online ISSN 2500-3259

ВАВИЛОВСКИЙ
ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ
VAVILOV JOURNAL OF GENETICS AND BREEDING

2026 • 30 • 2



Молекулярная и клеточная биология / Генетика и селекция растений / Генетическая инженерия / Медицинская генетика / Популяционная генетика / Биоинформатика и системная биология / Высокопроизводительное секвенирование / Цифровое фенотипирование

vavilov.elpub.ru
vavilovj-icg.ru
vavilov_journal@bionet.nsc.ru

2026 • 30 • 2

Online ISSN 2500-3259

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

Основан в 1997 г.
Периодичность 8 выпусков в год
doi 10.18699/vjgb-26-17

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. Аульченко – д-р биол. наук (Нидерланды)
О.С. Афанасенко – академик РАН, д-р биол. наук (Россия)
Д.А. Афонников – д-р биол. наук, доцент (Россия)
Л.И. Афтанас – академик РАН, д-р мед. наук (Россия)
Л.А. Беспалова – академик РАН, д-р с.-х. наук (Россия)
А. Бёрнер – д-р наук (Германия)
Н.П. Бондарь – канд. биол. наук (Россия)
С.А. Боринская – д-р биол. наук (Россия)
П.М. Бородин – д-р биол. наук, проф. (Россия)
А.В. Васильев – чл.-кор. РАН, д-р биол. наук (Россия)
М.И. Воевода – академик РАН, д-р мед. наук (Россия)
Т.А. Гавриленко – д-р биол. наук (Россия)
Н.Е. Грунтенко – д-р биол. наук (Россия)
С.А. Демаков – д-р биол. наук (Россия)
И.К. Захаров – д-р биол. наук, проф. (Россия)
И.А. Захаров-Гезехус – чл.-кор. РАН, д-р биол. наук (Россия)
С.Г. Инге-Вечтомов – академик РАН, д-р биол. наук (Россия)
А.В. Кильчевский – чл.-кор. НАНБ, д-р биол. наук (Беларусь)
А.М. Кудрявцев – чл.-кор. РАН, д-р биол. наук (Россия)
Д.М. Ларкин – канд. биол. наук (Великобритания)
С.А. Лашин – д-р биол. наук (Россия)
Ж. Ле Гуи – д-р наук (Франция)
И.Н. Лебедев – чл.-кор. РАН, д-р биол. наук, проф. (Россия)
Л.А. Лутова – д-р биол. наук, проф. (Россия)
Б. Люгтенберг – д-р наук, проф. (Нидерланды)
В.Ю. Макеев – чл.-кор. РАН, д-р физ.-мат. наук (Россия)
И.В. Максимов – д-р биол. наук (Россия)
Б.А. Малярчук – д-р биол. наук (Россия)
Ю.Г. Матушкин – канд. биол. наук (Россия)
В.И. Молодин – академик РАН, д-р ист. наук (Россия)
М.П. Мошкин – д-р биол. наук, проф. (Россия)
Л.Ю. Новикова – д-р с.-х. наук (Россия)
Е.К. Потокина – д-р биол. наук (Россия)
В.П. Пузырев – академик РАН, д-р мед. наук (Россия)
Д.В. Пышный – чл.-кор. РАН, д-р хим. наук (Россия)
Е.Ю. Рыкова – д-р биол. наук (Россия)
Е.А. Салина – чл.-кор. РАН, д-р биол. наук, проф. (Россия)
В.А. Степанов – академик РАН, д-р биол. наук (Россия)
И.А. Тихонович – академик РАН, д-р биол. наук (Россия)
Е.К. Хлесткина – чл.-кор. РАН, д-р биол. наук, проф. РАН (Россия)
Э.К. Хуснутдинова – д-р биол. наук, проф. (Россия)
М. Чен – д-р биол. наук (Китайская Народная Республика)
Е.В. Шахтшнейдер – д-р мед. наук (Россия)
С.В. Шестаков – академик РАН, д-р биол. наук (Россия)
С.В. Шеховцов – д-р биол. наук (Россия)
Н.С. Юдин – канд. биол. наук (Россия)
Н.К. Янковский – академик РАН, д-р биол. наук (Россия)

2026 • 30 • 2

Online ISSN 2500-3259

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

Founded in 1997
Publication frequency: 8 issues a year
doi 10.18699/vjgb-26-17

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. Bespalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia
N.P. Bondar, Cand. Sci. (Biology), Russia
S.A. Borinskaya, Dr. Sci. (Biology), Russia
P.M. Borodin, Professor, Dr. Sci. (Biology), Russia
A. Börner, Dr. Sci., Germany
M. Chen, Dr. Sci. (Biology), People's Republic of China
S.A. Demakov, Dr. Sci. (Biology), Russia
T.A. Gavrilenko, Dr. Sci. (Biology), Russia
N.E. Gruntenko, Dr. Sci. (Biology), Russia
S.G. Inge-Vechtomov, Full Member of the RAS, Dr. Sci. (Biology), Russia
E.K. Khlestkina, Corr. Member of the RAS, Professor of the RAS, Dr. Sci. (Biology), Russia
E.K. Khusnutdinova, Professor, Dr. Sci. (Biology), Russia
A.V. Kilchevsky, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology), Belarus
A.M. Kudryavtsev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia
D.M. Larkin, Cand. Sci. (Biology), Great Britain
S.A. Lashin, Dr. Sci. (Biology), Russia
J. Le Gouis, Dr. Sci., France
I.N. Lebedev, Corr. Member of the RAS, Professor, Dr. Sci. (Biology), Russia
B. Lugtenberg, Professor, Dr. Sci., Netherlands
L.A. Lutova, Professor, Dr. Sci. (Biology), Russia
V.Yu. Makeev, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.), Russia
I.V. Maksimov, Dr. Sci. (Biology), Russia
B.A. Malyarchuk, Dr. Sci. (Biology), Russia
Yu.G. Matushkin, Cand. Sci. (Biology), Russia
V.I. Molodin, Full Member of the RAS, Dr. Sci. (History), Russia
M.P. Moshkin, Professor, Dr. Sci. (Biology), Russia
L.Yu. Novikova, Dr. Sci. (Agricul.), Russia
E.K. Potokina, Dr. Sci. (Biology), Russia
V.P. Puzyrev, Full Member of the RAS, Dr. Sci. (Medicine), 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. Shakhtshneider, 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

СОДЕРЖАНИЕ • 2026 • 30 • 2

Молекулярная и клеточная биология

163 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Концепция природной реконструкции генома. Часть 4. Интеграция фрагментов экстраклеточной двуцепочечной ДНК в геном гемопоэтических стволовых клеток и формирование экстра-хромосомальных интермедиатов.
С.Г. Ошихмина, В.С. Рузанова, Г.С. Риттер, Е.В. Долгова, С.С. Кирикович, Е.В. Левитес, Я.Р. Ефремов, Т.В. Карамышева, А.С. Молодцева, Я.В. Райцина, О.С. Таранов, С.В. Сидоров, С.Д. Никонов, О.Ю. Леплина, А.А. Останин, Е.Р. Черных, Н.А. Колчанов, А.С. Проскурина, С.С. Богачев

181 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Наивные ИПСК человека, полученные при культивировании линии клеток ICGi022-A с праймированной плюрипотентностью в среде HENSM, способны к эффективной дифференцировке в эндотелиальные производные. *М.А. Арссан, А.И. Шевченко, С.М. Закиян, И.С. Захарова (на англ. языке)*

Генетика и селекция растений

194 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Молекулярно-генетическая изменчивость по ISSR-маркерам у *Caragana jubata* (Pall.) Poir. (Fabaceae) в горах Средней Азии и Южной Сибири. *Д.А. Кривенко, О.А. Чернышева, И.Н. Кубан, Е.В. Жмудь, И.В. Горбенко, А.А. Ачимова, О.Ю. Васильева, О.В. Дорогина*

205 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Изменчивость органельных геномов в коллекции раннеспелых сортов сои.
В.В. Александрович, М.Г. Синявская, О.П. Шатарнов, О.Г. Давыденко

212 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Анализ экспрессии генов *PR-1*, *PR-2*, *PR-10*, хитиназ и хитиназа-подобных белков в корнях гороха под влиянием салициловой кислоты и метилжасмоната.
А.М. Егорова

222 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Отбор стабильных мутантов риса с помощью линейных моделей со смешанными эффектами (LMM) и индексов стабильности. *П. Шарифи, А.А. Эбади, М.Т. Халладжян, Х. Аминпанах (на англ. языке)*

Генетическая инженерия

233 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Роль типа экспланта и применения селективного агента в первичной эффективности трансформации *Lens culinaris* Medik. *Т.В. Дюбенко, К.В. Смирнов, В.Е. Творогова (на англ. языке)*

241 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Получение и характеристика двойной нокаутной линии *Arabidopsis thaliana*, лишенной экспрессии *AOX1a* и *VTC2*.
Е.В. Гармаш, Е.С. Белых, К.В. Ядрихинский, Р.В. Малышев, И.О. Велегжанинов

Медицинская генетика

250 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Генетические корреляции между болезнями человека и плейотропные гены.
И.В. Зоркольцева, Н.М. Белоногова, А.В. Кириченко, Я.А. Цепилов, Т.И. Аксенович

259 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Новые регуляторные однонуклеотидные полиморфизмы (rSNPs) потенциально участвуют в регуляции глюконеогенеза в печени и вносят вклад в вариабельность АМФК-зависимых механизмов индивидуального ответа на метформин.
Е.Е. Корболина, И.С. Дамаров, Т.И. Меркулова (на англ. языке)

267 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Особенности коннективности дефолтной сети мозга в зависимости от полиморфизма гена транспортера серотонина (5-HTTLPR).
А.В. Бочаров, А.Н. Савостьянов, С.С. Таможников, А.Е. Сапрыгин, Д.А. Лебедкин, Е.А. Меркулова, Г.Г. Князев

Популяционная генетика

274 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Филогеография и таксономический статус видового комплекса *Formica picea* (Hymenoptera: Formicidae).
З.А. Жигульская, С.В. Шеховцов, С.В. Чеснокова, А.П. Бурнашева, А.А. Гурина, Р.Ю. Дудко, Т.В. Полубоярова, С.В. Решетников, Ю.Н. Сундуков, Д.И. Берман (на англ. языке)

Биоинформатика и системная биология

284 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Функциональная симметрия и воспроизводимость эволюционного процесса. *С.И. Барцев*

293 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Alembic: от разрозненных биологических данных к структурированным ресурсам.
И.В. Бездворных, К.И. Юдыцкий, Н.А. Черкасов, А.А. Самсонова, А.А. Канапин

Высокопроизводительное секвенирование

299 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Разработка и валидация программы PipeSeq для анализа данных секвенирования РНК на модели *Chlamydomonas reinhardtii*.
А.М. Нерезенко, П.А. Виролайнен, С.А. Тупицына, Е.М. Чекунова

311 ОБЗОР
Анализ вирома растений с помощью высокопроизводительного секвенирования: принципы и подходы.
Д.Д. Белкина, С.В. Виноградова

Цифровое фенотипирование

321 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Применимость инструмента для анализа временных рядов StatFaRmer при цифровом фенотипировании сои (*Glycine max*). *Д.С. Ульянов, А.А. Ульянова, А.А. Кочешкова, А.О. Блинков, А.В. Архипов, Я.С. Меглицкая, Н.Ю. Свистунова, Г.И. Карлов, М.Г. Дивашук*

330 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Описание морфологических характеристик колосьев пшеницы в базе данных SpikeDroidDB в виде цифрового паспорта. *Е.Г. Комышев, Ю.В. Кручинина, В.С. Коваль, А.А. Потешкина, Н.В. Петраш, В.В. Пискарев, Н.П. Гончаров, Д.А. Афонников*

© Сибирское отделение Российской академии наук, 2026
© Институт цитологии и генетики СО РАН, 2026
Вавиловский журнал генетики и селекции, 2026

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

CONTENTS • 2026 • 30 • 2

Molecular and cell biology

163 ORIGINAL ARTICLE
Concept of natural genome reconstruction. Part 4. Integration of extracellular double-stranded DNA fragments into the genome of hematopoietic stem cells and the formation of extrachromosomal intermediates. *S.G. Oshikhmina, V.S. Ruzanova, G.S. Ritter, E.V. Dolgova, S.S. Kirikovich, E.V. Levites, Y.R. Efremov, T.V. Karamysheva, A.S. Molodtseva, Y.V. Raitsina, O.S. Taranov, S.V. Sidorov, S.D. Nikonov, O.Y. Leplina, A.A. Ostanin, E.R. Chernykh, N.A. Kolchanov, A.S. Proskurina, S.S. Bogachev*

181 ORIGINAL ARTICLE
Naïve human iPSCs obtained by culturing the ICGi022-A cell line with primed pluripotency in HENSM medium efficiently differentiate into endothelial derivatives. *M.A. Arssan, A.I. Shevchenko, S.M. Zakian, I.S. Zakharova*

Plant genetics and breeding

194 ORIGINAL ARTICLE
Molecular genetic variability in coenopopulations of *Caragana jubata* (Pall.) Poir. (Fabaceae) in the mountains of Central Asia and Southern Siberia revealed using ISSR-markers. *D.A. Krivenko, O.A. Chernysheva, I.N. Kuban, E.V. Zhmud, I.V. Gorbenko, A.A. Achimova, O.Yu. Vasilyeva, O.V. Dorogina*

205 ORIGINAL ARTICLE
Variability of organelle genomes in a collection of early maturing soybean varieties. *V.V. Aleksandrovich, M.G. Siniauskaya, A.P. Shatarnov, O.G. Davydenko*

212 ORIGINAL ARTICLE
Analysis of *PR-1*, *PR-2*, *PR-10*, chitinases and chitinase-like proteins genes expression in pea roots under the action of salicylic acid and methyl jasmonate. *A.M. Egorova*

222 ORIGINAL ARTICLE
Selecting stable rice mutants with linear mixed models (LMM) and stability indexes. *P. Sharifi, A.A. Ebadi, M.T. Hallajian, H. Aminpanah*

Gene engineering

233 ORIGINAL ARTICLE
The role of explant type and selective agent application in the initial transformation rate of *Lens culinaris* Medik. *T.V. Dyubenko, K.V. Smirnov, V.E. Tvorogova*

241 ORIGINAL ARTICLE
Generation and characterization of a double-knockout *Arabidopsis thaliana* line lacking expression of *AOX1a* and *VTC2*. *E.V. Garmash, E.S. Belykh, K.V. Yadrikhinskiy, R.V. Malyshev, I.O. Velegzhaninov*

Medical genetics

250 ORIGINAL ARTICLE
Pleiotropic genes underlying genetic correlations across human diseases. *I.V. Zorkoltseva, N.M. Belonogova, A.V. Kirichenko, Y.A. Tsepilov, T.I. Axenovich*

259 ORIGINAL ARTICLE
Novel regulatory SNPs that can be activated due to metformin treatment may orchestrate liver gluconeogenesis and add to the variability in AMPK-dependent mechanisms of metformin response. *E.E. Korbolina, I.S. Damarov, T.I. Merkulova*

267 ORIGINAL ARTICLE
Features of connectivity of default mode network depending on polymorphism of serotonin transporter gene (5-HTTLPR). *A.V. Bocharov, A.N. Savostyanov, S.S. Tamozhnikov, A.E. Saprygin, D.A. Lebedkin, E.A. Merkulova, G.G. Knyazev*

Population genetics

274 ORIGINAL ARTICLE
Phylogeography and taxonomic status of the *Formica picea* complex (Hymenoptera: Formicidae). *Z.A. Zhigulskaya, S.V. Shekhovtsov, S.V. Chesnokova, A.P. Burnasheva, A.A. Gurina, R.Yu. Dudko, T.V. Poluboyarova, S.V. Reshetnikov, Yu.N. Sundukov, D.I. Berman*

Bioinformatics and systems biology

284 ORIGINAL ARTICLE
Functional symmetry and reproducibility of the evolutionary process.
S.I. Bartsev

293 ORIGINAL ARTICLE
Alembic: a framework for converting disparate biological data into structured resources. *I.V. Bezdvornykh, K.I. Yuditskiy, N.A. Cherkasov, A.A. Samsonova, A.A. Kanapin*

High-throughput sequencing

299 ORIGINAL ARTICLE
Development and validation of the PipeSeq program for RNA-seq data analysis in the *Chlamydomonas reinhardtii* as a model. *A.M. Nerezenko, P.A. Virolainen, S.A. Tupitsyna, E.M. Chekunova*

311 REVIEW
Plant virome analysis by high-throughput sequencing: concepts and approaches.
D.D. Belkina, S.V. Vinogradova

Digital phenotyping

321 ORIGINAL ARTICLE
Applicability of the StatFaRmer time series analysis tool in soybean (*Glycine max*) digital phenotyping. *D.S. Ulyanov, A.A. Ulyanova, A.A. Kocheshkova, A.O. Blinkov, A.V. Arkhipov, Ya.S. Meglitskaya, N.Yu. Svistunova, G.I. Karlov, M.G. Divashuk*

330 ORIGINAL ARTICLE
Description of morphological characteristics of wheat spike as a digital certificate in the SpikeDroidDB database.
E.G. Komyshev, Yu.V. Kruchinina, V.S. Koval, A.A. Poteshkina, N.V. Petrash, V.V. Piskarev, N.P. Goncharov, D.A. Afonnikov

© Siberian Branch RAS, 2026
© Institute of Cytology and Genetics, SB RAS, 2026
Vavilov Journal of Genetics and Breeding, 2026

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):163-180

doi 10.18699/vjgb-26-18

Концепция природной реконструкции генома.
Часть 4. Интеграция фрагментов экстраклеточной двуцепочечной ДНК в геном гемопоэтических стволовых клеток и формирование экстрахромосомальных интермедиатов

С.Г. Ошихмина [1, 2 *], В.С. Рузанова [1 *], Г.С. Риттер [1], Е.В. Долгова [1], С.С. Кирикович [1], Е.В. Левитес [1], Я.Р. Ефремов [1], Т.В. Карамышева[1], А.С. Молодцева[3], Я.В. Райцина[1, 2], О.С. Таранов [4], С.В. Сидоров [2, 5], С.Д. Никонов[6], О.Ю. Леплина [7], А.А. Останин [7], Е.Р. Черных [7], Н.А. Колчанов [1], А.С. Проскурина [1#], С.С. Богачев [1#]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный государственный исследовательский университет, Новосибирск, Россия
[3] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[4] Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р. п. Кольцово, Новосибирская область, Россия
[5] Городская клиническая больница № 1, Новосибирск, Россия
[6] Новосибирский научно-исследовательский институт туберкулеза, Новосибирск, Россия
[7] Научно-исследовательский институт фундаментальной и клинической иммунологии, Новосибирск, Россия
labmolbiol@mail.ru

Аннотация. Для оценки возможности интеграции в реципиентный геном гемопоэтических стволовых клеток экстраклеточных фрагментов двуцепочечной ДНК был сконструирован сложно составленный субстрат, состоящий из целого M13F-*AluI*-M13R фрагмента и двух его производных, появляющихся после гидролиза рестриктазами EcoRI и HindIII: M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII. В субстрате была представлена последовательность полилинкера плазмиды pBlueScript+, отсутствующая в геноме человека, которая обрамляла клонированный по сайту EcoRV *AluI* фрагмент человека. Клетки костного мозга человека были обработаны ДНК сконструированного сложно составленного субстрата, и с учетом времени репарации пангеномных одноцепочечных разрывов из клеток выросших колоний получены препараты метафазных пластинок. Проведенная FISH выявила специфические сигналы свечения. Одновременно ДНК, выделенная из колоний, полученных из клеток костного мозга, обработанных сложно составленным субстратом, была секвенирована. Проведено два раунда секвенирования: полногеномное и селективное после таргетной гибридизации на металлических бусах. Полученные результаты свидетельствуют, что гомологичный обмен между экстрахромосомальной и хромосомной ДНК возможен. Также возможна интеграция в геном по механизму однонитевого отжига, с участием микрогомологий. Обнаружены интермедиаты с одним концом фрагмента, интегрированным в геном по участку микрогомологии, и другим концом фрагмента, свободно свисающим в межхромосомное пространство. Проведена прямая оценка возможности интеграции TAMRA-меченых фрагментов двуцепочечной ДНК человека и *E. coli* в реципиентный геном на модели клеток костного мозга человека. Обнаружено, что специфические сигналы гомологичной ДНК распределены по телу хромосом (модель клеток костного мозга человека). Сигналы негомологичной ДНК *E. coli* преимущественно сконцентрированы в центромерных районах хромосом. Соотношение количества полученных ридов с элементами интеграции и сигналов FISH предполагало существование прочного взаимодействия экстраклеточных фрагментов и ДНК хромосом. В экспериментах показано, что линейная плазмидная ДНК после интернализации в гемопоэтические стволовые клетки формирует кольцо мономера. Интернализованная во внутриклеточное пространство экстраклеточная плазмидная ДНК выделяется совместно с ДНК хромосом после жестких процедур очистки и фракционирования. Этот факт предполагает существование прочного кольцевого ассоциата ДНК плазмиды и ДНК хромосом, сформированного без участия белкового каркаса в форме закольцованной хромосомной нити.

Ключевые слова: FISH; полногеномное секвенирование; интеграция в геном; экстрахромосомальные кольцевые структуры

Для цитирования: Ошихмина С.Г., Рузанова В.С., Риттер Г.С., Долгова Е.В., Кирикович С.С., Левитес Е.В., Ефремов Я.Р., Карамышева Т.В., Молодцева А.С., Райцина Я.В., Таранов О.С., Сидоров С.В., Никонов С.Д., Леплина О.Ю., Останин А.А., Черных Е.Р., Колчанов Н.А., Проскурина А.С., Богачев С.С. Концепция природной реконструкции генома. Часть 4. Интеграция фрагментов экстраклеточной двуцепочечной ДНК в геном гемопоэтических стволовых клеток и формирование экстрахромосомальных интермедиатов. *Вавиловский журнал генетики и селекции*. 2026;30(2):163-180. doi 10.18699/vjgb-26-18

Финансирование. Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации в Институте цитологии и генетики СО РАН (государственный бюджетный проект № FWNR-2026-0025) и при поддержке И.Н. Зайцевой и А.А. Пуртова.

© Ошихмина С.Г., Рузанова В.С., Риттер Г.С., Долгова Е.В., Кирикович С.С., Левитес Е.В., Ефремов Я.Р., Карамышева Т.В., Молодцева А.С., Райцина Я.В., Таранов О.С., Сидоров С.В., Никонов С.Д., Леплина О.Ю., Останин А.А., Черных Е.Р., Колчанов Н.А., Проскурина А.С., Богачев С.С., 2026

* # Авторы внесли равный вклад.

Контент доступен под лицензией Creative Commons Attribution 4.0

Concept of natural genome reconstruction.
Part 4. Integration of extracellular double-stranded DNA fragments into the genome of hematopoietic stem cells and the formation of extrachromosomal intermediates

S.G. Oshikhmina [1, 2*], V.S. Ruzanova [1*], G.S. Ritter [1], E.V. Dolgova [1], S.S. Kirikovich [1], E.V. Levites [1], Y.R. Efremov [1], T.V. Karamysheva[1], A.S. Molodtseva[3], Y.V. Raitsina[1, 2], O.S. Taranov [4], S.V. Sidorov [2, 5], S.D. Nikonov[6], O.Y. Leplina [7], A.A. Ostanin [7], E.R. Chernykh [7], N.A. Kolchanov [1], A.S. Proskurina [1#], S.S. Bogachev [1#]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[4] State Scientific Center of Virology and Biotechnology "Vector" of Rospotrebnadzor, Koltsovo, Novosibirsk region, Russia
[5] City Clinical Hospital No. 1, Novosibirsk, Russia
[6] Novosibirsk Tuberculosis Research Institute, Novosibirsk, Russia
[7] Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia
labmolbiol@mail.ru

Abstract. To assess the possibility of integrating extracellular double-stranded DNA fragments into the recipient genome of hematopoietic stem cells, a complex substrate was constructed consisting of the entire M13F-*AluI*-M13R fragment and its two restrictive derivatives, appearing after hydrolysis with restriction endonucleases EcoRI and HindIII: M13F-*AluI*-EcoRI and M13R-*AluI*-HindIII. The substrate contained a pBlueScript+ plasmid polylinker sequence, absent in the human genome, which framed the human *AluI* fragment cloned at the EcoRV site. Human bone marrow cells were treated with the DNA of the constructed complex substrate; taking into account the repair time of pangenomic single-strand breaks, preparations of metaphase plates were obtained. FISH revealed specific fluorescent signals. Simultaneously, DNA isolated from colonies obtained from bone marrow cells treated with a complex substrate was sequenced. Two rounds of sequencing were carried out: whole-genome and selective after targeted hybridization on metal beads. The results obtained indicate that homologous exchange between extrachromosomal and chromosomal DNA is possible. Integration into the genome via the single-strand annealing mechanism, involving microhomologies, is also possible. Intermediates were discovered that suggest the existence of an unusual integration into the genome at the nick of one end of the fragment and the other end of the fragment hanging freely into the interchromosomal space. A direct assessment of the possibility of integrating TAMRA-labeled fragments of fragmented human DNA and *E. coli* DNA into the genome of recipient cells was carried out using a human bone marrow cell model. The results obtained indicate that specific signals of homologous DNA are distributed throughout the chromosome body (human bone marrow cell model). Signals from nonhomologous *E. coli* DNA are predominantly concentrated in the centromeric regions of chromosomes. The ratio of the number of obtained reads with integration elements and FISH signals suggested the existence of a strong interaction between extracellular fragments and chromosomal DNA. Experiments have been conducted showing that linear plasmid DNA, after internalization into hematopoietic stem cells, forms a monomer ring. Internalized into the intracellular space, extracellular plasmid DNA is isolated together with chromosomal DNA after stringent purification and fractionation procedures. This fact suggests the existence of a strong ring associate of plasmid DNA and chromosome DNA formed without the participation of a protein framework in the form of a looped chromosomal strand.

Key words: FISH; whole genome sequencing; integration into the genome; extrachromosomal ring structures

For citation: Oshikhmina S.G., Ruzanova V.S., Ritter G.S., Dolgova E.V., Kirikovich S.S., Levites E.V., Efremov Y.R., Karamysheva T.V., Molodtseva A.S., Raitsina Y.V., Taranov O.S., Sidorov S.V., Nikonov S.D., Leplina O.Y., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Proskurina A.S. , Bogachev S.S. Concept of natural genome reconstruction. Part 4. Integration of extracellular double-stranded DNA fragments into the genome of hematopoietic stem cells and the formation of extrachromosomal intermediates. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):163-180. doi 10.18699/vjgb-26-18

Введение

В первых частях работы провозглашается и доказывается следующая концепция. Фрагменты экстраклеточной двуцепочечной ДНК (дцДНК) – молекулы линейной ДНК любого происхождения, размером от одного до нескольких нуклеосомных мономеров (200–2000 п.н. и более), доставляются в гемопоэтические стволовые клетки (ГСК) естественным природным механизмом. Интернализация обеспечивается общим положительным зарядом ГСК, который определяется специфическими белкам гликокаликса ГСК (Dolgova et al., 2012; Petrova et al., 2022; Ritter et al., 2022). В процессе мобилизации и интернализации участвуют белки, несущие в своей структуре гепарин-связывающий домен или кластеры положительно заряженных аминокислот. Доставленная в клетку дцДНК инициирует образование пангеномных одноцепочечных разрывов, которые, в свою очередь, служат триггером процесса терминальной дифференцировки ГСК (Рузанова и др., 2024).

Процесс терминальной дифференцировки связан с реорганизацией структуры хроматина высшего порядка (Рузанова и др., 2024). Он является причиной активации в клетке рекомбиногенной ситуации, ферментативная машина которой определяет возникновение и репарацию пангеномных одноцепочечных разрывов и становление

клетки на путь развития специфического гемопоэтического ростка (Jacobson et al., 1975; Farzaneh et al., 1982; Johnstone, Williams, 1982).

Рекомбиногенная ситуация представляет собой состояние клетки, в которой активированы механизмы метаболических превращений ДНК хроматина. К таким процессам относятся активация надзирающих киназ, модуляция клеточного цикла, репликативный стресс, активация репарационно-рекомбинационного механизма. В структуру последнего входит ферментативная машина механизма, а также ДНК субстраты, которые формируют метаболические интермедиаты ДНК ядра. В совокупности этот процесс для случая появления пангеномных одноцепочечных разрывов определяет восстановление непрерывности и корректность первичной последовательности ДНК хромосом на всем ее протяжении (Лихачева и др., 2008; Рузанова и др., 2024).

Фрагменты ДНК, попавшие в клетку, инициируют образование ников (Рузанова и др., 2024), индуцирующих рекомбиногенную ситуацию, и, таким образом, именно эти фрагменты служат триггером развития масштабных изменений в клетке. Молекулярная машина, инициирующая начало процесса, и причинно-следственные связи между факторами, запускающими механизм инициации, неизвестны.

История поведения внехромосомной, не вирусной ДНК в клетке начинается с работ 80-х годов прошлого века (Smith, Berg, 1984). В настоящее время эта тема не доминирует в исследованиях. Существует незначительное количество работ, характеризующих процессы, происходящие в клетке при попадании в нее фрагментов дцДНК, и описывающих, какие пути клеточного мониторинга при этом активируются.

В литературе приводятся сведения об активации систем контроля прогрессии клеточного цикла в ответ на появление в ядерном пространстве ДНК в форме синтетических олигонуклеотидов, ДНК SV40, апоптозных телец (Yakubov et al., 1989; Holmgren et al., 1999). Описываются молекулярные события, происходящие при интернализации плазмидных ДНК искусственными методами (Smith, Berg, 1984; Lin J. et al., 1985; Thomas et al., 1986). Известно несколько работ, свидетельствующих об активации контрольных точек клеточного цикла при интернализации в ядро синтетических олигонуклеотидов $(dA/dT)_{70}$. Наблюдаемая при этом активация молекулярной машины в отсутствие каких-либо повреждений геномной ДНК или аберрантной репликации свидетельствует о том, что структуры синтетических спаренных олигонуклеотидов, по-видимому, мимикрируют возникающие интермедиаты. Эти структуры могут быть представлены одноцепочечной ДНК, тупыми двуцепочечными концами, соединениями одноцепочечного участка и дуплекса и крестообразными формами. Предполагается, что при спаривании олигонуклеотидов образуются концы молекулы ДНК спаренных фрагментов, которые являются индукторами ATM/ATR-зависимого механизма контроля клеточного цикла, в котором участвуют такие регуляторы активности иерархических киназ, как RPA, RAD1, TopBP1, клапсин (Yoo et al., 2004, 2006; MacDougall et al., 2007; Zou, 2007). В работах, анализирующих молекулярные события, происходящие при интернализации ДНК, указывается на тот факт, что одноцепочечные фрагменты ДНК не вызывают активации контролирующих систем клетки (Kumagai, Dunphy, 2000). Проведенный анализ минимального количества линейных молекул, способных индуцировать остановку клеточного цикла, показал, что более 30 молекул, содержащих структуры оцДНК/дуплекс, активируют достаточный ответ, необходимый для определения уровня фосфорилирования Chk1 (MacDougall et al., 2007). Если в экстрахромосомальном пространстве ядра находится менее 30 линейных фрагментов ДНК, система мониторинга стрессовых ситуаций клетки остается спокойной. Это значит, что такое количество линейных молекул ДНК может попадать в ядро, не активируя систему иерархических киназ, и, возможно, является нормой транзитного присутствия экстраклеточных фрагментов, циркулирующих в плазме крови.

Известно, что локализованные в клетке линейные фрагменты экстраклеточной дцДНК сшиваются один с другим, формируя конкатемерные и кольцевые структуры размером до 10 т. п. н. (Perucho et al., 1980; Lin J. et al., 1985; Lin Y., Waldman, 2001; Rogachev et al., 2006; Potter et al., 2017, 2024). Также установлено, что находящиеся в ГСК линейные фрагменты плазмидной ДНК процессируются на глубину до 1 т. п. н. и лигируются в кольцо (Dolgova et al., 2013). Вместе это означает, что фрагменты экстраклеточной дцДНК, находящиеся в клетке, активируют репарационно-рекомбинационные процессы. Именно такой факт был экспериментально показан в работе (Pierandrei et al., 2016), где с использованием платформы RT2 Profiler PCR Array (SABioscience, Qiagen) был проведен анализ экспрессии 84 генов при трансфекции в клетку экстраклеточных фрагментов дцДНК. Оказалось, что доставленные в клетку фрагменты дцДНК активируют гены различных клеточных систем, включая как гены репарации и гомологичной рекомбинации ДНК, так и гены, регулирующие клеточный цикл и эпигенетические модификации. То есть в напряженное состояние приводятся основные молекулярные системы, регулирующие клеточный ДНК метаболизм.

Ники, так же как и двуцепочечные концы доставленных в ядро экстраклеточных фрагментов, являются триггером репарации и рекомбиногенной ситуации (Vriend, Krawczyk, 2017; Maizels, Davis, 2018). Механизмы и факторы, активируемые этими двумя триггерными структурами, различны, хотя и имеют пересекающиеся сигнальные пути (Рузанова и др., 2024). Таким образом, в клетке появляются два метаболита ДНК (ники в ДНК хромосом и двуцепочечные концы доставленных в ядро экстраклеточных фрагментов), запускающих однонаправленные репарационно-рекомбинационные процессы. Очевидно, что факторы обоих процессов будут вступать в определенные молекулярные взаимоотношения, о которых ничего не известно. Возможность такого сценария в рамках рассматриваемой концепции требует дальнейших глубоких исследований.

Интернализация фрагментов в клетку с использованием различных способов трансфекции представляет рутинную процедуру. Более того общепринят способ получения нокаутов, за который три независимых исследователя, Марио Капекки (Mario R. Capecchi), Оливер Смитис (Oliver Smithies), а также сэр Мартин Эванс (Martin J. Evans), в 2007 г. получили Нобелевскую премию. Подход основан на том, что фрагменты дцДНК в ядре будут находить своего гомолога и рекомбинировать с ним, внося нокаутную мутацию. То есть процесс естественного обмена генетической информацией между экстрахромосомальной ДНК и ДНК хроматина объективно существует и является имманентным свойством метаболической активности клетки, связанным со взаимодействием ДНК хроматина и внехромосомной ДНК (Perucho, Wigler, 1981; Kucherlapati et al., 1984; Smithies et al., 1985; Murnane et al., 1990; Hastings et al., 1993; Leung et al., 1997; Li et al., 2001; Lin Y., Waldman, 2001).

Фрагменты дцДНК, попавшие в ядро, становятся участниками событий, ими же и активированных. Появление в ядре рекомбиногенных ников и рекомбиногенных двуцепочечных концов с неизбежностью запускает рекомбинационные процессы, результатом которых, как мы предполагаем, будет фиксация в клетке новой генетической информации, содержащейся в интернализованных дцДНК фрагментах исходно внеклеточной локализации.

В настоящей части исследования была оценена возможность интеграции фрагментов экстраклеточной ДНК в геном ГСК и формирование кольцеобразных ассоциатов между ДНК фрагментов и ДНК хромосом.

Материалы и методы

Клетки костного мозга человека. В работе были использованы клетки криоконсервированного сепарата костного мозга больных лимфомой Ходжкина, предоставленного криобанком Научно-исследовательского института фундаментальной и клинической иммунологии (НИИФКИ). В Клинике иммунопатологии НИИФКИ в отделении гематологии с блоком трансплантации костного мозга проводится лечение больных гемобластозами высокодозной химиотерапией и трансплантацией аутологичных или аллогенных периферических ГСК. При заготовке периферических стволовых клеток наряду с основным продуктом «афереза» (который трансплантируется пациенту) заготавливается два-три образца (пробирки-спутники) сепарированных клеток для контроля качества продукта «афереза» и научных исследований. Такие образцы вместе с основным использовались в настоящей работе. Каждый образец сепарата костного мозга, включая образцы-спутники, сопровождается необходимым пакетом документов, включающим информированное согласие, протокол исследований материала костного мозга, протокол лечения, которые подписываются пациентом в соответствии с установленными нормами. После проведенного лечения образцы-спутники утилизируются согласно СанПиН или используются в научных целях. Документы, сопровождающие каждый забор материала костного мозга, хранятся в архиве криобанка НИИФКИ и могут быть востребованы по официальному запросу.

***AluI* двуцепочечный ДНК зонд.** Амплификацию *AluI* повтора человека (M13F-*AluI*-M13R фрагмент) проводили с помощью ПЦР. Матрица представляла собой ДНК *AluI* повтора, клонированного в плазмиде pUC19, включающего начало и конец тандемно повторяющихся последовательностей AluJ и AluY (NCBI: AC002400.1, 53494–53767) (Dolgova et al., 2012). Для амплификации использовали стандартные праймеры M13 (M13 for: 5′ GTAAAACGACGGCCAGT 3′, M13 rev: 5′ CAGGAAA CAGCTATGAC 3′) или специфические *AluI* праймеры (Dolgova et al., 2012). ПЦР фрагмент переосаждали 0.1 V NaAc 3 M, pH 5.2, и 1 V изопропанола в течение 10 мин при –20 °C с последующим центрифугированием. Осадок промывали в 70 % этаноле и растворяли в стерильной воде.

Препараты ДНК. Препарат hDNA^{gr} (реконструктор ДНК генома человека) представляет собой двуцепочечную ДНК человека в виде фрагментов размером 200–2000 п. н. hDNA^{gr} выделяли из плацент здоровых женщин, фрагментировали ультразвуковой дезинтеграцией до размеров 200–2000 п. н., депротеинизировали обработкой протеиназой К и подвергали фенол/хлороформной экстракции.

ДНК человека получали из плацент здоровых женщин аналогичным образом, без фрагментации.

ДНК *E. coli* выделяли стандартным способом с использованием лизоцима, ультразвукового разрушения клеток, депротеинизации протеиназой К и фенол/хлороформной экстракции.

Конструирование зонда для обработки клеток костного мозга. Для анализа возможной негомологичной интеграции и/или гомологичного обмена был сконструирован зонд, состоящий из *AluI* фрагмента, встроенного в полилинкер pBlueScript+ (pBS) по EcoRV сайту, и ограниченный последовательностями, простирающимися до сайтов посадки M13 праймеров и включающими эти сайты. В экспериментах по интернализации использовали смесь целого фрагмента M13F-*AluI*-M13R и двух его производных после гидролиза рестриктазами EcoRI и HindIII, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII.

Для анализа возможной прямой интеграции фрагментированная геномная ДНК человека и *E. coli* метилась нуклеотидом, содержащим флуорохром TAMRA, с добавлением праймера со случайной последовательностью (стат-праймера) в реакции амплификации с Taq-полимеразой (ООО «Медиген»).

Анализ интернализации TAMRA-меченого M13F-*AluI*-M13R ПЦР фрагмента в клетки карциномы Эрлиха производился как описано в статье (Ruzanova et al., 2022).

FISH проводили по стандартной процедуре с TAMRA-меченым зондом по методике, описанной в Приложении 1[1]. Результаты FISH анализировали на микроскопе AXIOSKOP 2 Plus (Zeiss, Германия). Для регистрации и обработки микроизображений использовали CCD-камеру (CoolCube 1, MetaSystems, Германия), набор фильт-

[1] Приложения 1 и 2 см. по адресу:
https://vavilovj-icg.ru/download/pict-2026-30/appx11.pdf

ров Set 49 (G 365, FT 395, BP 445/50) (Zeiss, Германия), ET-orange #1 (ET 546/22 x ET 590/33) (Chroma, США), программное обеспечение ISIS5 (MetaSystems, Германия). Количество проанализированных метафазных пластинок варьировало от 14 до 30 для разных групп.

Культивирование клеток костного мозга в метилцеллюлозной среде. Аспираты костного мозга получали из криобанка НИИФКИ. При стандартной процедуре разморозки до 90 % клеток костного мозга (включая CD34+ клетки, согласно FACS анализу) сохраняют свою жизнеспособность, что позволяет использовать этот материал без дополнительной верификации.

Клетки костного мозга человека, находящиеся в составе аспирата, отмывали средой RPMI и вводили в эксперимент. Контрольные клетки и образцы клеток, обработанные зондами, осаждали в течение 10 мин при 400g и ресуспендировали в среде IMDM+2 % FBS. Для количественного определения и анализа миелоидных предшественников обработанные клетки и клетки контрольного образца помещали в метилцеллюлозную среду MethoCult H4034 (Stem Cell Technologies). Культивирование клеток проводили в среднем 10–15 дней, в зависимости от целей эксперимента. Анализ и выделение клеток из метилцеллюлозной среды после культивации выполняли согласно инструкции производителя.

Таргетное секвенирование. Для секвенирования использовали ДНК, выделенную из клеток, обработанных сложно составленным зондом: M13F-*AluI*-M13R, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII. Обогащение библиотеки проводили путем гибридизации с биотинилированной ДНК полилинкера плазмиды pBS, иммобилизованной на Dynabeads® Streptavidin магнитных частицах (Life Technologies, США) по методу (Maricic et al., 2010). Секвенирование осуществляли на платформе Illumina MiSeq (TruSeq® DNA Sample Preparation Kit (Illumina)) с MiSeq Reagent Kit v2 (300 cycle).

Полногеномное секвенирование. Для секвенирования была взята ДНК, выделенная из клеток, обработанных сложно составленным зондом: M13F-*AluI*-M13R, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII. Секвенирование было произведено на приборе NovaSeq 6000 (Illumina) с длиной прочтения 2×150 п. н. Величина Q30 составила не менее 90 %. Для подготовки библиотек использовали набор “TruSeq Nano DNA Library Prep kit” (Illumina). Ультразвуковую фрагментацию ДНК выполняли на приборе ME220 Focused-ultrasonicator (Covaris). Качество ДНК-библиотеки определяли с помощью капиллярного электрофореза на приборе TapeStation 4200 (Agilent). Анализ полученных последовательностей проводился на платформах Ugene и Blastn. Первичные данные секвенирования хранятся в формате FASTQ и могут быть запрошены в индивидуальном порядке.

Формирование колец экстраклеточными фрагментами дцДНК, интернализованными ГСК и их ближайшими потомками. Схема эксперимента подробно представлена в разделе «Экспериментальное доказательство формирования колец длинными фрагментами дцДНК, интернализованными в ГСК, и закольцовывания ими нити ДНК хромосомы». Все экспериментальные процедуры выполнены согласно протоколам, описанным в практических пособиях (Маниатис и др., 1984; Glover, 1988).

Клетки костного мозга обрабатывали препаратом ДНК, отмывали от ДНК и инкубировали в течение 2 ч в CO_2-инкубаторе в полной питательной среде RPMI. Клетки собирали и отмывали физиологическим раствором. Солюбилизировали цитоплазматическую мембрану 0.2 % тритоном X-100, наслаивали на 10 % сахарозный градиент и центрифугировали при 500g 20 мин. Ядра собирали и отмывали физиологическим раствором. Затем ядра лизировали 1 % SDS, 50 мМ ЭДТА и депротеинизировали обработкой 200 мкг/мл протеиназы K при 58 °C 60 мин, после чего проводили фенол-хлороформную экстракцию. Полученный лизат наносили на 15–30 % NaCl ступенчатый (5 %) градиент и центрифугировали при ~300000g (L-8 центрифуга, SW 50.1 ротор, 47000 об/мин) в течение 2 ч при комнатной температуре. На схеме возле последней пробирки справа приведены маркеры молекулярных весов ДНК, располагающиеся в указанных процентных зонах градиента (Glover, 1988; Dolgova et al., 2013). В 7-й области градиента находится ДНК с размером более 40 т. п. н., которая формирует осадок на дне пробирки. ДНК супернатанта осаждали равным объемом изопропанола из 0.3 М ацетата натрия, pH 4.8, в течение ночи при –20 °C. Фракции 1–6 после центрифугирования и промывки от соли растворяли в 36 мкл H_2O. Осадок ДНК фракции 7 после промывки 70 % этанолом также растворяли в 36 мкл H_2O. На агарозный гель наносили по 12 мкл материала (1/3 от общего объема).

Был проведен анализ образца 7 (осадок на дне градиента). Qubit концентрация ДНК в образце 7 составила 1 нг/мкл. Суммарно получено 36 нг. Треть образца (12 нг) использовали для анализа градиента. Треть образца (12 нг) использовали в real-time ПЦР для анализа количества ДНК плазмиды pEGFP-N1 в области гена устойчивости к канамицину. Треть образца (12 нг) трансформировали в электрокомпетентные клетки DH-5 с титром 10^7–10^8.

В результате трансформации 12 нг фракции 7 было получено более 50 крупных колоний и около 5000 мелких. Во всех крупных колониях, из которых была выделена ДНК (8 колоний), присутствовал мономер pEGFP-N1. В мелких колониях ДНК плазмиды не содержалось. ДНК, выделенная из крупных колоний, гидролизовалась рестриктазами HindIII, EcoRI и HaeIII.

Результаты

Детекция синтетического ДНК субстрата методом FISH на метафазных пластинках, полученных из образцов клеток колоний

Как отмечено выше, колонии формируются в результате деления одной ГСК костного мозга. Следовательно, клетки одной колонии имеют одну и ту же генетику, соответствующую генетике исходной CD34+ ГСК (рис. 1). Если в результате обработки индуктором в исходной ГСК произойдут изменения в геноме, то эти изменения будут амплифицированы в индивидуальной колонии до состояния

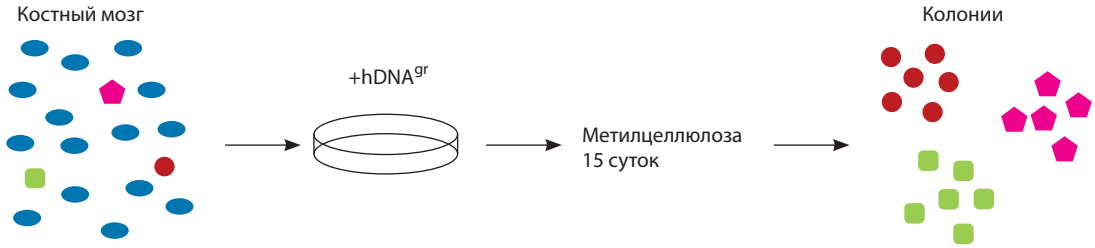


Рис. 1. Общая схема экспериментов.

возможности детекции такого события, что практически невозможно для индивидуальной CD34+ ГСК костного мозга (собственность ООО «ЭС ЛАБ ДИАГНОСТИК»).

Для выполнения таких экспериментов необходимо было сконструировать искусственный субстрат для обработки ГСК и меченный флуорохромом зонд для детекции возможной интеграции фрагментов искусственного субстрата. Были выбраны следующие варианты. Амплифицированный полилинкер плазмиды pBS был взят в качестве чужеродной ДНК, которая отсутствует в геноме человека. Этот фрагмент был мечен TAMRA флуорохромом и использовался в качестве зонда для гибридизации (pBSM13-TAMRA+). Одновременно для анализа были взяты сложно составленный субстрат, включающий *AluI* повтор, клонированный по сайту EcoRV, обрамленный полилинкером pBS, и два фрагмента, содержащих часть полилинкера pBS с прилегающим к нему *AluI* повтором и ограниченных EcoRI или HindIII сайтами рестрикции (M13F-*AluI*-M13R, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII). Включение TAMRA было проанализировано спектрофотометрически, электрофоретически и визуализировано в эксперименте по интернализации в клетки асцита карциномы Эрлиха (стандартная процедура оценки факта интернализации). Этот способ оценки мечения ДНК материала постоянно применяется в лаборатории. Результаты подготовительной работы приведены в Приложении 1.

Полученной конструкцией были обработаны клетки костного мозга человека. На 15-е сутки культивирования на метилцеллюлозе колонии были отмыты от среды, клетки обработаны колхицином и приготовлены препараты метафаз. Далее проведена FISH с TAMRA-меченым зондом pBSM13 (pBSM13-TAMRA+). Результаты гибридизации приведены на рис. 2, *A*. На препаратах детектируются множественные сигналы. Результаты свидетельствовали о возможной интеграции искусственного зонда в реципиентный геном. Также не исключался вариант прочной ассоциации экстраклеточных ДНК с телом хромосом. При анализе имиджей метафазных пластинок (до 40) мы не ставили задачи картирования TAMRA сигналов. В этой связи мы не добивались полной представленности кариотипа на изображениях. Основная цель экспериментов состояла в определении специфичности сигналов гибридизации, что подтверждалось анализом изображений.

Часть материала клеток, из которого были получены проанализированные метафазные пластинки, параллельно использовалась для таргетного и полногеномного секвенирования (выделенная из клеток суммарная ДНК).

Дополнительно были проведены эксперименты с меченным флуорохромом ДНК-зондом, позволяющие прямо детектировать присутствие меченого материала в ассоциации с хромосомами. Клетки костного мозга человека обрабатывались TAMRA-меченой ДНК человека и *E. coli* и высевались на метилцеллюлозу. Через 15 сут клетки отмывали, обрабатывали колхицином и анализировали метафазные пластинки (см. рис. 2, *B*). Было обнаружено несколько метафазных пластинок, несущих специфические сигналы. Оказалось, что в случае ДНК человека меченый материал детектируется во множестве сайтов на теле хромосом. В случае с бактериальной ДНК специфическая метка в большей степени сосредоточена в центромерных районах.

Во всех экспериментах с метафазными пластинками в качестве контрольных образцов использовались препараты метафаз, полученные из клеток колоний, выросших из образца клеток костного мозга, не обработанного ДНК субстратом.

Доказательство рекомбинационного взаимодействия искусственного зонда с геномом ГСК с использованием современных технологий секвенирования

Для доказательства рекомбинационного взаимодействия ДНК субстрата M13F-*AluI*-M13R, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII с геномом человеческих ГСК был взят тот же материал клеток, что и для приготовления метафаз и FISH. Была сконструирована библиотека и проведено два варианта секвенирования: полногеномное и секвенирование, где для увеличения числа целевых фрагментов выполнено обогащение библиотеки путем гибридизации с биотинилированной ДНК полилинкера плазмиды pBS, входящей в состав ДНК субстрата, иммобилизованной на Dynabeads® Streptavidin магнитных частицах (Life Technologies, США) по методу (Maricic et al., 2010). Полученные результаты были объединены в один раздел.

Схематическое изображение субстратов, взятых для обработки клеток, и субстратов, представленных в смеси, приведено на рис. 3, *A*. Таким образом, в смеси присутствовали следующие производные исходного субстрата: M13F-*AluI*-M13R, M13F-*AluI*-EcoRI, EcoRI-M13R, M13R-*AluI*-HindIII, HindIII-M13F. Нуклеотидные последовательности субстратов приведены в Приложении 2. Риды анализировались в прямой и обратной ориентации по обеим цепям.

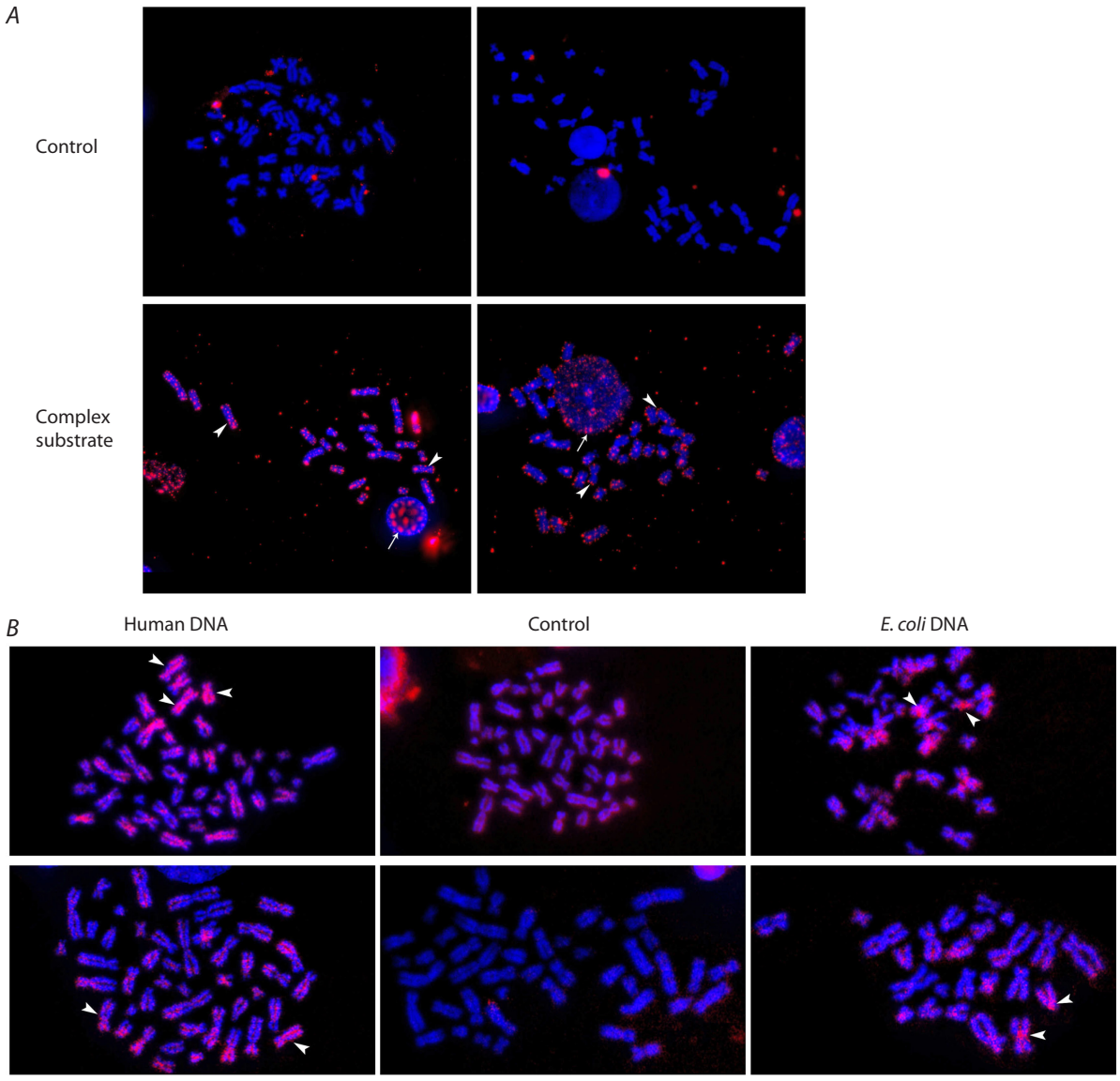


Рис. 2. Анализ присутствия специфических сигналов FISH после обработки клеток костного мозга выбранным зондом и специфических сигналов, детектируемых после прямого взаимодействия меченных флуорохромом дцДНК зондов и клеток костного мозга, на метафазных хромосомах.

A – FISH с зондом pBSM13-TAMRA+ метафазных хромосом, полученных после обработки колхицином клеток колоний, выросших из ГСК без обработки (Control) и обработанных в составе клеток костного мозга комплексным субстратом (M13F-*AluI*-M13R, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII). Стрелками указаны несколько сайтов гибридизации. *B* – распределение TAMRA-меченого материала на метафазных хромосомах, полученных из клеток колоний человека после обработки этих клеток TAMRA-мечеными ДНК человека и *E. coli*. Стрелками указаны множественные места локализации метки в образце, обработанном ДНК человека, и центромерные районы в образце, обработанном ДНК *E. coli*.

Анализ нуклеотидной последовательности выбранных клонов

Были выявлены риды с четко отсеченными концами, заканчивающимися на HindIII сайте, и последовательности праймера для M13for, что свидетельствует о присутствии в «свободном состоянии» фрагментов, входящих в исходный субстрат. Следует отметить, что короткие фрагменты субстрата содержали только часть полилинкера и последовательность праймера для M13 for. Этот факт может означать, что фрагменты сохранились в клетках в течение 15 сут культивирования (либо в одной покоящейся ГСК, либо разошлись по дочерним клеткам) и что фрагменты короче 100 п. н. не видимы для надзирающей ферментативной машины и не подверглись процессингу (сшивание в кольцо или конкатемеры, нуклеазная деградация любого типа) (см. Приложение 2).

Выделено 56 последовательностей с продолжением участка полилинкера плазмиды pBS за сайт рестрикции EcoRV. В большинстве из них присутствуют последовательности, фланкированные одними и теми же инвертированными повторами (TIR), примыкающими к обеим половинкам сайта клонирования *AluI* фрагмента EcoRV

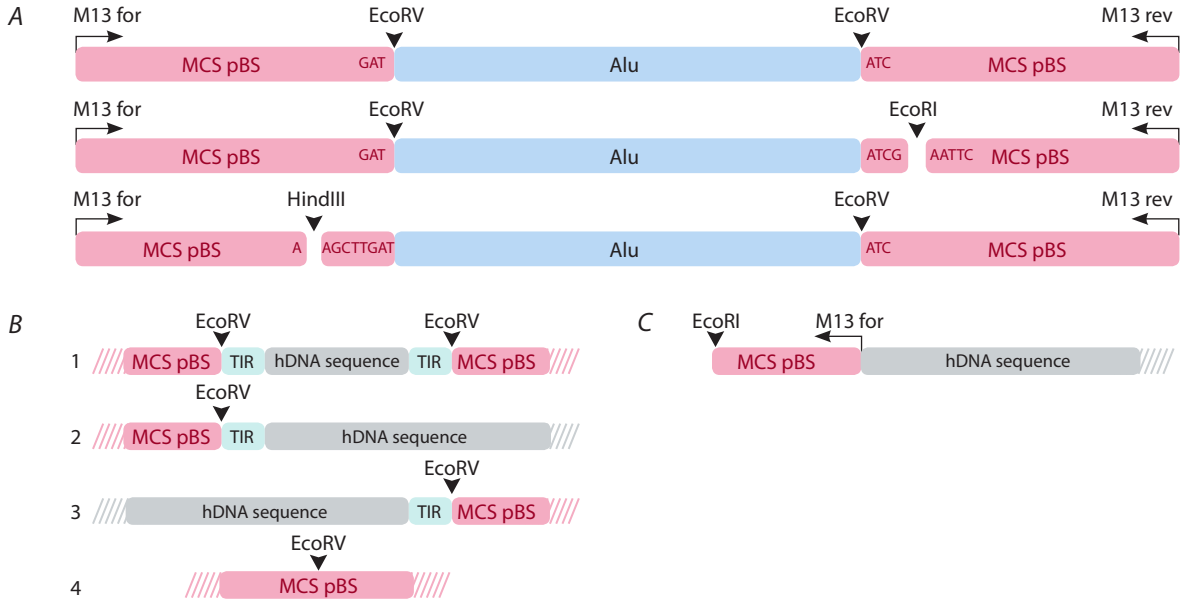


Рис. 3. Схематическое изображение используемого сложно составленного субстрата и нескольких его интермедиатов, возникших в результате индуцированных рекомбинационных процессов:

A – субстрат (смесь M13F-*AluI*-M13R, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII фрагментов), используемый для обработки клеток; *B*, *C* – последовательности, полученные при секвенировании.

(см. рис. 3, *B*1). В четырех ридах *AluI* повтор заменился на последовательность ДНК, состоящую из 84 нуклеотидов, ограниченную с двух сторон терминальными инвертированными повторами длиной 10 п. н. (TIR), примыкающими к половинкам сайта клонирования EcoRV (см. рис. 3, *B*2). Еще в 21 риде последовательность обрывается в направлении от левой половинки сохранившегося сайта клонирования *AluI* EcoRV, не достигает противоположной половинки сайта EcoRV и ограничена тем же инвертированным повтором TIR, прилегающим к сохранившемуся левому сайту EcoRV (см. рис. 3, *B*3). Этот 21 рид представлен тремя различными последовательностями, одна из которых составляет 37 п. н. Две другие имеют длину около 100 п. н. и содержат небольшие участки (5–7 п. н.) гомологии с *AluI* повтором из субстрата.

Одиннадцать ридов содержат интактный полноценный сайт EcoRV, без встройки. Это говорит о рекомбинации последовательности *AluI*, который клонирован по указанному сайту, и о восстановлении исходного EcoRV (см. рис. 3, *B*4).

Пятнадцать ридов совпадают с исходным субстратом, содержащим часть *AluI* повтора.

Обнаружены два рида с продолжительными (54 и 57 п. н.) участками геномной ДНК, объединенными с концом M13 for последовательности. В них полностью входит M13 for праймер и участок полилинкера, обрывающийся на остаточном сайте рестрикции HindIII (см. рис. 3, *C*).

Кроме того, выделены риды с внутренней рекомбинацией по сайту клонирования *AluI* EcoRV, риды с хаотично перетасованными участками полилинкера, в одном из которых выявляются сегменты ДНК несубстратного происхождения.

Экспериментальное доказательство формирования колец длинными фрагментами дцДНК, интернализованными в ГСК, и закольцовывания ими нити ДНК хромосомы

Видимое несоответствие количества флуоресцентных сигналов в экспериментах по FISH гибридизации, экспериментах по прямому анализу интеграции экстрахромосомных фрагментов в геном и результатов полногеномного секвенирования предполагало возможность существования механизма образования стабильных ассоциатов между фрагментами ДНК, локализованными в ядре, и нитями ДНК хромосомы. Поскольку масштабной (единичной) негомологичной интеграции не обнаружено, то было высказано предположение, что в данном случае задействован иной механизм образования стабильных комплексов. Одной их форм таких ассоциатов, которые не нуждаются в белковом креплении, являются закольцованные на нить хромосомы формы ДНК, которые могли быть сформированы доставленными в ядерное пространство фрагментами экстраклеточной дцДНК. Для проверки такой возможности было проведены эксперименты, в целом повторяющие наше исследование (Dolgova et al., 2013).

В экспериментах были проанализированы события и изменения на уровне молекулы, происходящие с субстратом – плазмидой pEGFP-N1, используемым дли обработки ГСК (рис. 4, *A*). В образце нативной плазмиды присутствовала димерная форма и не детектировалась мономерная форма. При трансформации 100 пг линеаризованной pEGFP-N1 (количество, соответствующее трансформации ДНК экспериментального образца со дна солевого градиента, – см. далее) электрокомпетентных клеток DH-5 не обнаружено трансформантов, что говорит об отсутствии кольцевой формы плазмиды.

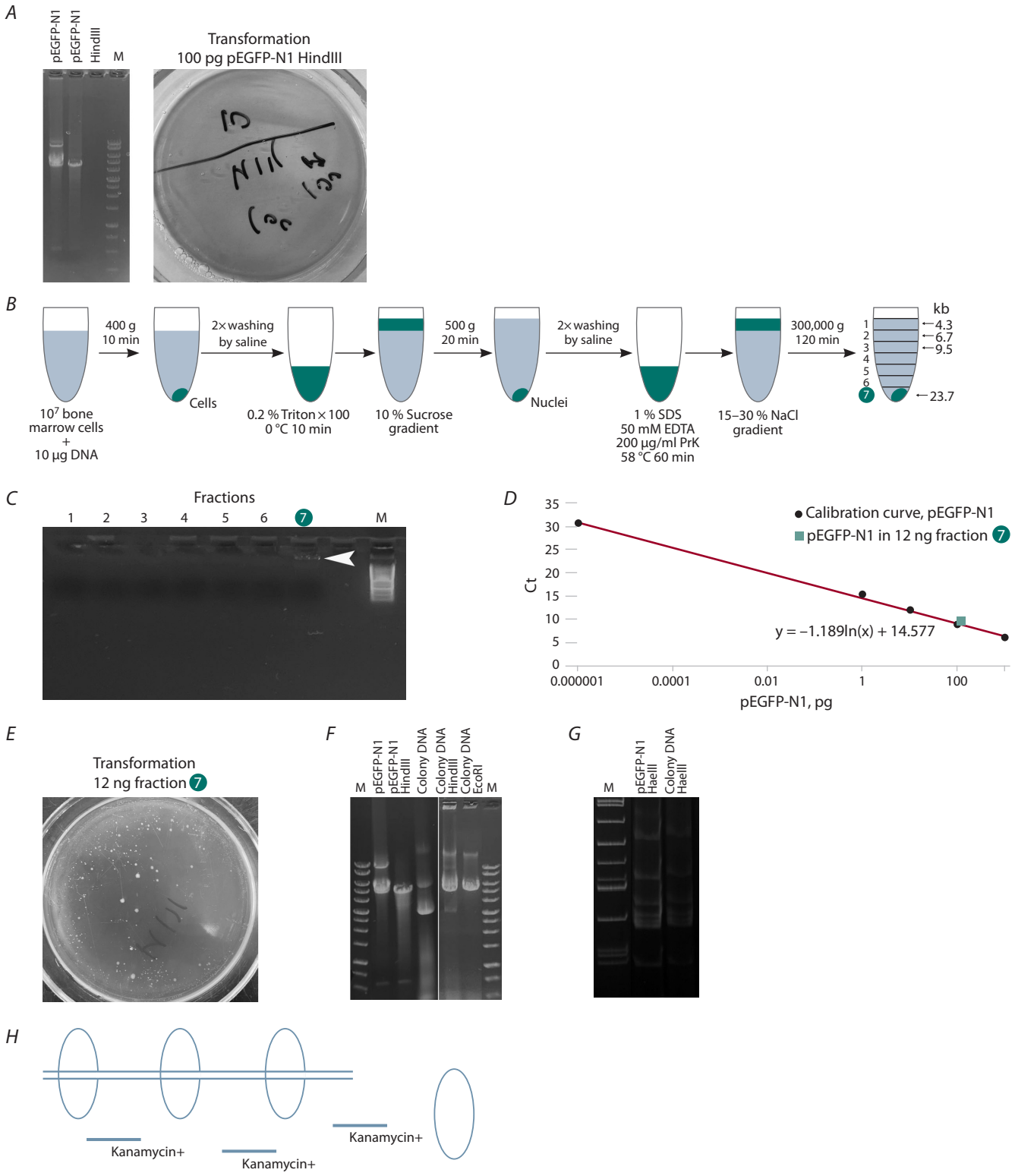


Рис. 4. Экспериментальное доказательство формирования колец длинными фрагментами дцДНК, интернализованными в ГСК, и закольцовывания ими нити ДНК хромосомы.

A – субстрат плазмиды pEGFP-N1, используемый для обработки ГСК. Электрофорез в агарозном геле нативной плазмиды pEGFP-N1, гидролизованной HindIII. Трансформация плазмиды pEGFP-N1, гидролизованной HindIII в *E. coli*, штамм DH-5. *B* – общая схема эксперимента; C – электрофорез образцов в 0.7 % агарозе; *D* – анализ количества ДНК плазмиды pEGFP-N1 в области гена устойчивости к канамицину; *E* – полученные в результате трансформации 12 нг фракции 7 более 50 крупных колоний и около 5000 мелких; *F*, *G* – агарозный и акриламидный гели с различными вариантами гидролизованной и негидролизованной ДНК плазмид: исходной pEGFP-N1 и выделенной из одного из трансформантов (крупная колония). *F* – ДНК, выделенная из крупных колоний, гидролизовалась по HindIII и EcoRI. Рестрикционный анализ свидетельствует о неповрежденном HindIII и EcoRI сайтах полилинкера. *G* – рестрикционный анализ свидетельствует о неповрежденном HaeIII паттерне мономера, а значит, об отсутствии внутриплазмидной рекомбинации. *H* – предполагаемое схематическое изображение закольцованной формы плазмиды, сформированной вокруг хромосомы, после ее попадания в клетку.

На рис. 4, *B* приведена общая схема эксперимента, ее подробное поэтапное описание дано в разделе «Материалы и методы» (Формирование колец экстраклеточными фрагментами дцДНК, интернализованными ГСК и их ближайшими потомками).

При электрофоретическом анализе фракций, полученных после высокоскоростного центрифугирования в солевом градиенте, только в последней фракции (7, дно пробирки) обнаруживается материал ДНК по свечению бромистого этидия (см. рис. 4, *C*).

Анализ количества ДНК плазмиды pEGFP-N1 в области гена устойчивости к канамицину свидетельствовал, что в 12 нг фракции 7 содержится ~100 пг ДНК, гомологичной указанному району плазмиды (см. рис. 4, *D*).

В результате трансформации 12 нг фракции 7 было получено более 50 крупных колоний и около 5000 мелких (см. рис. 4, *E*). Крупные колонии росли в жидкой среде на 50 мкг/мл канамицина и перевивались. Мелкие колонии давали импульс к росту в жидкой среде, и далее культура не развивалась. В мелких колониях ДНК плазмиды в форме мономера или димера не содержалось. Во всех крупных колониях, из которых была выделена ДНК (восемь колоний), присутствовал мономер pEGFP-N1. ДНК, выделенная из крупных колоний, обрабатывалась HindIII, EcoRI (см. рис. 4, *F*) и HaeIII (см. рис. 4, *G*) рестриктазами.

Совокупность полученных в этой серии экспериментов результатов свидетельствует о следующем:

- Клетки костного мозга обрабатывались плазмидной ДНК, линеаризованной по сайту HindIII. Поскольку исходная плазмида присутствует в форме димера, а в полученных трансформантах – в форме мономера, то можно считать, что в клетках костного мозга (ГСК) произошло восстановление кольцевой формы, линеаризованной по HindIII плазмиды.
- Рестрикционный анализ по HaeIII, HindIII и EcoRI подтверждает отсутствие внутриплазмидной рекомбинации (см. рис. 4, *F*, *G*).
- Выделение ДНК плазмид совместно с ДНК хромосом в жестких условиях фракционирования в градиенте NaCl после исчерпывающей обработки протеиназой K и фенол-хлороформной экстракции указывает на существование прочного комплекса/ассоциата между геномной и плазмидной ДНК.

Наиболее правдоподобным вариантом может быть закольцевание геномной ДНК кольцом плазмиды, образовавшимся в момент циклизации линейной формы pEGFP-N1 при ее попадании в клетку (см. рис. 4, *H*).

Полученное резкое несоответствие количества копий участка гена устойчивости к канамицину pEGFP-N1, количества трансформантов и количества выделенной ДНК можно объяснить следующим образом. Многочисленные деградированные плазмидные ДНК с участком гена устойчивости к канамицину могли после трансформации дать однократный синтез белка устойчивости к канамицину, что детектировалось как мелкие колонии. Именно эти фрагменты были оценены в количественной real-time ПЦР. При учете только крупных трансформантов на одну клетку, способную захватывать ДНК, приходится менее одной плазмидной ДНК. Возможно, в случае жизнеспособных трансформантов часть плазмидных колец, освободившихся от ДНК хромосом в процессе выделения, дала колонии. Другая часть осталась в ассоциации с ДНК хромосом и не дала трансформантов, что повлияло на точность количественной оценки.

Таким образом, полученные результаты совпадают с данными работы (Dolgova et al., 2013) и свидетельствуют, что в обоих исследованиях фрагменты дцДНК, интернализованные в ГСК, замыкаются в кольцо. По-видимому, часть колец в момент замыкания стохастически охватывает нить ДНК хромосомы, формируя комплекс, который существует длительное время.

Обсуждение

В рассматриваемой ситуации два типа ДНК структур, а именно ДНК хроматина и двуцепочечные экстрахромосомальные фрагменты, попали в зону активности рекомбиногенной ситуации в качестве рекомбинационных/рекомбиногенных партнеров. Существует серия публикаций, характеризующая экспериментально доказанные и гипотетические интермедиаты, возникающие при рекомбинации ДНК хроматина, ДНК хроматина в области одноцепочечного разрыва и ДНК экстрахромосомальных двуцепочечных фрагментов (Hastings et al., 1993; Leung et al., 1997; Cromie et al., 2001; Li et al., 2001). На рис. 5 приводятся схематические изображения таких интермедиатов. Рассматриваются два варианта взаимоотношений. В первом случае активная рекомбиногенная ситуация создает ферментную платформу рекомбинационного процесса, который возможен в любой точке генома при участии открытых концов двуцепочечных фрагментов (двойной реципрокный обмен между фрагментами и гомологичными им последовательностями хромосом) (см. рис. 5, *A*). Во втором случае рекомбиногенными являются концы нити ДНК с образованным ником (см. рис. 5, *B*).

В настоящей работе предпринята попытка оценить возможность интеграции экстраклеточного материала ДНК в геном с использованием двух подходов – FISH и пол-

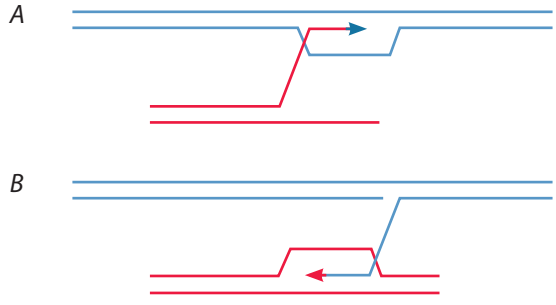


Рис. 5. Схематическое изображение рекомбинационных процессов (однонитевой отжиг, генная конверсия, спаривание микрогомологий), происходящих в клетке с участием геномной ДНК и доставленных в клетку фрагментов дцДНК.

A – рекомбинация за счет открытых двуцепочечных концов доставленных фрагментов; *B* – рекомбинация за счет одноцепочечного разрыва в геномной ДНК.

ногеномного секвенирования. Для этих целей был сконструирован сложно составленный зонд, состоящий из целого фрагмента, представленного центральным *AluI* фрагментом, обрамленным левой и правой частью полилинкера pBS, и нескольких его производных. Производные представляли собой части полного фрагмента субстрата после гидролиза EcoRI и HindIII рестриктазами. В смеси присутствовали пять различных элементов (см. рис. 3, *A*). Предполагалось, что если произойдет объединение ДНК зонда и ДНК хромосом, то с помощью FISH можно будет увидеть специфические сигналы на метафазных хромосомах.

Полученные структурированные в ядре сигналы гибридизации, как следовало из анализов имиджей, могли быть специфическими. Появление FISH сигнала предполагало возможность как гомологичного, так и негомологичного встраивания в геном молекул субстрата. Также было высказано еще одно предположение о возможном происхождении специфических сигналов. Оно было связано с попаданием внехромосомных структур в области активного хроматина, где формируются некие прочные хроматинизированные комплексы между внехромосомной ДНК и ДНК хромосом, сохраняющиеся до формирования метафаз (Møller et al., 2018; Wu et al., 2019; Zhu et al., 2021; Pecorino et al., 2022). Этот вопрос рассмотрен в последней части раздела «Обсуждение».

Для проверки гипотезы было проведено полногеномное и селективное секвенирование. Анализ полученных ридов выявил два варианта интермедиатов, свидетельствующих о рекомбинационных взаимоотношениях доставленных в ядро экстраклеточных фрагментов дцДНК и ДНК хромосом.

1. Обнаружено более 20 последовательностей, содержащих в себе участки геномной ДНК, заместившие собой *AluI* фрагмент четко по сайту клонирования EcoRV. Геномный сегмент ДНК распространяется вглубь от сайта клонирования *AluI* фрагмента EcoRV как с левой, так и с правой стороны ориентированного полилинкера или находится между двумя половинками сайта клонирования EcoRV. Во всех случаях этот геномный фрагмент примыкает к половинкам сайта клонирования EcoRV специфическими последовательностями, определяемыми как терминальный инвертированный повтор (TIR) длиной 10 п. н. В некоторых ридах фрагмент имеет гомологию с клонированным *AluI*, другие содержат небольшие участки (5–7 п. н.) гомологии с *AluI* повтором субстрата (рис. 6).

Известно, что TIR фланкирует последовательность мобильного генетического элемента hAT-5_RIr. hAT представляет собой суперсемейство ДНК-транспозонов, которые перемещаются в геноме за счет фермента транспозазы. Они присутствуют в геномах растений, грибов и животных. В геноме млекопитающих и человека они неактивны.

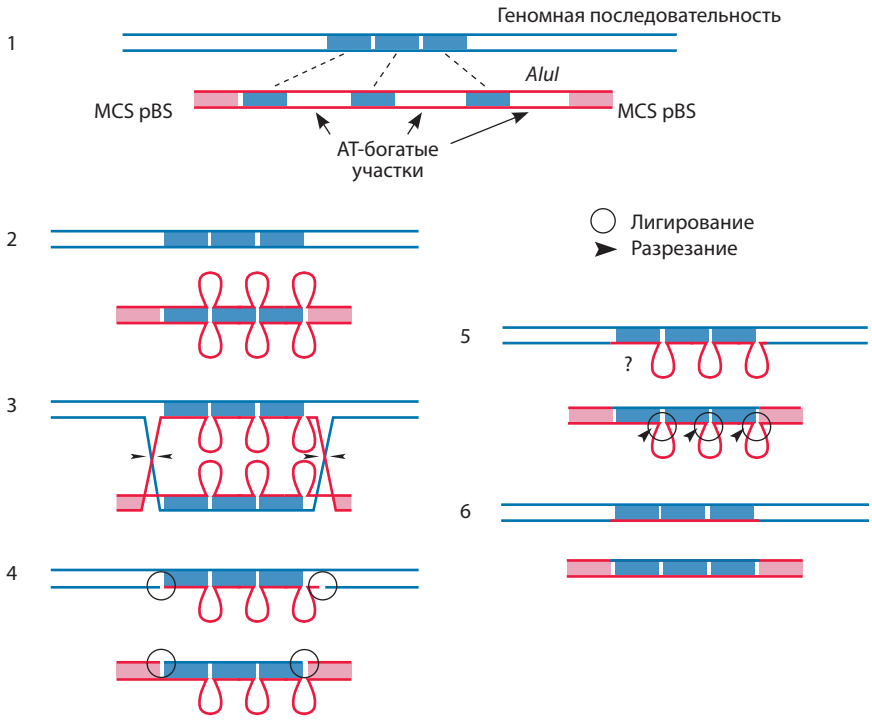


Рис. 6. Схематическое изображение возможных событий в процессе формирования одного из интермедиатов, выявленных при анализе ридов полногеномного секвенирования.

1 – геномная ДНК и доставленный в клетку субстрат (синим цветом выделены гомологичные участки); 2–4 – последовательные этапы рекомбинационных событий с учетом выпетливания AT-богатых негомологичных участков; 5, 6 – полученные в результате рекомбинации предполагаемые продукты, включая прочитанную при секвенировании последовательность.

Имеют размер около 2.5–5 т. п. н. и содержат терминальные инвертированные повторы. Согласно структурам конечных интермедиатов, полное замещение *AluI* повтора исходного субстрата на последовательность ДНК, ограниченную терминальными инвертированными повторами, происходило точно по флангам *AluI* повтора, который также является мобильным генетическим элементом генома человека (Hagan et al., 2003). Для одного из интермедиатов, в котором произошло замещение с сохранением обеих половинок сайта клонирования *AluI* EcoRV, четко прослеживается гомология с *AluI* субстратом.

Для рассматриваемой ситуации важен тот факт, что во всех проанализированных ридах центрально расположенная *AluI* нуклеотидная последовательность ДНК заместилась на некие последовательности из генома строго по сайту клонирования EcoRV. То есть произошла реципрокная гомологичная рекомбинация соответствующего района генома и *AluI* субстрата, при этом районы негомологичной ДНК плазмиды остались вне активности механизма рекомбинации. Не обнаружено ридов, в которых последовательности плазмидной ДНК были бы обрамлены последовательностями геномной ДНК, предполагающими негомологичное встраивание. Один из наиболее логичных вариантов объяснения возможных рекомбинационных событий для рида с сохранением полной структуры субстрата и полного замещения *AluI* последовательности ДНК на геномную последовательность приведен на рис. 6.

2. Выявлены два рида, где правая часть гидролизата (HindIII-праймер M13 for) исходного фрагмента находится в одной непрерывной последовательности с сегментом генома человека (54 и 57 п. н.), объединенных через участок микрогомологии M13 for и генома (рис. 7). При этом вторая часть обоих ридов заканчивается обрывом строго по HindIII сайту рестрикции, который является «конечной точкой» субстрата. Следует отметить, что участок геномной ДНК обоих ридов гомологичен на 95 %. Это означает, что интеграция произошла или в разных местах генома, или в одном и том же месте генома, но в разных клетках, где акцепторные последовательности гомологичны не на 100 %.

Предлагается следующее объяснение появления найденной структуры. В месте начала геномной последовательности интермедиата или в четырех буквах слева от начала геномной последовательности гибридного рида произошел одноцепочечный разрыв в геноме. В указанном сайте четыре нуклеотида геномной последовательности комплементарны (гомологичны) четырем нуклеотидам конца праймера M13 for. По механизму однонитевого отжига за счет комплементарного спаривания микрогомологии и формирования D-loop произошло достраивание по матрице полилинкера, т. е. слияние внехромосомной ДНК и ДНК генома (McVey, Lee, 2008; Hastings et al., 2009). Характерно, что оба рида заканчиваются строго по HindIII сайту рестрикции. Существует минимальная вероятность, что столь прецизионно, четко по HindIII, образовался разрыв исходной ДНК в результате озвучивания ультразвуком сразу для двух фрагментов. Этот факт в большей степени

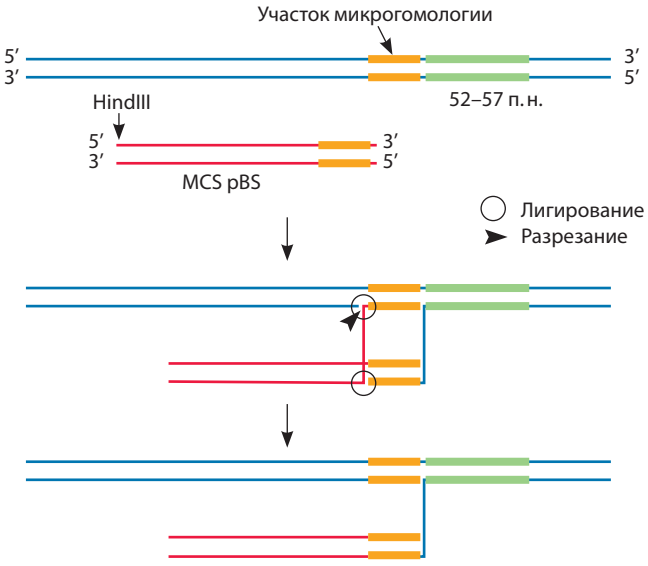


Рис. 7. Схематическое изображение возможных событий в процессе формирования интермедиатов, обнаруженных при анализе результатов полногеномного секвенирования, в которых участок геномной ДНК объединен с полилинкером плазмиды.

Предполагается, что происходит рекомбинация в месте одноцепочечного разрыва геномной ДНК по механизму однонитевого отжига за счет комплементарного спаривания микрогомологии. Синим цветом обозначена последовательность геномной ДНК, зеленым – участок геномной ДНК, детектируемый в двух полученных ридах, желтым – участок микрогомологии, красным – полилинкер pBS (MCS pBS).

означает, что хвосты с концевым HindIII не интегрированы в геном, а свободно свисают в пространстве ядра в течение 15 сут. Возможные интермедиаты указанного процесса приведены на рис. 7.

Таким образом, найдены интермедиаты, доказывающие возможность гомологичного обмена между экстраклеточными ДНК и ДНК хроматина.

Также найдены необычные структуры, которые с одного края ковалентно объединены с последовательностью реципиентного генома, тогда как второй край обрывался на сайте рестрикции, ограничивающем фрагмент субстрата (см. рис. 3, *C*). Вероятно, эта часть последовательности никак не объединена со второй нитью ДНК и «свисает» в пространство ядра или связана с ней белковым каркасом, что необходимо для компактизации хроматина в хромосому и дальнейших процессов клеточного цикла. Предполагается, что в указанных случаях объединение с геномной ДНК произошло по элементам микрогомологии между геномной ДНК и ДНК M13 праймера в сайте одноцепочечного разрыва.

Кроме того, выявлено несколько интермедиатов, свидетельствующих о произошедшей рекомбинации ДНК внутри субстрата. Обнаружены риды со множественными хаотичными перестановками субстратных последовательностей и последовательностей, не относящихся к субстрату, что также подтверждает активные рекомбинационные процессы, идущие в клетке.

Схематическое представление последовательности событий и варианты интермедиатов, возникающих при акте гомологичной рекомбинации

Интермедиат, возникший в результате рекомбинации с геномной ДНК внутреннего *AluI* повтора, клонированного по EcoRV сайту и обрамленного участками полилинкера, ограниченного праймерами к M13. Последовательность геномной ДНК, расположенная после первого терминального инвертированного повтора (TIR), гомологична сегментам *AluI* повтора, которые разделены АТ-богатыми участками (спейсерами) (см. рис. 6). Предположительно, *AluI* последовательность субстрата, содержащая спейсеры, могла сблизиться с гомологичным участком геномной ДНК за счет выпетливания АТ-богатых спейсерных последовательностей и линейного сближения участков *AluI* субстрата, гомологичных геномной ДНК. В предполагаемом варианте могла произойти гомологичная рекомбинация между геномной последовательностью и субстратом, которым обрабатывали клетки. При этом участок *AluI* повтора в экстрахромосомальной ДНК полностью заменился на фрагмент геномной ДНК. Такие события возможны как в месте вскрытой ДНК хроматина, так и в произвольно выбранном участке генома в результате реципрокного гомологичного обмена. Полученный результат свидетельствует об активном процессе гомологичной прецизионной рекомбинации между экстраклеточной ДНК и ДНК хроматина.

Часть полилинкера, обрывающаяся одним концом на сайте рестрикции HindIII и объединенная вторым концом с ДНК геномного происхождения за счет микрогомологии. Геном клетки, вступившей на путь дифференцировки, вскрывается пангеномными одноцепочечными разрывами, которые закрываются после 7 суток. По-видимому, появление найденных интермедиатов связано с существованием одноцепочечных разрывов именно в момент, когда в ядре присутствуют фрагменты экстраклеточного ДНК-зонда (см. рис. 7).

Возможны два варианта объяснения ситуации. Событие интеграции произошло на начальном этапе коммитирования, и интермедиаты возникли и существуют до формирования метафаз. Интермедиаты могли возникнуть во время или незадолго до выделения ДНК из клеток, на 15-е сутки культивирования. Данный вариант предполагает длительное сохранение в пространстве ядра коротких фрагментов субстрата. Как возникшие интермедиаты сохраняются в процессе дупликации и сохраняются ли вообще, непонятно. Если рассматривать вариант интеграции, то, вероятно, встраивание произошло после удвоения хромом на этих участках. Для конденсации хроматина и формирования метафазной хромосомы такие нарушения после прошедшей репликации, по-видимому, значения не имеют. Маловероятно, что полимераза проходит такие интермедиаты в ходе репликации, в таком случае они должны быть фиксированы специфическими комплексами.

Как было сказано выше, участок геномной ДНК обоих ридов гомологичен на 95 %. Это означает, что интеграция произошла или в разных местах генома, или в одном и том же месте генома, но в разных клетках, где акцепторные последовательности гомологичны не на 100 %.

Риды с множественными хаотичными перестановками субстратных последовательностей и последовательностей, не относящихся к субстрату. Схематическое изображение последовательностей со множественными хаотичными перестановками, полученных при секвенировании, приведено на рис. 8.

В последней серии экспериментов была проведена оценка возможности интеграции экстраклеточных фрагментов в геном прямой детекцией интегрированного фрагмента. Использовались два типа TAMRA-меченой ДНК, гомологичная ДНК человека и ДНК *E. coli*. Главным выводом экспериментов является тот факт, что в случае ДНК человека специфические сигналы детектируются по всему телу хромосом. В случае ДНК *E. coli* эти сигналы в большей степени локализованы в центромерном гетерохроматине. Если предположить, что произошла интеграция в геном, то мода этой интеграции для гомологичных и гетерологичных ДНК будет различной, а именно гомологичное встраивание человеческого субстрата и негомологичная интеграция ДНК *E. coli*. Центромерный гетерохроматин представляет собой балк (конгломерат) повторов, по массе соответствующих 1/3 массы хромосомы, создающих известную ее трехмерную архитектуру. Если произошла реальная интеграция в центромерный гетерохроматин, то в силу несопоставимости размеров и специфической организации α-сателлитов, попадание дополнительных ДНК в эту огромную массу ДНК, по-видимому, не отражается на функциональном состоянии центромерного локуса. Такая точка зрения может объяснить реальную интеграцию гетерологичной ДНК.

Противоречия между результатами FISH, прямой детекцией флуоресцентно-меченого зонда на метафазах и данными полногеномного секвенирования

При анализе результатов FISH, прямой детекции возможной интеграции флуоресцентно-меченого зонда в геном и данных полногеномного секвенирования в проведенных экспериментах с использованием зонда, состоящего из смеси фрагментов M13F-*AluI*-M13R, M13F-*AluI*-EcoRI и M13R-*AluI*-HindIII, обращает на себя внимание выраженное несоответствие количества сигналов гибридизации, или просто специфических флуоресцентных сигналов, и количества ридов, демонстрирующих реальное ковалентное объединение последовательности субстрата и последовательности хромосомы. Первое наиболее очевидное предположение состояло в том, что наблюдается артефакт метода. Тем не менее различные экспериментальные варианты, и в первую очередь структурированность сигналов в ядре, свидетельствовали о необычности и специфичности сигналов. В этой связи было принято решение провести дополнительные эксперименты, которые, как предполагалось, могут раскрыть суть механизма появления сигналов на метафазах.

Полученные результаты перевели наше внимание на явление, характеризующее присутствие в ядре большого

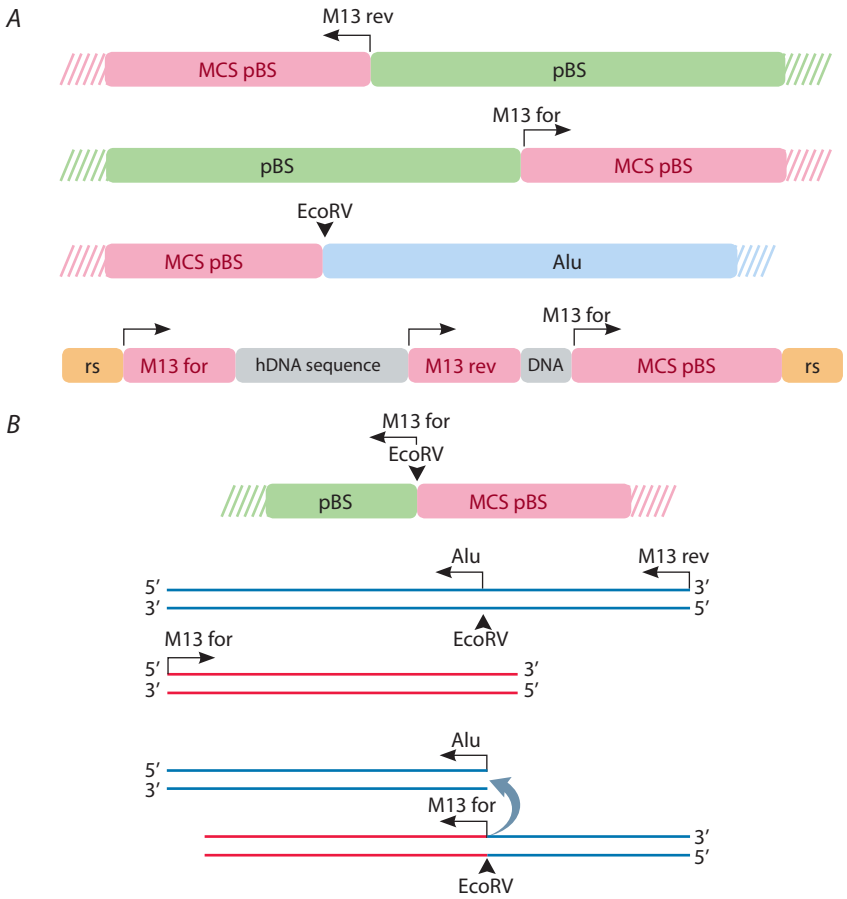


Рис. 8. Схематическое изображение последовательностей, полученных при секвенировании (*A*), и возможный вариант репаративного процесса (*B*).

количества кольцеобразной внеклеточной ДНК, которое активно дискутируется (Pecorino et al., 2022).

Вопрос о внехромосомной кольцевой ДНК естественного происхождения в опухолевых клетках различных типов опухолей начинается с обнаружения более 60 лет назад double minutes кольцевых экстрахромосомальных структур, не имеющих центромер и теломер (Spriggs et al., 1962). Отсутствие центромер приводит к случайному распределению внеклеточной ДНК при делении клеток, и закодированная в них информация передается неменделевским образом. Эти структуры формируются из хромосомного материала. Выделяют как минимум два типа таких структур. Один тип состоит из простых одиночных фрагментов ДНК, другой может быть химерным и содержать множество фрагментов из различных частей хромосом. Данный факт говорит о том, что выпавшие из хромосом в результате неопределенных событий фрагменты ДНК бессистемно лигируются в кольцо. Часть ДНК представляют собой структуры размером до 1 млн п. н., которые содержат множество генов и регуляторных областей. Другие, так называемые микро-внехромосомные кольцевые ДНК, находятся в пределе 100–2000 п. н. и характерны преимущественно для клеток здоровых тканей. В этих ДНК обнаруживают рибосомальные гены, ДНК мобильных элементов, кольцевые теломерные ДНК. В опухолевых клетках такие внехромосомные структуры могут реинтегрировать в геном, причем в «неродные» хромосомные области с гомогенным окрашиванием, содержащие амплифицированные гены (Pecorino et al., 2022). Было установлено, что внехромосомная кольцевая ДНК тесно связана с полимеразным комплексом (RNAPII) и, как предполагается, ассоциирована с активно транскрибирующимися доменами хромосом, с которыми образует фокусы концентрации хроматина (Wu et al., 2019; Zhu et al., 2021). Вероятно, ассоциаты внехромосомной кольцевой ДНК и ДНК хромосом стабильны и могут существовать в различные фазы клеточного цикла.

Аналогичные ДНК присутствуют в миоцитах и лейкоцитах человека как регулярно встречающиеся внехромосомные объекты и представляют собой общебиологическое явление (Møller et al., 2018).

В свете сказанного можно было бы предположить, что в нашем исследовании зонд, содержащий плазмидный таг, сформировал фокусы концентрации хроматина с хромосомами, не интегрировав в молекулу ДНК хромосомы. Эти ассоциаты сохранились до метафазы и дали множественные специфические сигналы гибридизации. Находится ли в соответствии с полученными данными

такое предположение? Возможна ли такая форма интермедиатов?

Было проведено полногеномное секвенирование. Результаты свидетельствуют, что гомологичная интеграция зонда возможна вследствие спаривания микрогомологий или гомологичной рекомбинации. Однако количество проанализированных ридов-интегратов не соответствует картине гибридизации, где детектируется множество специфических сигналов (как предполагалось первоначально, сайтов гибридизации) на метафазах, и, особенно, их структурированность в интерфазных ядрах. Также остались необъясненными результаты анализа прямой интеграции, где обнаружены множественные специфические сигналы при использовании флуоресцентно-меченого как гомологичного, так и гетерологичного зондов (см. рис. 2).

Количественная оценка данных секвенирования тоже не согласуется с картиной гибридизации. В секвенирование берется 40 нг ДНК, что соответствует ~4000 ядер. Суммарно при двух повторах полногеномного прочтения было обнаружено приблизительно по 50 ридов, несущих таг, для каждого сиквенса, т. е. на 80 клеток приходится одна молекула субстрата. Это также не соответствует картине гибридизации, где детектируется множество специфических сигналов (сигналов гибридизации) на метафазах и интерфазных ядрах (см. рис. 2).

Факты и их анализ предполагают формирование большого количества прочных ассоциатов, которые вместе с реальной интеграцией дают общую картину FISH. Почему многочисленные ассоциаты не детектируются при полногеномном секвенировании? Одно из предположений сводится к технической проблеме.

В эксперименте по прямому доказательству интернализации фрагментов дцДНК (9 теломерных повторов, 54 п. н.) в CD34 клетки человека и мыши было установлено, что дцДНК зонд доставляется как в CD34+, так и в CD34– клетки (Рузанова и др., 2024).

Оказалось, что в CD34+ клетках дцДНК субстрат, которым обрабатывали клетки, сохраняется в исходной форме в виде интактного фрагмента. При этом в небольшом количестве CD34– клеток доставленные фрагменты конкатемеризуются, и когда количество мономеров (54 п. н.) достигает 7 единиц (~350 п. н.), возможно, происходит их циклизация (Dolgova et al., 2013). По-видимому, CD34+ клетки, которые захватили фрагменты, не способны их процессировать до момента перехода в коммитированное состояние. Можно предположить два варианта событий. В первом случае CD34– клетки первого эшелона коммитирования, которые сохранили способность захватывать экстраклеточные ДНК, процессируют и циклизуют интернализованные мономеры. Во втором случае CD34+ клетки быстро, в процессе инкубации с экстраклеточным ДНК субстратом, переходят в терминальную дифференцировку в результате появления в клетке экстраклеточных фрагментов ДНК, теряют маркер CD34 и становятся способными процессировать депонированные внутри клетки фрагменты и формировать кольцеобразные структуры, что детектируется как свойство CD34+ клеток. Такие формы ДНК ассоциатов доходят до метафаз.

Техника полногеномного секвенирования на платформе Illumina предполагает фрагментацию геномной ДНК до фрагментов 100–200 п. н. Кольцо размером 100–300 п. н. с большой долей вероятности количественно не будет разрушено ультразвуком, и последовательности плазмидного тага (~200 п. н. в анализируемом эксперименте) не будут прочитаны, хотя на хромосомах они будут присутствовать в полном объеме. На наш взгляд, в этом предположении кроется причина противоречия результатов FISH и полногеномного секвенирования, а также прямой детекции возможной интеграции флуоресцентно-меченого зонда.

Сопоставление размера кольца, сформированного олигонуклеотидом (около 350 п. н., см. (Рузанова и др., 2024)), т. е. диаметра его охвата линейной молекулы интерфазной хромосомы, и размера полимеразного комплекса предполагает возможность свободного перемещения репликативной ферментативной машины сквозь кольцо в ходе репликации (толщина нити дцДНК ~2 нм, один виток дцДНК 10.5 п. н. ~ 3.4 нм, фрагмент 300 п. н. ~ 100 нм, диаметр кольца дцДНК 300 п. н. ~ 32 нм, полимеразный комплекс ~300 кДа ~ 20 нм) (https://www.dynamic-biosensors.com/project/list-of-protein-hydrodynamic-diameters/). Кольцо сохраняется до метафазы, где детектируется FISH.

Прямое доказательство формирования длинными фрагментами экстраклеточной дцДНК (pEGFP-N1 линеаризованная HindIII) кольца, охватывающего нить хромосомы

Высказанные сомнения и объективные противоречия подняли вопрос о природе сигналов на метафазных хромосомах, на который было необходимо ответить.

В нашей работе (Dolgova et al., 2013) на примере линеаризованной плазмиды pEGFP-N1 (4.7 т. п. н.) установлено, что если в ГСК (мышиная модель) индуцирована рекомбиногенная ситуация, спровоцированная появлением двуцепочечных разрывов при воздействии кросслинкирующего цитостатика циклофосфамида (феномен «окно смерти»), то линейные фрагменты плазмиды, интернализованные в эти клетки, претерпевают метаболические изменения. Они формируют конкатемеры, которые «зашиваются» в кольцо (до 4 мономеров в одном кольце). Часть фрагментов претерпевает концевой гидролиз глубиной до 1 т. п. н. и в такой форме присутствует в объеме ядра. Важным из результатов цитируемой работы является тот факт, что кольцевые формы плазмиды выделяются вместе с ДНК хромосом, находящейся на дне 15–30 % NaCl градиента (~300000g) после исчерпывающей обработки протеиназой K и фенольной экстракции. То есть плазмидная ДНК находится в контакте с ДНК хромосом, который сохраняется в условиях жесткого фракционирования без крепления к ней белковым каркасом. С позиции здравого смысла такое возможно или при интеграции плазмиды в состав хромосомы, или при формировании кольца вокруг нити хромосомы, что «заставляет» кольцо плазмиды седиментировать на дно солевого градиента вместе с геномной ДНК. В момент осмысления полученных результатов мы не придали значения этому факту.

В настоящем исследовании показано, что, аналогично данным цитируемой работы, в ГСК человека при попадании в нее экстраклеточных дцДНК фрагментов индуцируется рекомбиногенная ситуация, спровоцированная никами, которые эти фрагменты и инициировали. Сравнение работ (Рузанова и др., 2024) и (Dolgova et al., 2013) свидетельствует, что при развертывании двух рекомбиногенных ситуаций, имеющих различное происхождение (индуцированные двуцепочечными разрывами и никами), попавшие в клетку экстраклеточные фрагменты на первом этапе метаболических превращений претерпевают схожие изменения. Происходит их конкатемеризация (сшивка/лигирование по концам). Это могло означать, что все остальные превращения фрагментов в ядре аналогичны показанным в работе (Dolgova et al., 2013). То есть фрагменты так же конкатемеризуются и/или зашиваются в кольцо, и в процессе циклизации в стохастическом режиме нить ДНК хромосом может закольцовываться циклизованной плазмидой. Следовательно, можно предположить, что именно такие структуры детектируются во всех экспериментах, где проводили анализ флуоресцентных сигналов на метафазных хромосомах.

Для доказательства высказанного предположения была выполнена серия экспериментов, практически полностью повторяющих эксперименты из (Dolgova et al., 2013) (см. рис. 4). Единственным и принципиальным отличием (за исключением модели «мышь *vs* клетки костного мозга человека») было то, что в цитируемой работе мы опирались на развернутую «рекомбиногенную ситуацию», индуцированную двуцепочечными разрывами (межцепочечными сшивками), а в настоящем случае – на рекомбиногенную ситуацию, индуцированную никами. Были получены практически идентичные результаты. Плазмида в линейной форме попадает в ГСК, происходит ее циклизация по липкому концу (HindIII) и формируется кольцо. В процессе циклизации нить ДНК хромосомы закольцовывается плазмидой (при жестком фракционировании выделяется вместе с фракцией ДНК хромосом, находящейся на дне 15–30 % NaCl градиента). Такая структура длительно присутствует в ядре и вполне может детектироваться в составе метафазной хромосомы (см. рис. 4). Такое видение событий согласуется с данными полногеномного секвенирования, свидетельствующими об отсутствии негомологичной масштабной интеграции, которая давала бы множественные детектируемые специфические сигналы на метафазных хромосомах.

Что происходит с такими интермедиатами в процессе упаковки линейной хромосомы в метафазу и в процессе метафазного деления, неизвестно. Можно предположить, что или кольца рвутся физически, или существует ферментативная машина, осуществляющая линеаризацию кольца. В любом случае сохраненная таким образом генетическая информация может составлять некий информационный «резервуар», назначение которого неизвестно. Свободно расположенные кольца, сформированные фрагментами экстраклеточной дцДНК, могут являться аналогами колец, которые участвуют в механизме альтернативного удлинения теломер.

Заключение

Таким образом, во время появления пангеномных одноцепочечных разрывов (ников) в клетке активируются репарационно-рекомбинационные механизмы, индуцирующие рекомбинационные процессы между ДНК хроматина и экзогенной дцДНК. Важен тот факт, что не обнаружено ридов, демонстрирующих негомологичное (NHEJ) объединение экстраклеточного/экстрахромосомного ДНК субстрата и ДНК хромосом. Это свидетельствует о том, что во время данного общебиологического процесса возможно гомологичное изменение генома с минимальным риском негомологичных случайных, неконтролируемых, нарушающих функцию интеграций. По-видимому, экстраклеточные фрагменты ДНК, попавшие в ГСК, могут конкатемеризоваться, образовывать кольца как мономерной, так и конкатемерной формы, формировать кольцеобразные или иные пространственные ассоциаты с хромосомной ДНК, которые длительно присутствуют в клетке. Экстрахромосомальные кольца являются ДНК-матрицей, используемой клеткой в механизме альтернативного удлинения теломер. И наконец, экстраклеточные фрагменты дцДНК, находящиеся в клетке под давлением рекомбинационной ситуации, индуцированной как двуцепочечными разрывами, так и никами, претерпевают близкие метаболические превращения, что подразумевает общность и непротиворечивость двух процессов, которые «не конфликтуют» и могут дополнять один другой.

Список литературы / References

Лихачева А.С., Рогачев В.А., Николин В.П., Попова Н.А., Шилов А.Г., Себелева Т.Е., Струнки Д.Н., Черных Е.Р., Гельфгат Е.Л., Богачев С.С., Шурдов М.А. Участие экзогенной ДНК в молекулярных процессах, протекающих в соматической клетке. *Информационный вестник ВОГиС*. 2008;12(3):426-447
[Likhacheva A.S., Rogachev V.A., Nikolin V.P., Popova N.A., Shilov A.G., Sebeleva T.E., Strunkin D.N., Chernykh E.R., Gel'fgat E.L., Bogachev S.S., Shurdov M.A. Involvement of exogenous DNA in the molecular processes in somatic cell. *Informatsionnyy Vestnik VOGiS = Herald Vavilov Soc Genet Breed Sci*. 2008;12(3):426-447 (in Russian)]

Маниатис Т., Фрич Э., Сэмбрук Д. Методы генетической инженерии. Молекулярное клонирование. М.: Мир, 1984
[Maniatis T., Fritch E., Sambrook D. Methods of Genetic Engineering. Molecular Cloning. Moscow: Mir Publ., 1984 (in Russian)]

Рузанова В.С., Ошихмина С.Г., Проскурина А.С., Риттер Г.С., Кирикович С.С., Левитес Е.В., Ефремов Я.Р., Карамышева Т.В., Мещанинова М.И., Мамаев А.Л., Таранов О.С., Богачев А.С., Сидоров С.В., Никонов С.Д., Леплина О.Ю., Останин А.А., Черных Е.Р., Колчанов Н.А., Долгова Е.В., Богачев С.С. Концепция природной реконструкции генома. Часть 2. Влияние фрагментов экстраклеточной двуцепочечной ДНК на гемопоэтические стволовые клетки. *Вавиловский журнал генетики и селекции*. 2024;28(8):993-1007. doi 10.18699/vjgb-24-106
[Ruzanova V.S., Oshikhmina S.G., Proskurina A.S., Ritter G.S., Kirikovich S.S., Levites E.V., Efremov Y.R., Karamysheva T.V., Meschaninova M.I., Mamaev A.L., Taranov O.S., Bogachev A.S., Sidorov S.V., Nikonov S.D., Leplina O.Y., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Dolgova E.V., Bogachev S.S. A concept of natural genome reconstruction. Part 2. Effect of extracellular double-stranded DNA fragments on hematopoietic stem cells. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2024;28(8):993-1007. doi 10.18699/vjgb-24-106]

Cromie G.A., Connelly J.C., Leach D.R.F. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol Cell*. 2001;8(6):1163-1174. doi 10.1016/S1097-2765(01)00419-1

Dolgova E.V., Nikolin V.P., Popova N.A., Proskurina A.S., Orishenko K.E., Alyamkina E.A., Efremov Y.R., … Taranov O.S., Rogachev V.A., Sidorov S.V., Bogachev S.S., Shurdov M.A. Internalization of exogenous DNA into internal compartments of murine bone marrow cells. *Russ J Genet Appl Res*. 2012;2(6):440-452. doi 10.1134/S2079059712060056

Dolgova E.V., Efremov Y.R., Orishchenko K.E., Andrushkevich O.M., Alyamkina E.A., Proskurina A.S., Bayborodin S.I., … Omigov V.V., Minkevich A.M., Rogachev V.A., Bogachev S.S., Shurdov M.A. Delivery and processing of exogenous double-stranded DNA in mouse CD34+ hematopoietic progenitor cells and their cell cycle changes upon combined treatment with cyclophosphamide and double-stranded DNA. *Gene*. 2013;528(2):74-83. doi 10.1016/j.gene.2013.06.058

Farzaneh F., Zalin R., Brill D., Shall S. DNA strand breaks and ADP-ribosyl transferase activation during cell differentiation. *Nature*. 1982;300(5890):362-366. doi 10.1038/300362A0

Glover D.M. (Ed.) DNA Cloning. A Practical Approach. IRL Press, 1988

Hagan C.R., Sheffield R.F., Rudin C.M. Human Alu element retrotransposition induced by genotoxic stress. *Nat Genet*. 2003;35(3):219-220. doi 10.1038/ng1259

Hastings P.J., McGill C., Shafer B., Strathern J.N. Ends-in *vs*. ends-out recombination in yeast. *Genetics*. 1993;135(4):973-980. doi 10.1093/genetics/135.4.973

Hastings P.J., Ira G., Lupski J.R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet*. 2009;5(1):e1000327. doi 10.1371/journal.pgen.1000327

Holmgren L., Szeles A., Rajnavölgyi E., Folkman J., Klein G., Ernberg I., Falk K.I. Horizontal transfer of DNA by the uptake of apoptotic bodies. *Blood*. 1999;93(11):3956-3963. doi 10.1182/blood.V93.11.3956

Jacobson G.K., Pinon R., Esposito R.E., Esposito M.S. Single-strand scissions of chromosomal DNA during commitment to recombination at meiosis. *Proc Natl Acad Sci USA*. 1975;72(5):1887-1891. doi 10.1073/pnas.72.5.1887

Johnstone A.P., Williams G.T. Role of DNA breaks and ADP-ribosyl transferase activity in eukaryotic differentiation demonstrated in human lymphocytes. *Nature*. 1982;300(5890):368-370. doi 10.1038/300368A0

Kucherlapati R.S., Eves E.M., Song K.Y., Morse B.S., Smithies O. Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. *Proc Natl Acad Sci USA*. 1984;81(10):3153-3157. doi 10.1073/pnas.81.10.3153

Kumagai A., Dunphy W.G. Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in *Xenopus* egg extracts. *Mol Cell*. 2000;6(4):839-849. doi 10.1016/S1097-2765(05)00092-4

Leung W., Malkova A., Haber J.E. Gene targeting by linear duplex DNA frequently occurs by assimilation of a single strand that is subject to preferential mismatch correction. *Proc Natl Acad Sci USA*. 1997;94(13):6851-6856. doi 10.1073/pnas.94.13.6851

Li J., Read L.R., Baker M.D. The mechanism of mammalian gene replacement is consistent with the formation of long regions of heteroduplex DNA associated with two crossing-over events. *Mol Cell Biol*. 2001;21(2):501-510. doi 10.1128/mcb.21.2.501-510.2001

Lin J., Krishnaraj R., Kemp R.G. Exogenous ATP enhances calcium influx in intact thymocytes. *J Immunol*. 1985;135(5):3403-3410

Lin Y., Waldman A.S. Capture of DNA sequences at double-strand breaks in mammalian chromosomes. *Genetics*. 2001;158(4):1665-1674. doi 10.1093/genetics/158.4.1665

MacDougall C.A., Byun T.S., Van C., Yee M.C., Cimprich K.A. The structural determinants of checkpoint activation. *Genes Dev*. 2007;21(8):898-903. doi 10.1101/gad.1522607

Maizels N., Davis L. Initiation of homologous recombination at DNA nicks. *Nucleic Acids Res*. 2018;46(14):6962-6973. doi 10.1093/nar/gky588

Maricic T., Whitten M., Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. 2010;5(11):e14004. doi 10.1371/journal.pone.0014004

McVey M., Lee S.E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet*. 2008;24(11):529-538. doi 10.1016/j.tig.2008.08.007

Møller H.D., Mohiyuddin M., Prada-Luengo I., Sailani M.R., Halling J.F., Plomgaard P., Maretty L., Hansen A.J., Snyder M.P., Pilegaard H., Lam H.Y.K., Regenberg B. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun*. 2018;9(1):1069. doi 10.1038/S41467-018-03369-8

Murnane J.P., Yezzi M.J., Young B.R. Recombination events during integration of transfected DNA into normal human cells. *Nucleic Acids Res*. 1990;18(9):2733-2738. doi 10.1093/nar/18.9.2733

Pecorino L.T., Verhaak R.G.W., Henssen A., Mischel P.S. Extrachromosomal DNA (ecDNA): an origin of tumor heterogeneity, genomic remodeling, and drug resistance. *Biochem Soc Trans*. 2022;50(6):1911-1920. doi 10.1042/BST20221045

Perucho M., Hanahan D., Wigler M. Genetic and physical linkage of exogenous sequences in transformed cells. *Cell*. 1980;22(1):309-317. doi 10.1016/0092-8674(80)90178-6

Perucho M., Wigler M. Linkage and expression of foreign DNA in cultured animal cells. *Cold Spring Harb Symp Quant Biol*. 1981;45(Pt.2):829-838. doi 10.1101/sqb.1981.045.01.101

Petrova D.D., Dolgova E.V., Proskurina A.S., Ritter G.S., Ruzanova V.S., Efremov Y.R., Potter E.A., Kirikovich S.S., Levites E.V., Taranov O.S., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. The new general biological property of stem-like tumor cells (Part II: surface molecules, which belongs to distinctive groups with particular functions, form a unique pattern characteristic of a certain type of tumor stem-like cells). *Int J Mol Sci*. 2022;23(24):15800. doi 10.3390/ijms232415800

Pierandrei S., Luchetti A., Sanchez M., Novelli G., Sangiuolo F., Lucarelli M. The gene targeting approach of small fragment homologous replacement (SFHR) alters the expression patterns of DNA repair and cell cycle control genes. *Mol Ther Nucleic Acids*. 2016;5(4):e304. doi 10.1038/mtna.2016.2

Potter E.A., Dolgova E.V., Proskurina A.S., Efremov Y.R., Minkevich A.M., Rozanov A.S., Peltek S.E., … Baiborodin S.I., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. Gene expression profiling of tumor-initiating stem cells from mouse Krebs-2 carcinoma using a novel marker of poorly differentiated cells. *Oncotarget*. 2017;8(6):9425-9441. doi 10.18632/oncotarget.14116

Potter E.A., Dolgova E.V., Proskurina A.S., Ruzanova V.S., Efremov Y.R., Kirikovich S.S., Oshikhmina S.G., … Grivtsova L.U., Kolchanov N.A., Ostanin A.A., Chernykh E.R., Bogachev S.S. Stimulation of mouse hematopoietic stem cells by angiogenin and DNA preparations. *Braz J Med Biol Res.* 2024;57:e13072. doi 10.1590/1414-431X2024e13072

Ritter G.S., Dolgova E.V., Petrova D.D., Efremov Y.R., Proskurina A.S., Potter E.A., Ruzanova V.S., Kirikovich S.S., Levites E.V., Taranov O.S., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. The new general biological property of stem-like tumor cells. Part I. Peculiarities of the process of the double-stranded DNA fragments internalization into stem-like tumor cells. *Front Genet*. 2022;13:954395. doi 10.3389/fgene.2022.954395

Rogachev V.A., Likhacheva A., Vratskikh O., Mechetina L.V., Sebeleva T.E., Bogachev S.S., Yakubov L.A., Shurdov M.A. Qualitative and quantitative characteristics of the extracellular DNA delivered to the nucleus of a living cell. *Cancer Cell Int*. 2006;6:23. doi 10.1186/1475-2867-6-23

Ruzanova V., Proskurina A., Efremov Y., Kirikovich S., Ritter G., Levites E., Dolgova E., Potter E., Babaeva O., Sidorov S., Taranov O., Ostanin A., Chernykh E., Bogachev S. Chronometric administration of cyclophosphamide and a double-stranded DNA-mix at interstrand crosslinks repair timing, called "Karanahan" therapy, is highly efficient in a weakly immunogenic Lewis carcinoma model. *Pathol Oncol Res*. 2022;28:1610180. doi 10.3389/pore.2022.1610180

Smith A.J.H., Berg P. Homologous recombination between defective neo genes in mouse 3T6 cells. *Cold Spring Harb Symp Quant Biol*. 1984;49:171-181. doi 10.1101/sqb.1984.049.01.020

Smithies O., Gregg R.G., Boggs S.S., Koralewski M.A., Kucherlapati R.S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*. 1985;317(6034):230-234. doi 10.1038/317230A0

Spriggs A.I., Boddington M.M., Clarke C.M. Chromosomes of human cancer cells. *Br Med J*. 1962;2(5317):1431-1435. doi 10.1136/bmj.2.5317.1431

Thomas K.R., Folger K.R., Capecchi M.R. High frequency targeting of genes to specific sites in the mammalian genome. *Cell*. 1986;44(3):419-428. doi 10.1016/0092-8674(86)90463-0

Vriend L.E.M., Krawczyk P.M. Nick-initiated homologous recombination: protecting the genome, one strand at a time. *DNA Repair*. 2017;50:1-13. doi 10.1016/j.dnarep.2016.12.005

Wu S., Turner K.M., Nguyen N., Raviram R., Erb M., Santini J., Luebeck J., … Furnari F.B., Chang H.Y., Ren B., Bafna V., Mischel P.S. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature*. 2019;575(7784):699-703. doi 10.1038/S41586-019-1763-5

Yakubov L.A., Deeva E.A., Zarytova V.F., Ivanova E.M., Ryte A.S., Yurchenko L.V., Vlassov V.V. Mechanism of oligonucleotide uptake by cells: involvement of specific receptors? *Proc Natl Acad Sci USA*. 1989;86(17):6454-6458. doi 10.1073/pnas.86.17.6454

Yoo H.Y., Shevchenko A., Shevchenko A., Dunphy W.G. Mcm2 is a direct substrate of ATM and ATR during DNA damage and DNA replication checkpoint responses. *J Biol Chem*. 2004;279(51):53353-53364. doi 10.1074/jbc.M408026200

Yoo H.Y., Jeong S.Y., Dunphy W.G. Site-specific phosphorylation of a checkpoint mediator protein controls its responses to different DNA structures. *Genes Dev*. 2006;20(7):772-783. doi 10.1101/GAD.1398806

Zhu Y., Gujar A.D., Wong C.H., Tjong H., Ngan C.Y., Gong L., Chen Y.A., … Yi E., deCarvalho A.C., Ruan Y., Verhaak R.G.W., Wei C.L. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell*. 2021;39(5):694-707.e7. doi 10.1016/j.ccell.2021.03.006

Zou L. Single- and double-stranded DNA: building a trigger of ATR-mediated DNA damage response. *Genes Dev*. 2007;21(8):879-885. doi 10.1101/gad.1550307

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 04.07.2024. После доработки 09.04.2025. Принята к публикации 12.05.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):181-193

doi 10.18699/vjgb-26-19

Naïve human iPSCs obtained by culturing the ICGi022-A cell line with primed pluripotency in HENSM medium efficiently differentiate into endothelial derivatives

M.A. Arssan, A.I. Shevchenko, S.M. Zakian, I.S. Zakharova

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
arsanmhdamin@yandex.ru

Abstract. Naïve human pluripotent stem cells (PSCs) are a promising new tool in biomedical research. They provide access to the early embryonic development programmes and offer breakthrough solutions in regenerative medicine. However, the current inability to obtain long-term cultures of genetically and epigenetically stable naïve human PSC lines poses a challenge to their effective application in biomedicine. The recently proposed HENSM culture medium is claimed to enable the obtaining and long-term maintenance of naïve PSC lines. In this study, the potential of the HENSM medium for obtaining stable naïve human PSC lines was investigated. We successfully reset the primed induced pluripotent stem cell (iPSC) line ICGi022-A (K7-4Lf), derived from a healthy donor, to a naïve state using the HENSM medium. Naïve iPSCs grow in the form of dome-shaped colonies, both with and without a feeder layer of cells. The resulting cells retained expression of the key pluripotency factors and activated the naïve PSC transcriptional programme, including expression of endogenous retroviral elements, early epiblast marker genes and genes associated with totipotency. The naïve iPSC line was capable of differentiating into derivatives of the three primary germ layers, as well as producing trophoblast derivatives. Culturing naïve iPSCs in low-adhesion conditions resulted in the spontaneous formation of three-dimensional structures (blastoids) resembling early human blastocysts. The X chromosome, which was in an eroded inactive state in the original cell line, was reactivated in the naïve cells, but returned to its normal inactive state when the naïve cells were re-primed. Notably, naïve iPSCs demonstrated limited ability to directly differentiate into endothelial cells. However, their competence to give rise to mature endothelial derivatives was restored upon returning to the primed state, achieving comparable efficiency to the original primed iPSCs. Thus, the resulting naïve iPSC line has significant potential for studying the early stages of embryogenesis and for other biomedical applications, including disease modelling. However, the naïve ICGi022-A line proved to be karyotypically unstable during long-term cultivation using HENSM medium. As there is a risk of karyotypic aberrations during the maintenance of naïve PSCs, further improvement of the culture conditions is necessary to obtain reliable, karyotypically stable lines of naïve pluripotent cells.

Key words: primed and naïve human pluripotent stem cells; directed endothelial differentiation; hereditary disease cellular models

For citation: Arssan M.A., Shevchenko A.I., Zakian S.M., Zakharova I.S. Naïve human iPSCs obtained by culturing the ICGi022-A cell line with primed pluripotency in HENSM medium efficiently differentiate into endothelial derivatives. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):181-193. doi 10.18699/vjgb-26-19

Funding. The work was supported by the Russian Science Foundation grant No. 24-15-00346, https://rscf.ru/project/24-15-00346/

Acknowledgements. The immunefluorescent imaging was performed using resources of the Common Facilities Center of Microscopic Analysis of Biological Objects, ICG SB RAS (https://ckp.icgen.ru/ckpmabo/), supported by the State project of the Institute of Cytology and Genetics FWNR-2026–0024. M.A. Arssan is a member of Molecular Paleogenetics and Paleogenomics laboratory.

Ethical statement. This study did not involve human or animal subjects. The human iPSC line used in the study was obtained earlier in accordance with ethical considerations.

Наивные ИПСК человека, полученные при культивировании линии клеток ICGi022-A с праймированной плюрипотентностью в среде HENSM, способны к эффективной дифференцировке в эндотелиальные производные

М.А. Арссан, А.И. Шевченко, С.М. Закиян, И.С. Захарова

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
arsanmhdamin@yandex.ru

© Arssan M.A., Shevchenko A.I., Zakian S.M., Zakharova I.S., 2026
This work is licensed under a Creative Commons Attribution 4.0 License

Аннотация. Наивные плюрипотентные стволовые клетки (ПСК) человека – новый многообещающий инструмент биомедицинских исследований, открывающий доступ к ранним программам эмбрионального развития и прорывным решениям задач регенеративной медицины. Препятствием на пути к эффективному внедрению наивных ПСК в биомедицину является отсутствие возможности получать длительно культивируемые генетически и эпигенетически стабильные линии наивных ПСК. Ранее было заявлено, что культуральная среда HENSM позволяет индуцировать и длительно поддерживать линии наивных плюрипотентных стволовых клеток человека. В рамках данной работы мы проверили возможность получать стабильные линии наивных ПСК с использованием среды HENSM. С помощью среды HENSM мы успешно перевели линию ICGi022-A (K7-4Lf) праймированных индуцированных плюрипотентных стволовых клеток (ИПСК) здорового донора в наивное состояние. Наивные ИПСК растут в форме сфероподобных колоний как на питающем слое клеток, так и без него. Полученные клетки сохранили экспрессию ключевых факторов плюрипотентности и одновременно активировали транскрипционную программу наивных ПСК, включающую экспрессию эндогенных ретровирусных элементов, генов-маркеров раннего эпибласта и генов, ассоциированных с тотипотентностью. Наивная линия ИПСК была способна дифференцироваться в производные трех первичных зародышевых листков, а также давать производные трофобласта. При культивировании наивных ИПСК в низкоадгезивных условиях наблюдалось спонтанное формирование трехмерных структур – бластоидов, морфологически напоминающих ранние бластоцисты человека. Х-хромосома, имевшая нарушения в неактивном статусе в исходной линии клеток, реактивировалась в наивных клетках и восстанавливала нормальное неактивное состояние при повторном переводе наивных клеток в праймированное состояние. Важно отметить, что наивные ИПСК продемонстрировали низкую способность к прямой направленной дифференцировке в эндотелиоциты, однако их компетентность давать зрелые эндотелиальные производные восстанавливалась после их возврата к праймированному состоянию, достигая эффективности, сопоставимой с таковой у исходных праймированных ИПСК. Таким образом, полученная линия наивных ИПСК обладает значительным потенциалом для исследования ранних стадий эмбриогенеза и других биомедицинских приложений, включая моделирование заболеваний. Тем не менее при длительном культивировании с использованием среды HENSM наивная линия ICGi022-A оказалась кариотипически нестабильной. Поскольку риск кариотипических аберраций при поддержании наивных ПСК сохраняется, в дальнейшем для получения надежных, кариотипически стабильных линий наивных плюрипотентных клеток необходимо совершенствование условий культивирования.
Ключевые слова: праймированные и наивные плюрипотентные стволовые клетки человека; направленная эндотелиальная дифференцировка; клеточные модели наследственных заболеваний

Introduction

Cultures of human pluripotent stem cells (hPSCs) can reflect various stages of embryonic development. The majority of hPSC lines obtained and used to date are in a primed state, resembling the post-implantation epiblast. These cells are stable in long-term cultivation and are widely used in biomedical research and clinical trials. However, they exhibit limited developmental and differentiation pathways, demonstrating the epigenetic memory phenomenon that biases differentiation towards specific lineages (Stadtfeld, Hochedlinger, 2010; Gafni et al., 2013; Valamehr et al., 2014; Lee et al., 2017; Hu et al., 2020; Yu et al., 2021; Dekel et al., 2022). They cannot be used to model the earlier stages of embryonic development, such as X chromosome inactivation, and their chromatin is more closed, making them less accessible to genome editing by programmable nucleases. They also often demonstrate impaired X chromosome inactivation, termed erosion (Mekhoubad et al., 2012). In contrast, naïve hPSCs are in a more primitive developmental state that closely mirrors the pre-implantation epiblast. This gives them broader developmental plasticity than conventional (primed) hPSCs (Theunissen et al., 2016; Collier, Rugg-Gunn, 2018; Rostovskaya, 2022). This unique plasticity includes the ability to contribute to both the embryonic and extraembryonic lineages (Theunissen et al., 2014, 2016; Pham et al., 2022), enhanced proliferation and improved clonal expansion through tolerance of single-cell passaging. Greater amenability to genome editing and erased epigenetic memory through reduction of repressive chromatin marks also render naïve hPSCs a promising platform for advancing regenerative medicine, disease modelling and the study of early human development. Recent advances have demonstrated the feasibility of resetting primed hPSCs to the naïve state in defined media composed small molecules and growth factors (Chan et al., 2013; Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014; Duggal et al., 2015; Zimmerlin et al., 2016; Guo et al., 2017; Lee et al., 2017; Szczerbinska et al., 2019; Bayerl et al., 2021; Khan et al., 2021; Buckberry et al., 2023). Despite these advantages, deriving and maintaining naïve hPSCs in the long term remains challenging. A major concern is the genomic and epigenetic instability reported in naïve cultures, particularly under prolonged or suboptimal conditions. Chromosome aberrations, karyotype abnormalities, and variable X-chromosome reactivation status have all been observed (Theunissen et al., 2014, 2016; Fischer et al., 2025). These vulnerabilities underscore the need for well-defined, optimized culture systems.

The recently developed HENSM medium shows promise in supporting the long-term stable propagation of naïve hPSC lines while preserving the key molecular and functional features of naïve pluripotency (Bayerl et al., 2021). During the medium formulation, it was recognised that WNT signalling activity was the main cause of cell heterogeneity, imprinting loss, and chromosomal rearrangements in naïve hPSC cultures. Thus, the medium lacks WNT activators, such as GSK3 kinase inhibitors, and ensures almost complete downregulation of WNT signalling through tankyrase inhibition. The medium also contains LIF growth factor and MEK kinase inhibitor, two components that are common to all naïve conditions. LIF activates JAK/STAT3 signalling pathway, which is necessary for naïve pluripotency. The MEK inhibitor, together with

the PKC and SRC kinase inhibitors in the HENSM medium, downregulates the MEK/ERK cascade, which is responsible for primed pluripotency. The medium was tested on several ESC and iPSC cell lines. It was claimed that HENSM helps preserve genomic imprints and allows long-term naïve hPSC maintenance, which is crucial for accurate disease modelling and developmental studies. However, there are still no reports of long-term cultured naïve hPSCs obtained in the HENSM medium.

The aim of this study is to obtain stable naïve induced pluripotent stem cells (hiPSCs) in HENSM medium by deriving them from an existing primed hiPSC line, ICGi022-A (K7-4Lf), from a healthy donor (Malakhova et al., 2020). The primed K7-4Lf line is widely used for disease modeling, drug screening, and stem cell research due to its well-characterized genetic background and robust pluripotency (Malakhova et al., 2020; Klepikova et al., 2022; Ustyantseva et al., 2022; Zakharova et al., 2022, 2024; Grigor'eva et al., 2023, 2024; Kopytova et al., 2023; Sheveleva et al., 2023, 2024; Yarkova et al., 2024; Pavlova et al., 2024a, b; Rezvova et al., 2024; Nadtochy et al., 2025). Reprogramming K7-4Lf cells into a naïve state provides an opportunity to evaluate the developmental and differentiation capabilities of naïve pluripotency.

In this study, we reset primed hiPSCs into the naïve state and examined their developmental potential and differentiation abilities. We successfully derived naïve hiPSCs that exhibited key features of early embryonic cells, including spontaneous blastoids formation. Additionally, we investigated the ability of naïve hiPSCs to give rise to differentiated derivatives, with a particular focus on endothelial cells. Furthermore, we analysed the dynamics of X chromosome regulation, focusing on H3K27me3 and *XIST* RNA as indicators of epigenetic fidelity. Our findings emphasise the functional benefits and epigenetic challenges of naïve pluripotency, highlighting the necessity of improved resetting and culture protocols to maintain genomic and epigenetic stability.

Materials and methods

hiPSCs cultivation and differentiation. The human induced pluripotent stem cell line ICGi022-A (hPSCreg identifier: RRID:CVCL_ZE02) (Malakhova et al., 2020) with a karyotype of 46, XX was utilised in this study. The ICGi022-A line was derived and maintained under conditions typical of primed pluripotent cells on a layer of mitotically inactivated mouse embryonic fibroblasts. The medium was composed of DMEM/F12 (1:1), with the addition of 15 % serum replacement KnockOut SR (Thermo Fisher Scientific), 1 mM GlutaMax (Thermo Fisher Scientific), 1 % non-essential amino acid solution (NEAA) (Thermo Fisher Scientific), 50 U/ml penicillin and 50 μg/ml streptomycin (InvivoGen), as well as 0.25 mM 2-mercaptoethanol (Thermo Fisher Scientific) and 10 ng/ml bFGF (Sci-store).

Complete HENSM medium (Bayerl et al., 2021), for induction and maintenance of naïve state, consisted of a 1:1 mixture of Neurobasal and DMEM-F12 media, supplemented with 1 % N2, 1 % B27, 1 mM GlutaMAX, 1 % NEAA (all Thermo Fisher Scientific), 50 U/ml penicillin and 50 μg/ml streptomycin (InvivoGen), 0.2 % Geltrex (Thermo Fisher Scientific), 50 μg/ml L-ascorbic acid 2-phosphate (Sigma-Aldrich), 20 ng/ml recombinant LIF (Sci-store), and kinase inhibitors: 1 μM GSKi PD0325901, 2 μM TNKi XAV939, 2 μM PKCi Gö6983, 1.2 μM SRCi CGP77675 (all R&D) and 5 μM ROCKi thiazovivin (Stemolecule).

In order to expand cell cultures, colonies of both naïve and primed pluripotent cells were dissociated using a TrypLE enzyme (Thermo Fisher Scientific) and subsequently transferred into fresh medium containing 10 μM of the ROCKi thiazovivin (Stemolecule). The seeding ratio was established at 1:10 for primed iPSCs and at 1:3 for naïve iPSCs.

The spontaneous differentiation of naïve hiPSCs into derivatives of three primitive germ layers was performed in a monolayer. To achieve this, the cells were seeded at 30 % confluence on Matrigel (Corning), and the HENSM medium was replaced with a differentiation medium containing a 1:1 DMEM/F12 mixture (Gibco), 10 % fetal bovine serum (FBS) (Cell Technologies), 1 mM GlutaMax (Thermo Fisher Scientific), 1 % NEAA solution (Thermo Fisher Scientific), 50 U/ml penicillin and 50 μg/ml streptomycin (InvivoGen). The medium was changed every other day. The results of the spontaneous differentiation were analysed after 23 days using immunofluorescence.

The directed differentiation protocol for trophoblast derivatives started with the dissociation of confluent naïve IPSC colonies using TrypLE, followed by seeding at a density of 2×10^6 cells on a 10 cm^2 culture surface treated with collagen IV (Sigma-Aldrich) in HENSM medium. After two days, TSC medium was added to the cells as described in (Okae et al., 2018). The TSC medium comprises a 1:1 DMEM/F12 mixture (Gibco), 0.2 % FBS (Cell Technologies), 0.3 % BSA (Sigma-Aldrich), 0.1 mM 2-mercaptoethanol (Sigma-Aldrich), 1 % ITS-X (Gibco), 1.5 μM ascorbic acid (Sigma-Aldrich), 50 ng/ml EGF (Peprotech), 2 μM CHIR99021 and 0.5 μM A83-01, 1 μM SB431542 (all R&D), 0.8 mM valproic acid (Sigma-Aldrich), 5 μM thiazovivin (Stemolecule), 50 U/ml penicillin and 50 μg/ml streptomycin (InvivoGen). The medium was changed daily. On day 5 after the addition of the TSC medium, once the flattened colonies had reached 50–70 % confluence, they were expanded using TrypLE enzymatic disaggregation at a ratio of 1:4. Homogeneous trophoblast cell cultures were formed by the fifth passage. Thereafter, trophoblast cells were passaged with TrypLE every three to four days at a ratio of 1:4 to 1:6. At the tenth passage, the cells were plated on coverslips and stained with antibodies for specific markers.

Endothelial derivatives of hPSCs were obtained through mesodermal progenitor derivation and subsequent endothelial differentiation, according to the Gu protocol with modifications (Gu, 2018). First, the pluripotent cells were seeded onto a matrigel-treated surface (Corning) at a confluence of 60–70 %, in the original iPSC culture medium. The next day, differentiation in the mesodermal direction was started in RPMI 1640 medium containing 1 % B27 supplementation without insulin (all Thermo Fisher Scientific), 50 U/ml penicillin, 50 μg/ml streptomycin (InvivoGen), and the GSK3 kinase inhibitor CHIR99021 (R&D). The concentration of CHIR99021 in the differentiation medium was 6 μM for two days, after which

it was reduced to 3 μM for the following two days. The differentiating cells were then transferred to EGM-2 medium (Lonza), supplemented with 50 ng/ml VEGF, 25 ng/ml bFGF (both Sci-Store) and 10 μM SB431542 (R&D), for endothelial growth. For the next eight days, half of the endothelial medium were changed daily. The proportion of cells containing the CD31 marker of mature endothelial cells was determined using a FACSAria III device (BD, USA) after precipitating the cells with the appropriate antibodies (Supplementary Material 1)[1] and analysing 10^4 events. Unstained cells and cells incubated with fluorescently labelled mouse IgG1 were used as negative controls. The experiment was repeated three times. The statistical significance of the differences was measured using the Wilcoxon test with Bonferroni correction. Mature endothelial derivatives of hiPSCs were selected using magnetic sorting with antibodies to the human CD31 surface antigen (Miltenyi Biotec) and seeded onto a collagen IV-coated culture surface (Sigma-Aldrich) in EGM2 medium (Lonza).

Generation of human blastoids from naïve human iPSCs maintained in HENSM medium. For the generation of blastoids, naïve hiPSCs with 60–70 % confluence were dissociated into single cells using TrypLE Express at 37 °C for 3 minutes. The cells were then collected by centrifugation at 200g for 5 minutes and resuspended in fresh HENSM medium. The resulting cell suspension was incubated in a gelatinised tissue culture plate for 30 minutes at 37 °C in 5 % CO_2 to remove the feeder cells. The medium containing naïve human iPSCs was collected and passed through a 40 μm cell strainer. The AggreWell 400 (STEMCELL Technologies) was prepared by rinsing with an anti-adhesion solution (STEMCELL Technologies), followed by centrifugation at 2,000g for 5 minutes and incubation at room temperature for 10 minutes, according to the manufacturer›s instructions. After incubation, wells were washed once with HENSM medium, after which 0.5 ml of fresh HENSM containing 5 μM hiazovivin (Stemolecule) was added.

Approximately 30,000 cells (around 25 cells per microwell) were resuspended in 1 ml HENSM with 5 μM thiazovivin (Stemolecule) and seeded into one well of a prepared AggreWell 400 24-well plate. The plate was centrifuged at 200g for 1 minute and placed at 37 °C. Aggregates were formed within 12–16 hours. The HENSM medium was then changed to TDM medium and replaced every two days for six days. The TDM previously described for 5iLA-naïve hPSCs was slightly modified to adjust the HENSM protocols using the following: 1:1 (v/v) mixture of DMEM/F12 and Neurobasal medium, 0.5× N2 supplement, 0.5 % 0.5× GlutaMAX, 0.5× non-essential amino acids (all Thermo Fisher Scientific), 0.5× B27 supplement (Thermo Fisher Scientific), 0.1 mM β-mercaptoethanol (Sigma-Aldrich), 1 % ITS-X (Gibco), 0.5 % KnockOut SR (Thermo Fisher Scientific), 0.1 % FBS (Cell Technologies), 50 mg/ml BSA (Sigma-Aldrich), 50 U/ml penicillin and 50 μg/ml streptomycin (InvivoGen), 1 μM Gö6983, 1 μM PD0325901, 0.5 μM CGP77675, 0.5 μM A83-01, 1.25 μM CHIR99021, 0.5 μM SB431542 (all R&D), 25 ng/ml recombinant human LIF (Sci-Store), 10 ng/ml EGF (Peprotech), 0.75 μg/ml L-ascorbic acid, and 0.4 mM valproic acid (both from Sigma-Aldrich). After six days, the blastoids were isolated and treated with a 4 % formaldehyde fixative solution in PBS for immunostaining with OCT4 antibody.

RNA isolation, cDNA synthesis and semi-quantitative real-time PCR. RNA was isolated from primed and naïve iPSCs using a TRIzol reagent (Thermo Fisher Scientific) and a DNA-free kit (Thermo Fisher Scientific). cDNA was synthesised using a reverse transcriptase (M-MuLV) (Biolabmix) and a random hexamer primer (Thermo Fisher Scientific), according to the manufacturers' instructions. The relative gene expression levels of *NANOG*, *TFCP2L1*, *KLF17*, *TFE3*, *LTR7Y*, *DAZL* and *LEUTX* were determined by semi-quantitative real-time PCR using the ΔΔCt method, normalised to the levels of the housekeeping genes *ACTB*, *TFRC* and *B2M*. The experiment was performed in three biological and two technical replicates. The results were analysed using the qBase+ software (CellCarta, https://cellcarta.com/genomic-data-analysis/) and the generalised ΔΔCt method was employed, taking into account the reaction efficiency, which was calculated from the calibration curve results. The sequences of the oligonucleotides synthesised at Biosset (https://www.biosset.com/) are given in the Supplementary Material 2.

Immunofluorescent staining and RNA FISH. We used indirect immunofluorescence staining to detect surface antigens, transcription factors, and the epigenetic state of the X chromosome in hiPSCs, as well as markers of their differentiated derivatives. During cell preparation and antibody staining, we follow the protocol described previously (Vaskova et al., 2015; Zakharova et al., 2017, 2022). The cells were fixed in a 4 % formaldehyde solution for 10 minutes, permeabilised with a 0.5 % Triton X-100 solution (Sigma-Aldrich) for 30 minutes (this step was omitted for surface antigens), and blocked with a 1 % bovine serum albumin solution in 1x PBS. All procedures were performed at room temperature. Primary antibody incubation was conducted overnight at 4 °C. Precipitation between cells and fluorescently labelled secondary antibodies was performed in the dark at room temperature for one hour. Cell nuclei were stained with DAPI. The preparations were visualized and imaged using a Ti-E inverted fluorescence microscope (Nikon) and NIS Advanced Research software. Antibodies, their manufacturers, species of origin, and dilutions are listed in Supplementary Material 2.

To examine the epigenetic state of the X chromosome, immunofluorescent staining with antibodies to chromatin modification of the inactive X chromosome was combined with RNA FISH with probes to the RNA *XIST* (BAC clone RP11-256P2) and X-linked gene *HUWE1* (BAC clone RP11-975N19), according to the protocol given in (Vaskova et al., 2015). Images were acquired and analysed using a Nikon Ti-E inverted fluorescent microscope (Japan) and NIS Elements AR software. At least 300 nuclei were analysed in each experiment.

STR analysis. The short tandem repeat (STR) profile of the primed and naïve PSC line was determined by polymerase chain reaction with the COrDIS EXPERT 26 reagent kit (Genoanalytica, https://www.genoanalytica.ru/) according to the manufacturer's protocol, followed by amplicon separation

[1] Supplementary Materials 1–4 are available at:
https://vavilovj-icg.ru/download/pict-2026-30/appx12.pdf

on a 3130 Genetic Analyzer capillary electrophoresis instrument (Applied Biosystems). Electropherograms of the STR profiles obtained by PCR are available upon request from the authors. Table with STR results is provided in Supplementary Material 3.

Mycoplasma and episome detection. The absence of mycoplasma and episome contamination in the cells was detected by PCR. The primer sequences are provided in the Supplementary Material 2. Parameters for episome detection: 95 °C for 5 minutes, followed by 35 cycles of 95 °C for 15 seconds, 58 °C for 15 seconds and 72 °C for 20 seconds, then 72 °C for 5 minutes. Parameters for mycoplasma detection: 95 °C for 3 minutes, followed by 35 cycles of 95 °C for 15 seconds, 67 °C for 15 seconds and 72 °C for 20 seconds, then 72 °C for 5 minutes.

Karyotyping. Naïve hiPSCs were karyotyped at passages 12, 15 and 18. Metaphase chromosome spreads were obtained by methanol-acetic acid fixation, as previously described (Grigor'eva et al., 2024). The karyotype of naïve iPSCs was determined according to the International System of Human Cytogenetic Nomenclature.

Results

Several media have been proposed to date that allow primed human PSCs to be reset to a naïve state (Chan et al., 2013; Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014; Duggal et al., 2015; Zimmerlin et al., 2016; Guo et al., 2017; Lee et al., 2017; Szczerbinska et al., 2019; Bayerl et al., 2021; Khan et al., 2021; Buckberry et al., 2023). Among these, the HENSM medium has been shown to capture a naïve state that closely resembles the early pre-implantation epiblast (Bayerl et al., 2021). It has been claimed that naïve PSC lines obtained with the HENSM protocol are karyotypically and epigenetically stable during long-term cultivation. In this study, we aimed to return a primed hiPSC line to a naïve state using HENSM conditions. We used K7-4Lf iPSCs that were obtained and described previously (Malakhova et al., 2020). However, direct transferring K7-4Lf cells from a DMEM/F12 medium supplemented with KnockOut SR and FGF2 into a complete HENSM medium resulted in extensive cell death, preventing further progression of the resetting process. To overcome this barrier, we adopted a stepwise approach (Fig. 1*a*), first culturing the K7-4Lf cells in a primed cell medium supplemented with an SRC kinase inhibitor (CGP77675) and LIF. This pre-conditioning phase was lasted for three passages before the cells were transferred into a DMEM/F12-based HENSM naïve-induction medium that still omitted N2B27 supplements but contained KnockOut SR and FGF2, as well as the complete HENSM small molecule set and growth factor LIF. By the fourth passage in the naïve-inducing medium, the K7-4Lf colonies had begun to exhibit a transition towards the compact, dome-shaped morphology characteristic of naïve-like human pluripotent stem cells (Fig. 1*b*). At this stage, the cultures were transferred to complete N2B27-based HENSM medium supplemented with Geltrex. Under these conditions, the cells exhibited stable growth and could be propagated enzymatically using TrypLE in the presence of a ROCK inhibitor.

Following the establishment of the naïve culture, we removed FGF2 from the N2B27-based HENSM medium. While FGF2 supports initial cell survival during resetting, its continued presence can hinder the efficient acquisition of the naïve state. Notably, the resulting K7-4Lf cells maintained hallmark features of naïve pluripotency after FGF2 withdrawal at passage 6 of the resetting process, indicating that they had achieved FGF2-independent pluripotency. Furthermore, these cells were capable of forming compact, spherical colonies under feeder-free conditions when cultured on Matrigel-coated plates (Fig. 1*c*). However, during the early stages of resetting to the naïve state, we observed that K7-4Lf cells failed to maintain under feeder-free conditions, even in complete HENSM medium. This suggests a transient dependency on FGF2 and/or feeder support during the initial phases of resetting to the naïve state.

The resulting naïve K7-4Lf cells demonstrated robust stability under enzymatic passaging and could be cryopreserved and thawed without any apparent loss of viability or naïve morphology. The naïve K7-4Lf line maintained as mycoplasma free culture (Fig. 1*d*). It showed no signs of the episomes that had been used to derive the original K7-4Lf, indicating that resetting to a naïve state was free from the expression of exogenous reprogramming factors. We also demonstrated that the original cell line and its naïve derivative have identical STR profiles (see Supplementary Material 3).

Naïve K7-4Lf hiPSCs characterisation

The resulting naïve K7-4Lf cells demonstrated characteristics commonly associated with pluripotent cells. A histochemical assay revealed alkaline phosphatase activity (Fig. 1*e*), and immunostaining confirmed the presence of the transcription factors OCT4 and NANOG, as well as the conventional surface marker TRA-1-60 (Fig. 1*e*). Additionally, the cells expressed the naïve-specific markers CD75, KLF17, and TFE3 (Fig. 1*f*). Notably, TFE3 nuclear translocation is considered a pivotal event in the induction of naïve pluripotency and a defining feature of the naïve pluripotent state (Gafni et al., 2013; Mathieu et al., 2019). Functionally, these cells were able to differentiate into derivatives of all three germ layers (Fig. 1*g*), as well as into trophoblast cells that expressed surface markers typical of early placentation (Fig. 1*h*). Furthermore, naïve K7-4Lf cultures exhibited spontaneous self-organisation into blastoid-like structures under low-adhesion conditions (Fig. 1*i*), indicating that the cells possess developmental potency attributable to the naïve pluripotent state.

Furthermore, we analysed the expression of pluripotency-associated genes using RT-PCR. We focused specifically on how gene expression differs in naïve K7-4Lf cultures with and without FGF2. To monitor transcriptional dynamics over time, we analysed gene expression profiles at passages 6 and 11. Generally, we found that the naïve K7-4Lf line exhibited robust upregulation of naïve-specific transcription factors compared to the original primed cells (Fig. 2).

Previous studies have demonstrated that *NANOG* and *TFCP2L1* play a significant role in the acquisition and maintenance of the naïve pluripotent state (Theunissen et al., 2016; An et al., 2020). We found that *NANOG* expression increased

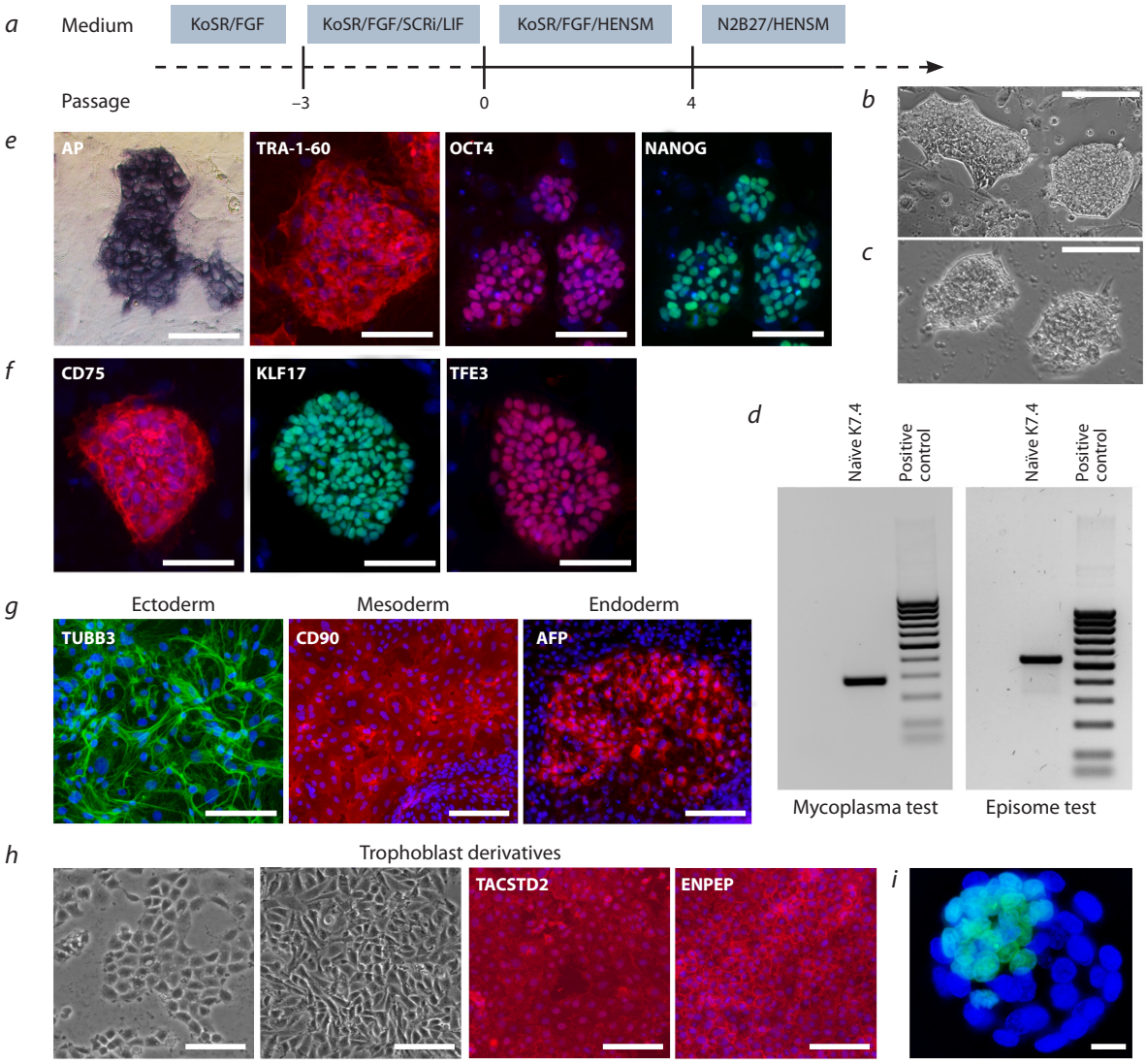


Fig. 1. K7-4Lf hiPSCs after induction of naïve pluripotency in HENSM medium.

a – schematic representation of the conversion stages from primed to naïve state; *b* – colony morphologies of naïve hiPSCs on a layer of mitotically inactivated mouse embryonic fibroblasts, Scale bar is 100 μm; *c* – colony morphologies of naïve hiPSCs on Matrigel-treated surface; Scale bar is 100 μm; *d* – absence of episomal vectors and mycoplasma contamination in naïve K7-4Lf line; *e* – common pluripotency markers: alkaline phosphotase (AP), TRA-1-60 (red), OCT4 (red), NANOG (green) detected in naïve hiPSCs line. Nuclei are stained with DAPI (blue). Scale bar is 100 μm; *f* – specific markers of naïve pluripotency: surface antigen CD75 (red), transcription factors KLF17 (green) and TFE3 (red). Nuclei are stained with DAPI (blue). Scale bar is 100 μm; *g* – ability of naïve iPSC line to differentiate into derivatives of three germ layers; *h* – morphology of early trophectoderm cells derived from naïve hiPSCs and immunostaining with antibodies to early placentation markers TACSTD2 (red) and ENPEP (red). Nuclei are stained with DAPI (blue). Scale bar is 100 μm; *i* – representative image of blastoid obtained by self-aggregation of naïve K7-4Lf iPSCs under low-adhesive conditions. Pluripotent cells are stained green with antibodies to OCT4. Nuclei are stained with DAPI. Scale bar is 10 μm.

substantially in FGF2-independent naïve K7-4Lf cells compared to those maintained with FGF2. Naïve K7-4Lf cultures also demonstrated concomitant *TFCP2L1* upregulation. In line with previous findings that FGF2 impairs the efficient resetting to naïve pluripotency, *TFCP2L1* levels were significantly higher in the FGF2-free condition. *NANOG* expression increased through passage 6, but then declined by passage 11. In contrast, *TFCP2L1* expression remained consistently high across both time points. The expression patterns of *KLF17* and *TFE3* mirrored the dynamic profile of *NANOG*, indicating coordinated regulation during the maturation of naïve identity.

In vivo studies of the early epiblast showed marked activity of transposable elements, particularly *HERVH* and *LTR7Y* (Szczerbinska et al., 2019). Mounting evidence implicates these endogenous retroviral elements as functional regulators of early human development. Consistent with this, we observed robust *LTR7Y* expression in naïve K7-4Lf cultures, with transcription levels progressively increasing over passages. This contrasted with the *NANOG* expression pattern, potentially indicating a transition into a developmental state beyond the *NANOG*-regulated phase of naïve pluripotency.

The expression of a *DAZL* gene, which is also known as a naïve hPSCs marker, was also examined. *DAZL* orchestrates

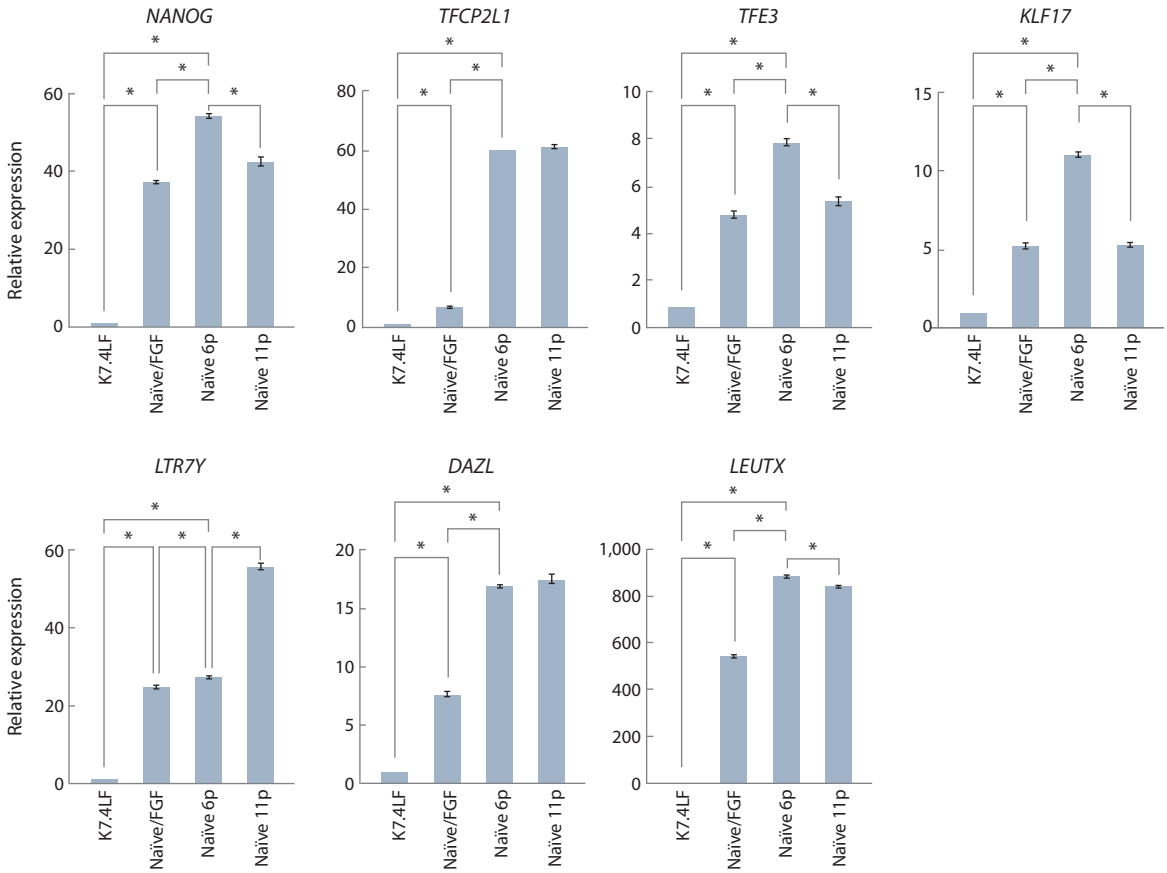


Fig. 2. Semi-quantitative analysis of *NANOG*, *TFCP2L1*, *TFE3*, *KLF17*, *LTR7Y*, *DAZL* and *LEUTX* gene expression in the original primed K7-4Lf line and its naïve derivatives cultured in KoSR/FGF/HENSM (Naïve/FGF) as well as in N2B27/HENSM (Naïve) at 6 and 11 passages.

Statistically significant differences ($p < 0.05$) are marked with an asterisk.

the initial stages of PGC commitment and is involved in the upregulation of TET1, which balances the level of 5-hydroxymethylcytosine (Welling et al., 2015). *DAZL* expression was elevated in naïve K7-4Lf hiPSCs and remained consistently higher in FGF2-independent naïve cultures. This finding suggests that the cells have acquired a molecular profile conducive to both germline differentiation competence and stable naïve pluripotency.

Strikingly, the naïve K7-4Lf cells also activated genes associated with a totipotent-like state. Specifically, we observed increased expression of the *LEUTX* gene, which encodes a chromatin protein involved in zygotic genome activation at the eight-cell stage (Mazid et al., 2022), with higher levels detected in FGF2-independent cultures. This is in line with our previous findings and suggests that the removal of FGF2 enables a more advanced reprogramming trajectory towards earlier embryonic-like states (Shevchenko et al., 2025).

In parallel with transcriptional analyses, we monitored chromosomal stability across passages. Karyotype analysis of naïve K7-4Lf cells at passage 12 revealed a normal chromosomal complement (Fig. 3). However, by passage 15, we detected signs of aneuploidy and polyploidy, indicating chromosomal instability during long-term cultivation. The subsequent karyotype analysis at passage 18 revealed a deletion in the short arm of chromosome 15 (Fig. 3) in more than 70 % of cells. This highlights the importance of ongoing genomic surveillance during long-term culture and suggests that medium composition and culture conditions can affect the karyotype stability in naïve hPSCs obtained and maintained in HENSM.

X chromosome epigenetics in primed, naïve and re-primed K7-4 iPSCs

In primed hPSC lines with two X chromosomes, one is active (Xa) and the other is inactive (Xi) due to X chromosome inactivation (XCI), which is complete by this developmental stage (Vallot et al., 2015; Disteche, 2016; Patel et al., 2017; Sahakyan et al., 2017). However, in most primed hPSC lines, the inactive state on Xi often becomes disrupted and undergoes erosion (Xe). XCI erosion is characterised by the loss of inactive chromatin marks and the partial reactivation of genes, preventing Xe from being re-inactivated during differentiation. This erosion is undesirable for hPSC application in biomedical research. Some naïve media, including HENSM, enable pluripotent cells to restore an inactive state on Xe when the cells are returned to a primed state and differentiated into germinal and somatic cells (Sahakyan et al., 2017; Vallot et al., 2017; An et al., 2020; Raposo et al., 2025). In this study, we evaluated the epigenetic state of the X chromosome in the

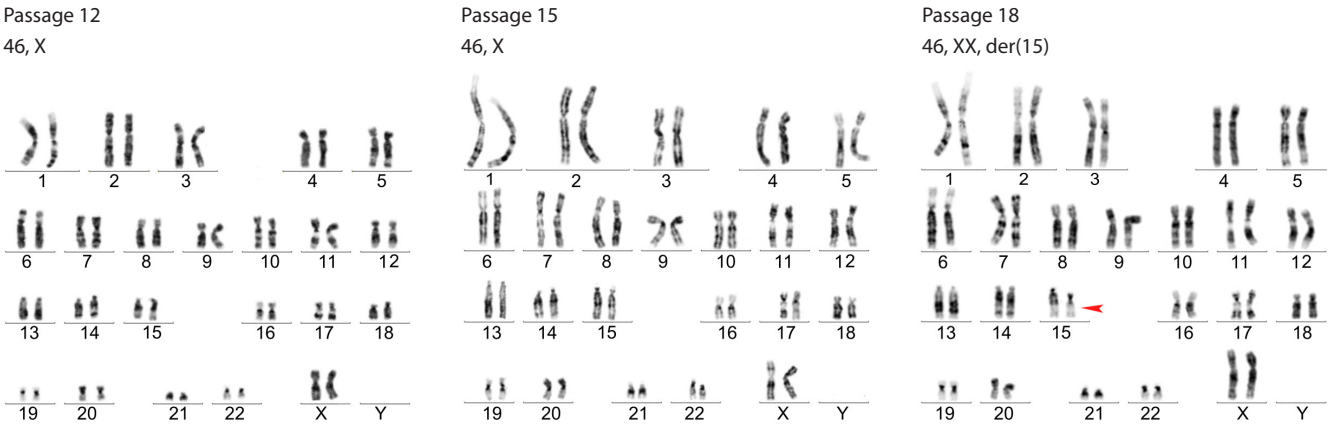


Fig. 3. Naïve hiPSCs K7-4Lf karyotyping at passages 12, 15 and 18.

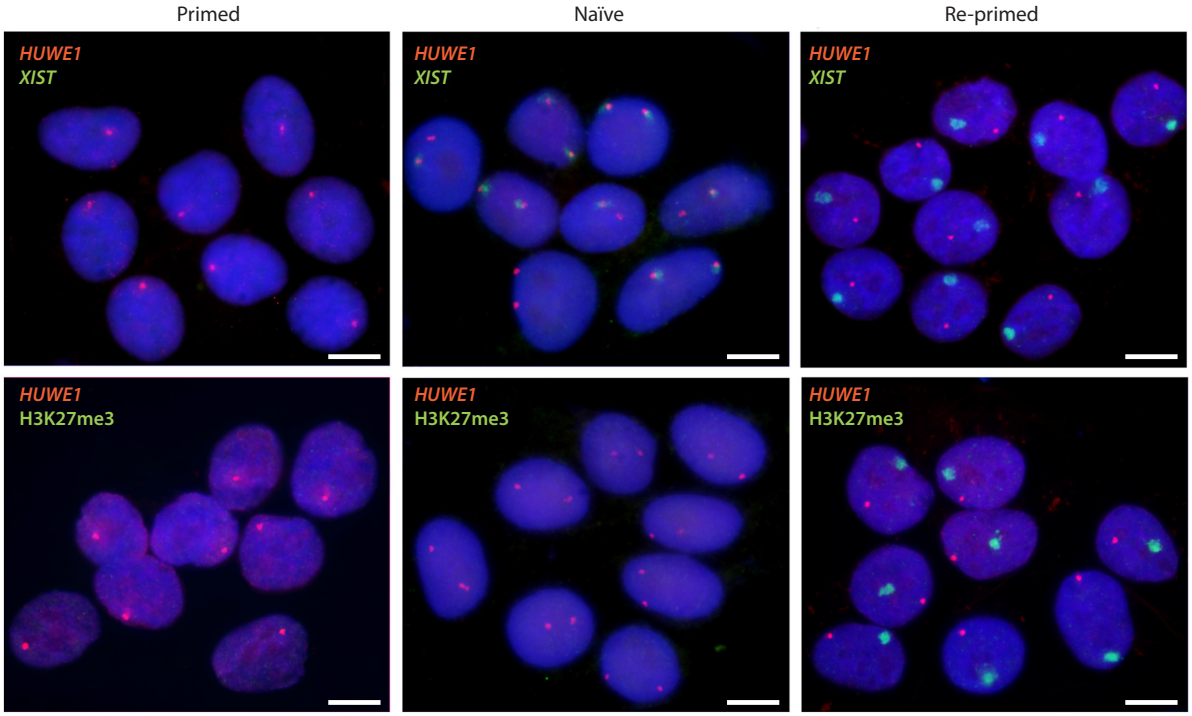


Fig. 4. X chromosome epigenetic state in the original primed, naïve and re-primed K7-4Lf hiPSCs, detected by RNA FISH with probes to RNA *XIST* (green) and *HUWE1* gene (red) as well as immunostaining to H3K27me3 histone mark (green). Nuclei are counterstained with DAPI (blue). Scale bar is 10 μm.

original primed hiPSC line K7-4Lf, focusing on the changes that occur when the cells are reset to the naïve state and when the naïve cells are returned to the primed state (re-primed).

Using RNA FISH and immunofluorescence, we showed that in the original primed K7-4Lf hiPSC line, 96.0 ± 1.2 % of nuclei lacked *XIST* RNA and H3K27me3 histone modification clouds, which are attributable to the Xi territory and are main XCI participants (Fig. 4). Thus, it can be stated that erosion on the inactive X chromosome is observed in the original K7-4Lf line. However, in primed K7-4Lf hiPSCs, the majority of cells (96.5 ± 2.1 %) exhibited just one *HUWE1* signal per nucleus, suggesting that it is transcribed on Xa and remains inactive on Xe. Reactivation of *HUWE1* on Xe is considered a late erosion event, suggesting that XCI erosion in the original K7-4Lf line is at an early stage (Vallot et al., 2015; Patel et al., 2017; Raposo et al., 2025).

Naïve K7-4Lf cells cultured in HENSM medium show biallelic expression of the X-linked gene *HUWE1* in 95.0 ± 1.6 % of nuclei. Biallelic expression of *XIST* RNA is also detected in some cells (27.5 ± 2.1 %). Accumulation of the inactive H3K27me3 modification on X chromosomes was not detected. This epigenetic state is typical of naïve hPSCs and reflects the intermediate stages of the X-chromosome reactivation that is characteristic of the early pluripotent state in humans (Vallot et al., 2015; Theunissen et al., 2016; Sahakyan et al., 2017; Bayerl et al., 2021; Khan et al., 2021).

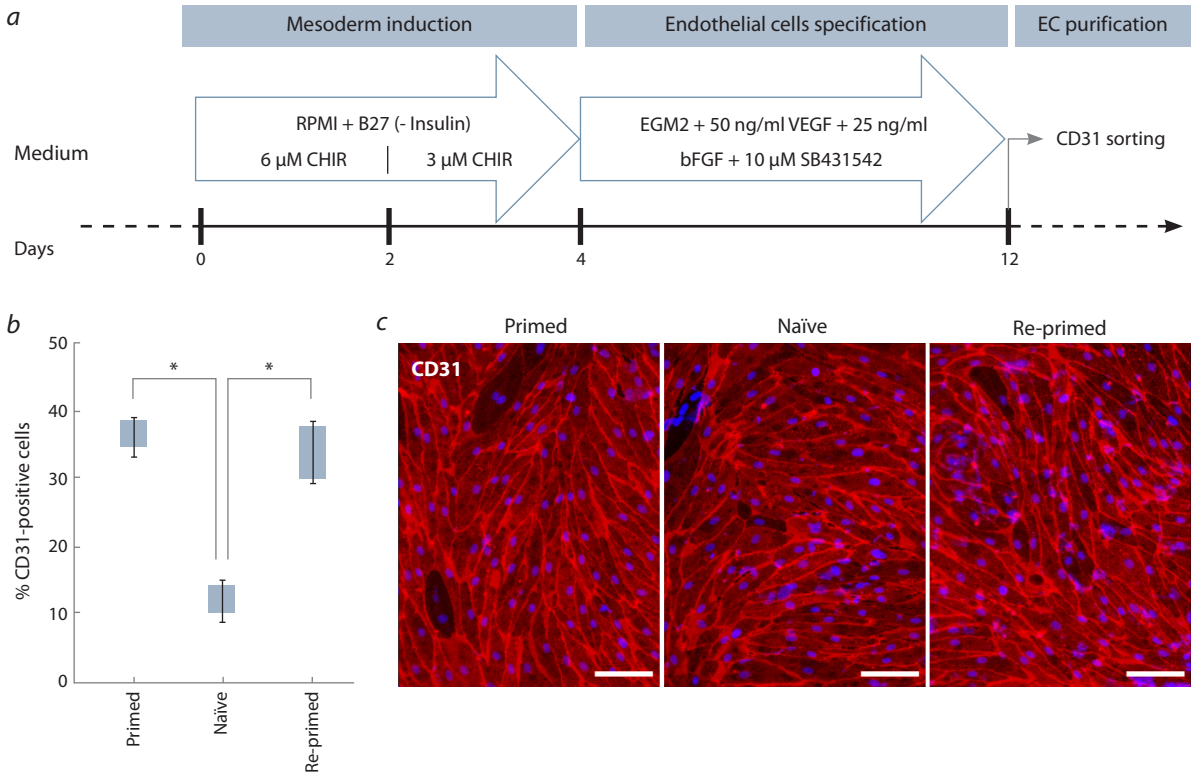


Fig. 5. Directed differentiation of original primed, naïve and re-primed K7-4Lf hiPSCs into endothelial derivatives.

a – schematic representation of the differentiation protocol; *b* – quantification of mature EC using CD31 (PECAM1) surface marker. Statistically significant differences ($p < 0.05$) are marked with an asterisk; *c* – hiPSC derived endothelial cells are stained with antibody to CD31 (red). Nuclei are counterstained with DAPI (blue). Scale bar is 100 μm.

In K7-4Lf cells derived from naïve hiPSCs after re-priming, we observed pattern of completed XCI (95.5 ± 1.3 % of nuclei). It appears as one distinct RNA *XIST* or H3K27me3 cloud corresponding to Xi and as one signal of X-linked gene *HUWE1* located apart from, corresponding to Xa (Fig. 4).

Thus, obtaining cells in naïve state from primed K7-4Lf hiPSC line in HENSM medium, and re-priming or differentiating them, allows us to get rid of erosion and obtain a culture with normal XCI.

Directed differentiation of naïve hiPSCs into endothelial derivatives

Although naïve pluripotent cells have an expanded differentiation potential, they may not differentiate efficiently into somatic lineages directly (Lee et al., 2017; Rostovskaya et al., 2019; Guo et al., 2021; Buckberry et al., 2023). To achieve efficient differentiation into specific somatic cell types, naïve hPSCs first need to be returned back to the primed state, after which their somatic derivatives can successfully be obtained according to protocols previously established for primed hPSCs.

In this study, we evaluated the effect of naïve pluripotency on the ability of hiPSCs to differentiate into endothelial-like cells (ECs), which had not been investigated previously. We compared the ability of the original primed K7-4Lf line, naïve and re-primed K7-4Lf cells to produce mature endothelial derivatives. Re-priming entails placing naïve K7-4Lf cultures in the medium used for maintaining primed K7-4Lf hiPSCs for 48 hours prior to directed EC differentiation.

We differentiated primed, naïve and re-primed K7-4Lf cells into ECs using a two-step directed protocol (James et al., 2010; Gu, 2018; Zakharova et al., 2024) (Fig. 5*a*). It comprises (1) mesoderm induction by WNT pathway activation and subsequent (2) endothelial specification using VEGF and FGF in EGM2 medium along with TGFβ pathway inhibition to suppress smooth muscle cell lineage. On day eight of endothelial specification, we used flow cytometry to quantify mature ECs for the endothelial marker CD31 (PECAM1) (Fig. 5*b*).

We found that attempts to directly differentiate K7-4Lf cells maintained in a naïve state resulted in poor induction of endothelial derivatives. This suggests that naïve hiPSCs have an impaired or incompatible background for endothelial specification. However, re-primed and primed K7-4Lf cells exhibited a significant proportion of mature endothelial cells, which did not differ significantly from each other and were comparable with endothelial specification experiments performed in other studies (Shevchenko et al., 2023; Zakharova et al., 2024). Thus, we found that re-priming renders the naïve K7-4Lf line competent to differentiate into endothelial cells at a level comparable to the original primed K7-4Lf cells (Fig. 5*b*).

In conclusion, endothelial cells obtained from all three K7-4Lf sources through this differentiation protocol by magnetic sorting were positive for CD31 (Fig. 5*c*), which is consistent with their identity as vascular endothelial cells.

Discussion

The rigorous selection of HENSM medium composition, taking into account the influence of its components on chromosomal, epigenetic and genetic stability, allowed optimal conditions for long-term cultivation of naïve hPSCs to be identified. This potentially paves the way for the mass production of stable naïve hPSC lines and their broader use in biomedical research.

In this study, we successfully reset the primed K7-4Lf hiPSC line to the naïve state using the HENSM medium. The characteristics of the resulting naïve iPSCs were summarised in the preliminary cell line passport (Supplementary Material 4). However, transferring primed hiPSCs directly into HENSM medium resulted in extensive cell loss, hindering efficient reprogramming and the establishment of stable naïve cultures. To address this issue, we optimised the induction protocol by introducing a transitional phase, culturing the hPSCs in SRC kinase inhibitor and LIF-supplemented medium prior to naïve induction. This intermediate step was crucial, as it conferred the necessary competence for naïve induction, reducing to minimum cell death and improving the efficiency and stability of the resetting process. The effectiveness of the transitional stage likely stems from its ability to reset the cellular signalling and epigenetic landscape, thereby priming the cells for subsequent naïve resetting. By modulating key pathways and reducing lineage priming, the transitional culture creates a more permissive environment for acquiring naïve pluripotency. Our optimised protocol, which included enzymatic passaging with TrypLE, consistently generated stable naïve hPSCs. These cells exhibited robust expression of pluripotency markers and significant upregulation of naïve-specific genes. Notably, the naïve hPSCs produced using our protocol exhibited the unique ability to form trophoblast cells and blastoid-like structures *in vitro*, highlighting their developmental potential and fidelity to the naïve state.

Our naïve K7-4Lf cells showed upregulation of *LEUTX*, which is a key marker of totipotent-like subpopulations that are active during early embryonic development. The increase in *LEUTX* expression following the withdrawal of FGF2 is consistent with our previous study (Shevchenko et al., 2025), which found that removing growth factors such as FGF2 increases the number of *LEUTX*-positive cells. This indicates a shift towards a totipotent-like transcriptional state within naïve hPSC cultures.

Naïve hPSCs demonstrate superior competence in generating neural stem cells, progenitors and primordial germ cell-like cells, compared to their primed counterparts (Irie, Surani, 2016; Kisa et al., 2017; Lee et al., 2017; Ozaki et al., 2022; Buckberry et al., 2023). This enhanced differentiation potential positions naïve hPSCs as a promising platform for various lineage specifications. However, naïve hPSCs appear to be limited in their ability to generate endothelial cells directly *in vitro*. This phenomenon may be explained by the absence of key transcription factors, such as OTX2 and ZIC2/3, which are essential for initiating gastrulation and subsequent somatic lineage specification. These genes are typically expressed in the post-implantation epiblast, a developmental stage that is mimicked by primed hPSCs (Warr et al., 2008; Di Giovannantonio et al., 2021; Ee et al., 2024; Hossain et al., 2024), but are absent in the naïve state. Consistent with this, we found that re-priming naïve hPSCs restores their ability to differentiate into endothelial cells. Therefore, the ability of naïve hPSCs to generate mesodermal and endodermal derivatives appears restricted unless the cells undergo a transient re-priming process (Guo et al., 2017; Lee et al., 2017; Liu et al., 2017; Buckberry et al., 2023).

The loss of nuclear *XIST* RNA and H3K27me3 foci in the original K7-4Lf line indicates erosion of the inactive X chromosome. Despite the fact that it occurs frequently, the erosion is often disregarded in stem cell research. XCI erosion leads to the aberrant expression of X-linked genes, which impairs the differentiation potential of hPSCs. This results in abnormal descendants that fail to accurately model physiological conditions (Anguera et al., 2012; Mekhoubad et al., 2012; D'Antonio-Chronowska et al., 2019; Motosugi et al., 2022). Importantly, XCI erosion persists after differentiation, compromising the reliability of disease models and drug screening platforms (Mekhoubad et al., 2012; Vallot et al., 2015; Patel et al., 2017). Differentiated derivatives with erosion may be non-functional and at risk of oncogenic transformation. Therefore, monitoring and preventing XCI erosion is essential for ensuring the fidelity of hPSC-derived models and clinical applications. Resetting to the naïve state enables reactivation of the X chromosome, but this is neither complete nor homogeneous, either between cell lines or between cells within a culture. Contrary to previous findings, a recent study has demonstrated that human pre-implantation epiblasts display two active X chromosomes marked by H3K27me3 (Alfeghaly et al., 2024). These observations suggest that generated naïve hPSCs exist in a metastable, transitional state that does not fully recapitulate the pre-implantation epiblast *in vivo* (Alfeghaly et al., 2024). Nevertheless, re-primed hiPSCs regained *XIST* RNA and H3K27me3 foci, indicating re-establishment of the normal inactive X chromosome state.

Chromosomal instability is a well-recognised challenge in naïve hPSC cultures. Previous reports indicate that naïve hPSCs typically retain a normal karyotype for up to nine passages when maintained in 5iLAF medium before abnormalities arise (Theunissen et al., 2014). In our study, naïve hiPSCs generated using our optimised protocol retained normal karyotypes for up to 15 passages, exceeding earlier benchmarks. However, chromosomal abnormalities were detected by passage 18. These results imply that, although our approach increases the period of genomic stability, the risk of karyotypic aberrations continues with prolonged culture. Ongoing monitoring and further refinement of culture conditions are necessary to ensure the genomic integrity of naïve hPSCs for research and therapeutic applications.

Conclusion

In summary, we successfully reset the primed K7-4Lf hiPSC line to a naïve pluripotent state, using these cells to explore the developmental and functional properties of human naïve pluripotency. Our findings demonstrate the potential of naïve reprogramming in accessing early developmental stages and modelling key aspects of embryogenesis by blastoid formation.

We identified the expression of totipotent-like genes within the naïve hPSC cultures, which reinforces the developmental plasticity of this state. In parallel, we demonstrated that re-priming can effectively recapitulate somatic differentiation competence. This enables the efficient generation of endothelial-like cells, thereby underscoring the developmental relevance of transitioning between pluripotent states. Naïve pluripotency opens up new opportunities to control and understand X chromosome reactivation and inactivation epigenetics in humans. However, this topic remains challenging and requires further study. Moving forward, the field must strive to develop more refined, stable and clinically translatable naïve hPSC systems that preserve the benefits of this state while mitigating its disadvantages. As progress continues, naïve hPSCs will remain pivotal in exploring human development and pioneering next-generation regenerative therapies.

References

Alfeghaly C., Castel G., Cazottes E., Moscatelli M., Moinard E., Casanova M., Boni J., … Boers J., Gribnau J., David L., Ouimette J.F., Rougeulle C. XIST dampens X chromosome activity in a SPEN-dependent manner during early human development. *Nat Struct Mol Biol*. 2024;31(10):1589-1600. doi 10.1038/s41594-024-01325-3

An C., Feng G., Zhang J., Cao S., Wang Y., Wang N., Lu F., Zhou Q., Wang H. Overcoming autocrine FGF signaling-induced heterogeneity in naïve human ESCs enables modeling of random X chromosome inactivation. *Cell Stem Cell*. 2020;27(3):482-497. doi 10.1016/j.stem.2020.06.002

Anguera M., Sadreyev R., Zhang Z., Szanto A., Payer B., Sheridan S., Kwok S., … Alvarez J., Gimelbrant A., Mitalipova M., Kirby J.E., Lee J.T. Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. *Cell Stem Cell*. 2012;11(1):75-90. doi 10.1016/j.stem.2012.03.008

Bayerl J., Ayyash M., Shani T., Manor Y., Gafni O., Massarwa R., Kalma Y., … Hanna S., Ben-Yosef D., Novershtern N., Viukov S., Hanna J.H. Principles of signaling pathway modulation for enhancing human naïve pluripotency induction. *Cell Stem Cell*. 2021;28(9):1549-1565.e12. doi 10.1016/j.stem.2021.04.001

Buckberry S., Liu X., Poppe D., Tan J.P., Sun G., Chen J., Nguyen T.V., … Breen J., Faulkner G.J., Nefzger C.M., Polo J.M., Lister R. Transient naïve reprogramming corrects hiPS cells functionally and epigenetically. *Nature*. 2023;620(7975):863-872. doi 10.1038/s41586-023-06424-7

Chan Y.S., Göke J., Ng J.H., Lu X., Gonzales K.A., Tan C.P., Tng W.Q., Hong Z.Z., Lim Y.S., Ng H.H. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell*. 2013;13(6):663-675. doi 10.1016/ j.stem.2013.11.015

Collier A., Rugg-Gunn P. Identifying human naïve pluripotent stem cells – evaluating state-specific reporter lines and cell-surface markers. *BioEssays*. 2018;40(5):1700239. doi 10.1002/bies.201700239

D'Antonio-Chronowska A., Donovan M.K.R., Young Greenwald W.W., Nguyen J.P., Fujita K., Hashem S., Matsui H., … Coulet F., Smith E.N., Adler E., D'Antonio M., Frazer K.A. Association of human iPSC gene signatures and X chromosome dosage with two distinct cardiac differentiation trajectories. *Stem Cell Rep*. 2019;13(5):924-938. doi 10.1016/j.stemcr.2019.09.011

Dekel C., Morey R., Hanna J., Laurent L., Ben-Yosef D., Amir H. Stabilization of hESCs in two distinct substates along the continuum of pluripotency. *iScience*. 2022;25(12):105469. doi 10.1016/j.isci.2022.105469

Di Giovannantonio L., Acampora D., Omodei D., Nigro V., Barba P., Barbieri E., Chambers I., Simeone A. Direct repression of *Nanog* and *Oct4* by OTX2 modulates the contribution of epiblast-derived cells to germline and somatic lineage. *Development*. 2021;148(10):dev199166. doi 10.1242/dev.199166

Disteche C.M. Dosage compensation of the sex chromosomes and autosomes. *Semin Cell Dev Biol*. 2016;56:9-18. doi 10.1016/j.semcdb.2016.04.013

Duggal G., Warrier S., Ghimire S., Broekaert D., Van der Jeught M., Lierman S., Deroo T., … Heijmans B.T., Deforce D., De Sutter P., De Sousa Lopes S.C., Heindryckx B. Alternative routes to induce naïve pluripotency in human embryonic stem cells. *Stem Cells*. 2015;33(9):2686-2698. doi 10.1002/stem.2071

Ee L.S., Medina-Cano D., Uyehara C.M., Schwarz C., Goetzler E., Salataj E., Polyzos A., Madhuranath S., Evans T., Hadjantonakis A.K., Apostolou E., Vierbuchen T., Stadtfeld M. Transcriptional remodeling by OTX2 directs specification and patterning of mammalian definitive endoderm. *bioRxiv*. 2024. doi 10.1101/2024.05.30.596630

Fischer L.A., Meyer B., Reyes M., Zemke J.E., Harrison J.K., Park K.M., Wang T., Jüppner H., Dietmann S., Theunissen T.W. Tracking and mitigating imprint erasure during induction of naïve human pluripotency at single-cell resolution. *Stem Cell Rep*. 2025;20(3):102419. doi 10.1016/j.stemcr.2025.102419

Gafni O., Weinberger L., Mansour A.A., Manor Y.S., Chomsky E., Ben-Yosef D., Kalma Y., … Amann-Zalcenstein D., Benjamin S., Amit I., Tanay A., Massarwa R., Novershtern N., Hanna J.H. Derivation of novel human ground state naïve pluripotent stem cells. *Nature*. 2013;504(7479):282-286. doi 10.1038/nature12745

Grigor'eva E.V., Kopytova A.E., Yarkova E.S., Pavlova S.V., Sorogina D.A., Malakhova A.A., Malankhanova T.B., Baydakova G.V., Zakharova E.Y., Medvedev S.P., Pchelina S.N., Zakian S.M. Biochemical characteristics of iPSC-derived dopaminergic neurons from N370S *GBA* variant carriers with and without Parkinson's disease. *Int J Mol Sci*. 2023;24(5):4437. doi 10.3390/ijms24054437

Grigor'eva E.V., Karapetyan L.V., Malakhova A.A., Medvedev S.P., Minina J.M., Hayrapetyan V.H., Vardanyan V.S., Zakian S.M., Arakelyan A., Zakharyan R. Generation of iPSCs from a patient with the M694V mutation in the *MEFV* gene associated with Familial Mediterranean fever and their differentiation into macrophages. *Int J Mol Sci*. 2024;25(11):6102. doi 10.3390/ijms25116102

Gu M. Efficient differentiation of human pluripotent stem cells to endothelial cells. *Curr Protoc Hum Genet*. 2018;98(1):e64. doi 10.1002/cphg.64

Guo G., von Meyenn F., Rostovskaya M., Clarke J., Dietmann S., Baker D., Sahakyan A., Myers S., Bertone P., Reik W., Plath K., Smith A. Epigenetic resetting of human pluripotency. *Development*. 2017;144(15):2748-2763. doi 10.1242/dev.146811

Guo G., Stirparo G.G., Strawbridge S.E., Spindlow D., Yang J., Clarke J., Dattani A., Yanagida A., Li M.A., Myers S., Özel B.N., Nichols J., Smith A. Human naïve epiblast cells possess unrestricted lineage potential. *Cell Stem Cell*. 2021;28(6):1040-1056. doi 10.1016/j.stem.2021.02.025

Hossain I., Priam P., Reynoso S.C., Sahni S., Zhang X.X., Côté L., Doumat J., Chik C., Fu T., Lessard J.A., Pastor W.A. ZIC2 and ZIC3 promote SWI/SNF recruitment to safeguard progression towards human primed pluripotency. *Nat Commun*. 2024;15(1):8539. doi 10.1038/s41467-024-52431-1

Hu Z., Li H., Jiang H., Ren Y., Yu X., Qiu J., Stablewski A., Zhang B., Buck M., Feng J. Transient inhibition of mTOR in human pluripotent stem cells enables robust formation of mouse-human chimeric embryos. *Sci Adv*. 2020;6(20):eaaz0298. doi 10.1126/sciadv.aaz0298

Irie N., Surani M.A. Efficient induction and isolation of human primordial germ cell-like cells from competent human pluripotent stem cells. In: Buszczak M. (Ed.) Germline Stem Cells. Methods in Molecular Biology. Vol. 1463. Humana Press, New York, 2016;217-226. doi 10.1007/978-1-4939-4017-2_16

James D., Nam H.S., Seandel M., Nolan D., Janovitz T., Tomishima M., Studer L., Lee G., Lyden D., Benezra R., Zaninovic N., Rosenwaks Z., Rabbany S.Y., Rafii S. Expansion and maintenance of human embryonic stem cell-derived endothelial cells by TGFβ inhi-

bition is Id1 dependent. *Nat Biotechnol*. 2010;28(2):161-166. doi 10.1038/nbt.1605

Khan S.A., Park K.M., Fischer L.A., Dong C., Lungjangwa T., Jimenez M., Casalena D., Chew B., Dietmann S., Auld D.S., Jaenisch R., Theunissen T.W. Probing the signaling requirements for naïve human pluripotency by high-throughput chemical screening. *Cell Rep*. 2021;35(11):109233. doi 10.1016/j.celrep.2021.109233

Kisa F., Shiozawa S., Oda K., Yoshimatsu S., Nakamura M., Koya I., Kawai K., Suzuki S., Okano H. Naïve-like ESRRB^{+} iPSCs with the capacity for rapid neural differentiation. *Stem Cell Rep*. 2017;9(6):1825-1838. doi 10.1016/j.stemcr.2017.10.008

Klepikova A., Nenasheva T., Sheveleva O., Protasova E., Antonov D., Gainullina A., Chikina E., Sakovnich O., Gerasimova T., Nikitina I., Shevalie D., Lyadova I. iPSC-derived macrophages: the differentiation protocol affects cell immune characteristics and differentiation trajectories. *Int J Mol Sci*. 2022;23(24):16087. doi 10.3390/ijms232416087

Kopytova A.E., Rychkov G.N., Cheblokov A.A., Grigor'eva E.V., Nikolaev M.A., Yarkova E.S., Sorogina D.A., … Bezrukikh V.A., Salogub G.N., Zakharova E.Y., Pchelina S.N., Emelyanov A.K. Potential binding sites of pharmacological chaperone NCGC00241607 on mutant β-glucocerebrosidase and its efficacy on patient-derived cell cultures in Gaucher and Parkinson's disease. *Int J Mol Sci*. 2023;24(10):9105. doi 10.3390/ijms24109105

Lee J.H., Laronde S., Collins T.J., Shapovalova Z., Tanasijevic B., McNicol J.D., Fiebig-Comyn A., Benoit Y.D., Lee J.B., Mitchell R.R., Bhatia M. Lineage-specific differentiation is influenced by state of human pluripotency. *Cell Rep*. 2017;19(1):20-35. doi 10.1016/j.celrep.2017.03.036

Liu X., Nefzger C.M., Rossello F.J., Chen J., Knaupp A.S., Firas J., Ford E., … Nilsson S.K., Schittenhelm R.B., Laslett A.L., Lister R., Polo J.M. Comprehensive characterization of distinct states of human naïve pluripotency generated by reprogramming. *Nat Methods*. 2017;14(11):1055-1062. doi 10.1038/nmeth.4436

Malakhova A.A., Grigor'eva E.V., Pavlova S.V., Malankhanova T.B., Valetdinova K.R., Vyatkin Y.V., Khabarova E.A., Rzaev J.A., Zakian S.M., Medvedev S.P. Generation of induced pluripotent stem cell lines ICGi021-A and ICGi022-A from peripheral blood mononuclear cells of two healthy individuals from Siberian population. *Stem Cell Res*. 2020;48:101952. doi 10.1016/j.scr.2020.101952

Mathieu J., Detraux D., Kuppers D., Wang Y., Cavanaugh C., Sidhu S., Levy S., … Hawkins R.D., Moon R.T., Ware C.B., Paddison P.J., Ruohola-Baker H. Folliculin regulates mTORC1/2 and WNT pathways in early human pluripotency. *Nat Commun*. 2019;10(1):632. doi 10.1038/s41467-018-08020-0

Mazid M.A., Ward C., Luo Z., Liu C., Li Y., Lai Y., Wu L., … Maxwell P.H., Xu X., Liu L., Li W., Esteban M.A. Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature*. 2022;605(7909):315-324. doi 10.1038/s41586-022-04625-0

Mekhoubad S., Bock C., De Boer A., Kiskinis E., Meissner A., Eggan K. Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell*. 2012;10(5):595-609. doi 10.1016/j.stem.2012.02.014

Motosugi N., Sugiyama A., Okada C., Otomo A., Umezawa A., Akutsu H., Hadano S., Fukuda A. De-erosion of X chromosome dosage compensation by the editing of *XIST* regulatory regions restores the differentiation potential in hPSCs. *Cell Rep Methods*. 2022;2(12):100352. doi 10.1016/j.crmeth.2022.100352

Nadtochy J.A., Medvedev S.P., Grigor'eva E.V., Pavlova S.V., Minina J.M., Chechushkov A.V., Malakhova A.A., Kovalenko L.V., Zakian S.M. Transgenic iPSC lines with genetically encoded MitoTimer to study mitochondrial biogenesis in dopaminergic neurons with tauopathy. *Biomedicines*. 2025;13(3):550. doi 10.3390/biomedicines13030550

Okae H., Toh H., Sato T., Hiura H., Takahashi S., Shirane K., Kabayama Y., Suyama M., Sasaki H., Arima T. Derivation of human trophoblast stem cells. *Cell Stem Cell*. 2018;22(1):50-63. doi 10.1016/j.stem.2017.11.004

Ozaki H., Suga H., Sakakibara M., Soen M., Miyake N., Miwata T., Taga S., … Iguchi G., Takahashi Y., Muguruma K., Inoue H., Arima H. Differentiation of human induced pluripotent stem cells into hypothalamic vasopressin neurons with minimal exogenous signals and partial conversion to the naïve state. *Sci Rep*. 2022;12(1):17381. doi 10.1038/s41598-022-22405-8

Patel S., Bonora G., Sahakyan A., Kim R., Chronis C., Langerman J., Fitz-Gibbon S., … Ardehali R., Pellegrini M., Lowry W.E., Clark A.T., Plath K. Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep*. 2017;18(1):54-67. doi 10.1016/j.celrep.2016.11.054

Pavlova S.V., Shulgina A.E., Minina J.M., Zakian S.M., Dementyeva E.V. Generation of isogenic iPSC lines for studying the effect of the p.N515del (c.1543_1545delAAC) variant on MYBPC3 function and hypertrophic cardiomyopathy pathogenesis. *Int J Mol Sci*. 2024a;25(23):12900. doi 10.3390/ijms252312900

Pavlova S.V., Shulgina A.E., Zakian S.M., Dementyeva E.V. Studying pathogenetic contribution of a variant of unknown significance, p.M659I (c.1977G > A) in MYH7, to the development of hypertrophic cardiomyopathy using CRISPR/Cas9-engineered isogenic induced pluripotent stem cells. *Int J Mol Sci.* 2024b;25(16):8695. doi 10.3390/ijms25168695

Pham T.X.A., Panda A., Kagawa H., To S.K., Ertekin C., Georgolopoulos G., van Knippenberg S.S.F.A., … Lluis F., David L., Rivron N., Balaton B.P., Pasque V. Modeling human extraembryonic mesoderm cells using naïve pluripotent stem cells. *Cell Stem Cell*. 2022;29(9):1346-1365.e10. doi 10.1016/j.stem.2022.08.001

Raposo A.C., Caldas P., Jeremias J., Arez M., Cazaux Mateus F., Barbosa P., Sousa-Luís R., … Mupo A., Eckersley-Maslin M., Casanova M., Grosso A.R., da Rocha S.T. Gene reactivation upon erosion of X chromosome inactivation in female hiPSCs is predictable yet variable and persists through differentiation. *Stem Cell Rep*. 2025;20(5):102472. doi 10.1016/j.stemcr.2025.102472

Rezvova M.A., Ovcharenko E.A., Klyshnikov K.Y., Glushkova T.V., Kostyunin A.E., Shishkova D.K., Matveeva V.G., Velikanova E.A., Shabaev A.R., Kudryavtseva Y.A. Electrospun bioresorbable polymer membranes for coronary artery stents. *Front Bioeng Biotechnol*. 2024;12:1440181. doi 10.3389/fbioe.2024.1440181

Rostovskaya M. Capacitation of human naïve pluripotent stem cells. In: Rugg-Gunn P. (Ed.) Human Naïve Pluripotent Stem Cells. Methods in Molecular Biology. Vol. 2416. Humana, New York, 2022;117-131. doi 10.1007/978-1-0716-1908-7_9

Rostovskaya M., Stirparo G., Smith A. Capacitation of human naïve pluripotent stem cells for multi-lineage differentiation. *Development*. 2019;146(7):dev172916. doi 10.1242/dev.172916

Sahakyan A., Kim R., Chronis C., Sabri S., Bonora G., Theunissen T., Kuoy E., Langerman J., Clark A., Jaenisch R., Plath K. Human naïve pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell*. 2017;20(1):87-101. doi 10.1016/j.stem.2016.10.006

Shevchenko A.I., Arssan A.M., Zakian S.M., Zakharova I.S. Chemokine CCL2 activates hypoxia response factors regulating pluripotency and directed endothelial differentiation of human pluripotent stem cells. *Russ J Dev Biol*. 2023;54(2):134-146. doi 10.1134/s1062360423020054

Shevchenko A.I., Arssan A.M., Zakharova I.S. Towards the generation of safe naïve human pluripotent cell lines. *Tomsk State Univ J Biol*. 2025;69:184-193. doi 10.17223/19988591/69/21 (in Russian)

Sheveleva O., Protasova E., Nenasheva T., Butorina N., Melnikova V., Gerasimova T., Sakovnich O., Kurinov A., Grigor'eva E., Medvedev S., Lyadova I. A model of iPSC-derived macrophages with *TNFAIP3* overexpression reveals the peculiarities of TNFAIP3 protein expression and function in human macrophages. *Int J Mol Sci* 2023;24(16):12868. doi 10.3390/ijms241612868

Sheveleva O., Protasova E., Grigor'eva E., Butorina N., Kuziaeva V., Antonov D., Melnikova V., Medvedev S., Lyadova I. The generation of genetically engineered human induced pluripotent stem cells overexpressing IFN-β for future experimental and clinically

oriented studies. *Int J Mol Sci*. 2024;25(22):12456. doi 10.3390/ijms252212456

Stadtfeld M., Hochedlinger K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev*. 2010;24(20):2239-2263. doi 10.1101/gad.1963910

Szczerbinska I., Gonzales K.A.U., Cukuroglu E., Ramli M.N.B., Lee B.P.G., Tan C.P., Wong C.K., Rancati G.I., Liang H., Göke J., Ng H.H., Chan Y.S. A chemically defined feeder-free system for the establishment and maintenance of the human naïve pluripotent state. *Stem Cell Rep*. 2019;13(4):612-626. doi 10.1016/j.stemcr.2019.08.005

Takashima Y., Guo G., Loos R., Nichols J., Ficz G., Krueger F., Oxley D., Santos F., Clarke J., Mansfield W., Reik W., Bertone P., Smith A. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*. 2014;158(6):1254-1269. doi 10.1016/j.cell.2014.08.029

Theunissen T.W., Powell B.E., Wang H., Mitalipova M., Faddah D.A., Reddy J., Fan Z.P., … Gao Q., Dawlaty M.M., Young R.A., Gray N.S., Jaenisch R. Systematic identification of culture conditions for induction and maintenance of naïve human pluripotency. *Cell Stem Cell*. 2014;15(4):471-487. doi 10.1016/j.stem.2014.07.002

Theunissen T.W., Friedli M., He Y., Planet E., O'Neil R.C., Markoulaki S., Pontis J., … Drotar J., Lungjangwa T., Trono D., Ecker J.R., Jaenisch R. Molecular criteria for defining the naïve human pluripotent state. *Cell Stem Cell*. 2016;19(4):502-515. doi 10.1016/j.stem.2016.06.011

Ustyantseva E., Pavlova S.V., Malakhova A.A., Ustyantsev K., Zakian S.M., Medvedev S.P. Oxidative stress monitoring in iPSC-derived motor neurons using genetically encoded biosensors of H_2O_2. *Sci Rep*. 2022;12(1):8928. doi 10.1038/s41598-022-12807-z

Valamehr B., Robinson M., Abujarour R., Rezner B., Vranceanu F., Le T., Medcalf A., Lee T., Fitch M., Robbins D., Flynn P. Platform for induction and maintenance of transgene-free hiPSCs resembling ground state pluripotent stem cells. *Stem Cell Rep*. 2014;2(3):366-381. doi 10.1016/j.stemcr.2014.01.014

Vallot C., Ouimette J.F., Makhlouf M., Féraud O., Pontis J., Côme J., Martinat C., Bennaceur-Griscelli A., Lalande M., Rougeulle C. Erosion of X chromosome inactivation in human pluripotent cells initiates with *XACT* coating and depends on a specific heterochromatin landscape. *Cell Stem Cell*. 2015;16(5):533-546. doi 10.1016/j.stem.2015.03.016

Vallot C., Patrat C., Collier A.J., Huret C., Casanova M., Liyakat Ali T.M., Tosolini M., Frydman N., Heard E., Rugg-Gunn P.J., Rougeulle C. *XACT* noncoding RNA competes with *XIST* in the control of X chromosome activity during human early development. *Cell Stem Cell*. 2017;20(1):102-111. doi 10.1016/j.stem.2016.10.014

Vaskova E.A., Medvedev S.P., Sorokina A.E., Nemudryy A.A., Elisaphenko E.A., Zakharova I.S., Shevchenko A.I., Kizilova E.A., Zhelezova A.I., Evshin I.S., Sharipov R.N. Transcriptome characteristics and X-chromosome inactivation status in cultured rat pluripotent stem cells. *Stem Cells Dev*. 2015;24(24):2912-2924. doi 10.1089/scd.2015.0204

Ware C.B., Nelson A.M., Mecham B., Hesson J., Zhou W., Jonlin E.C., Jimenez-Caliani A.J., … Blau C.A., Treuting P.M., Hawkins R.D., Cirulli V., Ruohola-Baker H. Derivation of naïve human embryonic stem cells. *Proc Natl Acad Sci USA*. 2014;111(12):4484-4489. doi 10.1073/pnas.1319738111

Warr N., Powles-Glover N., Chappell A., Robson J., Norris D., Arkell R. Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. *Hum Mol Gen*. 2008;17(19):2986-2996. doi 10.1093/hmg/ddn197

Welling M., Chen H.H., Muñoz J., Musheev M.U., Kester L., Junker J.P., Mischerikow N., Arbab M., Kuijk E., Silberstein L., Kharchenko P.V. DAZL regulates Tet1 translation in murine embryonic stem cells. *EMBO Rep*. 2015;16(7):791-802. doi 10.15252/embr.201540538

Yarkova E.S., Grigor'eva E.V., Medvedev S.P., Tarasevich D.A., Pavlova S.V., Valetdinova K.R., Minina J.M., Zakian S.M., Malakhova A.A. Detection of ER stress in iPSC-derived neurons carrying the p. N370S mutation in the *GBA1* gene. *Biomedicines*. 2024;12(4):744. doi 10.3390/biomedicines12040744

Yu L., Wei Y., Sun H.X., Mahdi A.K., Pinzon Arteaga C.A., Sakurai M., Schmitz D.A., … Okamura D., Mutto A.A., Gu Y., Ross P.J., Wu J. Derivation of intermediate pluripotent stem cells amenable to primordial germ cell specification. *Cell Stem Cell*. 2021;28(3):550-567. doi 10.1016/j.stem.2020.11.003

Zakharova I.S., Zhiven' M.K., Saaya S.B., Shevchenko A.I., Smirnova A.M., Strunov A., Karpenko A.A., Pokushalov E.A., Ivanova L.N., Makarevich P.I., Parfyonova Y.V. Endothelial and smooth muscle cells derived from human cardiac explants demonstrate angiogenic potential and suitable for design of cell-containing vascular grafts. *J Transl Med*. 2017;15(1):54. doi 10.1186/s12967-017-1156-1

Zakharova I., Saaya S., Shevchenko A., Stupnikova A., Zhiven' M., Laktionov P., Stepanova A., … Zavjalov E., Chernyavsky A., Romanov A., Karpenko A., Zakian S. Mitomycin-treated endothelial and smooth muscle cells suitable for safe tissue engineering approaches. *Front Bioengin Biotechnol*. 2022;10:772981. doi 10.3389/fbioe.2022.772981

Zakharova I.S., Shevchenko A.I., Arssan M.A., Sleptcov A.A., Nazarenko M.S., Zarubin A.A., Zheltysheva N.V., … Saaya S.B., Ezhov M.V., Kukharchuk V.V., Parfyonova Y.V., Zakian S.M. iPSC-derived endothelial cells reveal LDLR dysfunction and dysregulated gene expression profiles in familial hypercholesterolemia. *Int J Mol Sci*. 2024;25(2):689. doi 10.3390/ijms25020689

Zimmerlin L., Park T.S., Huo J.S., Verma K., Pather S.R., Talbot C.C., Jr., Agarwal J., … Cope L., Canto-Soler M.V., Friedman A.D., Baylin S.B., Zambidis E.T. Tankyrase inhibition promotes a stable human naïve pluripotent state with improved functionality. *Development*. 2016;143(23):4368-4380. doi 10.1242/dev.138982

Conflict of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received August 12, 2025. Revised October 20, 2025. Accepted November 6, 2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):194-204

doi 10.18699/vjgb-26-22

Молекулярно-генетическая изменчивость по ISSR-маркерам у *Caragana jubata* (Pall.) Poir. (Fabaceae) в горах Средней Азии и Южной Сибири

Д.А. Кривенко [1] ✉, О.А. Чернышева [1], И.Н. Кубан [2], Е.В. Жмудь [2], И.В. Горбенко [1], А.А. Ачимова [3], О.Ю. Васильева [2], О.В. Дорогина [2, 4]

[1] Сибирский институт физиологии и биохимии растений Сибирского отделения Российской академии наук, Иркутск, Россия
[2] Центральный сибирский ботанический сад Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Алтайский филиал Центрального сибирского ботанического сада Сибирского отделения Российской академии наук «Горно-Алтайский ботанический сад», Камлак, Республика Алтай, Россия
[4] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ krivenko.irk@gmail.com

Аннотация. Исследована генетическая структура по пяти ISSR-маркерам у 44 особей в выборках из восьми ценопопуляций (ЦП) *Caragana jubata* (Pall.) Poir. в горных условиях Центральной Азии – на Западном Памире и Внутреннем Тянь-Шане (Кыргызстан) и в Южной Сибири – на Алтае (Республика Алтай), Западном (Республика Тыва) и Восточном (Республика Бурятия) Саяне. Вид изучен в диапазоне географических расстояний, составляющих более 2.5 тыс. км и в диапазоне абсолютных высот 1570–3680 м. *Caragana jubata* занесена в Красные книги восьми субъектов Российской Федерации. Цель настоящей работы – выявление генетического разнообразия и гетерогенности в ЦП *C. jubata* в зависимости от их географической и высотной приуроченности в горах Центральной Азии и Южной Сибири. Исследования показали, что в ненарушенных местонахождениях изученные ЦП этого вида характеризовались высокими параметрами численности особей и занимали площадь более 100 м^2. Практически в каждой изученной нами выборке из ЦП *C. jubata* (кроме представителей из ЦП7) обнаружено наличие генотипов, содержащих уникальные фрагменты ДНК. Наибольшая доля (75 %) таких генотипов установлена в выборке из ЦП3 (Внутренний Тянь-Шань, хр. Тескей-Ала-Тоо, абсолютная высота 2550 м). Нами не найдено уникальных фрагментов в генотипах у особей из исследованной выборки в ЦП7 (Западный Саян, Республика Тыва). Возможная причина – антропогенное воздействие на растения в данном местонахождении. Выявленное преобладание внутрипопуляционной генетической изменчивости над межпопуляционной в выборках из восьми изученных ЦП *C. jubata* может свидетельствовать об устойчивости представителей этого вида в пределах исследованных нами частей ареала.

Ключевые слова: карагана гривастая; уязвимый вид; ценопопуляции; генетический полиморфизм; параметры генетического разнообразия

Для цитирования: Кривенко Д.А., Чернышева О.А., Кубан И.Н., Жмудь Е.В., Горбенко И.В., Ачимова А.А., Васильева О.Ю., Дорогина О.В. Молекулярно-генетическая изменчивость по ISSR-маркерам у *Caragana jubata* (Pall.) Poir. (Fabaceae) в горах Средней Азии и Южной Сибири. *Вавиловский журнал генетики и селекции*. 2026;30(2):194-204. doi 10.18699/vjgb-26-22.

Финансирование. Работа выполнена на средства, полученные при финансировании государственных заданий Центрального сибирского ботанического сада СО РАН «Анализ биоразнообразия, сохранение и восстановление редких и ресурсных видов растений с использованием экспериментальных методов», номер государственной регистрации: АААА-А21-121011290025-2 и Сибирского института физиологии и биохимии растений СО РАН «Изучение состава и динамики фиторазнообразия Байкальской Сибири в оригинальной информационно-аналитической среде», номер государственной регистрации: 125020905982-9, а также при финансовой поддержке проекта Министерства науки и высшего образования Российской Федерации № FSUS-2024-002.

Molecular genetic variability in coenopopulations of *Caragana jubata* (Pall.) Poir. (Fabaceae) in the mountains of Central Asia and Southern Siberia revealed using ISSR-markers

D.A. Krivenko [1] ✉, O.A. Chernysheva [1], I.N. Kuban [2], E.V. Zhmud [2], I.V. Gorbenko [1], A.A. Achimova [3], O.Yu. Vasilyeva [2], O.V. Dorogina [2, 4]

[1] Siberian Institute of Plant Physiology and Biochemistry of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
[2] Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Altai Branch of the Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences "Gorno-Altaian Botanical Garden", Kamlak, Republic of Altai, Russia
[4] Novosibirsk State University, Novosibirsk, Russia
✉ krivenko.irk@gmail.com

© Кривенко Д.А., Чернышева О.А., Кубан И.Н., Жмудь Е.В., Горбенко И.В., Ачимова А.А., Васильева О.Ю., Дорогина О.В., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

Abstract. The genetic structure was studied using five ISSR markers in 44 individuals of samples obtained from eight coenopopulations (CPs) of *Caragana jubata* (Pall.) Poir. in the mountainous conditions of Central Asia – in the Western Pamirs and Inner Tien Shan (Kyrgyzstan) and in Southern Siberia – in Altai (Republic of Altai), Western (Republic of Tyva) and Eastern (Republic of Buryatia) Sayan. We studied the species in a range of geographical distances of more than 2,500 km and in the range of absolute altitudes of 1,570–3,680 m. *Caragana jubata* is listed in the Red Data Books of eight subjects of the Russian Federation. The species population is declining, including due to anthropogenic impact. The aim of the current work is to identify genetic diversity and heterogeneity in *C. jubata* coenopopulations depending on their geographic and altitudinal confinement in the mountains of Central Asia and Southern Siberia. It was shown that in undisturbed locations, the studied CPs of this species were characterized by a high number of individuals and by the occupied area of more than 100 m^2. Almost every sample from the *C. jubata* CP studied by us (except for representatives from CP7) contained genotypes possessing unique DNA fragments. The highest proportion of such genotypes (75 %) was found in the sample from CP3 (Inner Tien Shan, Teskey-Ala-Too Ridge, absolute altitude 2,550 m). We did not find unique fragments in the genotypes of individuals from the studied sample of CP7 (Western Sayan, Republic of Tuva). Anthropogenic impact on plants at this location is a possible reason for that. The revealed predominance of intrapopulation genetic variability over interpopulation genetic variability in samples from eight *C. jubata* CPs studied by us may indicate the stability of representatives of this species within the parts of the range studied by us.

Key words: camel's tail caragana; vulnerable species; coenopopulations; genetic polymorphism; genetic diversity parameters

For citation: Krivenko D.A., Chernysheva O.A., Kuban I.N., Zhmud E.V., Gorbenko I.V., Achimova A.A., Vasilyeva O.Yu., Dorogina O.V. Molecular genetic variability in coenopopulations of *Caragana jubata* (Pall.) Poir. (Fabaceae) in the mountains of Central Asia and Southern Siberia revealed using ISSR-markers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):194-204. doi 10.18699/vjgb-26-22

Введение

Caragana jubata (Pall.) Poir. (карагана гривастая, или верблюжий хвост) – неогеновый (миоцен-плиоценовый) реликт, обособившийся от ближайших родственных видов более 4.8 млн лет назад (Zhang M.-L. et al., 2016). Это колючий однодомный, с обоеполыми цветками кустарник 0.3–2.5 м высотой. Светолюбивый, преимущественно высокогорный вид, мезопсихрофит, произрастает в лесном (от 1100 до 3000 м над уровнем моря – на Становом Нагорье и Тянь-Шане соответственно) и высокогорном поясах (до 4,700 м над уровнем моря – на Тибетском нагорье), в арктической тундре (до 500 м над уровнем моря – в Хараулахских горах), на каменистых россыпях и скалах, по берегам рек, под пологом разреженных лесов (Коропачинский, Встовская, 2012; Чурюлина, 2021; Ооржак и др., 2024). Современный ареал этого реликтового вида – паназиатский, дизъюнктивный: от Тибетского нагорья, Гималаев, Тянь-Шаня, Алтая, Саян и Станового нагорья до Хараулахских гор и Колымского нагорья (Churiulina, Bocharnikov, 2019). Оптимальными являются условия в верхней части пояса еловых лесов в Центральном Тянь-Шане, где растения хорошо облиственны, обильно плодоносят и достигают в высоту 2.0–2.5 м (Лысова, 1967).

Вид *C. jubata* занесен в Красные книги восьми субъектов Российской Федерации. В Красных книгах: Республики Алтай (2017), Магаданской области (2019) и Красноярского края (2022) ему присвоена 3 (R) категория статуса редкости – редкий вид с реликтовым дизъюнктивным ареалом; в Республиках Бурятия (2023) и Саха (Якутия) (2017), Иркутской области (2020), Забайкальском (2017) и Хабаровском (2019) краях – 2 (VU) категория статуса редкости – вид, сокращающийся в численности, в том числе в результате чрезмерного использования человеком.

Лимитирующими факторами распространения *C. jubata* могут служить изолированность популяций друг от друга и их низкая численность, а также требовательность к основному составу пород из-за произрастания преимущественно на известняках и карбонатных песчаниках. Кроме этого, угрозу популяциям этого вида создают экзогенные геоморфологические процессы – оползни и наводнения (Семенова, 2007; Коропачинский, Встовская, 2012). *C. jubata* широко используется в народной медицине Восточной Сибири, что приводит к бессистемным заготовкам сырья и вырубке кустарника для лекарственных целей, так как заготовки проводятся круглый год и растения заготавливаются целиком, включая корневую систему. Выпас скота также ведет к истощению природных популяций (Семенова, 2007; Парыгин, Худоногова, 2017; Чурюлина и др., 2020).

По данным исследования состава биологически активных веществ и фармакологической активности водного извлечения надземной части, доминирующая фракция биологически активных веществ *C. jubata* – полифенольные соединения (флавоноиды), преимущественно моно- и дигликозиды, являющиеся производными О-гидроксилированных флавонолов. Обнаружены антиоксидантная активность водных извлечений в условиях *in vitro*, их гепатопротекторная, противовирусная, противовоспалительная и ранозаживляющая активность, и при этом отмечено отсутствие токсичности. Поэтому водные извлечения рекомендованы для лечения заболеваний кожи и слизистых оболочек, связанных с бактериальными воспалительными процессами (Kakorin et al., 2020).

Отвары и спиртовые извлечения надземной части *C. jubata* также обладают гепатопротекторным, противоопухолевым, противовоспалительным и противовирусным свойствами (Какорин и др., 2018; Kakorin et al., 2018). Этанольный экстракт *C. jubata* применяется в традиционной китайской медицине для лечения гипертонии. В эксперименте у крыс он уменьшал размер инфаркта и сокращал отек мозга и улучшал неврологические показатели при индуцированном ишемическом инсульте (Wang et al., 2020; Zhao et al., 2023). Метанольные экстракты корней

проявили цитотоксическую активность против клеточных линий гепатоцеллюлярного рака (Yang et al., 2015).

Исследования генетической структуры популяций *C. jubata* крайне немногочисленны. Так, изучена внутривидовая дифференциация и филогеография этого вида на основе исследований секвенирования ядерных и хлоропластных маркеров у образцов с Тибета, Тянь-Шаня, Восточного Саяна и северо-восточного побережья Охотского моря (Zhang et al., 2016; Hantemirova et al., 2024). С помощью SSR-маркеров изучено влияние высотных градиентов на генетическую структуру альпийских растений *C. jubata* (Yao et al., 2025). Литературные источники по вопросам исследования закономерностей изменчивости по ISSR-маркерам в ценопопуляциях (ЦП) из географически удаленных местонахождений на разных абсолютных высотах, влияющих на генетическую дифференциацию ЦП *C. jubata*, нами не найдены.

В обзоре T. Ohsawa и I. Yuji (2008) обобщены данные, посвященные влиянию абсолютных высот произрастания травянистых многолетников и представителей дендрофлоры на внутрипопуляционный полиморфизм с использованием нейтральных молекулярных маркеров в региональных масштабах (порядка 1–10 км). Более крупные масштабы (> 102 км) они не рассматривали, поскольку большие горизонтальные расстояния, по мнению авторов, влияют на генетическую изменчивость не меньше, чем ограниченные вертикальные расстояния. Для оценки генетического полиморфизма были использованы изоферменты RAPD или SSR. Согласно выводам T. Ohsawa и I. Yuji, конкретный локус или аллель в определенном локусе может показывать четкую дифференциацию между популяциями на разных высотах в результате естественного отбора.

Показано, что в исследованиях генетической гетерогенности ЦП разных видов растений вдоль высотных градиентов генетическое разнообразие внутри ЦП варьируется по одному из четырех паттернов внутрипопуляционной генетической изменчивости, обычно представленных ожидаемой гетерозиготностью: L < M > H, L > M > H, L < M < H или L = M = H, где L, M и H представляют собой низкие, средние и большие высоты соответственно. В частности, в ряде работ продемонстрировано (Oyama et al., 1993; Yifru et al., 2006), что ЦП на промежуточных высотах имели более высокий уровень генетического разнообразия, чем расположенные в местообитаниях выше или ниже, так как географически центральные популяции, по мнению авторов, находятся в оптимальных условиях среды, тогда как периферийные популяции – в субоптимальных ситуациях. Например, у *Arabis serrata* Franch. & Sav. (резуха пильчатая) на градиенте абсолютных высот 1440–2400 м и *Triticum turgidum* subsp. *durum* (Desf.) Husn. (пшеница твердая) – 1500–3300 м, наиболее высокую внутрипопуляционную изменчивость наблюдали в средней части высотного диапазона по локусам ферментов и 29 nSSR-локусам соответственно.

X. Zhang с коллегами (2006) обнаружили, что высокогетерогенные условия окружающей среды, налагаемые высотными градиентами, могут влиять на нейтральные участки, тесно связанные с участком генома, находящимся под отбором. Даже в случаях, когда дифференциации не обнаружено, внутрипопуляционная генетическая изменчивость часто изменяется вдоль высотных градиентов.

Выявление статуса генетической изменчивости в зависимости от абсолютной высоты позволит определить дальнейшее направление исследований. В частности, эффективность отбора наиболее устойчивых форм при введении *C. jubata* в культуру существенно повысится при использовании материала для размножения у особей из наиболее генетически гетерогенных ЦП. Поэтому крайне важно исследовать генетическое разнообразие и структуру популяций дикорастущих уязвимых видов растений для их наиболее рационального применения (Liu et al., 2024; Yao et al., 2025).

Цель настоящей работы – выявить генетическое разнообразие и гетерогенность в ценопопуляциях *C. jubata* в зависимости от их географической и высотной приуроченности в горах Центральной Азии и Южной Сибири.

Материалы и методы

Материал для исследований собран нами в 2015, 2023–2024 гг. в ЦП из пяти регионов, расположенных в горных условиях: в Кыргызстане (Западный Памир – ЦП1, ЦП2 и Внутренний Тянь-Шань – ЦП3) и в трех субъектах Российской Федерации (Южная Сибирь): Республиках Алтай (ЦП4), Тыва (ЦП5–ЦП7) и Бурятия (ЦП8) на высотах 1575–3680 м над уровнем моря (рис. 1, Приложение 1)[1].

Для проведения генетического анализа собраны листья с 5–6 особей в каждой ЦП. ДНК выделена из листьев стандартными методами на основе цетилтриметиламмония бромида (ЦТАБ) (Набиева и др., 2020). Полимеразная цепная реакция (ПЦР) проведена на амплификаторе C1000 Touch (Bio-Rad, США) в объеме 25 мкл. Реакционная смесь содержала: 1.5 ед. Taq ДНК-полимеразы («Медиген», Россия); 2.7 мМ $MgCI_2$; 0.8 мМ ISSR-праймер (Medigen); раствор ДНК – 2 мкл; вода mQ H_2O – 2 мкл. Программа амплификации состояла из этапа денатурации ДНК в течение 90 с при 94 °C и 35 циклов, каждый из которых включал 40 с при 94 °C, 45 с – отжиг праймера и 90 с при 72 °C. Заключительный этап элонгации нуклеотидной цепи длился 5 мин при 72 °C.

Электрофоретическое разделение продуктов амплификации проводили в 1.5 % агарозном геле в 1x TAE буфере при напряжении 4 В/см. Для оценки длины ПЦР-продуктов использован ДНК-маркер 100 bp DNA Ladder (no stain) (BIORON GmbH, Германия).

Всего нами протестировано пять ISSR-праймеров, которые показали наибольшую эффективность в работе с редкими видами растений (Жмудь и др., 2023; Dorogina et al., 2023; Zhmud et al., 2024): HB12 ($(CAC)_3GC$), UBC834 ($(AG)_8YT$), UBC856 ($(AC)_8YA$), UBC857 ($(AC)_8YG$) и UBC881 ($(GGTG)_3$). Из них выбран UBC857, оказавшийся наиболее информативным (рис. 2, Приложение 2).

[1] Приложения 1 и 2 см. по адресу:
https://vavilovj-icg.ru/download/pict-2026-30/appx13.pdf

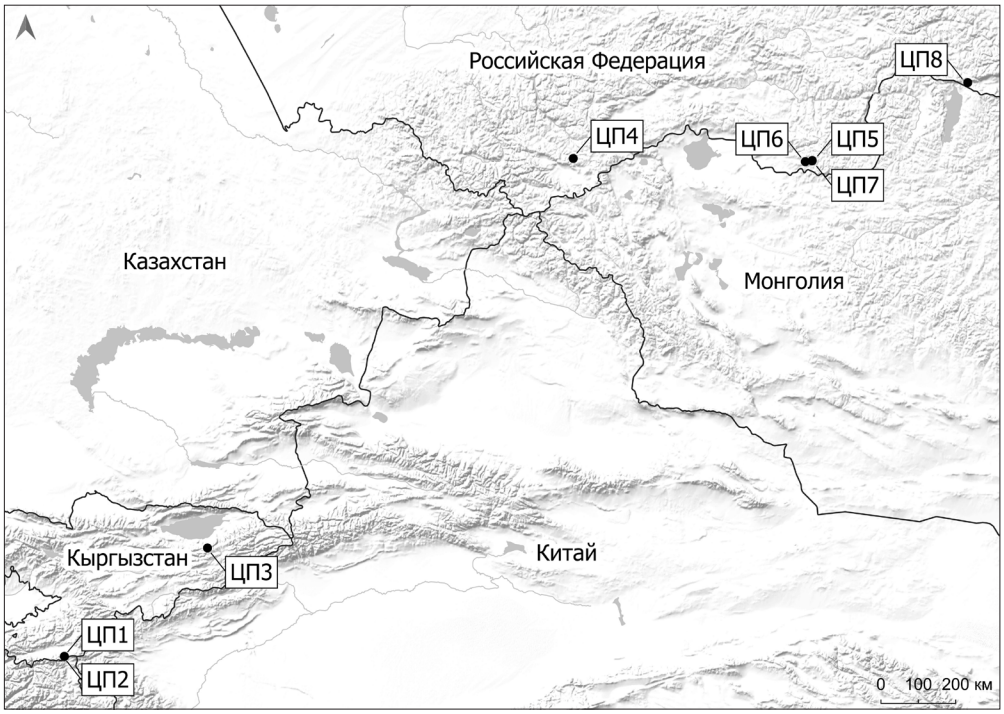


Рис. 1. Местонахождения исследованных особей *C. jubata* в Кыргызстане (ЦП1–ЦП3) и Российской Федерации: в Республиках Алтай (ЦП4), Тыва (ЦП5–ЦП7) и Бурятия (ЦП8).

Первичный анализ результатов исследования выполнен с использованием программного пакета и специализированного макроса GenAlEx6.1 для MS-Excel (Smouse et al., 2017). Дальнейшие анализ и визуализация результатов проведены с помощью среды для программирования R (R Core Team, 2024). Подготовку данных осуществляли с применением пакета ‘tidyverse’ (Wickham et al., 2019). Для построения дендрограммы задействовали дистанции Хэмминга и метод UPGMA, визуализация дерева проведена с помощью пакета ‘ggtree’ (Xu et al., 2022). Анализ соответствия выполнен с применением пакета ‘adegenet’ (Jombart, 2008) и использован для дополнительной визуализации генетической структуры исследованных ЦП. Этот анализ, подобно анализу главных компонент, позволяет снизить размерность данных и визуализировать их, однако предназначен для категориальных данных, использует хи-квадрат расстояния и больше подходит для генетических данных, где важны относительные частоты аллелей.

Анализ изоляции расстоянием проводили согласно методу, описанному L.S. Nørgaard с коллегами (2017), анализ изоляции высотой – по методу M.-M. Shi с коллегами (2011). Оба анализа выполнены с применением пакета ‘adegenet’ (Jombart, 2008), в качестве генетических дистанций использовали дистанции Жаккара, представление графиков выполняли с помощью пакета ‘ggpubr’ (Kassambara, 2023). Изоляция расстоянием – это феномен, описывающий паттерны генетических вариаций популяций, возникающие по причине пространственной ограниченности потока генов, и который определяется как снижение генетической схожести популяций с увеличением географических дистанций между ними (Wright, 1940). Изоляция высотой – подобный феномен, в котором в качестве пространственной компоненты используются разности высот популяций.

Результаты

Исследования показали, что изученные нами ЦП *C. jubata* характеризуются высокими параметрами численности особей и занимают площадь более 100 м^2 (см. Приложение 1). Так, в ЦП1 и ЦП2 выявлено более 200 особей, в ЦП3–ЦП6 – от 20 до 50. В основном они обитали в ненарушенных местонахождениях с высоким проективным покрытием травостоя, более 50 %. Исключение составила ЦП7, немногочисленные особи из которой (около 10 экземпляров) были подвержены антропогенному воздействию из-за прокладки дороги в ее местонахождении. В данной ЦП надземная часть особей выглядела поврежденной, и у них отсутствовали сформированные генеративные органы; также в этом местонахождении было невысокое покрытие травостоя (см. Приложение 1).

Молекулярно-генетические исследования в выборках из ЦП *C. jubata* показали, что число генотипов, содержащих уникальные фрагменты в исследованной выборке из всех ЦП, составило 18, а повторяющихся вариантов – 9 (табл. 1). В каждой из семи исследованных ЦП (ЦП1–ЦП6, ЦП8), кроме ЦП7, обнаружены генотипы с уникальными фрагментами, которые не отмечены у особей в других исследованных ЦП (см. табл. 1). В выборке у особей из ЦП7 выявлено только два варианта генотипов, идентичных вариантам генотипов особей 4.3 и 8.2, из ЦП4 и ЦП8 соответственно (см. табл. 1 и рис. 2).

Таблица 1. Аллели повторяющихся и обладающих уникальными фрагментами генотипов у особей из исследованных ЦП *C. jubata*

№ п/п	№ особи	Аллель	№ п/п	№ особи	Аллель
Варианты генотипов с уникальными фрагментами			22	3.2	–
1	1.4	011111011	23	3.3	–
2	1.6	010010011	24	1.3	010111011 (2)
3	2.2	010001011	25	1.5	–
4	2.3	000111011	26	1.1	010110011 (2)
5	2.4	001111111	27	2.1	–
6	3.4	110110010	28	5.3	010000100 (3)
7	3.5	110000010	29	5.4	–
8	3.6	000100111	30	5.5	–
9	4.1	010110100	31	4.2	010111100 (2)
10	4.4	000001100	32	5.6	–
11	4.5	011100100	33	7.1	110000000 (5)
12	5.1	110111000	34	7.2	–
13	5.2	110000100	35	7.3	–
14	6.1	001111000	36	7.5	–
15	6.6	010111100	37	8.2	–
16	8.3	010110000	38	6.2	010000100 (2)
17	8.4	110110000	39	6.3	–
18	8.5	110010000	40	6.4	010111000 (2)
Варианты генотипов с повторяющимися фрагментами (число)			41	8.1	–
19	1.2	010000011 (5)	42	4.3	010000000 (3)
20	2.5	–	43	6.5	–
21	3.1	–	44	7.4	–

Примечание. 1 – наличие или 0 – отсутствие аллелей.

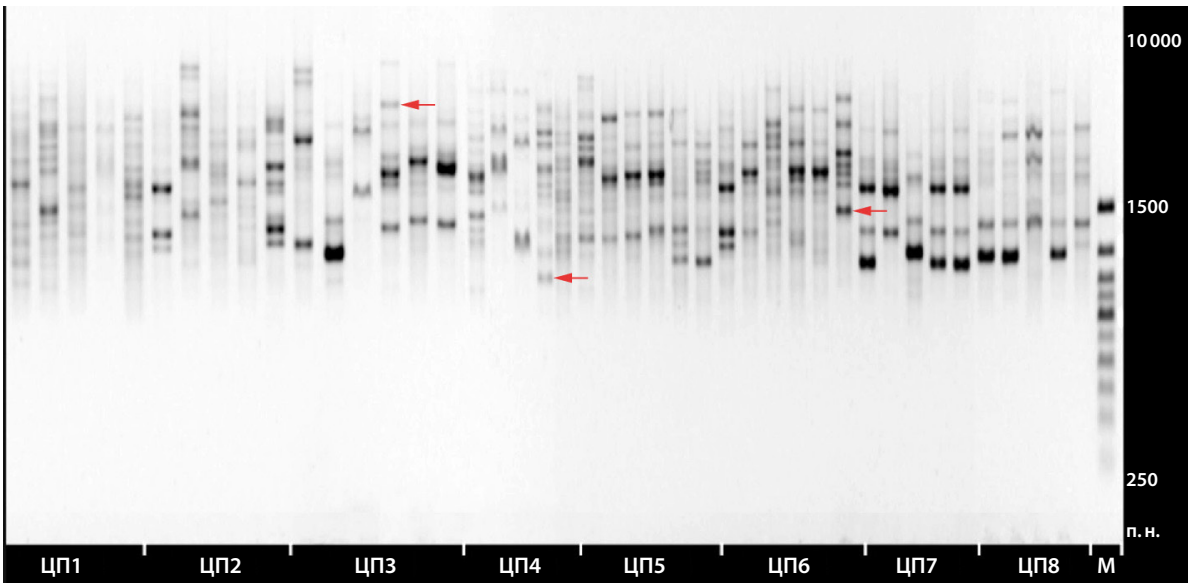


Рис. 2. Электрофореграмма продуктов амплификации фрагментов ДНК с праймером UBC857 в выборках из ЦП *C. jubata*.

ЦП1–ЦП8 – номера ЦП; М – маркер молекулярного веса; стрелками обозначены уникальные фрагменты генотипов.

Таблица 2. Параметры генетического разнообразия в выборках из ЦП *C. jubata*

ЦП	*N*	*Na*	*Ne*	*He*	*uHe*	*P*, %	*Nu/Ng*	*Nr*	Доля *Nu*, %
ЦП1	6	1.22 ± 0.28*	1.31 ± 0.14	0.17 ± 0.07	0.19 ± 0.08	44.4	2/5**	3	40
ЦП2	5	1.56 ± 0.24	1.44 ± 0.14	0.25 ± 0.07	0.28 ± 0.08	66.7	3/5	2	60
ЦП3	6	1.44 ± 0.29	1.35 ± 0.13	0.21 ± 0.07	0.23 ± 0.07	66.7	3/4	1	75
ЦП4	5	1.33 ± 0.33	1.46 ± 0.14	0.26 ± 0.07	0.29 ± 0.08	66.7	3/5	2	60
ЦП5	6	1.22 ± 0.32	1.29 ± 0.11	0.19 ± 0.06	0.20 ± 0.07	55.6	2/4	2	50
ЦП6	6	1.33 ± 0.33	1.44 ± 0.13	0.26 ± 0.07	0.28 ± 0.08	66.7	2/5	3	40
ЦП7	5	0.33 ± 0.24	1.11 ± 0.11	0.06 ± 0.05	0.06 ± 0.06	11.1	0/2	0	0
ЦП8	5	1.00 ± 0.33	1.33 ± 0.15	0.18 ± 0.08	0.20 ± 0.09	44.4	3/5	2	60

* Среднее значение ± стандартная ошибка; *N* – число образцов; *Na* – число аллелей; *Ne* – число эффективных аллелей; *He* – ожидаемая гетерозиготность; *uHe* – объективная ожидаемая гетерозиготность; *P* – доля полиморфных локусов; *Nu/Ng* – соотношение числа генотипов с уникальными фрагментами к числу вариантов генотипов в ЦП; *Nr* – число вариантов повторяющихся генотипов в каждой ЦП.
** В знаменателе приводится общее число вариантов генотипов в выборках из ЦП.

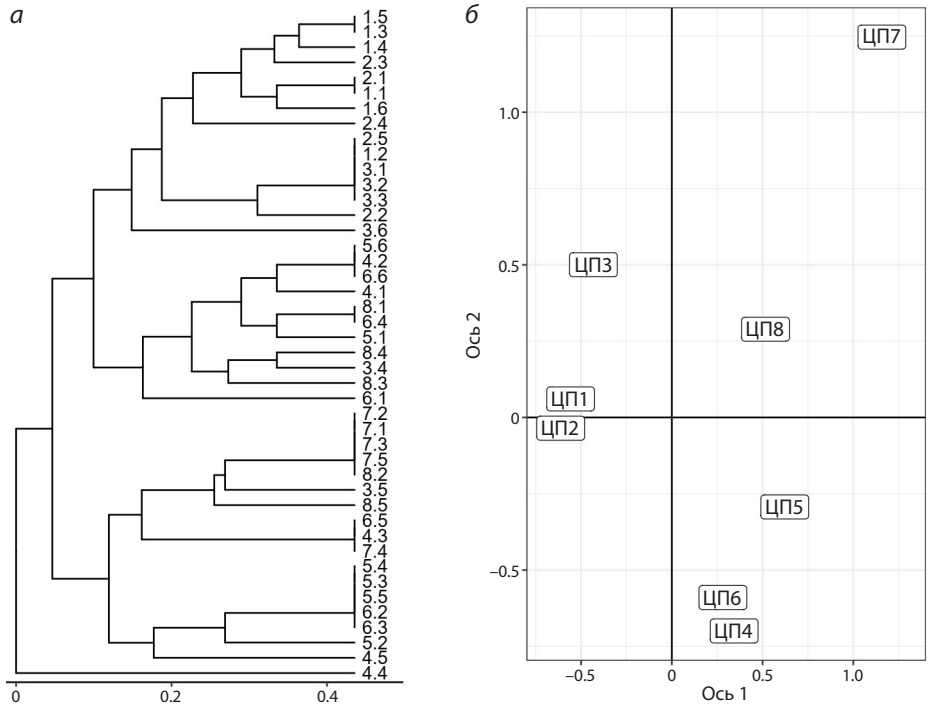


Рис. 3. Распределение *C. jubata* в группы ЦП по сходству генетической структуры.

а – дендрограмма UPGMA, построенная на основе данных по ISSR-спектрам с праймером UBC857; *б* – результат анализа соответствия. ЦП1 – особи 1.1–1.6; ЦП2 – особи 2.1–2.5; ЦП3 – особи 3.1–3.6; ЦП4 – особи 4.1–4.5; ЦП5 – особи 5.1–5.6; ЦП6 – особи 6.1–6.6; ЦП7 – особи 7.1–7.5; ЦП8 – особи 8.1–8.5.

Число вариантов генотипов у особей *C. jubata* в исследованных ЦП варьировало в диапазоне от 2 до 5 (табл. 2). Доля генотипов с уникальными фрагментами в исследованных выборках из ЦП варьировала от их отсутствия (в ЦП7) до 75 % (в ЦП3) (см. табл. 2 и рис. 2). Она рассчитана как отношение числа уникальных генотипов, содержащих уникальные фрагменты, к общему числу вариантов генотипов в каждой ЦП. Представители из ЦП3 произрастали в высокогорных условиях на абсолютной высоте 2550 м, т. е. занимали промежуточное географическое и высотное положение между памирскими и южносибирскими ЦП (см. Приложение 1).

Согласно данным анализа AMOVA, у образцов из восьми исследованных ЦП *C. jubata* межпопуляционная генетическая изменчивость более чем в 1.5 раза ниже внутрипопуляционной, 38 и 62 % соответственно.

Проведенное исследование, согласно кластерному анализу, показало наличие двух наиболее генетически отдаленных друг от друга групп особей – это выборки из ЦП1–ЦП2 и ЦП5–ЦП8. Они характеризуются невысоким генетическим сходством, что представлено на диаграмме распределения особей *C. jubata*, согласно их генотипическим особенностям (рис. 3), и находятся в основном в разных кладах. Соответственно со сходством генотипов

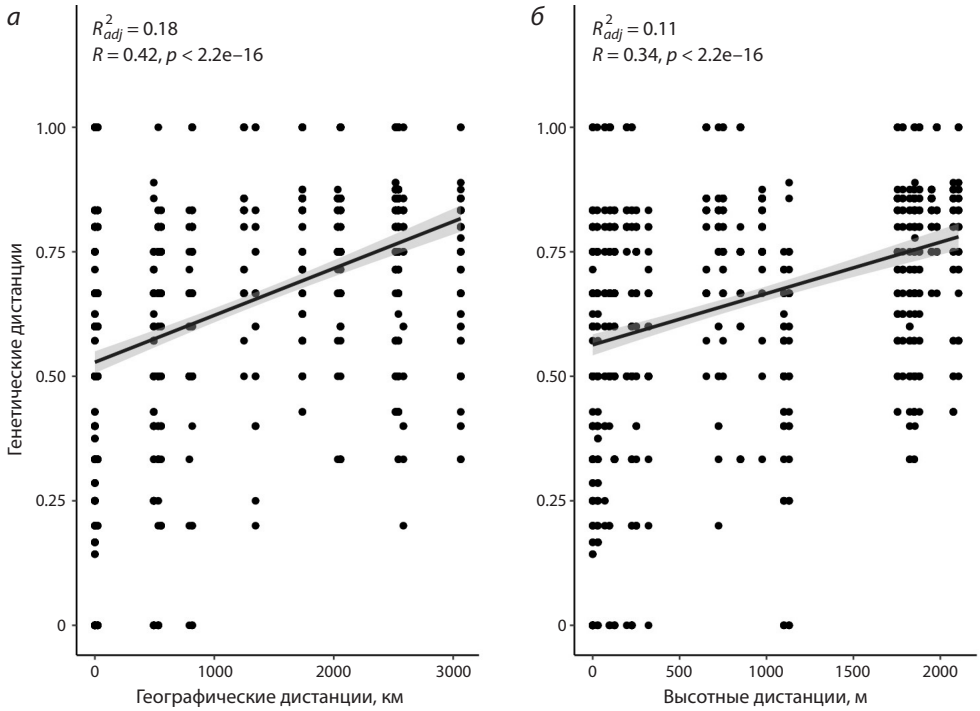


Рис. 4. Корреляции генетических дистанций ЦП *C. jubata* с географической удаленностью (*а*) и относительными высотами (*б*) их местонахождений.

Каждая точка представляет собой сравнение двух индивидуумов, характеризующееся генетической дистанцией (*Y*) и географической/высотной дистанцией (*X*).

исследованные особи *C. jubata* подразделены на два основных кластера, в основном в соответствии с их географической приуроченностью (см. рис. 3, *а*).

Согласно кластерному анализу, в верхней части сосредоточены преимущественно особи из горных районов Западного Памира и Внутреннего Тянь-Шаня (ЦП1–ЦП3). В эту же группу вошла часть южносибирских особей выборок из ЦП4–ЦП6 и ЦП8. В нижней части обособлены, как правило, южносибирские особи (ЦП4–ЦП8). Представители из ЦП7 располагаются отдельно, самыми близкими к ним являются группы особей из южносибирских ЦП. В пользу данного распределения также говорят результаты проведенного анализа соответствия (см. рис. 3, *б*).

По полученным нами данным, особи в крайних в географическом отношении ЦП произрастали на расстоянии более 2.5 тыс. км друг от друга. В результате проведенного статистического анализа данных в выборках из исследованных ЦП нами выявлены достоверные корреляции генетических дистанций с географическими и высотными, что свидетельствует об изоляции расстоянием и высотой, т. е. с увеличением географической и высотной удаленности в выборках у особей из данных ЦП возрастают и их генетические дистанции (рис. 4).

Обсуждение

Результаты секвенирования геномов представителей рода *Caragana* Fabr. показали, что центр его происхождения находится на Тибетском нагорье (Zhang M.-L. et al., 2016). В ксеротермический период плейстоцена *C. jubata* проник по горным системам на запад и север. Это подтверждается, в частности, данными его современного распространения, которое связано с горными районами: восточной частью Тибетского нагорья, северо-востоком Гималаев, Центральным и Южным Тянь-Шанем, Саянами и Становым нагорьем (Churiulina, Bocharnikov, 2019).

Нами проведены исследования *C. jubata* в выборках из ЦП, произраставших также в широком географическом диапазоне. Их местонахождения расположены в пяти горных регионах Центральной Азии – на Западном Памире, Внутреннем Тянь-Шане и в Южной Сибири – на Алтае, Западном и Восточном Саяне. Исследованы ЦП *C. jubata* на градиенте аблсолютных высот, составляющем более 2 тыс. м (см. Приложение 1) и в пределах географических расстояний более 2.5 тыс. км. Такие широкие географический и высотный диапазоны произрастания *C. jubata* соответствуют современному представлению, согласно которому у представителей *C. jubata* отмечена высокая экологическая неоднородность (Таловина, 2019).

В исследованном диапазоне высот более 2 тыс. м и расстояний 2.5. тыс. км в выборках у 44 представителей этого вида нами обнаружено 18 вариантов генотипов, обладающих уникальными фрагментами, что является достаточно высоким показателем для редкого вида растения. Ценность генетического разнообразия заключается в том, что вариация гена, нейтральная в настоящее время, может быть адаптивной впоследствии, при изменении условий среды (de Lafontaine et al., 2018). Популяция с более высоким генетическим разнообразием в таких условиях окружаю-

щей среды с большей вероятностью адаптируется к новой среде, чем популяция с меньшим разнообразием. Таким образом, сохранение генетического разнообразия вида может существенно повысить его шансы на выживание в течение продолжительного времени (Chung et al., 2023).

В горных условиях популяции растений, произрастающие на больших абсолютных высотах, зацветают позже, и период их цветения короче по сравнению с популяциями тех же видов, произрастающих на меньших абсолютных высотах (Reisch et al., 2005). Простое объяснение различий генетических вариаций среди популяций, расположенных на одинаковой высоте, но разделенных горным рельефом, заключается в затруднении разрастания их диаспор (Taberlet et al., 1998). Исследование с использованием палеоботанических и генетических данных для *Fagus sylvatica* L. (бук европейский) показало, что горные цепи в Европе не были географическими барьерами для этого вида, а, скорее, способствовали его распространению. Его расселение, напротив, было затруднено благодаря наличию больших равнин или долин (Magri et al., 2006). Из этого авторы пришли к выводу, что равнинные ландшафты могут стать препятствием для распространения диаспор, в том числе высокогорных видов.

Однако, как правило, редкие и эндемичные виды растений характеризуются значительно более низким генетическим разнообразием и, соответственно, более высокой гомозиготностью, чем виды, не являвшиеся таковыми (Hamabata et al., 2019). Генетическая гомогенизация популяций, особенно у редких видов растений, выражается в снижении гетерозиготности их особей. Популяции, лишенные генетической изменчивости, не способны эволюционировать в ответ на изменение условий окружающей среды и, таким образом, могут столкнуться с повышенным риском вымирания (Dorogina, Zhmud, 2020).

Кроме генотипов с уникальными фрагментами, у представителей *C. jubata* в исследованных ЦП нами обнаружено девять повторяющихся вариантов генотипов. Вариантами генотипов, аналогичными вариантам у особей из других ЦП, представлены особи из ЦП7. Отсутствие уникальных фрагментов в генотипах в выборке из данной ЦП может объясняться антропогенным воздействием на ее особи в местонахождении, где проводится прокладка автомобильной дороги. Подобную ситуацию мы отметили, например, у особей редкого в Сибири вида *Adonis villosa* Ledeb. (адонис пушистый), у 40 исследованных особей которого в шести ЦП в Республике Алтай выявлено только четыре варианта генотипов. В его местообитаниях также проводили дорожные работы и наблюдали повреждение растений и почвенного покрова (Zhmud et al., 2024).

Генетическая гомогенизация нами отмечена в местообитании одной из ЦП у редкого вида *Rhodiola rosea* L. (родиола розовая) в Республике Алтай, особи которой были повреждены из-за вытаптывания в месте летней стоянки и выпаса крупного рогатого скота (яков) (Dorogina et al., 2023). Возможно, сохранившиеся в ЦП7 особи *C. jubata* более устойчивы к повреждениям почвенного покрова, связанным в данном случае со строительными работами. Утрата генетического разнообразия особей в этой ЦП внушает опасения относительно ее устойчивого существования и в дальнейшем может привести к ее полному исчезновению.

При проведении кластерного анализа распределения наиболее полиморфных компонентов при ISSR-маркировании исследованные ЦП распределены в разные группы в соответствии с их генетической структурой. Ранее мы уже отмечали, что местонахождения исследованных нами ЦП *C. jubata* отличаются в крайних точках более чем на 2.5 тыс. км. Обнаружена статистически подтвержденная тенденция увеличения генетических расстояний между представителями из разных ЦП по мере их пространственного удаления друг от друга.

Нами также обнаружена тенденция увеличения доли уникальных генотипов, содержащих уникальные фрагменты, в ЦП *C. jubata*, произраставшей в горных условиях Тянь-Шаня на абсолютной высоте 2550 м – в средней части исследованного нами высотного диапазона, по сравнению с памирскими и южносибирскими ЦП. Согласно данным T. Ohsawa и I. Yuji (2008), такие популяции могут находиться в более благоприятных условиях среды, по сравнению с периферийными популяциями, произрастающими в субоптимальных ситуациях. По мнению этих авторов, изоляция горами играет важнейшую роль в диверсификации и эволюции видов.

Адаптация дикорастущих растений к стрессу при увеличении высоты произрастания – сложный процесс, который достигается за счет не только модификационной, но и генетической изменчивости (Сафаралихонов, Акназаров, 2021). Молекулярные исследования P.K. Bhardwaj с коллегами (2013) показали, что у *C. jubata* обнаружено преобладание генов, кодирующих шапероны, и генов, участвующих в росте и развитии при низкой температуре, которая является одним из главных сигналов, действующих на растения в высокогорьях. Согласно этим исследованиям, такие гены экспрессируются у *C. jubata* в естественной среде обитания. Их гомологи у *Arabidopsis thaliana* (L.) Heynh. (резушка Таля), *Glycine max* (L.) Merr. (соя культурная) и *Oryza sativa* L. (рис посевной) не демонстрируют схожей тенденции экспрессии генов при низкой температуре.

Конститутивная экспрессия и быстрая регуляция вышеуказанных генов предполагают способность *C. jubata* настраивать клеточный аппарат для поддержания роста и развития в своей экологической нише. Такая молекулярная и физиологическая пластичность позволяет *C. jubata* благополучно произрастать, в частности, в холодной высокогорной пустыне Гималаев (Bhardwaj et al., 2013). Обнаруженные нами тенденции дифференциации генетической структуры в исследованных ЦП *C. jubata*, вероятно, являются результатом адаптации представителей этого вида к различным эколого-географическим условиям.

По оценке Г.П. Семеновой (2007) *C. jubata* – среднеперспективный вид в условиях интродукции в лесостепной зоне Западной Сибири. Такие виды размножаются семенами или вегетативно, требуют определенного состава почв, полива, подбора участков с особыми условиями освещения, ежегодно цветут и плодоносят, но могут под-

мерзать, и в неблагоприятные годы их репродуктивная способность может быть ослаблена. Также *C. jubata* отнесен Г.П. Семеновой (2007) к долголетним растениям, так как продолжительность существования интродукционной популяции в местных условиях достигала 17 лет.

Обширный ареал этого вида и относительно высокое генетическое разнообразие позволяют предположить, что особи в выборках из большинства исследованных нами ЦП устойчивы и могут служить исходным материалом для создания перспективных интродукционных популяций.

Таким образом, при проведении статистического анализа генетической изменчивости выявлено достоверное влияние географической изоляции между генотипами, обусловленной большими расстояниями между исследованными ЦП *C. jubata*. Подтверждено также влияние высотной изоляции на генетические дистанции между особями разных ценопопуляций.

Заключение

Caragana jubata исследован нами в выборках из восьми ЦП, произраставших в широком диапазоне высот, составившем более 2.0 тыс. м (от 1570 до 3680 м), и географических расстояний (более 2.5 тыс. км) в горных условиях Памира, Тянь-Шаня и Южной Сибири. Проведенный ISSR-анализ полиморфных фрагментов ДНК позволил нам выявить тенденцию увеличения числа генотипов с уникальными фрагментами и доли полиморфных локусов у особей этого вида, произраставших в горах Тянь-Шаня на абсолютной высоте 2550 м, расположенной в серединной части исследованного высотного диапазона. Также нами обнаружена статистически подтвержденная тенденция увеличения генетических расстояний в выборках у представителей из разных ЦП по мере их пространственного удаления друг от друга.

Выявленное в результате исследований преобладание внутрипопуляционной генетической изменчивости над межпопуляционной может свидетельствовать об устойчивости данных ЦП в пределах исследованных частей ареала этого вида и об их хороших адаптивных способностях. Исключение составили особи выборки из ЦП7 (Республика Тыва), подвергавшиеся антропогенному воздействию, в которой генотипы с уникальными фрагментами нами не обнаружены.

Список литературы / References

Жмудь Е.В., Кубан И.Н., Ачимова А.А., Ямтыров М.Б., Дорогина О.В. Сравнительный анализ генетического полиморфизма в популяциях *Rhaponticum carthamoides* (Asteraceae) в Республике Алтай. *Труды по прикладной ботанике, генетике и селекции*. 2023;184(2):149-159. doi 10.30901/2227-8834-2023-2-149-159
[Zhmud E.V., Kuban I.N., Achimova A.A., Yamtyrov M.B., Dorogina O.V. Comparative analysis of genetic polymorphism in *Rhaponticum carthamoides* (Asteraceae) populations by ISSR markers in the Altai Republic. *Trudy po Prikladnoy Botanike, Genetike i Selektsii = Proceedings on Applied Botany, Genetics, and Breeding*. 2023; 184(2):149-159. doi 10.30901/2227-8834-2023-2-149-159 (in Russian)]

Какорин П.А., Бабенкова И.В., Теселкин Ю.О., Раменская Г.В., Демура Т.А., Кукес В.Г. Гепатопротекторная активность водного извлечения из побегов *Caragana jubata* (Pall.) Poir. на модели острого гепатита, индуцированного ацетаминофеном у крыс. *Биомедицинская химия*. 2018;64(3):241-246. doi 10.18097/PBMC20186403241
[Kakorin P.A., Babenkova I.V., Teselkin Yu.O., Ramenskaya G.V., Demura T.A., Kukes V.G. Hepatoprotective activity of aqueous extract from *Caragana jubata* (Pall.) Poir. shoots in the model of acute hepatitis induced by acetaminophen in rats. *Biomeditsinskaya Khimiya.* 2018;64(3):241-246. doi 10.18097/PBMC20186403241 (in Russian)]

Коропачинский И.Ю., Встовская Т.Н. Древесные растения Азиатской России. Новосибирск: Акад. изд-во «Гео», 2012
[Koropachinsky I.Yu., Vstovskaya T.N. Woody Plants of Asian Russia. Novosibirsk: Geo Publ., 2012 (in Russian)]

Красная книга Забайкальского края. Новосибирск: Дом мира, 2017
[The Red Book of Zabaykalskiy Kray. Novosibirsk: Dom Mira Publ., 2017 (in Russian)]

Красная книга Республики Алтай (растения). Горно-Алтайск: ООО «Горно-Алтайская типография», 2017
[The Red Book of Altay Republic. Plants. Gorno-Altaysk: Gorno-Altayskaya Tipografia Publ., 2017 (in Russian)]

Красная книга Республики Саха (Якутия). Т. 1. Редкие и находящиеся под угрозой исчезновения виды растений и грибов. М.: Изд-во «Реарт», 2017
[The Red Book of Sakha Republic (Yakutia). Vol. 1. Rare and Endangered Species of Plants and Fungi. Moscow: Reart Publ., 2017 (in Russian)]

Красная книга Магаданской области. Редкие и находящиеся под угрозой исчезновения виды животных, растений и грибов. Магадан: Охотник, 2019
[The Red Book of Magadan Oblast. Rare and Endangered Species of Animals, Plants and Fungi. Magadan: Okhotnik Publ., 2019 (in Russian)]

Красная книга Хабаровского края. Редкие и находящиеся под угрозой исчезновения виды растений, грибов и животных. Воронеж: ООО «Мир», 2019
[The Red Book of Khabarovsk Kray. Rare and Endangered Species of Plants, Fungi, and Animals. Voronezh: Mir Publ., 2019 (in Russian)]

Красная книга Иркутской области. Улан-Удэ: ПАО «Республиканская типография», 2020
[The Red Book of Irkutsk Oblast. Ulan-Ude, 2020 (in Russian)]

Красная книга Красноярского края. Т. 2. Редкие и находящиеся под угрозой исчезновения виды растений и грибов. Красноярск: СФУ, 2022
[The Red Book of Krasnoyarsk Kray. Vol. 2. Rare and Endangered Species of Plants and Fungi. Krasnoyarsk: Siberian Federal University, 2022 (in Russian)]

Красная книга Республики Бурятия. Редкие и находящиеся под угрозой исчезновения виды растений и грибов. Белгород: Константа, 2023
[The Red Book of Republic of Buryatia. Rare and Endangered Species of Plants and Fungi. Belgorod: Konstanta Publ., 2023 (in Russian)]

Лысова Н.В. Экология *Caragana jubata* (Pall.) Poir. в Центральном Тянь-Шане. *Бюллетень Главного ботанического сада АН СССР*. 1967;64:64-67
[Lysova N.V. Ecology of *Caragana jubata* (Pall.) Poir. in the Central Tien Shan. *Byulleten Glavnogo Botanicheskogo Sada = Bulletin of the Moscow Botanical Garden of the Academy of Sciences*. 1967;64: 64-67 (in Russian)]

Набиева А.Ю., Жмудь Е.В., Кубан И.Н., Дорогина О.В. Морфометрический и молекулярный анализ популяции *Cypripedium* × *ventricosum* (Orchidaceae) в Новосибирской области. *Ботанический журнал*. 2020;105(6):597-603. doi 10.31857/S0006813620060058
[Nabieva A.Yu., Zhmud E.V., Kuban I.N., Dorogina O.V. Morphometric and molecular analysis of *Cypripedium* × *ventricosum*

(Orchidaceae) population in the Novosibirsk Region. *Botanicheskii Zhurnal.* 2020;105(6):597-603. doi 10.31857/S0006813620060058 (in Russian)]

Ооржак У.С., Лопсан-Ендан А.Б., Доржу-оол Ю.Н. Биолого-экологические и химические особенности *Caragana jubata* (Pall.) Poir. *Вестник Тувинского государственного университета. Естественные и сельскохозяйственные науки.* 2024;1(5):16-24. doi 10.24411/2221-0458-2024-01-16-24
[Oorzhak U.S., Lopsan-Endan A.B., Dorzhu-ool Yu.N. Bioecological and chemical features of *Caragana jubata* (Pall.) Poir. *Vestnik Tuvinskogo Gosudarstvennogo Universiteta. Estestvennye i Selskokhozyajstvennye Nauki – Vestnik of of Tuvan State University*. *Natural and Agricultural Sciences.* 2024;1(5):16-24. doi 10.24411/2221-0458-2024-01-16-24 (in Russian)]

Парыгин И.А., Худоногова Е.Г. Растительные сообщества *Caragana jubata* (Pallas) Poiret юго-западных районов Бурятии. *Вестник Оренбургского государственного педагогического университета*. 2017;4(24):26-33
[Parygin I.A., Khudonogova E.G. Plant communities of *Caragana jubata* (Pallas) Poiret of the southwestern regions of Buryatia. *Vestnik Orenburgskogo Gosudarstvennogo Pedagogicheskogo Universiteta = Vestnik of Orenburg State Pedagogical University*. 2017; 4(24):26-33 (in Russian)]

Сафаралихонов А.Б., Акназаров О.А. Действие экологических факторов высокогорья Памира на жизнедеятельность растений. Душанбе: Дониш, 2021
[Safaralikhonov A.B., Aknazarov O.A. The Effect of Environmental Factors of the Pamir Highlands on Plant Life. Dushanbe: Donish Publ., 2021 (in Russian)]

Семенова Г.П. Редкие и исчезающие виды флоры Сибири: биология, охрана. Новосибирск: Акад. изд-во «Гео», 2007
[Semenova G.P. Rare and Endangered Species of Siberian Flora: Biology, Conservation. Novosibirsk: Geo Publ., 2007 (in Russian)]

Таловина Г.В. Дикие родичи культурных растений Магаданской области: перспективы сохранения генофонда. *Vavilovia*. 2019; 2(3):42-55. doi 10.30901/2658-3860-2019-3-42-55
[Talovina G.V. Crop Wild Relatives in the Magadan Province of Russia: Perspectives of Gene Pool Conservation. *Vavilovia*. 2019; 2(3):42-55. doi 10.30901/2658-3860-2019-3-42-55 (in Russian)]

Чурюлина А.Г. География редких видов (*Caragana jubata* (Pall.) Poir., *Betula lanata* (Regel) V. Vassil.) и сообществ с их участием в горных биомах Байкальского региона: Автореф. … канд. геогр. наук. М., 2021
[Churyulina A.G. Geography of rare species (*Caragana jubata* (Pall.) Poir., *Betula lanata* (Regel) V. Vassil.) and communities with their participation in the mountain biomes of the Baikal region: Cand. Sci. (Geogr.) Dissertation. Moscow, 2021 (in Russian)]

Чурюлина А.Г., Бочарников М.В., Огуреева Г.Н. География караганы гривастой (*Caragana jubata* (Pall.) Poir.) и ее фитоценотическая роль в растительном покрове гор. *Вестник Московского университета. Серия 5. География.* 2020;3:108-117
[Churyulina A.G., Bocharnikov M.V., Ogureeva G.N. Geography of *Caragana jubata* (Pall.) Poir. (Fabaceae) and its phytocenotic role in the mountain vegetation cover. *Vestnik Moskovskogo Universiteta. Seriya 5. Geografiya = Moscow University Bulletin. Series 5. Geography*. 2020;3:108-117 (in Russian)]

Bhardwaj P.K., Kapoor R., Mala D., Bhagwat G., Acharya V., Singh A.K., Vats S.K., Ahuja P.S., Kumar S. Braving the attitude of altitude: *Caragana jubata* at work in cold desert of Himalaya. *Sci Rep*. 2013;3:1022. doi 10.1038/srep01022

Chung M.Y., Mmerila J., Li J., Mao K., López-Pujo J., Tsumura Y., Chung M.G. Neutral and adaptive genetic diversity in plants: an overview. *Front Ecol Evol*. 2023;11:1116814. doi 10.3389/fevo.2023.1116814

Churiulina A.G., Bocharnikov M.V. *Caragana jubata* (Pall.) Poir. (Fabaceae), distribution mapping of rare relict species. *Bot Pacifica.* 2019;8(2):111-114. doi 10.17581/bp.2019.08207

de Lafontaine G., Napier J.D., Petit R.J., Hu F.S. Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology*. 2018;99(7):1530-1546. doi 10.1002/ecy.2382

Dorogina O.V., Zhmud E.V. Molecular-genetic methods in plant ecology. *Contemp Probl Ecol.* 2020;13(4):333-345. doi 10.1134/S1995425520040058

Dorogina O.V., Kuban I.N., Achimova A.A., Williams N., Lashchinskiy N.N., Zhmud E.V. Morphometric characteristics and genetic ISSR marker variability in *Rhodiola rosea* L. (Crassulaceae) in different ecological and geographic conditions in the Altai Republic. *Int J Molec Sci*. 2023;24(10):15224. doi 10.3390/ijms242015224

Hamabata T., Kinoshita G., Kurita K., Cao P.-L., Ito M., Murata J., Komaki Y., Isagi Y., Makino T. Endangered island endemic plants have vulnerable genomes. *Commun Biol*. 2019;2:244. doi 10.1038/s42003-019-0490-7

Hantemirova E.V., Khoreva M.G., Bessonova V.A. Genetic and ecological differentiation of *Caragana jubata* (Pall.) Poir. *Bot Pacifica*. 2024;13(2):15-22. doi 10.17581/bp.2024.13204

Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24(11):1403-1405. doi 10.1093/bioinformatics/btn129

Kakorin P.A., Tereshkina O.I., Ramenskaya G.V. Potential biological activity and chemical composition of *Caragana jubata* (Pall.) Poir. (review). *Pharm Chem J*. 2018;52(6):531-535. doi 10.1007/s11094-018-1854-x

Kakorin P.A., Fateeva T.V., Tereshkina O.I., Perova I.B., Ramenskaya G.V., Sologova S.S., Eller K.I. Antimicrobial activity of lyophilized aqueous extract from *Caragana jubata* (Pall.) Poir. *Pharm Chem J*. 2020;54(3):290-292. doi 10.1007/s11094-020-02193-z

Kassambara A. ggpubr: 'ggplot2' based publication ready plots. R package version 0.6.0. 2023. https://rpkgs.datanovia.com/ggpubr/

Liu C., Wang J., Ko Y.-Z., Shiao M.-S., Wang Y., Sun J., Yuan Q., Wang L., Chiang Y.-C., Guo L. Genetic diversities in wild and cultivated populations of the two closely-related medical plants species, *Tripterygium wilfordii* and *T. hypoglaucum* (Celastraceae). *BMC Plant Biol.* 2024;24(1):195. doi 10.1186/s12870-024-04826-x

Magri D., Vendramin G.G., Comps B., Dupanloup I., Geburek T., Gömöry D., Latałowa M., Litt T., Paule L., Roure J.M., Tantau I., Van Der Knaap W.O., Petit R.J., De Beaulieu J.-L. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol*. 2006;171(1):199-221. doi 10.1111/j.1469-8137.2006.01740.x

Nørgaard L.S., Mikkelsen D.M.G., Elmeros M., Chriél M., Madsen A.B., Nielsen J.L., Pertoldi C., Randi E., Fickel J., Brygida S., Ruiz-González A. Population genomics of the raccoon dog (*Nyctereutes procyonoides*) in Denmark: insights into invasion history and population development. *Biol Invasions*. 2017;19(5):1637-1652. doi 10.1007/s10530-017-1385-5

Ohsawa T., Yuji I. Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Glob Ecol Biogeogr*. 2008;17(2):152-163. doi 10.1111/j.1466-8238.2007.00357.x

Oyama K., Ito M., Yahara T., Ono M. Low genetic differentiation among populations of *Arabis serrata* (Brassicaceae) along an altitudinal gradient. *J Plant Res*. 1993;106(2):143-148. doi 10.1007/BF02344418

R Core Team. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2024. https://www.R-project.org/

Reisch C., Anke A., Röhl M. Molecular variation within and between ten populations of *Primula farinose* (Primulaceae) along an altitudinal gradient in the northern Alps. *Basic Appl Ecol*. 2005;6(1):35-45. doi 10.1016/j.baae.2004.09.004

Shi M.-M., Michalski S.G., Chen X.-Y., Durka W. Isolation by elevation: genetic structure at neutral and putatively non-neutral loci in a dominant tree of subtropical forests, *Castanopsis eyrie*. *PLoS One*. 2011;6(6):e21302. doi 10.1371/journal.pone.0021302

Smouse P.E., Banks S.C., Peakall R. Converting quadratic entropy to diversity: both animals and alleles are diverse, but some are more diverse than others. *PLoS One.* 2017;12(10):e0185499. doi 10.1371/journal.pone.0185499

Taberlet P., Fumagalli L., Wust-Saucy A.-G., Cosson J-F. Comparative phylogeography and postglacial colonization route in Europe. *Mol Ecol.* 1998;7(4):453-464. doi 10.1046/j.1365-294x.1998.00289.x

Wang Q., Wu X., Yang X., Zhang Y., Wang L., Li X., Qiu Y. Comprehensive quality evaluation of Lignum Caraganae and rapid discrimination of *Caragana jubata* and *Caragana changduensis* based on characteristic compound fingerprints by HPLC-UV and HPLC-MS/MS coupled with chemometrics analysis. *Phytochem Anal*. 2020;31(6):846-860. doi 10.1002/pca.2950

Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., François R., Grolemund G., … Takahashi K., Vaughan D., Wilke C., Woo K., Yutani H. Welcome to the Tidyverse. *J Open Source Software*. 2019;4(43):1686. doi 10.21105/joss.01686

Wright S. Breeding structure of populations in relation to speciation. *Am Nat.* 1940;74(752):232-248. doi 10.1086/280891

Xu S., Li L., Luo X., Chen M., Tang W., Zhan L., Dai Z., Lam T.T., Guan Y., Yu G. *Ggtree*: a serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta*. 2022;1(4):e56. doi 10.1002/imt2.56

Yang X.-Z., Lyu J.-N., Xu C., Lin Q.-X., Yang J., Wang C., Song P. Cytotoxic chemical constituents from *Caragana jubata* (Pall.) Poir. *J Yunnan Univ Nat Sci Ed*. 2015;37(1):134-139. doi 10.7540/j.ynu.20140364

Yao N., Gao J., Wang L., Ma Z., Zhang J., Ma F., Xu T. Development of SSR molecular markers and analysis of genetic diversity in *Caragana jubata* (Pall.) Poir. *Genet Resour Crop Evol*. 2025;72(7):8957-8972. doi 10.1007/s10722-025-02477-x

Yifru T., Hammer K., Huang X.Q., Röder M.S. Regional patterns of microsatellite diversity in Ethiopian tetraploid wheat accessions. *Plant Breed*. 2006;125(2):125-130. doi 10.1111/j.1439-0523.2006.01147.x

Zhang M.-L., Xiang X.-G., Xue J.-J., Sanderson S.C., Fritsch P.W. Himalayan uplift shaped biomes in Miocene temperate Asia: evidence from leguminous *Caragana*. *Sci Rep*. 2016;6:36528. doi 10.1038/srep36528

Zhang X., Korpelainen H., Li C. Microsatellite variation of *Quercus aquifolioides* populations at varying altitudes in the Wolong natural reserve of China. *Silva Fenn*. 2006;40(3):407-415. doi 10.14214/sf.326

Zhao D., Chen X., Wang R., Pang H., Wang J., Liu L. Determining the chemical profile of *Caragana jubata* (Pall.) Poir. by UPLC-QTOF-MS analysis and evaluating its anti-ischemic stroke effects. *J Ethnopharmacol*. 2023;309:116275. doi 10.1016/j.jep.2023.116275

Zhmud E.V., Kuban I.N., Achimova A.A., Papina O.N., Dorogina O.V. Variability of morphometric parameters in *Adonis villosa* Ledeb. (Ranunculaceae) and its genetic differentiation based on ISSR markers in the Altai Republic. *Contemp Probl Ecol.* 2024;17(4):557-565. doi 10.1134/S1995425524700355

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 12.03.2025. После доработки 06.10.2025. Принята к публикации 11.11.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):205-211

doi 10.18699/vjgb-26-23

Изменчивость органельных геномов в коллекции раннеспелых сортов сои

В.В. Александрович[1], М.Г. Синявская[1], О.П. Шатарнов[1], О.Г. Давыденко[2]

[1] Институт генетики и цитологии Национальной академии наук Беларуси, Минск, Республика Беларусь
[2] ООО «Соя-Север Ко», пос. Колодищи, Минская область, Республика Беларусь
valeria.alexandrovich@gmail.com

Аннотация. Изменчивость геномов клеточных органелл – хлоропластов и митохондрий – является немаловажной компонентой общей изменчивости генома растений. Получено большое количество данных о сравнительных особенностях организации последовательностей органельных ДНК для различных групп растений. В настоящей работе представлены новые оригинальные данные об изменчивости геномов митохондрий и хлоропластов у сои (*Glycine max* (L.) Merr.), важной хозяйственной и пищевой культуры, широко возделываемой на территории Центральной Европы, в том числе и в Республике Беларусь. Рабочей гипотезой нашего исследования изначально стало предположение: возможно, особенности изменчивости последовательности или структуры ДНК органелл сои определяют способность одних сортов выступать в роли лучших материнских родителей, а других – быть лучшими отцовскими формами. Получены новые полные нуклеотидные последовательности хлоропластного и митохондриального геномов 46 образцов культурной сои с применением метода секвенирования нового поколения (NGS) на платформе Illumina. Выполнено комплексное биоинформатическое сравнительное исследование внутривидовой изменчивости геномов органелл у 46 сортов сои разнообразного географического происхождения. Выявлены полиморфные локусы геномов. Данные об изменчивости ДНК верифицированы секвенированием по Сэнгеру. Спектр изменчивости органельных ДНК, исследованных методом полногеномного секвенирования сортов сои, представлен тремя гаплотипами хлоропластной ДНК (C1–C3) и пятью гаплотипами митохондриальной ДНК (M1–M5). Обнаружен сравнительно низкий уровень внутривидовой изменчивости геномов органелл у *G. max*. Хлоропластный геном сои обладал меньшим уровнем изменчивости последовательности, чем митохондриальный. Разработан набор ДНК-маркеров к полиморфным локусам геномов органелл, позволяющий дифференцировать сорта сои на плазматипы. Методом ПЦР с последующим секвенированием по Сэнгеру дополнительно изучено 90 образцов сои из коллекции. Низкий уровень внутривидовой изменчивости геномов органелл у *G. max* подтвержден на расширенной группе образцов. Большая часть коллекции была представлена тремя плазматипами – C1/M1, C2/M2 и C1/M3. 46 полных последовательностей хлоропластной ДНК помещено в NCBI GenBank. Гипотеза влияния ДНК органелл на комбинационную способность различных сортов в настоящее время не подтверждена. Требуются более детальное изучение механизмов ядерно-цитоплазматического взаимодействия, поиск ядерных маркеров, влияющих на экспрессию цитоплазматических генов.
Ключевые слова: соя; *Glycine max*; генетическая изменчивость; органеллы; хлоропласты; митохондрии; NGS

Для цитирования: Александрович В.В., Синявская М.Г., Шатарнов О.П., Давыденко О.Г. Изменчивость органельных геномов в коллекции раннеспелых сортов сои. *Вавиловский журнал генетики и селекции*. 2026;30(2):205-211. doi 10.18699/vjgb-26-23

Финансирование. Работа выполнена в рамках ГПНИ «Биотехнологии 2», 2021–2025 гг., подпрограмма «Геномика, эпигеномика, биоинформатика», задание 2.1.3.

Variability of organelle genomes in a collection of early maturing soybean varieties

V.V. Aleksandrovich[1], M.G. Siniauskaya[1], A.P. Shatarnov[1], O.G. Davydenko[2]

[1] The Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus
[2] Soya-North Co Ltd, Kolodishchi, Minsk oblast, Belarus
valeria.alexandrovich@gmail.com

Abstract. Variability of the genomes of cellular organelles (chloroplast and mitochondria) is an important component of the overall variability of the plant genome. A large amount of data has already been obtained on the comparative characteristics of the organization of organelle DNA sequences for different groups of plants. This paper presents new original data on the variability of mitochondrial and chloroplast genomes in soybean (*Glycine max* (L.) Merr.), a crop of great economic importance widely cultivated in Central Europe, including the Republic of Belarus. Initially, we supposed that the peculiarities of soybean organelle DNA sequence or organization promote certain soybean cultivars to be the best maternal and others, alternatively, the best paternal forms. As a result of the study, new complete nucleotide sequences of chloroplast and mitochondrial genomes of 46 soybean samples were obtained by the next generation

© Александрович В.В., Синявская М.Г., Шатарнов О.П., Давыденко О.Г., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

sequencing method (NGS) on the Illumina platform. A comprehensive bioinformatic comparative study of intraspecific organelle genome variability in 46 soybean varieties of diverse geographical origin was conducted. Polymorphic loci of genomes were discovered. Data on DNA variability were verified by Sanger sequencing. The spectrum of organelle DNA variability of cultivated soybean was represented by three chloroplast DNA haplotypes (C1–C3) and five mitochondrial DNA haplotypes (M1–M5). A comparatively low level of intraspecific variability of organelle genomes in *G. max* was revealed. The soybean chloroplast genome had a lower level of sequence variability than the mitochondrial genome. A set of DNA markers for polymorphic loci of organelle genomes was developed, allowing the differentiation of varieties of the studied group into plasmatypes. Additionally, 90 soybean samples from the collection were studied using PCR followed by Sanger sequencing. The low level of intraspecific variability of organelle genomes in *G. max* was confirmed on the extended group of samples. The majority of cultivars were represented by three plasmatypes – C1/M1, C2/M2 and C1/M3. 46 complete chloroplast DNA sequences have been deposited in NCBI GenBank. The hypothesis that organelle DNA influences the combining ability of different varieties has not yet been confirmed. A more detailed study of the mechanisms of nuclear-cytoplasmic interaction is required, as well as a search for nuclear markers that affect the expression of organelle genes.

Key words: soybean; *Glycine max*; genetic variability; organelles; chloroplasts; mitochondria; NGS

For citation: Aleksandrovich V.V., Siniauskaya M.G., Shatarnov A.P., Davydenko O.G. Variability of organelle genomes in a collection of early maturing soybean varieties. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2): 205-211. doi 10.18699/vjgb-26-23

Введение

Изучение особенностей организации и функционирования геномов клеточных органелл – одна из значимых областей современной генетики растений. За последние десятилетия множество исследований было посвящено данной теме с целью решения вопросов систематики, филогении, изучения ядерно-цитоплазматической коадаптации и взаимодействия геномов (Gualberto, Newton, 2017; Johnston, 2019; Tsunewaki et al., 2019). Геномы органелл важнейших сельскохозяйственных культур активно исследуются в мире (Siniauskaya et al., 2020; Hu et al., 2022; Yue et al., 2023).

Соя культурная (*Glycine max* (L.) Merr.) выращивается в условиях Беларуси уже не один год. В 1980-х годах зародилась отечественная школа генетики и селекции этой культуры на базе лаборатории нехромосомной наследственности Института генетики и цитологии Национальной академии наук Беларуси. Основное направление селекции сои в Беларуси – создание раннеспелых и ультрараннеспелых сортов с достаточно высокими урожайностью и содержанием белка в зерне.

Подбор исходного материала для гибридизации – ключевое звено в селекции любой культуры. Известно, что сорта могут различаться по сортообразующей способности (продукции новых гибридов и сортов). Поэтому необходимы анализ родословных и поиск таких сортов. На основании изучения родословных гибридов, сортов, создаваемых другими исследователями, а также собственных данных, нами выделены сорта сои, являющиеся наиболее перспективными материнскими или отцовскими родителями, которые наиболее часто используются в скрещивании. Учитывая строго материнское наследование органелл у сои, можно предположить, что наблюдаемая дифференциация сортов по способности давать успешные новые комбинации генов у гибридов определяется ядерно-цитоплазматическими взаимодействиями, возможно, особенностями организации и экспрессии органельных геномов.

Наличие коллекций исходного селекционного материала растений с широким варьированием по ряду хозяйственно ценных признаков – необходимая предпосылка создания новых, перспективных сортов. Изучение фенотипической, генетической изменчивости коллекций является одним из базовых основополагающих направлений современной селекции, в том числе и маркер-ассоциированной.

Первые молекулярно-генетические исследования по изучению изменчивости органельных геномов культурной и дикой сои проводились в 1980-х годах методом анализа полиморфизма длин рестрикционных фрагментов (ПДРФ). Классической в то время была классификация линий культурной и дикорастущей сои на три гаплотипа хлоропластной ДНК (хпДНК) и пять гаплотипов митохондриальной (мтДНК) (Shoemaker et al., 1986). Метод ПДРФ использовался для оценки разнообразия органельных геномов линий дикорастущей сои (*Glycine soja*) (Abe et al., 1999), сортов сои, возделываемых на территории Китая (Shimamoto et al., 1998).

Возможность более детального изучения разнообразия геномов органелл растений, в том числе и сои, другим методом – по изменчивости длины микросателлитных повторов (SSR) в хпДНК – показана W. Powell с коллегами (1995). Применение такого подхода позволило расширить представления об изменчивости хпДНК сои, получить качественно новые данные. Так, в работе (Xu et al., 2002) путем исследования шести хлоропластных микросателлитов выделено 52 гаплотипа *G. soja* и 8 гаплотипов культурной сои, при этом 75 % линий *G. max* относилось к одному самому распространенному гаплотипу.

Работы по изучению изменчивости генома хлоропластов сои были начаты в Институте генетики и цитологии НАН Беларуси в 2000-х годах. ПЦР-ПДРФ исследование коллекции имеющихся в то время в лаборатории нехромосомной наследственности 60 сортов позволило обнаружить три плазматипа среди них, в зависимости от отсутствия либо присутствия в хпДНК сайтов узнавания рестриктаз *Eco*RI и *Cla*I. Большинство сортов (из более 60 изученных) имели дополнительный *Cla*I сайт, три линии – дополнительный сайт *Eco*RI, хпДНК двух линий не имела сайтов рестрикции *Cla*I и *Eco*RI (Синявская и др., 2004).

Исследованы также микросателлитные повторы хпДНК; проведена оценка разнообразия коллекций образцов сои различного происхождения (Aksyonova et al., 2007).

Результаты ряда авторских коллективов по анализу молекулярного разнообразия органельных геномов неизменно демонстрировали, что дикорастущая соя обладает более высоким уровнем цитоплазматической изменчивости, чем ее культурный сородич, – большинство сортов *G. max* было отнесено к одному либо двум основным гаплотипам (Shimamoto, 2001; Xu et al., 2002). Сорта китайской селекции были более генетически разнообразны, чем сорта, выращиваемые на территории Северной Америки и Европы (Yue et al., 2023). Низкий уровень полиморфизма органельных геномов у сои может объясняться многими факторами: однородительским наследованием органелл у сои по материнской линии, что исключает возможность рекомбинации геномов от двух родителей; особенностями размножения культурной сои, например строгим самоопылением.

Изучение особенностей структуры геномов органелл позволяет получить информацию, которую можно применить для рассмотрения и предположения о возможных микроэволюционных изменениях внутри рода *Glycine*. Общепринятой считается теория монофилетического происхождения культурной сои от предковой формы *G. soja*. Данная гипотеза выдвинута на основе исследования ядерных маркеров, полногеномного секвенирования и панелей ядерных SNP (Jeong et al., 2019). Однако в нескольких работах (Xu et al., 2002; Fang et al., 2016) была предложена теория повторяющихся процессов доместикации сои. На происхождение современных культурных сортов от нескольких изначальных материнских линий указывают особенности геномов хлоропластов и митохондрий, в частности, у большинства сортов геномы органелл относятся к одному из двух основных плазматипов. C. Fang с коллегами (2016) разделяют все гаплотипы хпДНК на две группы и предполагают, что в процессе доместикации сои было задействовано несколько материнских линий дикорастущей сои, что дало начало двум группам плазматипов у современных сортов.

Развитие методов секвенирования нового поколения (NGS) позволило получить принципиально новую информацию об организации органельных геномов растений, в том числе и сои, их последовательностях и изменчивости. Y. Yue с коллегами (2023) провели анализ данных NGS более 2000 линий сои из общедоступных баз данных и обнаружили, что 69.2 % всех культивируемых сортов сои относится к одному плазматипу – CT1/MT1, 18.1 % – к плазматипу CT2/MT2, что соотносится с данными более ранних исследований.

Полногеномное секвенирование (платформа Illumina) было ранее успешно применено нами для изучения и оценки разнообразия хп- и мтДНК ячменя (род *Hordeum*) (Siniauskaya et al., 2020); разработан подход к анализу результатов полногеномного секвенирования органельной ДНК. В настоящем исследовании нами поставлена цель – методом NGS (Illumina) получить новые данные по полногеномным последовательностям органельных ДНК, оценить их уровень изменчивости, взяв в качестве выборки образцы из коллекции раннеспелых и ультрараннеспелых сортов лаборатории нехромосомной наследственности Института генетики и цитологии НАН Беларуси.

Материалы и методы

Выделение ДНК. Материалом исследования служили образцы органельной и тотальной ДНК 46 раннеспелых сортов сои из коллекции Института генетики и цитологии, включая сорта белорусской селекции (полный список приведен в Приложении)[1]. Выбранные сорта являлись хорошими материнскими или отцовскими формами по многолетним данным (неопубликованные данные).

Для выделения органелл по модифицированной методике S.O. Triboush с коллегами (1998) использовали первые молодые листья 7–10-дневных проростков сои. Изолированные органеллы подвергали лизису; органельную ДНК очищали фенол-хлороформом (стандартный протокол). Качество препарата органельной ДНК проверяли методом ПДРФ с последующим анализом в 0.8 % агарозном геле по методике S.O. Triboush с коллегами (1998).

Выделение тотальной ДНК проводили методом фенол-хлороформной экстракции из листьев растений, проращиваемых в теплицах, либо из первых листьев 7-дневных этиолированных проростков сои (стандартный протокол).

Проведение NGS. ДНК органелл исследована методом парноконцевого секвенирования на приборе Illumina MiSeq с использованием набора реагентов MiSeq Reagent Kit v3 (600-cycle) (Illumina Inc., США) в соответствии с рекомендациями производителя. Для приготовления ДНК-библиотеки применяли наборы Illumina DNA Prep (M) Tagmentation (24 samples, IPB) и Nextera XT Index Kit v2 Set A (96 indexes, 384 samples) (Illumina Inc., США) в соответствии с рекомендациями производителя.

Анализ данных NGS. Данные полногеномного секвенирования обработаны согласно разработанному ранее алгоритму (Ермакович и др., 2020), который включал в себя: выравнивание прочтений на референсные последовательности хлоропластного и митохондриального геномов, конвертацию в файлы формата bam и их сортировку, генерацию VCF-файлов и их фильтрацию. Файлы выравнивания прочтений (bam) визуализировались с помощью Unipro Ugene и IGV. В качестве референсных геномов использованы сборки хлоропластного генома сорта Bragg (идентификационный номер GenBank MW357264) и митохондриального генома сорта Aiganhuang (NC020455).

Верификация данных NGS. Для выявленных после анализа данных полногеномного секвенирования полиморфных локусов органельных геномов был сконструирован набор праймеров для верификации сайтов изменчивости секвенированием по Сэнгеру.

Полиморфные участки ДНК амплифицировали отдельно в 15 мкл реакционной смеси, содержащей 30–40 нг ДНК образца, 7.5 мкл 2х смеси реагентов ArtMix (ООО «АртБиоТех», Республика Беларусь), по 1 мкл соответствующих праймеров (5 пмоль/мкл) (ОДО «Праймтех»,

[1] Приложение см. по адресу:
https://vavilovj-icg.ru/download/pict-2026-30/appx14.pdf

Республика Беларусь). ПЦР проводили на амплификаторе C1000 (Bio-Rad Laboratories, Inc., США) по следующему протоколу: 5 мин при 95 °C, затем 30 циклов, каждый из которых включал денатурацию при 95 °C в течение 30 с, отжиг праймеров в течение 30 с и элонгацию при 72 °C в течение 25–80 с, после чего следовала финальная элонгация при 72 °C в течение 5 мин. Продукт ПЦР идентифицировали в 1.5 % агарозном геле.

Секвенирование по Сэнгеру исследуемых образцов проведено на генетическом анализаторе 3500 Applied Biosystems (Thermo Fisher Scientific Inc., США) с применением набора BrilliantDye Terminator v3.1 Cycle Sequencing Kit (NimaGen B.V., Нидерланды) согласно рекомендациям производителя. Результаты секвенирования анализировали с помощью программ Chromas и FinchTV путем сопоставления полученных последовательностей ДНК с референсным геномом *G. max* (сорта Bragg и Aiganhuang) из NCBI GenBank.

Филогенетический анализ выполняли путем создания presence/absence матрицы на основании наличия либо отсутствия отличий в хп- и мтДНК с использованием программы Mesquite (http://www.mesquiteproject.org), расчета матрицы дистанций и построения дендрограммы в программе phylip (Felsenstein, 1989) на основе алгоритма neighbor-joining. Визуализацию дендрограмм осуществляли в онлайн сервисе Iroki (Moore, 2020).

Сборка хлоропластных геномов. Полные нуклеотидные последовательности хпДНК всех 46 сортов, взятых для исследования, получены путем ручного редактирования референсной последовательности хлоропластного генома Bragg с учетом обнаруженных отличий.

Результаты и обсуждение

Проведено NGS геномов хлоропластов и митохондрий 46 сортов культурной сои для изучения их внутривидовой изменчивости. Полученные после NGS прочтения были обработаны и выровнены на референсные последовательности хлоропластного генома сорта Bragg и митохондриального генома сорта Aiganhuang для получения VCF-файлов, содержащих информацию об отличиях хп и мт геномов изученных сортов от референса. В результате сравнительного анализа VCF-файлов идентифицированы полиморфные локусы хлоропластной и митохондриальной ДНК.

Изменчивость генома хлоропластов

В хпДНК обнаружено девять полиморфных сайтов. Межсортовая изменчивость хпДНК сои была представлена пятью однонуклеотидными заменами (SNP) и четырьмя районами микросателлитных повторов. Четыре из пяти идентифицированных SNP находились в кодирующих областях генов *atpB*, *rps4*, *accD* и *rps3*; все они являлись синонимичными. Практически все найденные SNP и повторы располагались в большом однокопийном регионе (LSC) и только один SNP – в малом однокопийном регионе (SSC).

В ряде работ по поиску межвидовой, межродовой изменчивости геномов хлоропластов других культур определены «горячие точки» изменчивости хпДНК, такие как *ccsA-ndhD*, *trnH-psbA*, *ndhG-ndhI*, *rps18-rpl20*, *rps15-ycf1*, *psbZ-trnG-trnS*, *trnK-rps16*, *trnD-trnY*, *trnW-trnP*, *rpl33-rps18*, *petG-trnW*, *atpB-rbcL* и *rpl32-trnL* (Iram et al., 2019; Mehmood et al., 2020). Нами не обнаружены отличия в данных районах, кроме SSR-локуса в межгенной последовательности *atpB-rbcL*, по-видимому, по причине таксономического сходства исследуемых сортов сои.

Из четырех обнаруженных нами полиморфных микросателлитных локусов хпДНК SSR-локус в позиции 6967 (T_{11}–T_{10}) соответствует ранее описанному локусу gmcp4 (Xu et al., 2002). Остальные микросателлиты – впервые описанные и могут быть использованы для выявления генетической изменчивости хлоропластного генома у сои.

Однонуклеотидная замена в позиции 82035 в кодирующей последовательности гена *rps3* затрагивает сайт рестрикции *Cla*I и описана ранее как полиморфный локус, по которому сорта сои дифференцируются на два типа методом ПДРФ (Kanazawa et al., 1998). Этот же SNP в гене *rps3* включен Y. Yue с коллегами в число репрезентативных (позиция 82028 в пластидном геноме Zhonghuang 13). Данная замена, по классификации Y. Yue с коллегами, разделяет все сорта на две группы: первая группа (1848 линий) характеризуется наличием альтернативного аллеля по локусу 82028 (гаплотипы CT1, CT4, CT7, CT22, CT28, CT33 и CT44), вторая группа имеет референсный аллель (остальные 732 линии) (Yue et al., 2023).

В результате проведенного нами исследования получены нуклеотидные последовательности хлоропластных геномов 46 сортов сои, которые опубликованы в NCBI GenBank под идентификационными номерами OQ148707–OQ148730, OR834463–OR834484.

Изменчивость генома митохондрий

Сравнительный анализ последовательностей митохондриального генома, полученных на основании полногеномного секвенирования для 46 сортов сои, позволил обнаружить 15 полиморфных локусов: 3 SNP в некодирующих областях (межгенных районах), 8 полиморфных SSR-локусов, 3 инделя (INDEL – вставка/делеция) и 1 инверсия. Только 2 точки из 15 находились в предполагаемых кодирующих областях: одна мутация в микросателлитном повторе *orf110c* (позиция 294 729 п. н. в референсном геноме сорта Aiganhuang), инверсия – в *orf160b* (позиция 205 470).

Нами выявлена изменчивость в межгенных районах *atp6-1–trnK*, *rps3-orf114a*, *orf100c-orf136b*, *trnD-orf114b*, *orf151-orf261*, *atp9-trnM*, в интроне *nad4* митохондриального генома. Аналогичные данные получены Y. Yue с коллегами в 2023 г., при этом полиморфные позиции были отнесены к числу репрезентативных (маркерных). Это позиции 451199, 284000, 312950, 489992, 256743, 295084 и 484505 п. н. в мтДНК сорта Zhonghuang 13 соответственно вышеуказанным районам. Данные полиморфные локусы позволяют дифференцировать гаплотипы MT1 и MT2 по классификации Y. Yue с коллегами (2023).

Для подтверждения обнаруженных отличий в хп и мт геномной ДНК был сконструирован набор праймеров к сайтам изменчивости геномов и проведено изучение секвенированием по Сэнгеру данных полиморфных локу-

сов у 46 образцов. Праймеры разработаны для изучения наиболее важных маркерных точек изменчивости геномов, по которым происходила дифференциация отдельных гаплотипов хп- и мтДНК сортов сои: позиции 6967, 75657 и 116598 хлоропластного, 100092, 197000, 205470, 248977 митохондриального геномов. Также разработаны комбинации праймеров для исследования наиболее спорных либо трудно интерпретируемых при анализе bam и VCF-файлов районов генома: SSR-локус 51525 хпДНК, индели в позициях 158807 и 321983 мтДНК. Все найденные между сортами различия были подтверждены.

Генетическое разнообразие органельных геномов сои

В соответствии с изменчивостью хлоропластного генома исследованные сорта были разделены на три гаплотипа: 40 сортов (87 %) отнесены к гаплотипу I; 4 сорта (Китросса, Любаша, Оресса и Воронежская 31) – к гаплотипу II; сорта Легенда и Щара – к гаплотипу III. Обнаруженные гаплотипы были обозначены нами как C1–C3.

На основании 16 полиморфных локусов митохондриальной ДНК выделено пять гаплотипов, обозначенных нами как M1–M5. 10 инделей и 4 SNP имеются у сортов Любаша, Китросса, Оресса и Воронежская 31 (гаплотип M2), 1 индель – у родственных сортов Василиса, Амазонка, McCall, Птичь и Сахара (гаплотип M3). Также сорта Optimus (M5) и Злата (M4) отличаются от других сортов инверсией и одним инделем соответственно. Большинство сортов (76 %) отнесено к гаплотипу M1. В табл. 1 и 2 представлены выявленные в органельных геномах отличия и их распределение между различными гаплотипами хп- и мтДНК.

Исходя из полученных нами данных исследование любого из сортов изученной коллекции сои по микросателлитным повторам № 1 и 5 позволяет различать типы хпДНК C1, C2 и C3. Их можно рекомендовать для использования в работах по изучению генетического разнообразия в различных коллекциях сои по геному хлоропластов без привлечения иных хпДНК-маркеров.

Дополнительно для 90 сортов из коллекции проведено секвенирование по Сэнгеру с помощью праймеров, дифференцирующих соответствующие гаплотипы хп- и мтДНК. Согласно обобщенным результатам полногеномного секвенирования и секвенирования по Сэнгеру среди изученных сортов нашей коллекции (так называемого северного экотипа) идентифицируется шесть плазматипов органельной ДНК: C1/M1, C2/M2, C1/M3, C1/M4, C1/M5 и C3/M1. 65.4 % сортов сои (89 из 136 изученных) отно-

Таблица 1. Полиморфные локусы хлоропластной ДНК сои и выявленные на их основе гаплотипы

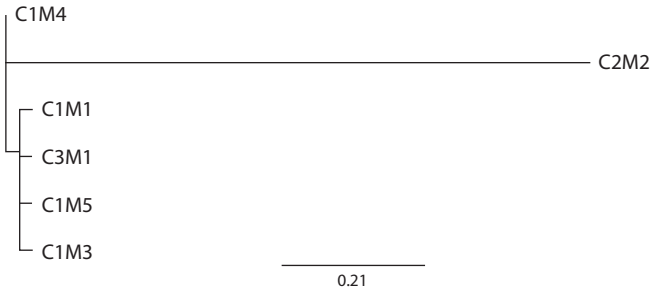
№ п/п	Полиморфная точка хпДНК (относительно MW357264.1)	Гаплотип		
		C1	C2	C3
1	6967 (*rbcL-atpB*)	$(T)_{11}$	$(T)_{11}$	$(T)_{10}$
2	8408 (*atpB*)	G	A	G
3	15507 (*rps4*)	T	A	T
4	38504 (интрон *rpoC1*)	$(A)_4$	$(A)_5$	$(A)_4$
5	51525 (*atpA-trnR*)	$(A)_{19}$	$(A)_{13}$	$(A)_{19}$
6	57873 (*accD*)	C	A	C
7	75657 (интрон *petD*)	$(T)_{14}$	$(T)_{15}$	$(T)_{14}$
8	82035 (*rps3*)	G	T	G
9	116598 (интрон *ndhA*)	T	A	T

Таблица 2. Полиморфные локусы митохондриальной ДНК сои и выявленные на их основе гаплотипы

№ п/п	Полиморфная точка мтДНК (относительно NC020455.1)	Гаплотип				
		M1	M2	M3	M4	M5
1	66782 (*atp6-1-trnK*)	$(GCTTC)_2$	GCTTC	$(GCTTC)_2$	$(GCTTC)_2$	$(GCTTC)_2$
2	238088 (*orf151-orf261*)	$(GCTTC)_2$	GCTTC	$(GCTTC)_2$	$(GCTTC)_2$	$(GCTTC)_2$
3	294604 (*orf105b-orf110c*)	$(GCTTC)_2$	GCTTC	$(GCTTC)_2$	$(GCTTC)_2$	$(GCTTC)_2$
4	321983 (*rps3-orf114a*)	$(TATAA)_3$	$(TATAA)_2$	$(TATAA)_3$	$(TATAA)_3$	$(TATAA)_3$
5	350927 (*orf100c-orf136b*)	T	TATAAA	T	T	T
6	100092 (интрон *nad4*)	GCTCG	$(GCTCG)_2$	GCTCG	GCTCG	GCTCG
7	105562 (*trnD-orf114b*)	G	T	G	G	G
8	197000 (*orf113-orf160b*)	ACCTT	TA	ACCTT	ACCTT	ACCTT
9	221263 (*trnfM-3-orf151*)	C	A	C	C	C
10	238213 (*orf151-orf261*)	TGTAGT	$(TGTAGT)_2$	TGTAGT	TGTAGT	TGTAGT
11	294729 (*orf110c*)	TGTAGT	$(TGTAGT)_2$	TGTAGT	TGTAGT	TGTAGT
12	333070 (*atp9-trnM*)	A	T	A	A	A
13	248977 (*ccmFn-rps4*)	T	TTCAC	T	TTCAC	T
14	158807 (*nad4L-1-nad6*)	TGCCTA	TGCCTA	$(TGCCTA)_2$	TGCCTA	TGCCTA
15	205470 (*orf160b*)	GA	GA	GA	GA	TC

Таблица 3. Дифференциация сортов сои на плазматипы согласно результатам NGS и секвенирования по Сэнгеру

Плазматип	Сорта сои
C1/M1	Припять, Соер 6, Соер 7, Снежок, Антрацит, Аннушка, Мезенка, Ясельда, Adoc, Venus, Gallard, Вилия, Устя, Maple Donovan, Лира, Виктория, Самер 1, Самер 2 и др.
C2/M2	Ореса, Китросса, Любаша, Лотаро, Хэ Нун 95, Dornburger Stamm, Славянка, Волма, Воронежская 31, Violetta, Добрынь, Олвия, Раніца, Рось, Селекта 101, Селена, Памелла Саатбай, Цзинь Юань, Хава, Иней
C1/M3	Сахара, Амазонка, Василиса, УСХИ-6, Tundra, McCall, Птичь, Viola, AC Colombe, НМ 648, Анилин, Барбаро, Белгородская 6, ЕС Композитор, ЕС Сенатор, ЕС Профессор, Люмария, ОАК Купер, РЖТ Руслана, Сиберия, Угра, Суедина, Glasier
C1/M4	Злата, Шарм, Вилана
C1/M5	Optimus
C3/M1	Легенда, Щара



Дендрограмма филогенетических отношений в исследуемой коллекции сортов сои.

сится к гаплотипу C1/M1. 20 (14.7 %) и 23 (16.9 %) сорта имеют плазматипы C2/M2 и C1/M3 соответственно. Три из 136 сортов (2.2 %) относятся к плазматипу C1/M4. Плазматипы C1/M5 и C3/M1 являются наиболее редкими и включают один (0.7 %) и два (1.4 %) сорта соответственно (табл. 3).

По результатам филогенетического анализа обнаруженные плазматипы распределяются на две клады, одну из которых составляют плазматипы C1/M1, C1/M3, C1/M4, C1/M5 и C3/M1, наиболее схожие между собой по последовательности органельных геномов. Гаплотип C2/M2 составляет вторую кладу и имеет наибольшее количество отличий от гаплотипа C1/M1 – 8 полиморфизмов хпДНК и 13 полиморфизмов мтДНК (см. рисунок).

Учитывая, что преобладающее большинство исследованных нами сортов сои имеет плазматип C1/M1, можно считать его основным в коллекции. Плазматипы C1/M3, C1/M4, C1/M5 и C3/M1 – вероятнее всего, производные от него, так как они отличаются от C1/M1 всего одной полиморфной позицией в хлоропластном (C3) либо митохондриальном геномах (M2–M5).

Идентифицированные нами плазматипы C1/M1 и C2/M2 соотносятся с обнаруженными Y. Yue с коллегами в 2023 г. плазматипами CT1/MT1 и CT2/MT2. Согласно их исследованию, данные плазматипы наиболее распространены среди сортов культурной сои – 69.2 и 18.1 % соответственно (Yue et al., 2023). В нашем исследовании мы подтвердили, что плазматип CT1/MT1 (C1/M1 по нашей классификации) – превалирующий и среди сортов белорусской селекции, а также обнаружили редкие отличия, не описанные другими исследователями ранее. В соответствии с этим мы выделили пять дополнительных к основному плазматипов, из которых C1/M3 является более распространенным, чем C2/M2 в нашей коллекции сортов. Это, предположительно, объясняется географо-экологическими особенностями Беларуси, подбором сортов, которые по ряду других признаков наиболее эффективно можно возделывать в Беларуси.

К сожалению, основополагающая гипотеза нашего исследования о возможных особенностях геномов органелл, по которым различаются лучшие материнские и отцовские формы сои, пока не подтверждается, так как и те, и другие сорта имеют сходные гаплотипы хп- и мтДНК. Для выяснения этого вопроса требуется более глубокое изучение тонких механизмов взаимодействия ядра и цитоплазмы.

Заключение

Сравнительный анализ данных NGS у 46 сортов сои раннеспелой группы позволил выявить полиморфные локусы хлоропластной и митохондриальной ДНК, оценить внутривидовое разнообразие ДНК органелл сортов сои разнообразного географического происхождения.

Обнаружено девять отличий в хлоропластном и 15 отличий в митохондриальном геномах культурной сои. Весь спектр изменчивости органельных ДНК сои представлен тремя плазматипами хлоропластной ДНК, пятью плазматипами митохондриальной ДНК. Комбинация изменчивости по обоим органельным геномам позволила выделить шесть плазматипов органельных ДНК и дифференцировать сорта коллекции в соответствии с ними. С помощью разработанных на основании анализа NGS данных ДНК-маркеров к полиморфным локусам геномов органелл проведено изучение изменчивости плазмона у 90 сортов из коллекции лаборатории, отличающихся широким спектром происхождения и чувствительностью к длине дня. 46 полных последовательностей хлоропластной ДНК сои помещено в NCBI GenBank под идентификационными номерами OQ148707–OQ148730, OR834463–OR834484.

Полученные нами результаты и анализ родословных сортов подтверждают данные других исследователей о том, что гаплотипы C1 и M1 – основные среди сортов, возделываемых на территории Северной Америки, Европы и стран СНГ. Плазматип C1/M1 с наибольшей частотой обнаружен и в коллекции исследованных нами образцов сои. Гаплотипы C2 хпДНК и M2 мтДНК являются характерными для современных сортов сои китайской селекции. Интересно, что плазматип C2/M2 достаточно часто встречался в нашей коллекции у ряда образцов, имеющих исконно китайские материнские формы в своей родословной.

Выделение шести плазматипов служит показателем низкого уровня генетического разнообразия в изученной коллекции сортов сои. Исследованная нами группа раннеспелых и ультрараннеспелых сортов представляет собой эволюционно молодую и весьма узкую группу сортов (это сорта северного экотипа). С учетом направленной селекции данных сортов на нейтральность к фотопериодизму низкий уровень генетического разнообразия вполне ожидаем в настолько ограниченной группе линий сои (Rosenzweig et al., 2003).

Выявление источников цитоплазматической изменчивости, сортов с отличными от большинства гаплотипами хпДНК и мтДНК и интрогрессия нового генетического материала в селекционный процесс необходимы для плодотворного развития селекции новых сортов сои для северных широт.

Список литературы / References

Ермакович А.Е., Синявская М.Г., Панкратов В.С., Левданский О.Д., Голоенко И.М., Шимкевич А.М., Луханина Н.В., Давыденко О.Г. Оценка изменчивости хлоропластных и митохондриальных геномов ячменя методом NGS-анализа органельных смесей. Известия Национальной академии наук Беларуси. Серия биологических наук. 2020;65(3):358-364. doi 10.29235/1029-8940-2020-65-3-358-364

[Yermakovich A.E., Siniauskaya M.G., Pankratov V.S., Liaudanski A.D., Halayenka I.M., Shymkevich A.M., Lukhanina N.V., Davydenko O.G. Barley chloroplast and mitochondrial genomes diversity evaluation by NGS of the organelle DNA mixtures. *Vestsi Natsyyanal'nai Akademii Navuk Belarusi. Seriya Biyalagichnykh Navuk = Proceedings of the National Academy of Sciences of Belarus. Biological Series*. 2020;65(3):358-364. doi 10.29235/1029-8940-2020-65-3-358-364 (in Russian)]

Синявская М.Г., Вигуро А.В., Розенцвейг В.Е., Голоенко Д.В. Изменчивость хлоропластного генома у сои. В: Международная научная конференция «Молекулярная генетика, геномика и биотехнология» (Минск, Беларусь, 24-26 ноября 2004 г.). Минск, 2004;346-347

[Siniauskaya M.G., Viguro A.V., Rosenzweig V.E., Goloenko D.V. The variability of chloroplast genome of soybean. In: International Scientific Conference on Molecular Genetics, Genomics and Biotechnology (November 24-26, 2004, Minsk, Belarus). Minsk, 2004; 346-347 (in Russian)]

Abe J., Hasegawa A., Fukushi H., Mikami T., Ohara M., Shimamoto Y. Introgression between wild and cultivated soybeans of Japan revealed by RFLP analysis for chloroplast DNAs. *Econ Bot*. 1999; 53:285-291. doi 10.1007/BF02866640

Aksyonova E.A., Rosenzweig V.E., Goloenko D.V., Davydenko O.G. The analysis of belarusian soybean cultivars diversity using nuclear and chloroplast SSR markers. In: Computational Phylogenetics and Molecular Systematics "CPMS 2007". Moskow, 2007

Fang C., Ma Y., Yuan L., Wang Z., Yang R., Zhou Z., Liu T., Tian Z. Chloroplast DNA underwent independent selection from nuclear genes during soybean domestication and improvement. *J Genet Genomics*. 2016;43(4):217-221. doi 10.1016/j.jgg.2016.01.005

Felsenstein J. PHYLIP: Phylogeny Inference Package (Version 3.2). *Cladistics*. 1989;5:164-166.

Gualberto J., Newton K. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol*. 2017;68:224-252. doi 10.1146/annurev-arplant-043015-112232

Hu Y., Sun Y., Zhu Q., Fan L., Li J. Poaceae chloroplast genome sequencing: great leap forward in recent ten years. *Curr Genomics*. 2022;23(6):369-384. doi 10.2174/1389202924666221201140603

Iram S., Hayat M.Q., Tahir M., Gul A., Abdullah, Ahmed I. Chloroplast genome sequence of *Artemisia scoparia*: comparative analyses and screening of mutational hotspots. *Plants*. 2019;8(11):476. doi 10.3390/plants8110476

Jeong S.C., Moon J.K., Park S.K., Kim M.S., Lee K., Lee S.R., Jeong N., Choi M.S., Kim N., Kang S.T., Park E. Genetic diversity patterns and domestication origin of soybean. *Theor Appl Genet*. 2019;132(4):1179-1193. doi 10.1007/s00122-018-3271-7

Johnston I.G. Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. *Mol Plant*. 2019;12(6):764-783. doi 10.1016/j.molp.2018.11.002

Kanazawa A., Tozuka A., Shimamoto Y. Sequence variation of chloroplast DNA that involves *Eco*RI and *Cla*I restriction site polymorphisms in soybean. *Genes Genet Syst*. 1998;73(2):111-119. doi 10.1266/ggs.73.111

Mehmood F., Shahzadi I., Waseem S., Mirza B., Ahmed I., Waheed M.T. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics*. 2020;112(1):581-591. doi 10.1016/j.ygeno.2019.04.010

Moore R.M., Harrison A.O., McAllister S.M., Polson S.W., Wommack K.E. Iroki: automatic customization and visualization of phylogenetic trees. *PeerJ*. 2020;8:e8584. doi 10.7717/peerj.8584

Powell W., Morgante M., Andre C., McNicol J.W., Machray G.C., Doyle J.J., Tingey S.V., Rafalski J.A. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol.* 1995;5(9):1023-1029. doi 10.1016/s0960-9822(95)00206-5

Rosenzweig V.E., Goloenko D.V., Davydenko O.G., Shablinskaya O.V., Swiecicki W. Breeding strategies for early soybeans in Belarus. *Plant Breeding*. 2003;122(5):456-458. doi 10.1046/j.1439-0523.2003.00874.x

Shimamoto Y. Polymorphism and phylogeny of soybean based on chloroplast and mitochondrial DNA analysis. *Jpn Agric Res Q*. 2001; 35(2):79-84. doi 10.6090/jarq.35.79

Shimamoto Y., Fukushi H., Abe J., Kanazawa A., Gai J., Gao Z., Xu D. RFLPs of chloroplast and mitochondrial DNA in wild soybean, *Glycine soja*, growing in China. *Genet Resour Crop Evol*. 1998;45: 433-439. doi 10.1023/A:1008693603526

Shoemaker R.C., Hatfield P.M., Palmer R.G., Atherly A.G. Chloroplast DNA variation in the genus *Glycine* subgenus *Soja*. *J Hered*. 1986;77(1):26-30. doi 10.1093/oxfordjournals.jhered.a110161

Siniauskaya M.G., Makarevich A.M., Goloenko I.M., Pankratov V.S., Liaudanski A.D., Danilenko N.G., Lukhanina N.V., Shimkevich A.M., Davydenko O.G. The study of organelle DNA variability in alloplasmic barley lines in the NGS era. *Vavilov J Genet Breed*. 2020;24(1):12-19. doi 10.18699/VJ19.589

Triboush S.O., Danilenko N.G., Davydenko O.G. A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. *Plant Mol Biol Rep*. 1998;16:183-189. doi 10.1023/A:1007487806583

Tsunewaki K., Mori N., Takumi S. Experimental evolutionary studies on the genetic autonomy of the cytoplasmic genome "plasmon" in the *Triticum* (wheat)–*Aegilops* complex. *Proc Natl Acad Sci USA*. 2019;116(8):3082-3090. doi 10.1073/pnas.1817037116

Xu D., Abe J., Gai J., Shimamoto Y. Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theor Appl Genet*. 2002;105:645-653. doi 10.1007/s00122-002-0972-7

Yue Y., Li J., Sun X., Li Z., Jiang B. Polymorphism analysis of the chloroplast and mitochondrial genomes in soybean. *BMC Plant Biol*. 2023;23(1):15. doi 10.1186/s12870-022-04028-3

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 18.06.2025. После доработки 20.09.2025. Принята к публикации 24.09.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):212-221

doi 10.18699/vjgb-26-24

Анализ экспрессии генов *PR-1*, *PR-2*, *PR-10*, хитиназ и хитиназа-подобных белков в корнях гороха под влиянием салициловой кислоты и метилжасмоната

А.М. Егорова

Казанский институт биохимии и биофизики – структурное подразделение Федерального исследовательского центра «Казанский научный центр Российской академии наук», Казань, Россия
egorova@kibb.knc.ru

Аннотация. В реализации защитного ответа растений принимают участие стрессовые фитогормоны – салициловая и жасмоновая кислоты. Повышение содержания этих соединений при инфицировании растений патогенами приводит к активации сигнальных путей, в конечном итоге изменяющих экспрессию генов и синтез белков, среди которых выделяется группа индуцируемых, связанных с патогенезом белков (PR). Для оценки активации этих сигнальных путей используются фитогормон-зависимые, так называемые маркерные гены. В настоящей работе выявлены и охарактеризованы экспрессирующиеся в корнях гороха гены маркерных для салицилатного сигналинга белков PR-1. Проанализировано влияние салициловой кислоты и метилжасмоната на их экспрессию. Показано, что в корнях гороха ген *Psat1g156240*, кодирующий PR-1, входит в число наиболее значительно экспрессирующихся генов. Салициловая кислота не вызывала изменение экспрессии этого гена, в то же время он индуцировался метилжасмонатом через 24 ч. Анализ экспрессии других генов, кодирующих белки PR-1, показал, что салицилат не влиял на их экспрессию через 24 и 72 ч. При анализе экспрессии генов хитиназ и хитиназа-подобных белков первые не проявляли специфичность ответа на действие салицилата и метилжасмоната, за исключением гена *Psat1g131280*, экспрессия которого повышалась через 24 и 72 ч при действии метилжасмоната. Гены хитиназа-подобных белков, *Psat1g147600*, *Psat1g147560*, *Psat1g149120*, практически не экспрессировались в корнях гороха контрольного варианта и специфично индуцировались салициловой кислотой. Гены β-1,3-глюканаз не индуцировались исследованными фитогормонами в корнях. Полученные результаты позволили определить конкретные гены, кодирующие хитиназа-подобные белки, экспрессия которых индуцируется салициловой кислотой. Эти гены могут быть использованы для оценки активации салицилат-зависимого сигнального пути в корнях гороха.

Ключевые слова: маркерный ген; PR-белки; сигналинг; салициловая кислота; жасмоновая кислота; хитиназа-подобные белки; фитоиммунитет

Для цитирования: Егорова А.М. Анализ экспрессии генов *PR-1*, *PR-2*, *PR-10*, хитиназ и хитиназа-подобных белков в корнях гороха под влиянием салициловой кислоты и метилжасмоната. *Вавиловский журнал генетики и селекции*. 2026;30(2):212-221. doi 10.18699/vjgb-26-24

Финансирование. Исследование выполнено за счет гранта РНФ и Академии наук Республики Татарстан по проекту № 24-26-20123.

Analysis of *PR-1*, *PR-2*, *PR-10*, chitinases and chitinase-like proteins genes expression in pea roots under the action of salicylic acid and methyl jasmonate

A.M. Egorova

Kazan Institute of Biochemistry and Biophysics of Kazan Scientific Center of the Russian Academy of Sciences, Kazan, Russia
egorova@kibb.knc.ru

Abstract. Stress phytohormones – salicylic and jasmonic acids – participate in the plant defense response. The increase in the content of these compounds during plant infection by pathogens leads to the activation of signaling pathways, ultimately resulting in changes in gene expression and protein synthesis, including a group of pathogenesis-related (PR) proteins. Phytohormone-dependent so-called marker genes, are used to assess the activation of these signaling pathways. In this study, *PR-1* genes, which are markers for salicylic acid signaling and expressed in pea roots, were identified and characterized, and the effects of salicylic acid and methyl jasmonate on their expression were analyzed. It was shown that in pea roots, PR-1, encoded by the *Psat1g156240* gene, is among the most significantly expressed genes in the control. Salicylic acid did not cause a change in the expression of this gene; however, it was induced by methyl jasmonate after 24 h. Analysis of the expression of other genes encoding PR-1 proteins showed that

© Егорова А.М., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

salicylate had no effect on their expression after 24 and 72 h. Analysis of the expression of genes encoding chitinases and chitinase-like proteins showed that the former do not exhibit specificity in response to the salicylate and methyl jasmonate, except for the *Psat1g131280*, the expression of which increased after 24 and 72 h of treatment with methyl jasmonate. Genes *Psat1g147600*, *Psat1g147560*, *Psat1g149120* encoding chitinase-like proteins, were barely expressed in pea roots in control, and were specifically induced by salicylic acid. β-1,3-glucanase genes were not induced in roots by the studied phytohormones. The obtained results allowed to reveal specific genes including chitinase-like proteins, the expression of which is salicylate-inducible. These genes can be used for the assessment of the activation of the salicylate-dependent signaling pathway in pea roots.

Key words: marker gene; PR proteins; signaling; salicylic acid; jasmonic acid, chitinase-like proteins; phytoimmunity

For citation: Egorova A.M. Analysis of *PR-1*, *PR-2*, *PR-10*, chitinases and chitinase-like proteins genes expression in pea roots under the action of salicylic acid and methyl jasmonate. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):212-221. doi 10.18699/vjgb-26-24

Введение

В течение жизни растения постоянно подвергаются воздействию патогенов и неблагоприятных факторов среды. Фитоиммунитет позволяет им справляться с атакой патогенов, преодолевать эти неблагоприятные условия и выживать. Кроме того, растения выработали широкий ряд защитных механизмов в ответ на действие стрессоров и атаку патогенов, включая патоген-специфичные гены устойчивости, антимикробные соединения, такие как фитоалексины, индуцируемые патогенами защитные белки (de Wit, 2007).

Салициловая (СК) и жасмоновая кислоты (ЖК) являются фитогормонами, участвующими в реализации защитного ответа растений (Ding P., Ding Y., 2020; Wang et al., 2021). Считается, что СК играет роль в защите от биотрофных и гемибиотрофных патогенов, питающихся живыми тканями растений, тогда как ЖК – главным образом в защите от некротрофных патогенов и травоядных насекомых (Glazebrook, 2005). Атака растений патогенами приводит к образованию сигнальных молекул и активации сигнальных путей, среди которых ключевым является один из вышеупомянутых факторов иммунитета растений (Ding L.N. et al., 2022). Активация сигнального каскада вызывает изменение экспрессии генов и синтеза белков. Эти белки либо сами обладают антипатогенными свойствами, либо участвуют в синтезе защитных соединений.

При изучении ответа растений на действие различных факторов часто необходима оценка активации того или иного сигнального пути для выяснения особенностей реализации этого ответа. Для этого, как правило, используют анализ активации экспрессии так называемых маркерных генов. В основном эти гены охарактеризованы на модельных растениях, таких как арабидопсис, табак, рис, и преимущественно на надземных органах. Но не всегда результаты, полученные на модельных растениях, могут быть безоговорочно распространены на другие виды, в том числе важные сельскохозяйственные культуры. Так, на примере *Brassica rapa* было показано, что широко используемые маркерные гены салицилат- и жасмонат/этилен-зависимых сигнальных путей не проявляют специфичность экспрессии в ответ на действие данных гормонов и уровень их экспрессии различается в побегах и корнях (Papadopoulou et al., 2018). В другой работе на растениях *Capsicum annuum* продемонстрировано, что маркерный для салицилатного сигналинга ген *PR-1* является более подходящим кандидатом для МеЖК/этилен-зависимого пути, но не для СК-пути (Perez-Aranda et al., 2024). В растениях *Brachypodium distachyon* анализ экспрессии генов *BdPR1-4* и *BdPR1-5* регулировался как СК, так и МеЖК (Kouzai et al., 2016).

В постгеномную эру благодаря развитию омиксных технологий стало возможно анализировать большие массивы данных, включая данные по экспрессии генов и синтезу белков. Имеющийся пул транскриптомных данных позволяет проанализировать полный спектр экспрессирующихся в конкретном органе генов, как конститутивных, так и изменяющихся под влиянием различных условий. Выявление и идентификация таких генов, играющих роль в адаптации растений, может быть полезна при селекции устойчивых сортов (Tyagi et al., 2022).

В реализации защитного ответа растений важную роль играют белки, индуцируемые патогенами и обладающие антипатогенным действием (PR-белки). В зависимости от особенностей их функционирования они подразделяются на 17 классов. Многие из PR-белков, наряду с индукцией патогенами, активируются стрессовыми фитогормонами (van Loon et al., 2006). Салицилат-зависимые гены имеют большое значение в реализации защитного ответа растений (Maleck et al., 2000; Blanco et al., 2009). Одним из наиболее известных примеров маркерного гена для салицилат-зависимого сигналинга является *PR-1* (Breen et al., 2017). Салициловой кислотой регулируются и другие представители *PR*-генов – β-1,3-глюканазы (*PR-2*) (Hennig et al., 1993; Jayaraj et al., 2004) и хитиназы (*PR-3*, *4*, *8*, *11*) (Davis et al., 2002; Тарчевский и др., 2010).

В последние годы большое внимание исследователи уделяют процессам, происходящим в ризосфере, взаимодействию корней растений с патогенами, симбиотическими бактериями и другими обитателями почвы. Часто от этих взаимоотношений зависит урожай, поскольку как поражение посевов патогенами корней, так и эффективность образования симбиозов (в отношении бобовых) с клубеньковыми бактериями оказывает значительное влияние на их рост и урожайность (Solomon et al., 2024). Кроме того, корни играют важную роль в жизни растений не только в качестве «добытчиков» воды и минеральных веществ, но и как участники защитного ответа. Некоторые соединения, обладающие антипатогенными свойствами, такие как алкалоиды и терпеноиды, синтезируются в корнях и транспортируются в надземные органы. В дополне-

ние к этому корни реализуют свои специфичные защитные механизмы (Erb, 2012).

Ранее в своей работе мы столкнулись с различием протеомного ответа листьев и корней гороха на действие экзогенной СК, что также может отражать разницу в стратегии их защитного ответа. В листьях гороха СК приводила к значительному повышению содержания ряда PR-белков – хитиназ, β-1,3-глюканаз (Тарчевский и др., 2010), тогда как в корнях наиболее существенно повышалось содержание хитиназа-подобных белков, относящихся к гликозидгидролазам семейства 18 (GH18) (Egorova, 2025). Естественно, что разные условия среды роста надземных органов и корней оказывают влияние на особенности их ответа на то или иное воздействие. Поэтому выявление особенностей ответа корней, в том числе на атаку патогенов, важно, поскольку может быть использовано при селекции устойчивых к почвенным патогенам сортов и для разработки агробиотехнологических подходов для защиты посевов конкретных культур от почвенных патогенов. При изучении реализации защитного ответа корней и оценки активации того или иного сигнального пути необходимо использовать адекватные маркерные гены. В этой связи поиск, характеристика и особенности экспрессии этих генов, специфичность в отношении того или иного сигнального соединения представляются актуальной задачей.

В настоящей работе были выявлены и охарактеризованы гены *PR-1*, хитиназ и хитиназа-подобных белков, β-1,3-глюканаз и *PR-10* в корнях *Pisum sativum*. Проанализирован уровень экспрессии этих генов в контроле, а также при действии СК и метилжасмоната (МеЖК). Экспрессию генов оценивали методом сравнительного транскриптомного анализа контрольных и обработанных СК и МеЖК корней гороха.

Материал и методы

Выращивание и обработка растений. Семена гороха (*P. sativum* L.) сорта Тан стерилизовали в 0.05 % растворе перманганата калия в течение 20 мин, затем тщательно промывали дистиллированной водой. Проращивали на влажной фильтровальной бумаге в течение 3 сут в темноте. Проростки выращивали на ¼ среды Хогленда–Арнона при 16-часовом световом дне, 25 °С, 8-дневные проростки использовали для дальнейшего эксперимента. В опытном варианте в среду роста растений добавляли СК в концентрации 50 мкМ (Sigma, США) или МеЖК в концентрации 20 мкМ (Serva, Германия), которые растворяли в этиловом спирте (500 мкл). Контролем служили проростки, выращенные на ¼ среды Хогленда–Арнона, в которую добавляли рассчитанное с учетом разбавления количество этилового спирта. Через 24 и 72 ч корни проростков гороха фиксировали в жидком азоте.

Транскриптомный анализ. Тотальную РНК выделяли из 200 мкг корней гороха с использованием реактива ExtractRNA («Евроген», Россия), по рекомендациям производителя. Концентрацию и качество РНК определяли с помощью нанофотометра Implen NanoPhotometer (Implen, GmbH) и в 1 % агарозном гель-электрофорезе. Получение библиотек кДНК, высокоэффективное секвенирование и биоинформационный анализ полученных транскриптомных данных были выполнены компанией Novogene (Пекин, КНР).

Качество РНК было проверено на системе Agilent Nano 6000 Assay Kit of the Bioanalyzer 2100 (Agilent, США). мРНК была очищена от тотальной РНК с помощью поли-Т олигоконъюгированных магнитных частиц. Фрагментацию проводили с использованием двухвалентных катионов при повышенной температуре в X5 First Strand Synthesis Reaction Buffer (Thermo Fisher, США). Первая цепь кДНК была синтезирована со случайным гексапраймером и M-MuLV обратной транскриптазой (Thermo Fhisher, США). Вторую цепь кДНК синтезировали при участии ДНК полимеразы I и РНКазы H. После аденилирования 3′-концов фрагментов ДНК адаптер со структурой шпильки был лигирован для подготовки к гибридизации. Для отбора фрагментов кДНК предпочтительной длины (370–420 п. н.) фрагменты библиотеки были очищены с помощью системы AMPure XP system (Beckman Coulter, Беверли, США). После проведения ПЦР с универсальными праймерами продукты были очищены и качество библиотеки оценено на системе Agilent Bioanalyzer 2100 (Agilent, США).

После генерации кластеров cBot с набором TruSeq PE Cluster Kit v3-cBot-HS (Illumina, США) библиотека была секвенирована на платформе Illumina NovaSeq 6000 (Illumina, США) и получены парные прочтения длиной 150 п. н. Сырые данные были обработаны на языке perl. На этом этапе после удаления прочтений, содержащих адаптеры, поли-N-содержащие последовательности и прочтения плохого качества, были получены чистые данные и просчитано содержание Q20, Q30 и GC. Для последующего анализа использовались очищенные данные высокого качества. В качестве референсного был выбран геном *Pisum sativum*_v1a (Kreplak et al., 2019) (https://urgi.versailles.inra.fr/Species/Pisum), индексирование референсного генома и выравнивание полученных чистых прочтений с ним выполнено с помощью Hisat2 v2.0.5.

Подсчет числа прочтений, сопоставленных с каждым геном, проводили в featureCounts v1.5.0-p3. FPKM каждого гена был просчитан на основе длины гена и количества прочтений, совпавших с этим геном. FPKM (число прочтений на тысячу нуклеотидов на миллион картированных прочтений) учитывает влияние глубины секвенирования и длины гена на количество прочтений одновременно (Trapnell et al., 2010).

Анализ дифференциальной экспрессии генов осуществляли в пакете DESeq2 R (1.20.0) (Love et al., 2014). Полученные значения p были скорректированы с помощью подхода Бенджамини–Хохберга для контроля частоты ложных совпадений. Гены со значением $p \leq 0.05$ и значением log2 FC ≥ 1 считались дифференциально экспрессируемыми. В каждом варианте было взято минимум три биологические повторности.

Экспрессию генов в данной работе оценивали с помощью значения FPKM, поскольку не все проанализи-

рованные гены входили в число дифференциально экспрессирующихся. На рисунках представлены средние значения экспрессии, выраженные в FPKM, и стандартные отклонения. Дифференциально экспрессирующиеся гены отмечены звездочкой (*). Данные, полученные в результате секвенирования, доступны в NCBI (BioRroject PRJNA1358375).

Биоинформационный анализ. Поиск BLAST по предсказанным последовательностям генов и белков гороха проводили на основе данных генома гороха *P. sativum*_v.1a (Kreplak et al., 2019) в базе EnsemblePlants (https://plants.ensembl.org/Pisum_sativum/Info/Index). Множественное выравнивание белковых последовательностей осуществляли в программе MEGA v. 10.2.6 (Kumar et al., 2018). Полные аминокислотные последовательности белков для множественного выравнивания были получены из баз данных UniProt и NCBI. Доменная организация белков проанализирована и предсказана с использованием Conserved Domains Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Результаты

Проведен транскриптомный анализ трех групп корней проростков гороха: контрольные растения и растения под влиянием СК (24 и 72 ч) и МеЖК (24 и 72 ч). Каждый вариант содержал от 36 до 57 млн прочтений. Количество, качество и характеристики полученных данных приведены в Приложении[1].

Выравнивание полученных прочтений с референсным геномом гороха *P. sativum*_v1a (Kreplak et al., 2019) позволило выявить около 20 тыс. транскриптов. Через 24 ч СК вызывала изменение экспрессии 1476 генов: у 635 генов экспрессия повышалась, у 841 – понижалась. Через 72 ч действия СК изменялась экспрессия 2420 генов: экспрессия 977 генов повышалась, 1443 – понижалась. При действии МеЖК через 24 ч изменялась экспрессия 3384 генов: у 1546 повышалась и у 1838 генов понижалась. Через 72 ч МеЖК приводил к изменению экспрессии 5462 генов: экспрессия 2218 генов повышалась, 3244 – понижалась.

Нами выполнен анализ экспрессии в корнях гороха генов PR-1, хитиназ и хитиназа-подобных белков, β-1,3-глюканаз и PR-10 для оценки возможности их использования в качестве маркерных для салицилатного сигналинга и выявления конкретных изоформ, индуцируемых именно СК. Выбор генов этих групп белков обусловлен тем, что для PR-1, хитиназ, β-1,3-глюканаз и PR-10 показана индукция СК (Davis et al., 2002; Li et al., 2005; Spoel et al., 2007; Xu et al., 2024) и антипатогенная активность (van Loon et al., 2006), тогда как хитиназа-подобные белки значительно индуцируются в корнях гороха СК (Egorova, 2025). Оценка экспрессии выявленных генов проводилась под влиянием СК и МеЖК, поскольку они участвуют в защитном ответе растений и определяют характер реализуемой защитной стратегии.

[1] Приложение см. по адресу: https://vavilovj-icg.ru/download/pict-2026-30/appx15.xlsx

Экспрессия генов *PR-1* в корнях гороха

PR-1 относится к семейству обогащенных цистеином секреторных белков (CAP). В геноме гороха к этому семейству предположительно относится 12 последовательностей. Поиск BLAST в геномной базе данных гороха с использованием охарактеризованных последовательностей PR-1 из табака (P09042) и арабидопсиса (AT2G14610) позволил выявить сходные с ними последовательности. Анализ экспрессии выявленных в геноме гороха последовательностей PR-1 генов показал, что в корнях гороха контрольного варианта экспрессируются девять генов CAP (рис. 1, *а*). Экспрессия генов *Psat7g104840*, *Psat6g069320* и *Psat1g211920* была низкой (<10 FPKM) и далее в этой работе не обсуждается.

Для характеристики белков гороха было проведено множественное выравнивание аминокислотных последовательностей белков, кодируемых экспрессирующимися в корнях гороха генами *PR-1*, с охарактеризованными белками из арабидопсиса и табака. Так, белок, кодируемый наиболее значительно экспрессирующимся геном *Psat1g156240*, имеет 60 % идентичности аминокислотной последовательности с белком PR-1 P09042 *Nicotiana tabacum* и AtPR1 (AT2G14610) арабидопсиса. Белок, кодируемый геном *Psat1g213200*, имеет 66.4 % идентичности с PR1A_TOBAC (P08299) и 65 % с AtPR1. CAP белки характеризуются наличием консервативных CAP мотивов (CAP1–CAP4) и CAPE пептида (Han et al., 2023). Два PR-1 белка, кодируемых генами *Psat1g211480* и *Psat0s4615g0040*, имеют усеченный CAPE пептид и не содержат CAP2 мотива (см. рис. 1, *б*), тогда как другие консервативные мотивы присутствуют в их последовательности.

В контрольном варианте уровень экспрессии генов *Psat1g156240* и *Psat1g213200* самый высокий по сравнению с другими генами *PR-1*. Было оценено влияние СК и МеЖК на экспрессию генов *PR-1* (см. рис. 1, *а*). Экспрессия гена *Psat1g156240* регулировалась МеЖК и повышалась через 24 ч. СК не вызывала достоверного изменения экспрессии этого гена. Экспрессия гена *Psat1g213200* достоверно не изменялась под влиянием СК, но снижалась под влиянием МеЖК (см. рис. 1, *в*). Уровень экспрессии генов *Psat1g211480* и *Psat0s4615g0040* в контроле был ниже по сравнению с двумя предыдущими генами и не изменялся ни при действии СК, ни при действии МеЖК в изученных временных точках. Экспрессия других CAP генов – *Psat0ss9363g0160*, *Psat4g090120*, *Psat2g163240*, *Psat1g211840* и *Psat0ss9363g0200* – в контрольном варианте была низкой, не изменялась под влиянием СК, но снижалась при действии МеЖК (см. рис. 1, *а*).

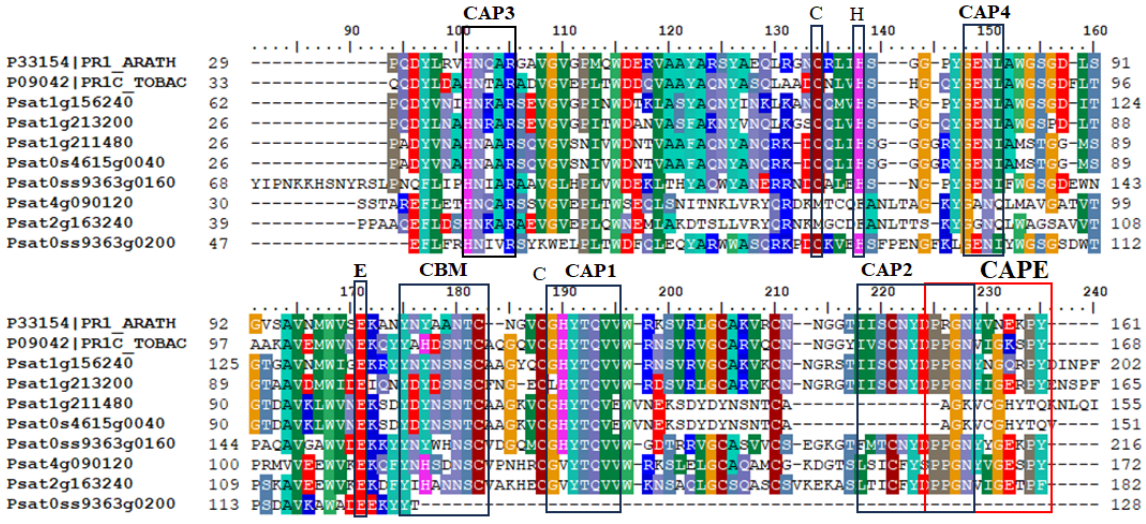
Экспрессия генов хитиназ и хитиназа-подобных белков в корнях гороха

Хитиназы (EC 3.2.1.14) относятся к гликозил-гидролазам семейств 18 (GH18) и 19 (GH19). Гены хитиназ *Psat3g016800*, *Psat1g084840*, *Psat2g145680*, *Psat3g016400*, *Psat5g291720*, *Psat2g128240*, *Psat4g175600*, *Psat4g175720*, *Psat0s794g0040*, *Psat5g256760*, *Psat5g298520*,

а

Ген	Контр	СК 24 ч	СК 72 ч	МеЖК 24 ч	МеЖК 72 ч
Psat1g156240	2473	1929	3891	8786*	7743
Psat1g213200	984	493	1423	231*	86*
Psat1g211480	10	145	101	8	14
Psat0s4615g0040	1	26	19	1	1
Psat0ss9363g0160	49	18	72	9	11*
Psat4g090120	27	5	34	5*	9
Psat2g163240	13	11	18	5	3*
Psat1g211840	11	6	22	1*	1
Psat0ss9363g0200	13	8	15	5	2*

б



в

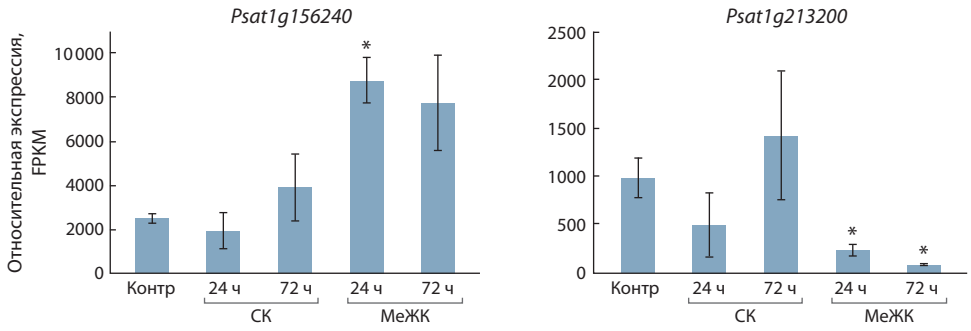


Рис. 1. Обогащенные цистеином секреторные белки, гены которых экспрессируются в корнях гороха. *а* – экспрессирующиеся в корнях гороха CAP белки (PF 00188) в контрольном варианте (Контр) и под влиянием СК (50 мкм) и МеЖК (20 мкМ) в течение 24 и 72 ч. Цветовая шкала представляет наиболее сильно (красный цвет) и слабо (синий) экспрессируемые гены, уровень экспрессии выражен в FPKM; *б* – множественное выравнивание аминокислотных последовательностей PR-1 белков арабидопсиса (P33154), табака (P09042) и последовательностей CAP белков, экспрессирующихся в корнях гороха. Черными рамками показаны консервативные CAP1–CAP4 домены, консервативные цистеин, гистидин, глютаминовая кислота, кавеолин-связывающий мотив (CBM), ответственный за связывание со стеролами. CAPE пептид выделен красной рамкой; *в* – дифференциально-экспрессирующиеся CAP гены в корнях гороха при действии СК и МеЖК.

Здесь и на рис. 2–4: представлены средние значения ± стандартные отклонения трех-пяти биологических повторностей; * $p < 0.05$ и значение log2 FC ≥ 1 – значимые различия экспрессии генов в вариантах СК и МеЖК по сравнению с контролем.

Psat1g169040, *Psat4g171240*, *Psat4g177760*, *Psat5g020240*, *Psat6g166120*, *Psat6g194800* не экспрессируются в корнях, или уровень их экспрессии низкий (<10 FPKM), в связи с чем далее в данной работе не обсуждаются (см. Приложение).

Анализ транскриптомных данных показал, что в корнях гороха экспрессируются шесть генов хитиназ, кодирующих белки семейства GH18, и два гена семейства GH19, которые, в свою очередь, принадлежат к разным группам PR-белков (рис. 2, *а*). Все экспрессирующиеся в корнях гороха хитиназы GH19 содержат лизоцим-подобный домен, включая наиболее значительно экспрессируемый ген *Psat5g000800* (см. рис. 2, *а*). Ни СК, ни МеЖК не вызывали достоверного изменения экспрессии этого гена.

а

	Ген	Контр	СК 24 ч	СК 72 ч	МеЖК 24 ч	МеЖК 72 ч	Суперсемейство	pfam	Консервативный домен	Группа PR
GH19	*Psat5g000800*	586	1500	827	721	1367	Lysozyme-like domain	PF00182	ChtBD1_GH19	PR-3
	Psat4g103520	9	23	26	13	3	Lysozyme-like domain	PF00182	Glyco_hydro_19	PR-3
GH18	*Psat3g005200*	57	119	73	23	11*	Glycoside hydrolase	PF00704	GH18_hevamine_XipI_class_III	PR-8
	Psat6g194720	35	31	49	37	36	Glycoside hydrolase	PF00704	GH18_hevamine_XipI_class_III	PR-8
	Psat6g194680	7	29	132	22	64	Glycoside hydrolase	PF00704	GH18_hevamine_XipI_class_III	PR-8
	Psat1g131280	13	16	152	2805*	1482*	Glycoside hydrolase	PF00704	GH18_hevamine_XipI_class_III	PR-8
	Psat0s5923g0040	20	35*	20	38*	25	Glycoside hydrolase	PF00704	GH18_plant_chitinase_class_V	PR-11
	Psat3g007040	142	125	119	157	118	Endochitinase-like	PF00187	Barwin domain	PR-4

б

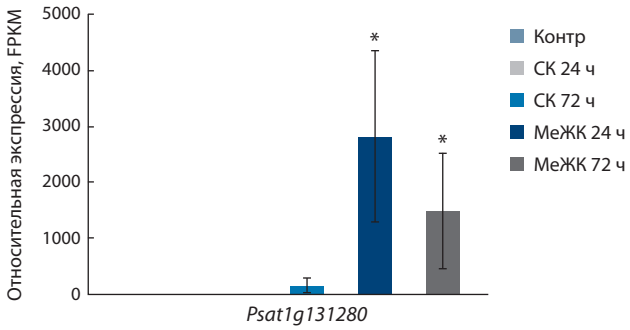


Рис. 2. Влияние салициловой кислоты и метилжасмоната на экспрессию генов хитиназ в корнях гороха. *а* – экспрессирующиеся в корнях гороха гены хитиназ GH18 и GH19 в контрольном варианте (Контр) и под влиянием СК (50 мкм) и МеЖК (20 мкМ) в течение 24 и 72 ч. Цветовая шкала представляет наиболее сильно (красный цвет) и слабо (синий) экспрессируемые гены, уровень экспрессии выражен в FPKM; *б* – экспрессия гена *Psat1g131280* в корнях гороха при действии СК и МеЖК.

Экспрессия второго гена представителей хитиназ GH19, *Psat4g103520*, значительно ниже по сравнению с уровнем экспрессии первого гена и тоже не изменялась под влиянием СК и МеЖК.

Уровень экспрессии генов семейства GH18 в контрольном варианте ниже по сравнению с представителями GH19. Хитиназы группы PR-3 содержат гевеин-подобный домен. Экспрессия генов *Psat3g005200* и *Psat6g194720*, кодирующих кислые эндохитиназы, не изменялась при действии салицилата, но экспрессия *Psat3g005200* значимо снижалась под влиянием МеЖК через 72 ч. Экспрессия гена *Psat6g194680* не изменялась ни при действии СК, ни МеЖК (см. рис. 2, *а*). Экспрессия гена *Psat1g131280*, аннотированного как гевеамин А, существенно повышалась под влиянием МеЖК через 24 и 72 ч (см. рис. 2). Экспрессия *Psat0s5923g0040*, принадлежащего GH18, незначительно повышалась при воздействии СК и МеЖК через 24 ч (см. рис. 2, *а*).

Другой представитель GH18 семейства белков, продукт гена *Psat3g007040*, наряду с хитин-связывающим доменом содержит домен барвина. Белки с этим доменом относят к PR-4 белкам. Ни СК, ни МеЖК не оказывали влияния на экспрессию этого представителя хитиназ гороха (см. рис. 2, *а*).

В предыдущей работе (Egorova, 2025) нами была показана значительная индукция в корнях гороха салицилатом группы белков, относящихся к семейству GH18. Идентифицированные белки кодируются генами *Psat1g147560*, *Psat1g147600*, *Psat1g149120* и *Psat1g148600*. Анализ хитиназной активности этих белков показал, что они не способны разрушать хитин и относятся к хитиназа-подобным белкам. Проведенный транскриптомный анализ влияния на экспрессию этих генов СК позволил выявить значительное повышение экспрессии всех четырех генов при действии СК, тогда как в контроле экспрессия была на низком уровне (Egorova, 2025). В настоящей работе было проанализировано действие МеЖК, который не оказывал влияния на экспрессию всех четырех проанализированных генов (рис. 3).

Экспрессия генов *PR-2* и *PR-10* в корнях гороха

В листьях гороха ранее нами была выявлена индукция β-1,3-глюканаз (PR-2) и белка устойчивости к болезни (PR-10) при действии СК (Тарчевский и др., 2010). Протеомный анализ белков корней гороха, обработанных СК, показал повышение содержания белков устойчивости к болезни и АБК-зависимых белков, относящихся к группе PR-10, но мы не обнаружили изменения содержания β-1,3-глюканаз (Тарчевский и др., 2010). В геноме гороха был проведен поиск генов, кодирующих эти белки, и на основе транскриптомных данных проанализирована их экспрессия. В этой части работы мы сделали акцент лишь на наиболее значительно экспрессирующихся в корнях гороха генах (>100 FPKM), поскольку у генов с более низким уровнем экспрессии не происходило достоверного изменения при действии фитогормонов (см. Приложение).

В геноме гороха было выявлено более 50 генов β-1,3-глюканаз, из которых в корнях на разном уровне экспрессировалось примерно 20 генов (см. Приложение). Ген *Psat1g161880* экспрессировался в контрольном варианте на

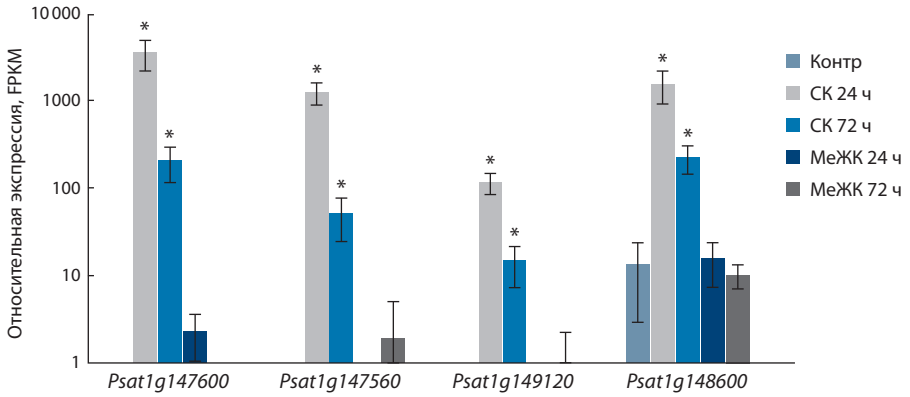


Рис. 3. Экспрессия генов хитиназа-подобных белков при действии СК (50 мкМ) и МеЖК (20 мкМ) в корнях гороха в течение 24 и 72 ч (FPKM).

		Контр	СК 24 ч	СК 72 ч	МеЖК 24 ч	МеЖК 72 ч	Аннотация
PR-2	*Psat1g161880*	522	310	460	135	174	Glucan endo-1,3-beta-glucosidase
	Psat3g165080	83	40	124	59	39	Glucan endo-1,3-beta-glucosidase 11
	Psat7g179640	3	145*	27	2	9	Glucan endo-1,3-beta-glucosidase
PR-10	*Psat4g180800*	4654	4696	5324	4000	3280	Pathogenesis-related protein Bet v I family
	Psat1g156920	2405	2555	2784	1744	2675	ABR17
	Psat1g157240	567	1247	1627*	422	579	Disease resistanse response protein Pi176
	Psat1g157360	362	1065	841	327	293	Disease resistanse response protein Pi49
	Psat1g157200	118	418	252	198	48	Disease resistanse response protein Pi49

Рис. 4. Список некоторых наиболее значительно (>100 FPKM) экспрессирующихся в корнях гороха генов *PR-2* (β-1,3-глюканаз) и *PR-10* в контрольном варианте (Контр) и под влиянием СК (50 мкм) и МеЖК (20 мкМ) в течение 24 и 72 ч. Цветовая шкала представляет наиболее сильно (красный цвет) и слабо (синий) экспрессируемые гены, уровень экспрессии выражен в FPKM.

высоком уровне, и его экспрессия не изменялась при действии СК и МеЖК (рис. 4). Экспрессия гена *Psat3g165080* не зависела от СК и МеЖК. Ген *Psat7g179640* практически не экспрессировался в контроле, но активировался СК через 24 ч (см. рис. 4).

В корнях гороха экспрессируется более 10 генов PR-10 белков. Именно представитель этой группы белков, кодируемый геном *Psat4g180800*, является наиболее сильно экспрессируемым в корнях гороха. Ни СК, ни МеЖК не оказывали значительного влияния на экспрессию этого гена. Ген *Psat1g156920*, продукт которого аннотирован как АБК-зависимый белок 17, тоже экспрессировался на высоком уровне, и его экспрессия не изменялась под влиянием изученных фитогормонов. Через 72 ч СК вызывала повышение экспрессии гена *Psat1g157240*, продукт которого аннотирован как белок устойчивости к болезни (см. рис. 4).

Обсуждение

При работе с немодельными растениями, к которым относится основная часть сельскохозяйственных видов, возникают расхождения с результатами, установленными на модельных видах растений. В частности, данные о маркерных генах, активация экспрессии которых связана с включением того или иного фитогормон-зависимого сигнального пути, получены именно на модельных растениях арабидопсиса, табака и др. Особенно это важно при изучении защитных механизмов растений, поскольку полученные данные могут быть использованы при разработке подходов для защиты этих видов и при селекции устойчивых сортов.

Основной объект наших исследований – горох. Ранее нами изучено влияние факторов фитоиммунитета – салициловой и жасмоновой кислот – на белковый спектр корней, выявлены наиболее значительные изменения содержания белков (Тарчевский и др., 2010; Яковлева и др., 2013; Egorova, 2025). В настоящей работе проведен анализ экспрессии некоторых классов PR-генов и генов хитиназа-подобных белков для оценки их ответа на действие салицилата и МеЖК, поскольку мы столкнулись с необходимостью выбора адекватного маркерного гена для оценки активации СК-зависимого пути в корнях гороха. Выбор времени обработки фитогормонами связан с тем, что при изучении протеомов через 72 ч мы наблюдали изменения белкового спектра, кроме того, нас интересовали процессы, происходящие на начальных этапах активации фитогормон-зависимых защитных механизмов.

В первую очередь была проанализирована экспрессия общепринятого СК-зависимого маркерного гена *PR-1* и его паралогов в корнях гороха. Впервые белок PR-1 был выделен из листьев табака, инфицированных ВТМ, его

количество могло составлять до 2 % от общего содержания белков инфицированных листьев табака (Alexander et al., 1993). Но не все изоформы PR-1 белков индуцировались при инфицировании. В ряде ранних работ продемонстрировано, что PR-1 белки принимают участие в процессах роста и развития растений, не связанных с патогенезом, – они накапливались в стареющих листьях растений (Fraser, 1981), чашелистиках развивающихся цветов (Lotan et al., 1989). Сверхэкспрессия *PR-1* приводит к повышению устойчивости к грибам (Sarowar et al., 2005) и бактериям (Kiba et al., 2007), у трансгенов повышалась устойчивость к патогенам (Ali et al., 2018). Защитные функции PR-1 белков могут быть связаны со способностью взаимодействовать со стеролами и с присутствием в их структуре CAPE пептида (Breen et al., 2017; Han, 2023). Было показано, что CAPE пептид индуцируется в растениях механическим повреждением и обработкой МеЖК, отщепляется от PR-1 белков с N-конца по консервативной CNYx последовательности и активирует экспрессию генов ингибиторов протеаз, *PR-1*, *PR-7*, *ERF-5* (Chen et al., 2014). Наиболее значительно экспрессируемые в корнях гороха гены PR-1 белков *Psat1g156240* и *Psat1g213200* содержат CAPE пептид и последовательность CNYx (см. рис. 1, *б*). Экспрессия гена *Psat1g156240* достоверно повышалась при действии МеЖК через 24 ч (см. рис. 1, *в*). Именно этот ген исследователи используют в качестве маркерного СК-зависимого в листьях гороха (Barba-Espín et al., 2011). Полученные нами результаты показывают, что по крайней мере в корнях гороха *Psat1g156240* не может быть использован для оценки активации салицилатного сигналинга. В то же время белки, кодируемые генами *Psat1g211480* и *Psat0s4615g0040*, содержат усеченный CAPE пептид и не содержат последовательность CNYx. В корнях гороха эти гены экспрессируются на низком уровне и не меняют интенсивность экспрессии при действии СК в изученных нами временных точках (см. рис. 1, *а*, *б*). Возможно, активация генов происходит раньше 24 ч. Активация *PR-1* генов МеЖК и другими воздействиями была продемонстрирована и ранее. Активация *AtPRB1* гена МеЖК/этиленом была показана на корнях арабидопсиса (Santamaria et al., 2001). В рисе изоформы *PR-1* генов по-разному отвечали на действие СК и МеЖК, кроме того, наблюдалась разница в уровне их экспрессии в зависимости от органа растения (Mitsuhara et al., 2008).

Хитиназы (EC 3.2.1.14) относятся к гликозил-гидролазам, главным образом семейств 18 и 19, разрушают хитин, являющийся компонентом клеточных стенок многих патогенных микроорганизмов (Davies, Henrissat, 1995). Поскольку в листьях гороха хитиназы были одними из наиболее значительно индуцируемых белков при действии СК (Тарчевский и др., 2010), мы проанализировали экспрессию генов хитиназ в корнях. Помимо того, индукция салицилатом транскриптов хитиназ была показана у сосны (Davis et al., 2002).

Согласно транскриптомным данным, в корнях гороха ген *Psat5g000800* экспрессируется на высоком уровне, кодирует белок, относящийся к семейству GH19, и характеризуется наличием лизоцим-подобного домена. Экспрессия этого гена не чувствительна к действию СК и МеЖК (см. рис. 2, *а*). Предполагается, что хитиназы GH19 играют важную роль в защите от почвенных патогенов, секретируются в ризосферу (De-la-Peña et al., 2010). Число экспрессирующихся в корнях гороха хитиназ, кодирующих белки семейства GH18, больше по сравнению с GH19, а уровень их экспрессии ниже по сравнению с генами, кодирующими хитиназы GH19. Кроме того, и они не проявляют специфичность регуляции изученными фитогормонами, за исключением гена *Psat1g131280*, экспрессия которого достоверно повышалась при действии МеЖК (см. рис. 2). Полученные данные показывают, что экспрессирующиеся в корнях гороха изоформы хитиназ не могут быть использованы для оценки активации СК-зависимого пути.

Наряду с генами, кодирующими активные хитиназы, в геномах растений присутствуют гены, кодирующие неактивные хитиназа-подобные белки. В ходе эволюционного процесса эти белки утратили способность расщеплять хитин, в том числе из-за аминокислотных замен в каталитически важных доменах и в связи с особенностями формирования третичной структуры (Kesari et al., 2015). Мы выявили значительную активацию экспрессии четырех генов хитиназа-подобных белков в корнях гороха под действием СК (рис. 3). Причем эти гены являлись салицилат-зависимыми и не изменялись под влиянием МеЖК, что делает их хорошими кандидатами для использования в качестве маркеров СК-зависимого пути в корнях гороха, за исключением гена *Psat1g148600*. Мы показали, что содержание белка, кодируемого геном *Psat1g148600*, повышается при действии донора NO – нитропруссида Na (Egorova, Tarchevsky, 2024).

Растительные β-1,3-глюканазы работают совместно с хитиназами, катализируя деградацию компонентов клеточных стенок микроорганизмов β-1,3-глюканов и хитина (Jongedijk et al., 1995; Dos Santos, Franco, 2023). В корнях гороха наиболее значительно экспрессируется по одному гену β-1,3-глюканаз и хитиназ; уровень экспрессии этих генов схож. Если некоторые гены, кодирующие хитиназы, индуцируются СК и МеЖК, то β-1,3-глюканаза, кодируемая геном *Psat7g179640*, индуцируется только СК. Экспрессия других генов β-1,3-глюканаз существенно не изменялась ни при действии СК, ни МеЖК.

Как уже упоминалось выше, *Psat4g180800*, кодирующий PR-10 белок, демонстрирует самый высокий уровень экспрессии среди генов в корнях гороха: СК и МеЖК не вызывали изменения его экспрессии. Другие гены, кодирующие PR-10 белки, главным образом нечувствительны к этим фитогормонам, и их экспрессия не изменяется под их влиянием.

Одним из признаков PR-белков является их индукция при инфицировании патогенами или под влиянием стрессовых фитогормонов. Анализ транскриптомных данных и полученные нами ранее протеомные данные показывают, что в корнях гороха экспрессия и содержание некоторых представителей PR-белков высокие – гены, кодирующие некоторые PR-белки, входят в число наиболее значительно экспрессируемых генов в контроле. Среди них ген PR-1 – *Psat1g156240*, гены PR-10 белков – *Psat4g180800*

и *Psat1g156920*. Это может быть связано с тем, что корни находятся в постоянном контакте с обитателями ризосферы, среди которых присутствуют и патогенные организмы. Имеющийся в корнях пул белков с антипатогенной активностью позволяет им быть подготовленными к возможному инфицированию, а при активации салицилат-зависимого защитного ответа происходит активация экспрессии и синтеза хитиназа-подобных белков, выступающих, по всей вероятности, в качестве дополнительного механизма защитного ответа корней гороха.

Заключение

Проведенный в настоящей работе транскриптомный анализ базовой и индуцируемой экспрессии некоторых генов PR-белков в корнях гороха показал, что при выборе маркерных генов для оценки активации салицилатного сигналинга следует обращать внимание на особенности их экспрессии в конкретном органе. Корни гороха отличаются высоким уровнем экспрессии генов и содержания антипатогенных белков. Транскриптомный анализ показал, что наиболее значительно экспрессируемые в корнях гороха гены PR-1 не могут быть использованы для оценки активации салицилатного сигналинга. Хитиназы и β-1,3-глюканазы, индуцируемые СК в листьях гороха, не проявляют специфичность в ответ на ее действие в корнях. Полученные данные демонстрируют разницу ответа листьев и корней на действие стрессовых фитогормонов. На примере растений гороха установлено, что гены салицилат-индуцируемых в листьях антипатогенных белков в корнях экспрессируются на высоком уровне в нормальных условиях, тогда как к салицилат-индуцируемым относятся гены, кодирующие хитиназа-подобные белки, которые также обладают антипатогенной активностью (неопубликованные данные). Кроме того, гены хитиназа-подобных белков практически не экспрессируются в контроле и индуцируются именно салицилатом, что является преимуществом при анализе экспрессии этих генов с целью определения активации салицилат-зависимого сигнального пути.

Список литературы / References

Тарчевский И.А., Яковлева В.Г., Егорова А.М. Салицилат-индуцированная модификация протеомов у растений (обзор). *Прикладная биохимия и микробиология*. 2010;46(3):263-275 [Tarchevsky I.A., Yakovleva V.G., Egorova A.M. Salicylate-induced modification of plant proteomes (review). *Appl Biochem Microbiol*. 2010;46(3):241-252. doi 10.1134/S0003683810030026]

Яковлева В.Г., Егорова А.М., Тарчевский И.А. Протеомный анализ влияния метилжасмоната на корни проростков гороха. *Доклады Академии наук*. 2013;449(2):236-239. doi 10.7868/s0869565213080288 [Yakovleva V.G., Egorova A.M., Tarchevsky I.A. Proteomic analysis of the effect of methyl jasmonate on pea seedling roots. *Dokl Biochem Biophys*. 2013;449:90-93. doi 10.1134/S1607672913020105]

Alexander D., Goodman R.M., Gut-Rella M., Glascock C., Weymann K., Friedrich L., Maddox D., Ahl-Goy P., Luntz T., Ward E. Increased tolerance to two oomycete pathogens in transgenic tobacco expressing pathogenesis-related protein 1a. *Proc Natl Acad Sci USA*. 1993;90(15):7327-7331. doi 10.1073/pnas.90.15.7327

Ali S., Ganai B.A., Kamili A.N., Bhat A.A., Mir Z.A., Bhat J.A., Tyagi A., Islam S.T., Mushtaq M., Yadav P., Rawat S., Grower A. Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance. *Microbiol Res*. 2018;212-213:29-37. doi 10.1016/j.micres.2018.04.008

Barba-Espín G., Clemente-Moreno M.J., Alvarez S., García-Legaz M.F., Hernández J.A., Díaz-Vivancos P. Salicylic acid negatively affects the response to salt stress in pea plants. *Plant Biol* (*Stuttg*). 2011;13(6):909-917. doi 10.1111/j.1438-8677.2011.00461.x

Blanco F., Salinas P., Cecchini N.M., Jordana X., Van Hummelen P., Alvarez M.E., Holuigue L. Early genomic responses to salicylic acid in Arabidopsis. *Plant Mol Biol*. 2009;70:79-102. doi 10.1007/s11103-009-9458-1

Breen S., Williams S.J., Outram M., Kobe B., Solomon P.S. Emerging insights into the functions of pathogenesis-related protein 1. *Trends Plant Sci*. 2017;22(10):871-879. doi 10.1016/j.tplants.2017.06.013

Chen Y.L., Lee C.Y., Cheng K.T., Chang W.H., Huang R.N., Nam H.G., Chen Y.R. Quantitative peptidomics study reveals that a wound-induced peptide from PR-1 regulates immune signaling in tomato. *Plant Cell*. 2014;26(10):4135-4148. doi 10.1105/tpc.114.131185

Davies G., Henrissat B. Structures and mechanisms of glycosyl hydrolases. *Structure*. 1995;3:853-859. doi 10.1016/S0969-2126(01)00220-9

Davis J.M., Wu H., Cooke J.E., Reed J.M., Luce K.S., Michler C.H. Pathogen challenge, salicylic acid, and jasmonic acid regulate expression of chitinase gene homologs in pine. *Mol Plant Microbe Interact*. 2002;15(4):380-387. doi 10.1094/MPMI.2002.15.4.380

De-la-Peña C., Badri D.V., Lei Z., Watson B.S., Brandão M.M., Silva-Filho M.C., Sumner L.W., Vivanco J.M. Root secretion of defense-related proteins is development-dependent and correlated with flowering time. *J Biol Chem*. 2010;285(40):30654-30665. doi 10.1074/jbc.M110.119040

de Wit P.J.G.M. How plants recognize pathogens and defend themselves. *Cell Mol Life Sci*. 2007;64:2726-2732. doi 10.1007/s00018-007-7284-7

Ding L.N., Li Y.T., Wu Y.Z., Li T., Geng R., Cao J., Zhang W., Tan X.L. Plant disease resistance-related signaling pathways: recent progress and future prospects. *Int J Mol Sci*. 2022;23(24):16200. doi 10.3390/ijms232416200

Ding P., Ding Y. Stories of salicylic acid: a plant defense hormone. *Trends Plant Sci*. 2020;25(6):549-565. doi 10.1016/j.tplants.2020.01.004

Dos Santos C., Franco O.L. Pathogenesis-related proteins (PRs) with enzyme activity activating plant defense responses. *Plants* (*Basel*). 2023;12(11):2226. doi 10.3390/plants12112226

Egorova A.M. Unexpected response of pea roots to the salicylic acid: induction of chitinase-like proteins compared to chitinases. *J Plant Growth Regul*. 2025;44:3164-3178. doi 10.1007/s00344-024-11607-1

Egorova A.M., Tarchevsky I.A. Proteome analysis of the nitric oxide donor effect on *Pisum sativum* L. roots. *Appl Biochem Microbiol*. 2024;60(6):1292-1300. doi 10.1134/S0003683824604694

Erb M. The role of roots in plant defence. In: Mérillon J., Ramawat K. (Eds) Plant Defence: Biological Control. Progress in Biological Control. Vol. 12. Springer, Dordrecht, 2012;291-309. doi 10.1007/978-94-007-1933-0_12

Fraser R.S.S. Evidence for the occurrence of the "pathogenesis-related" proteins in leaves of healthy tobacco plants during flowering. *Physiol Plant Pathol*. 1981;19(1):69-76. doi 10.1016/S0048-4059(81)80009-4

Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol*. 2005;43:205-227. doi 10.1146/annurev.phyto.43.040204

Han Z., Xiong D., Schneiter R., Tian C. The function of plant PR1 and other members of the CAP protein superfamily in plant-pathogen interactions. *Mol Plant Pathol*. 2023;24(6):651-668. doi 10.1111/mpp.13320

Hennig J., Dewey R.E., Cutt J.R., Klessig D.F. Pathogen, salicylic acid and developmental dependent expression of a β-1,3-glucanase/GUS gene fusion in transgenic tobacco plants. *Plant J*. 1993;4(3):481-493. doi 10.1046/j.1365-313x.1993.04030481.x

Jayaraj J., Muthukrishnan S., Liang G., Velazhahan R. Jasmonic acid and salicylic acid induce accumulation of β-1,3-glucanase and

thaumatin-like proteins in wheat and enhance resistance against *Stagonospora nodorum*. *Biologia Plantarum*. 2004;48:425-430. doi 10.1023/B:BIOP.0000041097.03177.2d

Jongedijk E., Tigelaar H., van Roekel J.S.C. Bres-Vloemans S.A., Dekker I., van den Elzen P.J.M., Cornelissen B.J.C., Melchers L.S. Synergistic activity of chitinases and β-1,3-glucanases enhances fungal resistance in transgenic tomato plants. *Euphytica*. 1995;85:173-180. doi 10.1007/BF00023946

Kesari P., Patil D.N., Kumar P., Tomar S., Sharma A.K., Kumar P. Structural and functional evolution of chitinase-like proteins from plants. *Proteomics*. 2015;15:1693-1705. doi 10.1002/pmic.201400421

Kiba A., Nishihara M., Nakatsuka T., Yamamura S. Pathogenesis-related protein 1 homologue is an antifungal protein in *Wasabia japonica* leaves and confers resistance to *Botrytis cinerea* in transgenic tobacco. *Plant Biotechnol*. 2007;24:247-253. doi 10.5511/plantbiotechnology.24.247

Kouzai Y., Kimura M., Yamanaka Y., Watanabe M., Matsui H., Yamamoto M., Ichinose Y., Toyoda K., Onda Y., Mochida K., Noutoshi Y. Expression profiling of marker genes responsive to the defence-associated phytohormones salicylic acid, jasmonic acid and ethylene in *Brachypodium distachyon*. *BMC Plant Biol*. 2016;16:59. doi 10.1186/s12870-016-0749-9

Kreplak J., Madoui M.A., Capal P., Novák P., Labadie K., Aubert G., Bayer P.E., … Lichtenzveig J., Macas J., Doležel J., Wincker P., Burstin J. A reference genome for pea provides insight into legume genome evolution. *Nat Genet.* 2019;51:1411-1422. doi 10.1038/s41588-019-0480-1

Kumar S., Stecher G., Li M., Knyaz 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

Li Y.F., Zhu R., Xu P. Activation of the gene promoter of barley β-1,3-glucanase isoenzyme GIII is salicylic acid (SA)-dependent in transgenic rice plants. *J Plant Res*. 2005;118:215-221. doi 10.1007/s10265-005-0213-7

Lotan T., Ori N., Fluhr R. Pathogenesis-related proteins are developmentally regulated in tobacco flowers. *Plant Cell.* 1989;1(9):881-887. doi 10.1105/tpc.1.9.881

Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. doi 10.1186/ s13059-014-0550-8

Maleck K., Levine A., Eulgem T., Morgan A., Schmid J., Lawton K.A., Dang J.L., Dietrich R.A. The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nat Genet.* 2000;26:403-410. doi 10.1038/82521

Mitsuhara I., Iwai T., Seo S., Yanagawa Y., Kawahigasi H., Hirose S., Ohkawa Y., Ohashi Y. Characteristic expression of twelve rice *PR1* family genes in response to pathogen infection, wounding, and defense-related signal compounds (121/180). *Mol Genet Genomics*. 2008;279(4):415-427. doi 10.1007/s00438-008-0322-9

Papadopoulou G.V., Maedicke A., Grosser K., van Dam N.M., Martínez-Medina A. Defence signalling marker gene responses to hormonal elicitation differ between roots and shoots. *AoB Plants*. 2018;10(3):ply031. doi 10.1093/aobpla/ply031

Perez-Aranda A., Loera-Muro A., Caamal-Chan M.G. Expression analysis of defense signaling marker genes in *Capsicum annuum* in response to phytohormones elicitation. *Mol Biol Rep*. 2024;52(1):9. doi 10.1007/s11033-024-10071-0

Santamaria M., Thomson C.J., Read N.D., Loake G.J. The promoter of a basic *PR1*-like gene, *AtPRB1*, from *Arabidopsis* establishes an organ-specific expression pattern and responsiveness to ethylene and methyl jasmonate. *Plant Mol Biol.* 2001;47:641-652. doi 10.1023/A:1012410009930

Sarowar S., Kim Y.J., Kim E.N., Kim K.D., Hwang B.K., Islam R., Shin J.S. Overexpression of a pepper basic pathogenesis-related protein 1 gene in tobacco plants enhances resistance to heavy metal and pathogen stresses. *Plant Cell Rep*. 2005;24:216-224. doi 10.1007/s00299-005-0928-x

Solomon W., Janda T., Molnár Z. Unveiling the significance of rhizosphere: implications for plant growth, stress response, and sustainable agriculture. *Plant Physiol Biochem*. 2024;206:108290. doi 10.1016/j.plaphy.2023.108290

Spoel S.H., Johnson J.S., Dong X. Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. *Proc Natl Acad Sci USA*. 2007;104(47):18842-18847. doi 10.1073/pnas.0708139104

Trapnell C., Williams B.A., Pertea G., Davis A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J., Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511-515. doi 10.1038/nbt.1621

Tyagi P., Singh D., Mathur S., Singh A., Ranjan R. Upcoming progress of transcriptomics studies on plants: an overview. *Front Plant Sci*. 2022;13:1030890. doi 10.3389/fpls.2022

van Loon L.C., Rep M., Pieterse C.M. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol*. 2006;44:135-162. doi 10.1146/annurev.phyto.44.070505.143425

Wang Y., Mostafa S., Zeng W., Jin B. Function and mechanism of jasmonic acid in plant responses to abiotic and biotic stresses. *Int J Mol Sci.* 2021;22(16):8568. doi 10.3390/ijms22168568

Xu L.N., Jiang X.R., Lin J.X., Li K., Javed T., Zhao J.-Y., Gao S.-J. Pathogenesis-related protein 10 family genes involved in sugarcane responses to biotic stressors and salicylic acid. *J Plant Growth Regul*. 2004;43:3907-3919. doi 10.1007/s00344-024-11371-2

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию 20.03.2025. После доработки 09.11.2025. Принята к публикации 11.11.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):222-232

doi 10.18699/vjgb-26-25

Selecting stable rice mutants with linear mixed models (LMM) and stability indexes

P. Sharifi [1], A.A. Ebadi[2], M.T. Hallajian[3], H. Aminpanah[1]

[1] Department of Agronomy, Ra. C., Islamic Azad University, Rasht, Iran
[2] Rice Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Rasht, Iran
[3] Nuclear Science and Technology Research Institute of Iran, Tehran, Iran
peyman.sharifi@iau.ac.ir

Abstract. Mutation serves as a pivotal source of diversity in plant breeding. This study focused on identifying stable rice mutant lines. Fourteen rice mutant lines, along with four conventional cultivars, were evaluated in a randomized complete block design with three replicates across three Iranian locations (Rasht, ChaparSar, and Fars province) during two growing seasons (2015, 2016). All statistical analyses were performed using the 'metan' (multi-environment trial analysis) R package. Single-environment ANOVA indicated significant genotypic effects for all traits. Likelihood ratio tests (LRTs) confirmed significant environment and genotype-by-environment interaction (GEI) effects for all traits. The first three principal components (PCs) captured 68.13, 14.46, and 9.76 % of the GEI variation, respectively. Heatmap visualization of yield performance and WAASB (weighted average of absolute scores based on best linear unbiased prediction, BLUP) highlighted genotypes G3, G9, G6, G12, and G5 as both high-yielding and stable. Multi-trait stability index (MTSI) analysis, designed to reveal genotypic strengths and weaknesses, selected only genotypes G7, G5, and G1. The top five genotypes based on the harmonic mean of the relative performance of genotypic values (HMRPGV) were G5, G12, G7, G2, and G1. The superior performance of certain mutants demonstrates that mutation has effectively generated significant genetic diversity. Notably, genotypes G12, G5, and G9 exhibited a clear advantage over the other genotypes and warrant consideration for selection or cultivar release; however, only G5 was selected based on all traits in the MTSI index and could therefore undergo selection or cultivar introduction processes.

Key words: genotype-by-environment interaction; MTSI; mutation breeding; simultaneous selection index; WAASB

For citation: Sharifi P., Ebadi A.A., Hallajian M.T., Aminpanah H. Selecting stable rice mutants with linear mixed models (LMM) and stability indexes. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):222-232. doi 10.18699/vjgb-26-25

Funding. The project was supported by the Iranian Rice Research Institute 03-04-0455-9411.

Authors' contribution. MTH treated parental seeds with gamma rays. AAE planned the research, carried the lead role in managing the research and collecting the data. PS analyzed the data and wrote the manuscript with contribution from all the authors.

Отбор стабильных мутантов риса с помощью линейных моделей со смешанными эффектами (LMM) и индексов стабильности

П. Шарифи [1], А.А. Эбади[2], М.Т. Халладжян[3], Х. Аминпанах[1]

[1] Департамент агрономии, Исламский университет Азад, Решт, Иран
[2] Иранский исследовательский институт риса, Организация по сельскохозяйственным исследованиям, образованию и распространению знаний (AREEO), Решт, Иран
[3] Научно-исследовательский институт ядерных наук и технологий Ирана, Тегеран, Иран
peyman.sharifi@iau.ac.ir

Аннотация. Мутационные изменения служат основным источником разнообразия в селекции растений. Данное исследование было сосредоточено на выявлении стабильных мутантных линий риса. Четырнадцать мутантных линий и четыре обычных сорта риса были оценены в рандомизированном посеве с тремя повторениями в трех регионах Ирана (Решт, Чапарсар и провинция Фарс) в течение двух вегетационных сезонов 2015 и 2016 гг. Все статистические анализы выполнены с использованием пакета R 'metan' (multi-environment trial analysis). Однофакторный ANOVA показал значимые генотипические эффекты для всех признаков. Тесты отношения правдоподобия (LRT) подтвердили значимые эффекты среды и взаимодействия генотип–среда (GEI) для всех признаков. Первые три главные компоненты определяли 68.13, 14.46 и 9.76 % вариаций GEI соответственно. Визуализация тепловой карты для выраженности урожайности и WAASB (взвешенное среднее абсолютных баллов на основе наилучшего линейного несмещенного прогноза, BLUP) позволила выделить генотипы G3, G9, G6, G12 и G5 как высокоурожайные и стабильные. Анализ с помощью индекса мультипризнаковой стабильности

© Sharifi P., Ebadi A.A., Hallajian M.T., Aminpanah H., 2026
This work is licensed under a Creative Commons Attribution 4.0 License

(MTSI), разработанный для выявления сильных и слабых сторон генотипов, отобрал только генотипы G7, G5 и G1. С использованием параметра гармонического среднего относительной производительности генотипических значений (HMRPGV) были отмечены пять лучших генотипов – G5, G12, G7, G2 и G1. Высокая выраженность признаков у некоторых мутантов демонстрирует, что мутационные изменения могут эффективно создавать значительное генетическое разнообразие. В частности, генотипы G12, G5 и G9 имели явное преимущество перед другими генотипами и могут быть использованы для последующей селекции или для создания сорта. Однако только генотип G5 был отобран с учетом индекса MTSI на основе всех признаков и, следовательно, может быть использован в дальнейшем селекционном процессе или для создания сорта.
Ключевые слова: взаимодействие генотип–среда; MTSI; мутационная селекция; индекс одновременного отбора; WAASB

Introduction

Rice stands as a critical cereal crop, serving as a primary food source for over half of the global population. It assumes a vital role in ensuring food and nutritional security, while concurrently contributing to poverty alleviation (Khush, 2005). Crop production is susceptible to abiotic stresses, including drought and salinity, necessitating the evaluation of genotypes across diverse environments. This approach facilitates the identification of adaptable genotypes suitable for broad or specific environmental conditions (Sharifi, 2020).

Mutation is a key source of generating variation in plant breeding programs, allowing for the development of novel traits or the improvement of existing ones. This process relies on alterations within the genetic structure, leading to new heritable traits that can spread within plant populations (Shu, 2012). Mutation breeding is often used to develop cultivars that are more tolerant to biotic and abiotic stresses, enhance quality, or alter other plant characteristics (Ebadi et al., 2019). This strategy is a useful complement to existing germplasm and can help in the development of new cultivars. As plant phenotypes are affected by genotype-by-environment interaction (GEI), stability analysis is essential to understand the performance of genotypes across different environments and guide breeders in selecting superior genotypes (Sharifi et al., 2017). Consequently, yield stability and adaptability analyses in multi-environment trials are a critical component of genotype evaluation programs before a new breeding line is released as a commercial variety (Kang, 1988). This demonstrates the plant's ability to maintain yield potential despite environmental fluctuations (Yan et al., 2000).

Stability analysis has become an important part of research and plant breeding. Various statistical models, including univariate and multivariate methods, are used to assess the stability and adaptability of genotypes. The simultaneous selection index named as grain yield in a single non-parametric index (GSI), which is based on the sum of ranks for yield stability and performance, is often used as a criterion for stability analysis (Farshadfar, 2008). Yan (2016) noted that mean performance and stability are not always equally important when evaluating genotypes. Stability is more of a concern when genotypes are tested in only a few environments. However, when genotypes are adequately tested, information about their stability may already be reflected in their mean performance, because highly unstable genotypes – those that perform very well in some environments but poorly in others – will not be among the top performers across all environments. Newer indices for evaluating stability based on linear mixed models (LMMs) include WAASB (weighted average of absolute scores based on best linear unbiased prediction (BLUP)) and WAASBY (WAASB stability index and yield performance, for simultaneous selection based on yield performance and stability) (Olivoto et al., 2019b). This approach combines additive main effects and multiplicative interaction models (AMMI), graphical tools with the predictive precision of BLUP (Olivoto et al., 2019b). The relative importance of mean performance and stability can vary when making selection decisions, which can change how genotypes are ranked (Yan, 2016). This explains why many breeders and variety evaluation organizations base their decisions only on mean performance, yet still achieve significant breeding progress (Yan, 2016). Given the importance of the WAASBY index, it can be used for all stability indices instead of the GSI (Farshadfar, 2008), allowing for the identification of the most superior genotypes.

In addition to high yield, which remains the primary breeding objective, a cultivar must meet minimum requirements for every trait that is significant to growers, processors, and end-users (Yan, 2021). Consequently, when data on multiple traits are available, the multi-trait stability index (MTSI) has several valuable applications for the simultaneous selection of mean performance and stability. Indeed, the greatest challenge in plant breeding is to combine all desirable traits within a single genotype, as key breeding objectives are often negatively correlated due to genetic linkage or pleiotropy (Yan, 2021). These indices are used in the stability analysis of rice (Sharifi et al., 2021), lentil (Karimizadeh et al., 2020), forage (Santos, Marza, 2020), barley (Ahakpaz et al., 2021), and wheat (Verma, Singh, 2020). An alternative analysis for stability studies based on mixed models is the harmonic mean of genotypic values and of the relative performance of genotypic values (HMRPGV), which provides information on stability, adaptability, and yield performance of genotypes in the same unit and scale as the evaluated trait (Resende, 2007). Selecting genotypes with the highest values of the harmonic mean of genotypic values (HMGV), relative performance of genotypic values (RPGV), and HMRPGV allows for simultaneous selection for yield performance and stability. This approach is used to evaluate grain yield stability in rice (Colombari-Filho et al., 2013), wheat (Coan et al., 2018; Verma, Singh, 2020), and corn (Rodovalho et al., 2015).

Several studies have assessed GEI in rice mutant lines. Dushyanthakumar and Shadadashari (2007) found significant effects of genotypes, environment, and their interactions when evaluating the yield stability of 15 rice mutants across nine environments. They used regression coefficients and

deviations from regression to identify stable mutants. Dewi and Dwimahyani (2019) evaluated 12 mutant rice genotypes at 16 locations and estimated yield stability using regression lines. Rahayu (2020) used AMMI analysis to differentiate stable mutant rice lines from 25 genotypes. Relatively few studies have used linear mixed models for rice stability analysis. Donoso-Ñanculao et al. (2015) analyzed 10 rice genotypes using BLUP and identified superior genotypes. Balestre et al. (2010) also used BLUP to analyze the stability of rice genotypes and found that phenotypic means had lower predictive potential compared to BLUP.

The objectives of this study were to (1) evaluate GEI in rice genotypes across diverse environments in Iran, and (2) identify stable and high-yielding mutant lines.

Materials and methods

Plant material and experimental design

Fourteen rice mutant lines, along with their parental cultivars (Hashemi, Tarom, and Khazar) and the Gilaneh variety as a control (Table 1), were evaluated in a randomized complete block design (RCBD) with three replications. The experiment was conducted across three geographically distinct research stations in Iran: Rasht, ChaparSar, and Fars province. Data were collected over two years (2015, 2016). In both years, sowing took place in April, and transplanting was carried out in May at the 4–5 leaf stage. Harvesting occurred in August in both years as well.

The rice mutant lines originated from seeds of the parental cultivars subjected to 300 Gray of gamma irradiation from a Cobalt-60 source. Irradiation was performed at the Iranian Agricultural Atomic Research Institute in Tehran in 2009, with seed moisture content maintained at eight percent. Following irradiation, the M1 generation was grown in the field. Surviving plants were self-pollinated, and seeds from the M2 generation were collected. This self-pollination process was repeated through the M5 generation to achieve homozygosity.

Prior to sowing, seeds of all experimental genotypes (mutant lines and cultivars) were surface-sterilized with a 20 % sodium hypochlorite solution for 10 minutes, followed by rinsing with sterile distilled water. This procedure was carried out in May of each year. Seeds were then placed in Petri dishes containing moistened filter paper to initiate germination. Germination occurred in a controlled environment at 35 °C and 70 % humidity. Seedlings were subsequently transplanted into 120 cm×60 cm boxes within greenhouses, ensuring suitable conditions for transplant production.

Twenty-five-day-old seedlings (approximately 30 cm tall) were transplanted to the main field at a planting density of 25 plants per square meter. Each plot measured two square meters. Data were collected from ten randomly selected plants within each plot. The following traits were measured: plant height, tiller number, panicle length, number of filled grains per panicle, number of unfilled grains per panicle, fertility percentage, and hundred-grain weight. Grain yield was determined on a per-square-meter basis.

Table 1. The origin/pedigree of rice genotypes used in this experiment

Genotype code	Genotype	Pedigree
G1	TM6-230-VE-7-5-1	Parent: Tarom
G2	TM6-230-VE-8-4-1	
G3	TM6-250-10-7-1	
G4	TM6-B-2-1-E	
G5	TM6-B-7-1	
G6	TM6-B-19-2	
G7	HM5-250-E-1-1	Parent: Hashemi
G8	HM5-250-E-3-2	
G9	HM5-250-6-6	
G10	HM5-250-7-6	
G11	HM5-300-12-1	
G12	HM5-300-3-1	
G13	HM5-300-5-1	
G14	KM5-200-4-2-E	Parent: Khazar
G15	Khazar	IR36/TNAU4756
G16	Hashemi	Iranian landrace
G17	Tarom	Iranian landrace
G18	Gilaneh	Saleh × Abjibuji

Statistical analysis

To assess genotypic stability and adaptability, the harmonic mean of genotypic values (HMGV) and the relative performance of genotypic values (RPGV) were calculated. The harmonic mean of the relative performance of genotypic values (HMRPGV) was employed for the simultaneous evaluation of stability, adaptability, and yield. These indices were computed using the following formulas (Resende, 2007):

$$\mathrm{HMGV}_i = \frac{\mathrm{E}}{\sum_{j=1}^{\mathrm{E}} \frac{1}{\mathrm{GV}_{ij}}}, \quad (1)$$

$$\mathrm{RPGV}_i = \frac{1}{\mathrm{E}} \sum_{j=1}^{\mathrm{E}} \frac{\mathrm{GV}_{ij}}{\mu_j}, \quad (2)$$

$$\mathrm{HMRPGV}_i = \frac{\mathrm{E}}{\sum_{j=1}^{\mathrm{E}} \frac{1}{\mathrm{GV}_{ij}/\mu_j}}, \quad (3)$$

where μ_j represents the general mean for the j-th environment; E denotes the number of environments; and GV_{ij}: $u_j+g_i+ge_{ij}$ indicate the genotypic value of the i-th genotype in the j-th environment. Here, u_j is the mean of the j-th environment, while g_i and ge_{ij} are the BLUP values of the i-th genotype and the interaction between the i-th genotype and j-th environment, respectively.

The weighted average of absolute scores from the singular value decomposition (SVD) of BLUP-estimated interaction effects (WAASB_i) was computed using equation (4) (Olivoto et al., 2019a):

$$\mathrm{WAASB}_i = \frac{\sum_{k=1}^{p} |\mathrm{IPCA}_{ik} \times \mathrm{EP}_k|}{\sum_{k=1}^{p} \mathrm{EP}_k}, \quad (4)$$

where IPCA_{ik} is the score of the i-th genotype (or environment) in the k-th IPCA, and EP_k is the value of the explained variance by the k-th IPCA. The k-th IPCA is the k-th interaction principal

Table 2. Evaluation of significance of factors for random effects (E and GEI) by LRT (χ^2) and for fixed effect (G) by ANOVA and estimation of variance components by REML

SOV	Plant height, cm	Tiller number	Panicle length, cm	Number of filled grains per panicle	Number of unfilled grains per panicle	Fertility percentage	Hundred-grain weight, g	Grain yield, kg·ha^{-1}
				Mean Square				
Genotype	102.7**	9.39**	8.64**	59.31**	40.13**	19.76**	0.049**	224817**
				LRT (χ^2)				
Environment	109.5**	39.63**	81.07**	102.1**	0.0000014^{ns}	2.11^{ns}	95.98**	71.75**
Gen:Env	384.3**	240.9**	184.9**	540.8**	931.3**	0.089**	379.2**	577.6**
				Variance components				
REML				Estimates (%)				
σ^2_E	592.1 (94.11)	5.19 (45.64)	9.96 (80.42)	421.9 (79.57)	0.00000016 (0.00000017)	3.62 (6.57)	0.047 (75.73)	306144 (52.25)
$\sigma^2_{E/R}$	0.40 (0.06)	0.19 (1.73)	0.012 (0.09)	0.14 (0.03)	0.00000021 (0.00000023)	0 (0)	0 (0)	1181 (0.20)
σ^2_i	30.14 (4.79)	4.23 (37.20)	1.52 (12.30)	96.48 (18.20)	87.66 (96.76)	49.66 (90.17)	0.012 (19.75)	252401 (43.08)
σ^2_e	6.49 (1.03)	1.75 (15.43)	0.88 (7.18)	11.68 (2.20)	2.93 (3.23)	1.79 (3.26)	0.0028 (4.51)	26182 (4.46)
σ^2_p	629.2	11.38	12.39	530.2	90.60	55.07	0.063	585907

Note. The numbers in parentheses indicate the percentage of variance accounted for by the phenotypic variance. LRT, likelihood ratio test. σ^2_E, environmental variance; $\sigma^2_{E/R}$, Block-within-Environment variance (or Environment/Replication variance); σ^2_i, Genotype × Environment interaction variance; σ^2_e, residual variance; σ^2_p, phenotypic variance.
ns, non-significant, *, **: significant at 5 % and 1 % probability levels, respectively.

component axis (IPCA) in AMMI, used to compute WAASB. The genotype with the lowest WAASB value is considered the most stable (Olivoto et al., 2019a).

The WAASBY index, which allows simultaneous selection based on grain yield and stability, was derived using the following formula (Olivoto et al., 2019a):

$$\text{WAASBY}_i = \frac{(rG_i \times \theta_Y) + (rW_i \times \theta_S)}{\theta_Y + \theta_S}. \quad (5)$$

Where WAASBY_i is the superiority index for the i-th genotype, and θ_Y and θ_S are the weights for seed yield and stability (WAASB), respectively. The rescaled values (0–100) for the response variable (rG_i) and WAASB (rW_i) were calculated as follows:

$$rG_i = \frac{100-0}{G_{max}-G_{min}} \times (G_i - G_{max}) + 100, \quad (6)$$

$$rW_i = \frac{0-100}{W_{max}-W_{min}} \times (W_i - W_{max}) + 0. \quad (7)$$

The multi-trait stability index (MTSI) for the i-th genotype was calculated as follows (Olivoto et al., 2019b):

$$\text{MTSI}_i = \left[\sum_{j=1}^{f} (F_{ij} - F_j)^2\right]^{0.5}. \quad (8)$$

Where, F_{ij} is the j-th score of the i-th genotype, and the ideotype (F_j) is an ideal reference genotype with the most desirable values for each trait, f represents the number of traits (variables) included in the analysis. The ideotype has the highest WAASBY (100) for all analyzed variables (Olivoto et al., 2019b). The genotype with the lowest MTSI is then closer to the ideotype and therefore presents a high mean performance and stability (MPE) for all analyzed variables.

Selection differential (SD), selection gain (SG), and percentage of selection differential (% SD) represent the difference between the mean of selected genotypes (X) and the entire population, the expected genetic progress, and the relative selection intensity, respectively.

$$\text{SD} = \overline{X}_S - \overline{X}_0, \quad (9)$$

$$\text{SG} = h^2 \times \text{SD}, \quad (10)$$

$$\%\ \text{SD} = \frac{\text{SD}}{\overline{X}_0} \times 100. \quad (11)$$

$\overline{X}_S$: mean of selected genotypes; $\overline{X}_0$: mean of the entire population; h^2: heritability of the trait.

The selection differential for mean performance and the WAASBY index was calculated for each trait considering a selection intensity of 15 %.

All statistical analyses were performed using the 'metan' (multi-environment trial analysis) R package (Olivoto, Lúcio, 2020). An example of statistical analysis using R can be seen in the Supplementary file[1].

Results

Analysis of variance, estimation of variance components and predicted grain yield

Single-environment ANOVA indicated significant genotypic effects for plant height, tiller number, panicle length, number of filled grains per panicle, number of unfilled grains per panicle, fertility percentage, hundred-grain weight, and grain yield. Likelihood ratio tests (LRTs) confirmed significant environment and genotype-by-environment interaction (GEI) effects for all traits. Analysis of variance also indicated a significant effect of genotype on all traits (Table 2).

[1] Supplementary file is available at:
https://vavilovj-icg.ru/download/pict-2026-30/appx16.pdf

Principal component analysis and biplot interpretation

The first three principal components (PCs) captured 68.13, 14.46, and 9.76 % of the GEI variation, respectively (Fig. 1). A type III biplot (WAASB values *vs* grain yield, Fig. 2 and Fig. 3) was used to visualize genotypic stability.

Genotypes G14 and G15, along with environments E3 and E6, positioned in the first quadrant, exhibited below-average grain yield and high WAASB values, indicating instability and a substantial contribution to GEI. Conversely, genotypes G2, G7, and G8, and environments E1, E4, and E5, in the second quadrant, showed above-average grain yield but were also unstable with significant GEI contributions. Environments E3, E4, and E6 displayed the highest WAASB values, indicating greater discriminating ability. Genotypes G4, G6, G9–G11, G13, G16, and G17, located in the third quadrant, demonstrated low yield and high stability (low WAASB). The most desirable genotypes – G1, G3, G5, G12, and G18 – occupied the fourth quadrant, characterized by high yield and stability. Environment E2, also in this quadrant, showed high yield performance but low WAASB values, indicating poor genotypic discrimination ability (Fig. 2).

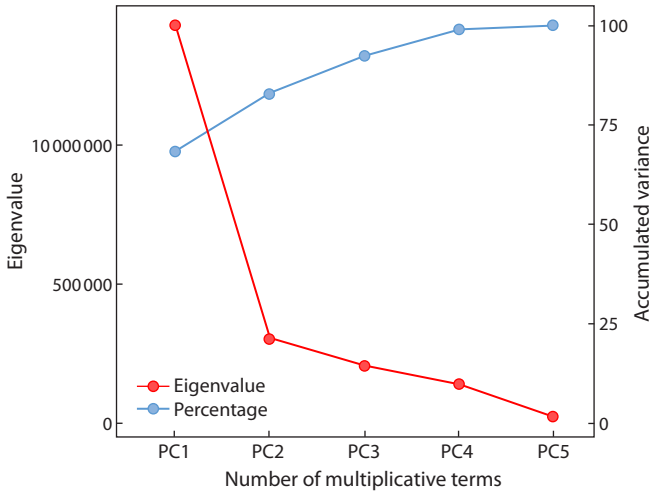


Fig. 1. Eigenvalues of the BLUP_GEI matrix for grain yield of 18 rice genotypes across six environments.

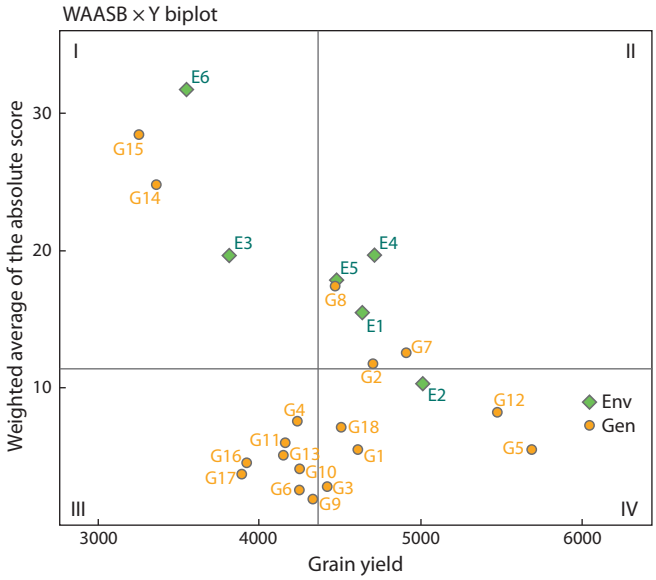


Fig. 2. Biplot of grain yield *vs* WAASB of 18 rice genotypes evaluated over six environments (combinations of two cultivation years in three locations).

Genotype ranking based on stability and performance weights

Figure 3 illustrates genotype rankings based on varying weights assigned to WAASB (stability) and grain yield. Ranking of 18 rice genotypes was based on the WAASBY index, illustrating variations in genotype performance under different weighting schemes for stability (WAASB) versus grain yield (BLUPs). The weights (θ) range from 0 (prioritizing yield alone) to 1 (prioritizing stability alone), with θ = 0.5 representing equal emphasis on both traits. Ranking based solely on stability (WAASB weight = 1) identified genotypes G6, G9, G3, G10, G13, and G17 as most stable. Ranking based solely on yield (yield weight = 1) identified G1, G2, G5, G7, and G12 as the highest-yielding genotypes.

Hierarchical clustering revealed distinct genotype groupings based on stability and yield: (1) G4, G8, G14, and G15 were the least desirable; (2) G10, G11, G13, G16, and G17 were stable but low-yielding; (3) G1, G2, G7, and G18 were high-yielding but unstable; and (4) G3, G5, G6, G9, and G12 (particularly G5) were both high-yielding and stable.

Factor analysis, multi-trait stability index (MTSI) and genotype selection

Factor analysis of the WAASBY index retained two principal components (eigenvalue > 1), explaining 71.5 % of the total variation (Table 3). Factorial loadings after varimax rotation and communalities obtained in the factor analysis are represented in Table 4. The eight traits were grouped into two factors: FA1 (agronomic and plant stature traits including plant height (PH), filled grains per panicle (FG), unfilled grains per panicle (UNFG), fertility percentage (FP), hundred-grain weight (HGW), grain yield (GY)), and FA2 (panicle-related traits contain tiller number (TN) and panicle length (PL)). While primarily associated with FA1, FG, UNFG, and GY also exhibited high scores on FA2 (Table 5).

Ranking based on MTSI identified G7 (MTSI = 1.97), G5 (MTSI = 2.07), and G1 (MTSI = 2.24) as superior genotypes (Fig. 4, cutoff at MTSI = 2.24). The cutoff value (MTSI = 2.24) was determined following the methodology of Olivoto et al. (2019b), where the Multi-Trait Stability Index (MTSI) ranks genotypes based on their stability and multi-trait performance. This threshold represents the upper limit of the top-performing genotypes (G1, G5, G7), beyond which stability and performance decline significantly. The selection of this cutoff aligns with standard MTSI applications, where natural breaks in the index distribution or percentile-based thresholds are used to distinguish superior genotypes. The genotype with the lowest MTSI is then closer to the ideotype and therefore presents a high MPE for all analyzed variables (Olivoto et al., 2019b).

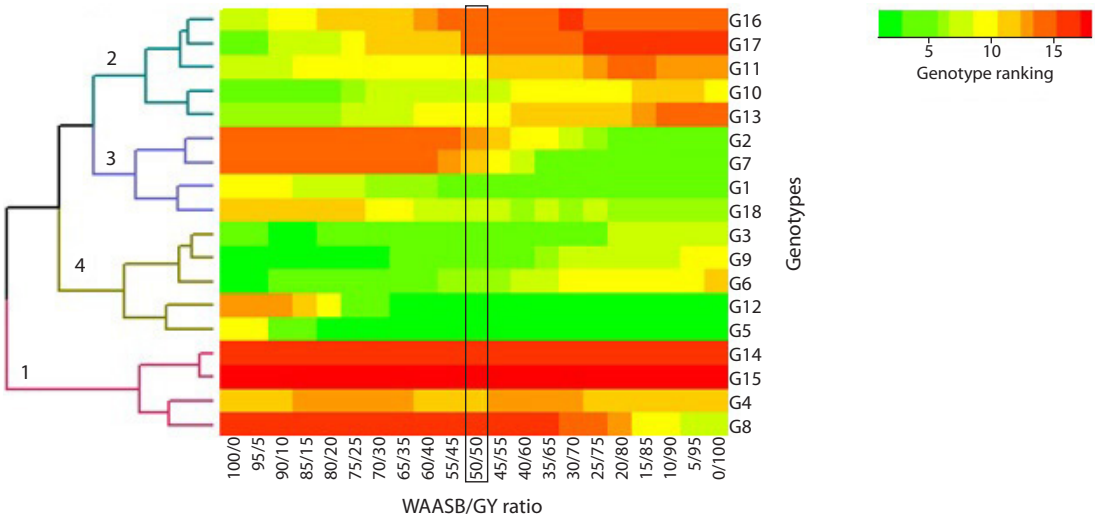


Fig. 3. Ranks of 18 rice genotypes with different weights for stability and grain yield.

Table 3. Eigenvalues and explained variance obtained in the factor analysis

PC	Eigenvalues	Variance, %	Cum. variance, %
PC1	4.61	57.6	57.6
PC2	1.11	13.9	71.5
PC3	0.997	12.5	84.0
PC4	0.636	7.95	91.9
PC5	0.343	4.29	96.2
PC6	0.22	2.75	99.0
PC7	0.073	0.913	99.9
PC8	0.00898	0.112	100.0

Table 4. Factorial loadings after varimax rotation, and communalities obtained in the factor analysis

Traits	FA1	FA2	Communality	Uniquenesses
Plant height (PH)	–0.908	0.133	0.841	0.159
Tiller number (TN)	–0.152	–0.872	0.784	0.216
Panicle length (PL)	–0.0767	–0.670	0.455	0.545
Filled grains per panicle (FG)	–0.570	–0.330	0.434	0.566
Unfilled grains per panicle (UNFG)	–0.717	–0.623	0.902	0.0981
Fertility percentage (FP)	–0.713	–0.639	0.917	0.0829
Hundred-grain weight (HGW)	–0.728	–0.229	0.582	0.418
Grain yield (GY)	–0.653	–0.617	0.807	0.193

Table 5. The related traits and genotypes for each factor

Factor	Related traits to FA	Related genotypes to FA
FA1	PH, FG, UNFG, FP, HGW, GY	G18, G2, G4, G5, G6, G1, G10, G14, G15
FA2	TN, PL	G7, G8, G9, G11, G12, G13, G17, G3

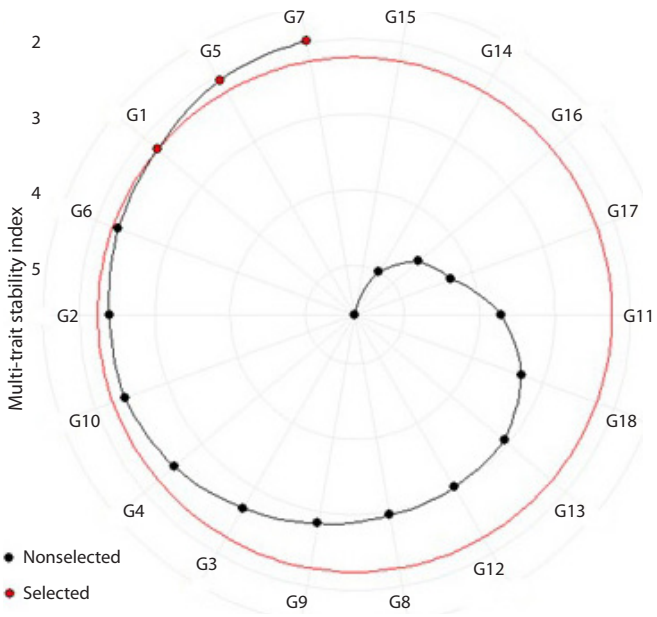


Fig. 4. Genotype ranking and genotypes selected for MTSI.

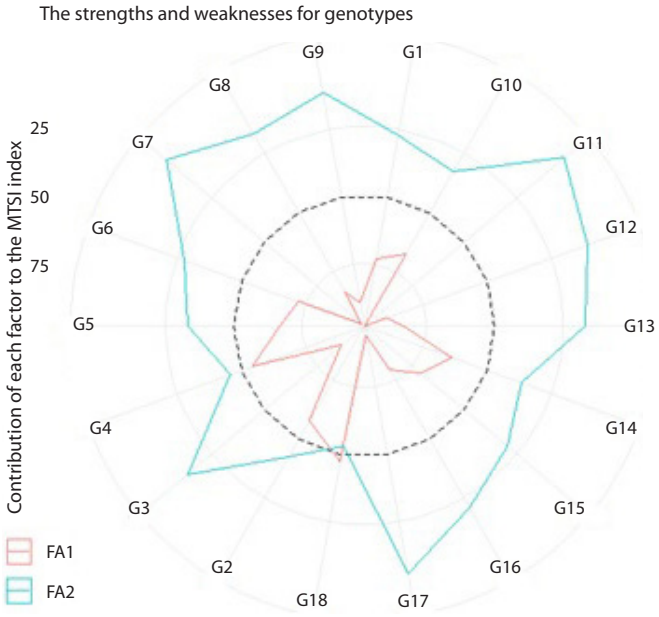


Fig. 5. The strengths and weaknesses of genotypes are shown as the proportion of each factor on the computed multi-trait stability index (MTSI) of all genotypes.

The smaller the proportion explained by a factor (closer to the external edge), the closer the traits within that factor are to the ideotype. FA1: PH, FG, UNFG, FP, HGW and GY; and FA2: TN and PL.

The MTSI resulted in positive selection differentials (SD) for WAASBY of all traits, ranging from 13 % (TN) to 49.5 % (HGW) (Table 6). Negative SDs for mean performance were observed for PH and UNFG, while positive SDs were noted for the remaining traits. This led to negative selection gains (SG) for PH (–0.35 %) and UNFG (–47.5 %), and positive SG for FG, FP, HGW, GY, TN, and PL (0.996 % ≤ SG ≤ ≤ 16.2 %). These results indicate that the selected genotypes exhibited improved genotypic stability compared to the original population, and indicate that MTSI is effective for genetic breeding.

Visualization of strengths and weaknesses

Figure 5 visualizes the strengths and weaknesses of each genotype. The contribution of each factor (FA1 and FA2) to the MTSI is ranked for each genotype, with the most contributing factor near the plot center and the least contributing factor near the plot edge (Olivoto et al., 2019b). Genotypes associated with a specific factor demonstrate high mean performance and stability (low WAASB) for the traits primarily influenced by that factor. Ten genotypes (G1, G2, G4, G5, G6, G10, G14, G15, G16 and G18) showed strengths related to FA1 (Fig. 5), indicating desirable combinations of high mean performance and stability for FA1-related traits: reduced PH and UNFG, and increased FG, FP, HGW, and GY. Eight genotypes (G3, G7, G8, G9, G11, G12, G13, and G17) exhibited strengths related to FA2 (Fig. 5). For G5, the second-ranked genotype by MTSI, significant FA1 contribution confirms its favorable

Table 6. Selection gains for mean performance of 18 rice genotypes obtained in a multi-trait selection with the multi-trait stability index (MTSI)

Traits	Factor	Goal	Mean performance (BLUP)						WAASBY index			
			Xo	Xs	SD	% SD	SG	SG %	Xo	Xs	SD	% SD
Plant height (PH)	FA1	Decrease	123	123.437	–0.437	–0.354	–0.409	–0.332	55.7	64.2	8.49	15.2
Filled grains per panicle (FG)		Increase	91.7	94.2	2.49	2.72	2	2.18	44.4	53.9	9.56	21.6
Unfilled grains per panicle (UNFG)		Decrease	13.5	7.1	–6.43	–47.5	–5.95	–43.9	83.3	95.8	12.5	14.9
Fertility percentage (FP)		Increase	88	92.7	4.63	5.26	4.21	4.78	78.1	93.7	15.6	19.9
Hundred-grain weight (HGW)			2.55	2.8	0.246	9.64	0.23	9.09	57.2	85.7	28.5	49.5
Grain yield (GY)			4360	5065	705	16.2	623	14.3	59.9	75.1	15.3	25.5
Tiller number (TN)	FA2	Increase	16.6	16.8	0.166	0.996	0.13	0.81	61.7	69.8	8.06	13
Panicle length (PL)			25.8	26.2	0.469	1.82	0.42	1.63	51.7	61.6	9.86	19.1

Note. Xo: mean performance (BLUP) and WAASBY index of the original population; Xs: mean performance (BLUP) and WAASBY index of the selected genotypes; SD: selection differential; % SD: percentage of selection differential; SG: selection gain.

Table 7. Ranking of the genotypes in all environments evaluated for adaptability parameters of genotypic values for the grain yield of rice genotypes evaluated in six environments

GEN	Grain yield, kg·ha^{-1}	HMGV†	HMGV_order	RPGV	RPGV*$\bar{\mu}$	RPGV_order	HMRPGV	HMRPGV*$\bar{\mu}$	HMRPGV_order
G1	4607	4537	5	1.058	4612	5	1.054	4597	5
G2	4699	4640	4	1.084	4727	4	1.073	4680	4
G3	4418	4333	8	1.012	4414	8	1.009	4401	8
G4	4227	4096	12	0.965	4207	12	0.959	4183	12
G5	5684	5652	1	1.313	5725	1	1.306	5696	1
G6	4237	4181	10	0.974	4247	10	0.972	4236	10
G7	4904	4884	3	1.138	4962	3	1.124	4901	3
G8	4466	4430	7	1.039	4531	6	1.017	4436	7
G9	4323	4276	9	0.994	4333	9	0.993	4330	9
G10	4240	4164	11	0.973	4241	11	0.969	4225	11
G11	4157	4023	14	0.948	4134	14	0.942	4105	14
G12	5470	5448	2	1.266	5519	2	1.257	5481	2
G13	4148	4061	13	0.950	4144	13	0.947	4130	13
G14	3353	2365	17	0.742	3236	17	0.589	2568	17
G15	3252	2288	18	0.716	3121	18	0.568	2477	18
G16	3908	3845	15	0.898	3917	15	0.894	3897	15
G17	3883	3832	16	0.892	3890	16	0.891	3882	16
G18	4504	4433	6	1.037	4521	7	1.028	4481	6

Note. † Performance genetic value (RPGV); RPGV*$\bar{\mu}$: stability of genotypic values (HMGV), adaptability and stability of genotypic values (HMRPGV); HMRPGV*$\bar{\mu}$: general mean of all environments; the parameters are expressed in kg·ha^{-1}, except for RPGV and HMRPGV.

combination of reduced PH and UNFG, and increased FG, FP, HGW, and GY. This genotype exhibited lower plant height and unfilled grains per panicle, and higher tiller number, hundred-grain weight, and grain yield, compared to the overall average. Genotype G12 showed higher tiller number, fertility percentage, and grain yield, along with lower plant height and unfilled grains per panicle. While identified as superior by other indices, G12 did not rank highly using MTSI. This discrepancy is likely due to the low selection gain for traits in FA1, specifically hundred-grain weight.

Genotypic stability and adaptability assessment using HMRPGV

Based on the Harmonic Mean of Relative Performance of Genotypic Values (HMRPGV), the top five genotypes for stability and adaptability, relative to check varieties, were G1, G2, G7, G12, and G5. The HMRPGV multiplied by the overall mean (HMRPGV*$\bar{\mu}$) for these genotypes were 4597, 4680, 4901, 5481, and 5696 kg·ha^{-1}, respectively (Table 7). Selecting these genotypes would result in a 16.35 % increase in grain yield compared to the overall mean (4360 kg·ha^{-1}). This selection, based on stability, adaptability, and yield, highlights the positive response of these genotypes to improved environmental conditions and their consistent performance across diverse environments.

Discussion

A broad genetic base is crucial for successful rice breeding programs, and induced mutation represents a valuable technique for expanding genetic diversity (Cheema, 2006). The significant effect of genotype observed across all environments, as indicated by simple analysis of variance, underscores the genetic diversity present within the experimental material. This result suggests that mutation has effectively contributed to a reasonable level of diversity within the genetic material under investigation. This finding aligns with the work of other researchers who have assessed the stability of rice mutant lines and demonstrated a significant impact of mutation in generating diversity with respect to grain yield and other important traits, particularly yield components (Dushyanthakumar, Shadadashari, 2007; Donoso-Ñanculao et al., 2015; Dewi, Dwimahyani, 2019; Rahayu, 2020).

Experimental trials designed to evaluate yield performance, stability, and adaptability are essential for characterizing genetic materials such as mutant lines before they are released as new breeding lines or varieties (Shu, 2012). Conversely, the observed variation in genotype grain yield across different environments highlights the need to carefully determine the yield performance and stability of genotypes in a range of environmental conditions (Ebadi et al., 2019). Therefore, it is crucial to investigate the yield stability of rice mutant lines and

to identify those lines that exhibit both high yield potential and stable performance across diverse environmental conditions. The likelihood ratio test (LRT) conducted in this study revealed significant effects of both the environment and genotype-by-environment interaction (GEI). Similar results have been reported by other researchers conducting multi-environment trials with rice (Bose et al., 2011; Akter et al., 2015; Sharifi et al., 2017; Rahayu, 2020). Variations in environmental conditions from year to year and from location to location contribute to a pronounced GEI effect, which underscores the role of genetic factors in influencing genotype performance across different environmental conditions (Dia et al., 2016).

The restricted maximum likelihood (REML) analysis conducted in this study indicated that environmental variance contributed more significantly to the phenotypic variance observed for traits such as plant height, number of tillers, panicle length, number of filled grains per panicle, hundred-grain weight, and grain yield. In contrast, GEI had a greater contribution to the observed variance in the number of unfilled grains per panicle and fertility percentage. The first two principal components derived from the GEI analysis accounted for 68.13 and 14.46 % of the total genotype-by-environment interaction variation, respectively. In line with this finding, other researchers have acknowledged the substantial contribution of the first two principal components in explaining genotype-by-environment interaction effects on rice grain yield (Nayak et al., 2008; Akter et al., 2015; Allahgholipour, 2017; Rahayu, 2020).

Genotype-by-environment interaction reduces the predictability of genotype performance in target environments based on observations made in test environments (Yan et al., 2011). Numerous procedures and statistical methods have been developed to quantify genotype by environment interaction through the evaluation of genotypes across multiple environments. In addition to conventional multivariate methods like AMMI and GGE biplot analysis, the best linear unbiased predictor (BLUP) method has also been recommended for evaluating genotypes in different environments and for quantifying genotypic stability. Several researchers have employed the BLUP method to assess the stability of rice genotypes (Balestre et al., 2010; Donoso-Ñanculao et al., 2015). The use of restricted maximum likelihood (REML) for estimating variance components offers significant flexibility in analysis (Patterson, Thompson, 1971) and can effectively handle complex data structures (Searle et al., 1992).

The WAASBY index (Olivoto et al., 2019a) facilitates the weighting between mean performance and stability (MPE) in genotype evaluation. The primary advantage of this procedure is its integration of AMMI's graphical tools (Gauch, 2013) with the predictive accuracy of BLUP for stability analysis (Piepho, 1994; van Eeuwijk et al., 2016). Consequently, the WAASB approach, which proposes the SVD (singular value decomposition) of a two-way table with BLUPs for GEI interaction (instead of the residual of the additive model as in standard AMMI), allows for a graphical representation of a random GEI structure. The WAASB index accounts for more than one (stability based on IPCA1) or two (AMMI-stability value) interaction principal component axes. In the present research, the first three principal components accounted for over 82 % of the GEI variation in grain yield, making the WAASB and WAASBY indices more effective in identifying superior genotypes. On the other hand, the variation of certain genotypes is explained by more IPCAs, making indices such as WAASB necessary, which is calculated using additional principal components. Consequently, we utilized this index and identified genotypes G1, G3, G5, G12, and G18 as superior (Fig. 2). The biplot based on this index was also employed to assess the discriminating ability of environments, revealing that the environments in the first two sections of biplot type III (Fig. 2), particularly E6, exhibited the highest WAASB value and strong genotype discrimination capability. However, the genotypes in these two quadrants with a significant contribution to the GEI were unstable. In the fourth quadrant of this biplot, genotypes G18, G1, G3, G5, and G12 emerged as the most desirable, demonstrating high yield stability and performance. The ranking of rice genotypes with varying weights for grain yield performance and stability in Figure 3 illustrated the diversity among rice mutant lines. In this heatmap, scenarios with different weights of yield stability and performance were simulated to display changes in genotype ranking. The grain yield of mutant lines G5, G12, G2, G7, and G1 surpassed that of parental varieties and Gilaneh, the control variety.

The HMRPGV method evaluates grain yield performance, stability, and adaptability simultaneously within a genotypic context (Resende, 2007). The HMRPGV of 1.306 for G5 indicated a 30.6 % increase in grain yield over the general mean. It appears that yield performance is prioritized over stability in the HMRPGV, and the selection of genotypes is based more on yield performance (as compared to Fig. 3) than on stability.

When data on several traits are available, the multi-trait stability index (MTSI) has many useful applications for simultaneous selection of mean performance and stability. This index is an MTSI-based on factor analysis, designed for simultaneous selection using both fixed and mixed effect models for multiple traits (Olivoto et al., 2019b). The factor-analytic variance-covariance structure accommodates different variances across environments and covariances between pairs of environments by approximating a completely unstructured covariance matrix through factor analysis. Therefore, it is reasonable to assume that the model will fit better than a standard mixed effect model. This factor analysis of WAASBY, which is the first step in computing the MTSI index, revealed that the eight traits were grouped into two factors. A Euclidean distance was then used to compute the distance between the genotypes' scores and the ideotype's score. This plot (Fig. 4) displays the genotype ranking in ascending order for the MTSI index. The selected genotypes using MTSI, including G7, G5, and G1, are shown in red, and the spiral depicts the cutpoint according to the selection pressure. A higher weight for mean performance was assigned to tiller number, panicle length, filled grains per panicle, fertility percentage, hundred grain weight, and grain yield. It is preferable to select highly productive genotypes that may not perform well in terms of broad stability and then consider the adaptability of these genotypes to specific environments to make better recommendations. The MTSI index enables simultaneous selection of genotypes based on performance and stability across all traits, making it a unique

index (Olivoto et al., 2019a). This index has been used for the selection of superior genotypes in *Solanum melongena* (Koundinya et al., 2019), *Brassica* spp. (Bocianowski et al., 2019), chickpea (Sellami et al., 2021), maize (Olivoto et al., 2021), and rice (Sharifi et al., 2021).

MTSI serves as an effective tool for identifying the strengths and weaknesses of genotypes and selecting those with desired mean performance and stability (Olivoto et al., 2021). This approach enabled the identification of genotypes that combine the desired mean performance and stability (MPE) for important traits such as grain yield (GY), hundred-grain weight (HGW), fertility percentage (FP), and tiller number (TN) (Fig. 5).

Conclusion

The significant effect of genotype indicated that the mutation process has successfully created a reasonable level of diversity in the genetic material. The first three principal components accounted for 92.35 % of the GEI variation with respect to grain yield, and the WAASB index was used to evaluate the stability of the genotypes. Biplot type III analysis revealed that genotypes G18, G1, G3, G5, and G12 were both high-yielding and stable across the tested environments. The heatmap visualization of yield performance and WAASB stability scores indicated that genotypes G3, G9, G6, G12, and G5 were highly productive and stable. The MTSI (multi-trait stability index) identified genotypes G7, G5, and G1 as the selected genotypes based on the overall stability and performance across all measured traits. According to the HMRPGV index, G5, G12, G7, G2, and G1 were identified as the top-performing genotypes. In conclusion, two genotypes, G12 and G5, exhibited a significant advantage over all other genotypes; however, only G5 was selected based on all traits in the MTSI index and could therefore undergo selection or cultivar introduction processes.

References

Ahakpaz F., Abdi H., Neyestani E., Hesami A., Mohammadi B., Nader Mahmoudi K., Abedi-Asl G., Jazayeri Noshabadih M.R., Ahakpaz F. Genotype-by-environment interaction analysis for grain yield of barley genotypes under dryland conditions and the role of monthly rainfall. *Agric Water Manage.* 2021;245. doi 10.1016/j.agwat.2020.106665

Akter A., Hasan M.J., Kulsum M.U., Rahman M.H., Paul A.K., Lipi L.F., Akter S. Genotype × environment interaction and yield stability analysis in hybrid rice (*Oryza sativa* L.) by AMMI biplot. *Bangladesh Rice J.* 2015;19(2):83-90. doi.org/10.3329/brj.v19i2.28168

Allahgholipour M. Analysis of grain yield stability of new rice (*Oryza sativa* L.) genotypes originated from Iranian local cultivars. *Iranian J Crop Sci*. 2017;18(4):288-301. (in Persian with English abstract).

Balestre M., dos Santos V.B., Soares A.A., Reis M.S. Stability and adaptability of upland rice genotypes. *Crop Breed Appl Biotech*. 2010;10:357-363. doi 10.1590/S1984-70332010000400011

Bocianowski J., Niemann J., Nowosad K. Genotype-by-environment interaction for seed quality traits in interspecific cross-derived *Brassica* lines using additive main effects and multiplicative interaction model. *Euphytica*. 2019;215(1):7. doi 10.1007/s10681-018-2328-7

Bose L.K., Jambhulkar N.N., Pande K., Singh O.N. Use of AMMI and other stability statistics in the simultaneous selection of rice genotypes for yield and stability under direct-seeded conditions. *Chil J Agric Res*. 2011;74(1):3-9. doi 10.4067/S0718-58392014000100001

Cheema A.A. Mutation breeding for rice improvement in Pakistan: achievements and impact. *Plant Mut Rep*. 2006;1(1):36-39. https://www-pub.iaea.org/MTCD/publications/PDF/Newsletters/PMR-01-01.pdf

Coan M.M.D., Marchioro V.S., Franco F.A., Pinto R.J.B., Scapim C.A., Baldissera J.N.C. Determination of genotypic stability and adaptability in wheat genotypes using mixed statistical models. *J Agric Sci Tech*. 2018;20:1525-1540. https://jast.modares.ac.ir/article-23-20190-en.pdf

Colombari-Filho J.M., Resende M.D.V., de Morais O.P., Castro A.P., Guimaraes E.L., Pereira J.M., Utumi M.M., Breseghello F. Upland rice breeding in Brazil: a simultaneous genotypic evaluation of stability, adaptability and grain yield. *Euphytica.* 2013;192:117-129. doi 10.1007/s10681-013-0922-2

Dewi A.K., Dwimahyani I. Grain yield stability analysis of Jembar local rice mutant lines generated from mutation breeding. *IOP Conf Ser Earth Environ Sci*. 2019;230:012112. doi 10.1088/1755-1315/230/1/012112

Dia M., Wehner T.C., Arellano C. Analysis of genotype × environment interaction (G×E) using SAS programming. *Agron J.* 2016;108:1838-1852. doi 10.2134/agronj2016.02.0085

Donoso-Ñanculao G., Paredes M., Becerra V., Arrepol C., Balzarini M. GGE biplot analysis of multi-environment yield trials of rice produced in a temperate climate. *Agric Res*. 2015;76(2):152-157. doi 10.4067/S0718-58392016000200003

Dushyanthakumar B.M., Shadadashari Y.G. Stability analysis of P.U. Belliyappa local rice mutants. *Karnataka J Agric Sci*. 2007;20(4):724-726. http://14.139.155.167/test5/index.php/kjas/article/viewFile/1032/1024

Ebadi A.A., Hallajian M.T., Kordrostami M. The genetic variation and stability analysis of rice mutant lines using AMMI model under normal and drought stress conditions. *Genetika.* 2019;51(2):687-699. doi 10.2298/GENSR1902687E

Farshadfar E. Incorporation of AMMI stability value and grain yield in a single non-parametric index (GSI) in bread wheat. *Pakistan J Biol Sci*. 2008;11(14):1791-1796. doi 10.3923/pjbs.2008.1791.1796

Gauch H.G. A simple protocol for AMMI analysis of yield trials. *Crop Sci*. 2013;53(5):1860-1869. doi 10.2135/cropsci2013.04.0241

Kang M.S. A rank-sum method for selecting high-yielding, stable corn genotypes. *Cereal Res Comm*. 1988;16:113-115

Karimizadeh R., Pezeshkpour P., Barzali M., Mehraban A., Sharifi P. Evaluation the mean performance and stability of lentil genotypes by combining features of AMMI and BLUP techniques. *J Crop Breed*. 2020;12(36):160-170. (In Persian with English abstract). doi 10.52547/jcb.12.36.160

Khush G.S. What it will take to feed 5.0 billion rice consumers in 2030. *Plant Mol Biol*. 2005;59:1-6. doi 10.1007/s11103-005-2159-5

Koundinya A.V.V., Pandit M.K., Ramesh D., Mishra P. Phenotypic stability of eggplant for yield and quality through AMMI, GGE and cluster analyses. *Sci Hort*. 2019;247:216-223. doi 10.1016/j.scienta.2018.12.019

Nayak D., Bose L.K., Singh S., Nayak P. Additive main effects and multiplicative interaction analysis of host-pathogen relationship in rice-bacterial blight pathosystems. *Plant Path J*. 2008;24(3):337-351

Olivoto T., Lúcio A.D. Metan: an R package for multi-environment trial analysis. *Meth Ecol Evol*. 2020;11:783-789. doi 10.1111/2041-210X.13384

Olivoto T., Lúcio A.D.C., da Silva J.A.G., Marchioro V.S., de Souza V.Q., Jost E. Mean performance and stability in multi-environment trials I: combining features of AMMI and BLUP techniques. *Agron J.* 2019a;111(6):2949-2960. doi 10.2134/agronj2019.03.0220

Olivoto T., Lúcio A.D.C., da Silva J.A.G., Sari B.G., Diel M.I. Mean performance and stability in multi-environment trials II: selection based on multiple traits. *Agron J*. 2019b;111(6):2961-2969. doi 10.2134/agronj2019.03.0221

Olivoto T., Nardino M., Meira D., Meier C., Follmann D.N., de Souza V.Q., Konflanz V.A., Baretta D. Multi-trait selection for mean per-

formance and stability in maize. *Agron J.* 2021;113:3968-3974. doi 10.1002/agj2.20741

Patterson H.D., Thompson R. Recovery of interblock information when block sizes are unequal. *Biometrics*. 1971;58:545-554. doi 10.2307/2334389

Piepho H.P. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theor Appl Genet*. 1994;89(5):647-654. doi 10.1007/BF00222462

Rahayu S. Yield stability analysis of rice mutant lines using AMMI method. *J Phys Conf Ser.* 2020;1436:012019. doi 10.1088/1742-6596/1436/1/012019

Resende M.D.V. Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético. Embrapa Florestas, Colombo, Brazil, 2007

Rodovalho M.A., Coan M.M.D., Scapim C.A., Pinto R.J.B., Contreras-Soto R.I. Comparison of HMRPGV, Lin and Binns's and Annichiarico's methods for maize hybrid selection for high and stable yield. *Maydica.* 2015;60(1):1-7

Santos F., Marza F. Selection of forage oat genotypes through GGE Biplot and BLUP. *bioRxiv*. 2020. doi 10.1101/2020.03.10.986422

Searle S.R., Casella G., McCulloch C.E. Variance Components. John Wiley and Sons, New York, USA, 1992

Sellami M.H., Lavini A., Pulvento C. Phenotypic and quality traits of chickpea genotypes under rainfed conditions in South Italy. *Agronomy*. 2021;11:962. doi 10.3390/agronomy11050962

Sharifi P. Evolution, domestication, breeding methods and the latest breeding findings in rice. Agric Natural Res Engining Org of Iran. 2020. (in Persian)

Sharifi P., Aminpanah H., Erfani R., Mohaddesi A., Abbasian A. Evaluation of genotype×environment interaction in rice based on AMMI model in Iran. *Rice Sci*. 2017;24(3):173-180. doi 10.1016/j.rsci.2017.02.001

Sharifi P., Erfani A., Abbasian A., Mohaddesi A. Stability of some of rice genotypes based on WAASB and MTSI indices. *Iranian J Gen Plant Breed*. 2021;9(2):1-11. doi 10.30479/IJGPB.2021.14432.1283

Shu Q.Y. Plant Mutation and Biotechnology. CABI, 2012

van Eeuwijk F.A., Bustos-Korts D.V., Malosetti M. What should students in plant breeding know about the statistical aspects of genotype × environment interactions? *Crop Sci*. 2016;56(5):2119-2140. doi 10.2135/cropsci2015.06.0375

Verma A., Singh G.P. Simultaneous application of AMMI measures and yield for stability analysis of wheat genotypes evaluated under irrigated late sown conditions of Central Zone of India. *J Appl Nat Sci*. 2020;12(4):541-549. doi 10.31018/jans.v12i4.2391

Yan W. Analysis and handling of G × E in a practical breeding program. *Crop Sci*. 2016;56:2106-2118. doi 10.2135/cropsci2015.06.0336

Yan W. A systematic narration of some key concepts and procedures in plant breeding. *Front Plant Sci*. 2021;12:724517. doi 10.3389/fpls.2021.724517

Yan W., Hunt L.A., Sheny Q., Szlavnics Z. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci*. 2000;40:597-605 doi 10.2135/cropsci2000.403597x

Yan W., Pageau D., Frégeau-Reid J., Durand J. Assessing the representativeness and repeatability of test locations for genotype evaluation. *Crop Sci*. 2011;51:1603-1610. doi 10.2135/cropsci2011.01.0016

Data availability statement. The data that support this study cannot be publicly shared due to ethical or privacy reasons and may be shared upon reasonable request to the corresponding author if appropriate.

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

Received March 11, 2025. Revised August 8, 2025. Accepted August 8, 2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):233-240

doi 10.18699/vjgb-26-26

The role of explant type and selective agent application in the initial transformation rate of *Lens culinaris* Medik.

T.V. Dyubenko [1], K.V. Smirnov [2, 3], V.E. Tvorogova [2, 4]

[1] Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
[2] Saint Petersburg State University, St. Petersburg, Russia
[3] All-Russia Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia
[4] Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russia

krubaza@mail.ru

Abstract. *Lens culinaris* Medik. (lentil) is an agronomically important leguminous species, but its genome modification is rarely used for obtaining new varieties, probably due to a low efficiency of transformation protocols. Development of universal, genotype-independent protocols for obtaining transgenic plants usually relies, among other factors, on the possibility of obtaining a substantial number of transgenic cells *in vitro*. This study aimed to adapt a previously developed *Agrobacterium*-mediated transformation protocol, used for a related legume, for the production of transgenic callus tissue in *L. culinaris*. We used two different markers of transgenic tissue, beta-glucuronidase and green fluorescent protein, to find an optimal type of explant for obtaining transgenic tissue in lentil. We also evaluated the impact of hygromycin, a common selective agent, on the amount of transgenic tissue in developing transformed explants of *L. culinaris*. According to our results, the transformation protocol commonly used for *Medicago truncatula* Gaertn. leaf explants can also be applied for obtaining transgenic calli from *L. culinaris* shoot apices. Explants from shoot apices demonstrated higher initial transformation rate in comparison with explants from roots, stems and leaves. Moreover, explants of different types, which were cultivated on medium without hygromycin, developed significantly fewer calli expressing reporter genes than those grown on hygromycin-containing medium, confirming that hygromycin may be used as an effective selection agent for lentil. During our analysis, we noticed GUS-like staining in calli which didn't contain plasmids for *GUS* gene expression. This can be explained with so-called intrinsic GUS-like activity, which was described in previous research. These data can be used for further development of effective and universal *L. culinaris* transformation and genome editing protocols.

Key words: lentil; hygromycin; GUS; GFP; transformation; *Lens culinaris*

For citation: Dyubenko T.V., Smirnov K.V., Tvorogova V.E. The role of explant type and selective agent application in the initial transformation rate of *Lens culinaris* Medik. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):233-240. doi 10.18699/vjgb-26-26

Funding. The study is performed in the framework of the State Assignments for FRC VIR No. FGEM-2022-0002 "Identification of the possibilities of the gene pool of legumes to optimize their selection and diversification of use in various sectors of the national economy". The research was also funded by the Sirius University of Science and Technology project: PBB-RND-2243.

Acknowledgements. Authors are grateful to Veronika Y. Simonova for the help with literature analysis.

Роль типа экспланта и применения селективного агента в первичной эффективности трансформации *Lens culinaris* Medik.

Т.В. Дюбенко [1], К.В. Смирнов [2, 3], В.Е. Творогова [2, 4]

[1] Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия
[2] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
[3] Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии, Пушкин, Санкт-Петербург, Россия
[4] Научно-технологический университет «Сириус», федеральная территория «Сириус», Краснодарский край, Россия

krubaza@mail.ru

Аннотация. *Lens culinaris* Medik. (чечевица) – важная сельскохозяйственная культура из семейства бобовых, но модификация ее генома редко используется для создания новых сортов. Вероятно, это связано с низкой эффективностью существующих протоколов трансформации. Как правило, разработка универсальных, независимых от генотипа протоколов получения трансгенных растений зависит, среди прочего, от возможности образования существенного количества трансгенных клеток *in vitro*. В настоящем исследовании мы предприняли попытку адаптировать разработанный ранее для другого бобового протокол агробактериальной трансформации для получения трансгенной каллусной ткани у *L. culinaris*. Чтобы оценить эффективность трансформации, мы выбрали две независимые репортерные системы: бета-глюкуронидазу и зеленый флуоресцентный белок. С помощью этих систем мы нашли оптимальный тип экспланта для создания каллусной ткани чечевицы, экспрессирующей рекомбинантную ДНК. Мы также оценили влияние гигромицина, одного из распространенных селективных агентов, на количество

© Dyubenko T.V., Smirnov K.V., Tvorogova V.E., 2026
This work is licensed under a Creative Commons Attribution 4.0 License

трансгенной ткани в развивающихся трансформированных эксплантах *L. culinaris*. Согласно нашим результатам, протокол трансформации, который обычно применяется для эксплантов листьев *Medicago truncatula* Gaertn., может быть использован и для получения трансгенных каллусов из верхушек побегов *L. culinaris*. Экспланты из апексов побегов продемонстрировали более высокую первичную эффективность трансформации по сравнению с эксплантами из корней, стеблей и листьев. Кроме того, экспланты разных типов, культивируемые на среде без гигромицина, образовывали значительно меньше каллусов, экспрессирующих репортерные гены, по сравнению с эксплантами, выращиваемыми на среде с гигромицином. Это подтверждает возможность использования гигромицина в качестве эффективного селективного агента для чечевицы. В ходе нашего анализа мы также обнаружили GUS-подобное окрашивание в каллусах, не содержащих плазмид для экспрессии гена *GUS*, что можно объяснить так называемой внутренней GUS-подобной активностью, которая была описана в предыдущих исследованиях. Эти данные могут быть полезны для дальнейшей разработки эффективных и универсальных протоколов трансформации и редактирования генома *L. culinaris*.

Ключевые слова: чечевица; гигромицин; GUS; GFP; трансформация; *Lens culinaris*

Introduction

Lentils (*Lens culinaris* Medik.) are one of the oldest domesticated crops (Erskine et al., 2009), the cultivation history of which dates back over 8,000 years (Liber et al., 2021). Lentil production has doubled since the beginning of the 21st century (Kaale et al., 2023), coinciding with a rising societal interest in plant-based meat alternatives and growing trends towards more sustainable food production (Runte et al., 2024). Being a leguminous plant, *L. culinaris* is used both as a nitrogen-fixing crop improving the soil fertility and as a source of protein for human nutrition. The level of carbohydrates, including starch and dietary fiber, is also quite high in lentils, making them a valuable crop among other legumes. Furthermore, biochemical properties, specific for different lentil cultivars, provide a wide range of lentil applications in food production (Kaale et al., 2023).

Despite many features, which make lentils a valuable crop, its production and usage are related with diverse obstacles, including undesirable traits common for most legumes and also drawbacks which are specific for this culture. The features which require alleviation in lentil, include high content of different anti-nutrients, such as phytate, as well as low levels of some amino acids (Joshi et al., 2017), pod shattering (Cao et al., 2024), etc.

Advancements in next-generation sequencing led to the discovery of many genetic markers and genes related to different desirable and undesirable traits in lentils (Kumar et al., 2023). These data can be widely used for selection and breeding of lentil cultivars with specific useful features. However, currently this information may rarely be used for genome modification of lentils, because, similarly to many legumes, the transformation and regeneration effectiveness of this species is quite low. There are only a few studies on *L. culinaris* transformation, and, to our knowledge, all of them report on low effectiveness of obtaining transgenic lentil plants (Gulati et al., 2002; Celikkol Akcay et al., 2009; Khatib et al., 2011; Chopra et al., 2012; Das et al., 2012, 2019; Polowick, Yan, 2023), which doesn't exceed 7 percents (i. e. 7 transgenic plants per 100 explants) for individual cultivars. Moreover, these protocols are developed for specific lentil varieties and, therefore, their effectiveness may depend greatly on genotype transformed.

The transformation effectiveness depends both on the success of foreign DNA introduction into plant cells and on further fate of the transformed cell, i. e. its ability to give rise to a new plant. Consequently, these parameters are influenced by many factors accompanying transformation, including the method of transformation, explant type, cultivation medium and the method of transformant selection. A significant part of plant transformation protocols includes the stage of obtaining transgenic callus. At this stage, modified cells proliferate and some of them acquire the capacity for regeneration. Therefore, the efficiency of transgenic callus formation plays an important role in overall transformation efficiency. Moreover, maximization of transgenic tissue yield for a plant species facilitates the screening for morphogenic regulators, i. e. genes that are able to induce plant regeneration from transgenic tissue and can be used for effective cultivar-independent transformation (Bakulin et al., 2025).

In the current study, we made an attempt to find an optimal type of explant for obtaining transgenic callus tissue in lentil. We also evaluated the impact of hygromycin, a common selective agent, on the amount of transgenic tissue in developing transformed explants of *L. culinaris*. We built our transformation protocol on the highly effective transformation system used for *Medicago truncatula*, a well-studied model leguminous plant (Cosson et al., 2006). This system includes the formation of a significant amount of transgenic callus tissue. Moreover, as a model legume, *M. truncatula* provides an efficient platform for identifying key morphogenic regulators (Yakovleva et al., 2024), which can subsequently be validated in *L. culinaris.* The optimized protocol presented in this study provides a foundation for such validation experiments.

Materials and methods

Plant material and bacterial strains. *L. culinaris* Ekaterininskaya zelenosemyannaya cultivar (k-2942) seeds from the collection of the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR) were used in the study. For cloning, the *Escherichia coli* DH10B strain was used. For *Agrobacterium*-mediated transformation, the *Rhizobium radiobacter* (*Agrobacterium tumefaciens*) AGL1 strain was used. *E. coli* and *A. tumefaciens* cultivation and transformation were performed as described in (Potsenkovskaia et al., 2022).

Plasmids used in the study. For *L. culinaris* transformation, two plasmids were used: pMDC32_GUS (Tvorogova et al., 2019) with the beta-glucuronidase (GUS) gene and pMDC32_eGFPer obtained in this study. To construct the pMDC32_eGFPer plasmid, the *eGFPer* gene, encoding the enhanced green fluorescent protein (GFP) with endoplasmic re-

ticulum localization signal, was amplified from the pB7WG2D vector (Karimi et al., 2002) with primers containing attB1 and attB2 sites (forward primer: GGGGACAAGTTTGT ACAAAAAAGCAGGCTTCATGGTGAAGACTAATC TTTTTCTC; reverse primer: GGGGACCACTTTGTACA AGAAAGCTGGGTCTTACAGCTCGTCCTTCTTG) and cloned into the pDONR207 entry vector (Invitrogen) and then to the pMDC32 vector (Curtis, Grossniklaus, 2003) using Gateway technology (Invitrogen).

***Agrobacterium*-mediated transformation.** For explants preparation, seeds of *L. culinaris* were incubated in concentrated sulfuric acid for 7 minutes, rinsed with tap water 7 times, then incubated in 5–7 % sodium hypochlorite solution (the "Belizna" bleach) for 8 minutes and rinsed with sterile water 10 times in a laminar hood. After sterilization, seeds were put on modified Fahraeus medium (Fåhraeus, 1957; Tvorogova et al., 2019) and left to germinate for 12 days at 21–24 °C, 16 h (light)/8 h (dark) photoperiod. After that, germinated seedlings were used to prepare explants. Root explants (about 10–15 mm length) were obtained from root tips, shoot apex explants (about 10–15 mm length) were taken from the main and/or lateral shoots. Stem explants (about 3–8 mm length) were obtained from stem fragments between the first and the second leaves of the seedlings. For leaf explants, one or two first true leaves were taken. Seedling growth was asynchronous, and the explant size depended on the seedling development stage.

For *Agrobacterium*-mediated transformation itself, a modified protocol of Cosson et al. (2006) for *M. truncatula* Gaertn. was used. Agrobacterial strain carrying either pMDC32_GUS or pMDC32_eGFPer plasmids was recovered from glycerol stock (500 µl of 50 % glycerol mixed with 500 µl of overnight stationary-phase culture) stored at –80 °C and cultured on solid YEP medium (Table S1)[1] containing kanamycin (50 mg/l) and rifampicin (40 mg/l). Kanamycin resistance was provided by the pMDC32_GUS or pMDC32_eGFPer plasmid, whereas rifampicin resistance was characteristic of the AGL1 strain. On such medium, bacteria were cultivated at 30 °C for 1–2 days. Then, a small amount of bacteria from the Petri dish was sown to the liquid YEP with the same concentrations of antibiotics, in which they were grown overnight in a thermoshaker at 30 °C and 200 rpm. On the next day, 2 ml of bacterial culture were transferred to 50 ml of AB-MES medium (Wu et al., 2014, Table S1) with the addition of 40 mg/l (200 µM) acetosyringone, kanamycin (50 mg/l) and rifampicin (40 mg/l). Bacteria were cultivated in this medium at 30 °C and 200 rpm for about 3 hours. Then, they were centrifuged (4,000g, 15 minutes, room temperature) and resuspended in liquid ABM-SH medium (Table S1) with 4 mg/l (18 µM) 2,4-dichlorophenoxyacetic acid (2,4-D), 0.5 mg/l (2.22 µM) 6-benzylaminopurine (BAP), and 40 mg/l (200 µM) acetosyringone to the final OD600 0.1. ABM-SH represented a mixture of 0.5X AB-MES components and 0.5X modified liquid PCI-4 medium (Hoffmann et al., 1997) components, similarly to the infiltration medium from the AGROBEST protocol for *Arabidopsis thaliana* (L.) Heynh. (Wu et al., 2014).

[1] Supplementary Table S1 and Figures S1, S2 are available at: https://vavilovj-icg.ru/download/pict-2026-30/appx17.pdf

Prepared explants were incubated in *Agrobacterium* suspension in ABM-SH medium for 15 minutes with slow gentle agitation. After the inoculation stage, explants were blotted on sterile filter paper, then put on solid ABM-SH medium (Table S1) and cultivated in the dark at 21–24 °C for about 36 hours. After cocultivation, explants were put on solid PCI-4 medium (Hoffmann et al., 1997, Table S1) for callus formation with 4 mg/l (18 µM) 2,4-D, 0.5 mg/l (2.22 µM) BAP, and with the addition of cefotaxime (250 mg/l), to eliminate agrobacteria. In some experiments, hygromycin (10 mg/l), necessary for the selection of transgenic plant cells, was added to the medium. On this medium, explants were cultivated for 44 days, with transfer on the fresh medium every 14–16 days. At the end of cultivation, explants were either stained with X-Gluc and analysed with stereomicroscope to detect *GUS* expression, or analysed with the fluorescent stereomicroscope to detect GFP fluorescence.

GUS staining. For GUS staining, X-Gluc substrate (Sisco Research Laboratories Pvt. Ltd.) was used. Plant tissues were collected in NT buffer (100 mM Tris-HCl, 50 mM NaCl) and then transferred to X-Gluc buffer (100 mM Tris-HCl, 50 mM NaCl, 0.2 mM $K_3[Fe(CN)_6]$, 2.5 mM X-Gluc), where they were incubated at 37 °C overnight. After staining, tissues were transferred to 70 % ethanol. Photos of single explants were taken with a ZEISS SteREO Discovery.V12 stereomicroscope. Photos of Petri dishes with explants were taken with the Vilber Fusion-FX6.Edge imaging system. For microscopic images, pieces of callus tissue were squashed and examined under a Leica DM500 microscope.

Statistical analysis and software. For evaluation of explant type and hygromycin presence impact, 20–28 explants per group were used. Statistical significance of differences was evaluated with Fisher's exact test with Bonferroni adjustment for multiple comparisons, if necessary. Statistical analysis of results and diagram drawing were performed in the R studio using the vcd (Meyer et al., 2006, 2024) and dplyr (Wickham et al., 2025) packages. Final figures were created using the Inkscape software.

Results

Evaluation of transgenic tissue development from different explant types using the GFP marker

We decided to build our transformation protocol on the transformation system used for *M. truncatula* (Cosson et al., 2006), in which transgenic plants are developed from transformed leaf explants through somatic embryogenesis. To find the optimal variant of explant type for transformation, we performed *Agrobacterium*-mediated transformation of root, stem, leaf and shoot apex explants with plasmid pMDC32_eGFPer for *eGFPer* overexpression. After cocultivation, explants were transferred to the callus-inducing medium with or without hygromycin. The concentrations of plant growth regulators and cefotaxime for agrobacteria elimination were identical to the ones used for *M. truncatula*, but the concentration of hygromycin was lower (10 mg/l instead of 25 mg/l used for *M. truncatula*), because 10 mg/l concentration is successfully used for other legumes to obtain transgenic tissue (Olhoft et

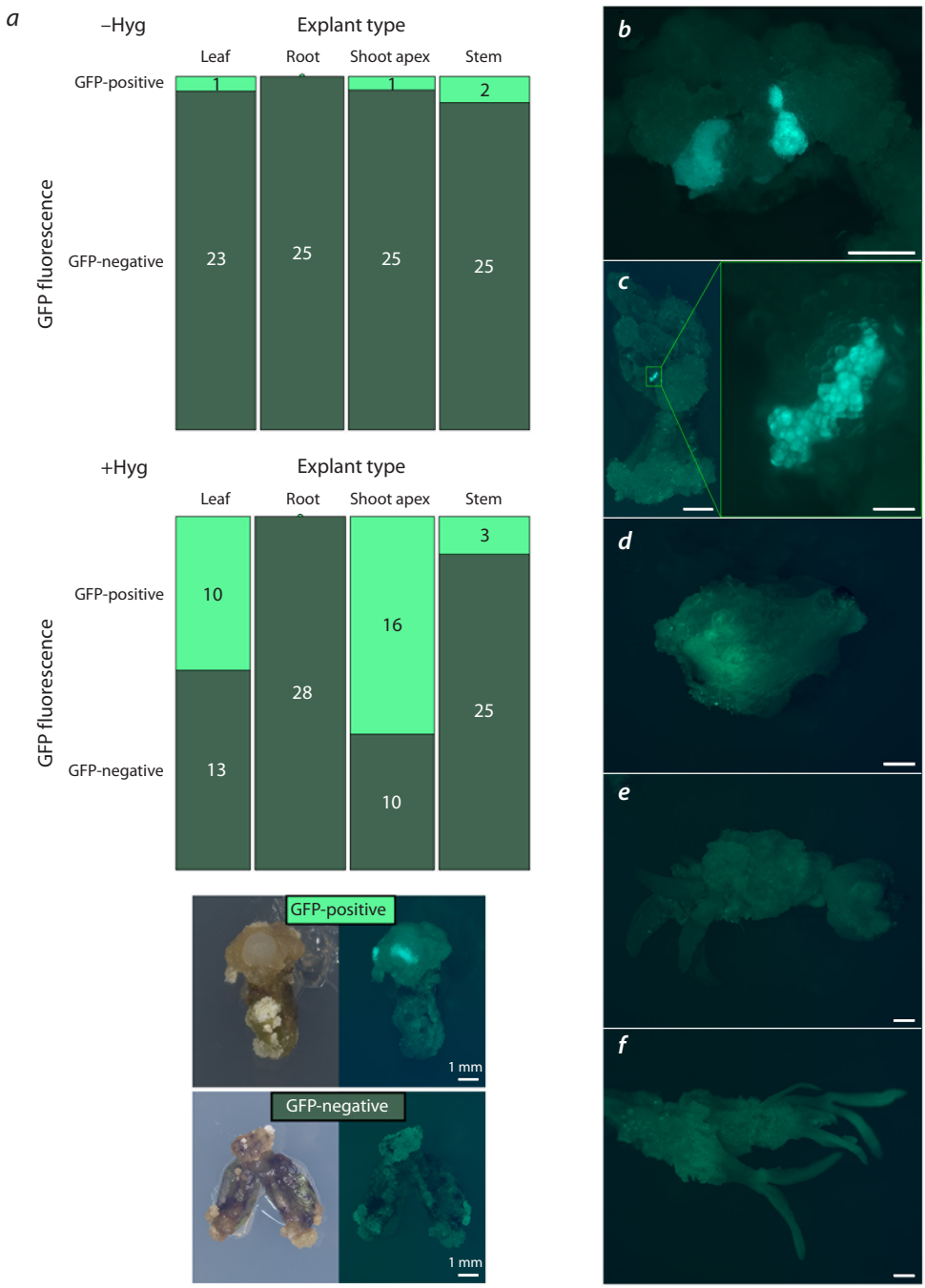


Fig. 1. *a*, Mosaic plots demonstrating the number of calli with GFP fluorescence (GFP-positive) and without GFP fluorescence (GFP-negative) developed from different types of explants, on the medium without hygromycin (above) or on the medium with addition of 10 mg/l hygromycin (below). In a legend below, the photos of explants were taken in white light (left) and fluorescent (right) light. *b*–*d*, Calli with GFP fluorescence. *e*, *f*, Calli without detectable GFP fluorescence, obtained after explant transformation with the pMDC32_GFP (*e*) or pMDC32_GUS (*f*) plasmids.

Size bar is 1 mm for *b*–*f* and 100 μm for the inset in *c*.

al., 2003; Li et al., 2023; Tisseyre et al., 2024). After 46 days of cultivation on this callus-inducing medium, calli were developed from all types of explants both on the medium with hygromycin and hygromycin-free medium (Figs. S1, S2). We analysed GFP fluorescence to evaluate the effectiveness of transformation.

To avoid false positive results, we also analysed the green fluorescence in the explants transformed with pMDC32_GUS construction, which didn't contain the *GFP* gene. Bright GFP fluorescence in separate tissue areas (Fig. 1*b*, *c*), as well as less bright diffuse fluorescence (Fig. 1*d*), were easily distinguishable from autofluorescence (Fig. 1*e*, *f*), which was observed in control pMDC32_GUS explants. It is worth noting that some fluorescent areas with weak signal, as well as autofluorescent areas, can only be seen with high magnification (60× and more). Therefore, for robust comparative analysis, it was

crucial to evaluate the presence of fluorescence for all explants without changing magnification and other visualization settings on the stereomicroscope.

According to our results, on the medium supplemented with 10 mg/l hygromycin, the type of explant had an impact on the frequency of fluorescent calli (Fisher's exact test, *p*-value 4.359e–08). The highest frequency of transformed calli with GFP fluorescence was detected for explants from shoot apices (Fig. 1*a*). Leaf explants also demonstrated a rather high frequency of transformation, whereas only 3 out of 28 stem explants developed fluorescent calli. Calli from the root explants didn't demonstrate any transgenic fluorescent tissue at all (Fig. 1*a*). Overall, shoot apex explants formed significantly more fluorescent calli in comparison with explants from roots or stems (Fisher's exact test with Bonferroni adjustment for multiple comparisons, *p*-values 1.51e–06 and 8.46e–04, respectively), whereas differences between explants from shoot apices and leaves were not statistically significant.

At the same time, explants of all types, cultivated on medium without hygromycin, developed significantly fewer fluorescent calli than those grown on hygromycin-containing medium (Fisher's exact test, *p*-value 2.09e–06), confirming that hygromycin may be used as an effective selection agent for lentil (Fig. 1*a*).

Evaluation of transgenic tissue development from different explant types using GUS marker

We performed a similar experiment with transformation of different explant types using GUS, another transgenic tissue marker. Explants were transformed with the pMDC32_GUS plasmid (Tvorogova et al., 2019) using the protocol mentioned above. After 46 days of cultivation (starting from the day of transformation), GUS activity was evaluated in explants. Usage of GUS allowed us to detect single stained cells (at about 80× magnification on stereomicroscope), which we were not able to visualize with the GFP marker. The results of transgenic tissue detection were similar to those obtained using the GFP marker. On the media with hygromycin, the highest transformation efficiency was shown for shoot apex explants, followed by leaf explants. Both root and stem explants demonstrated a low transformation rate (Fig. 2). In total, shoot apex explants formed significantly more stained calli in comparison with explants from leaves, roots or stems (Fisher's exact test with Bonferroni adjustment for multiple comparisons,

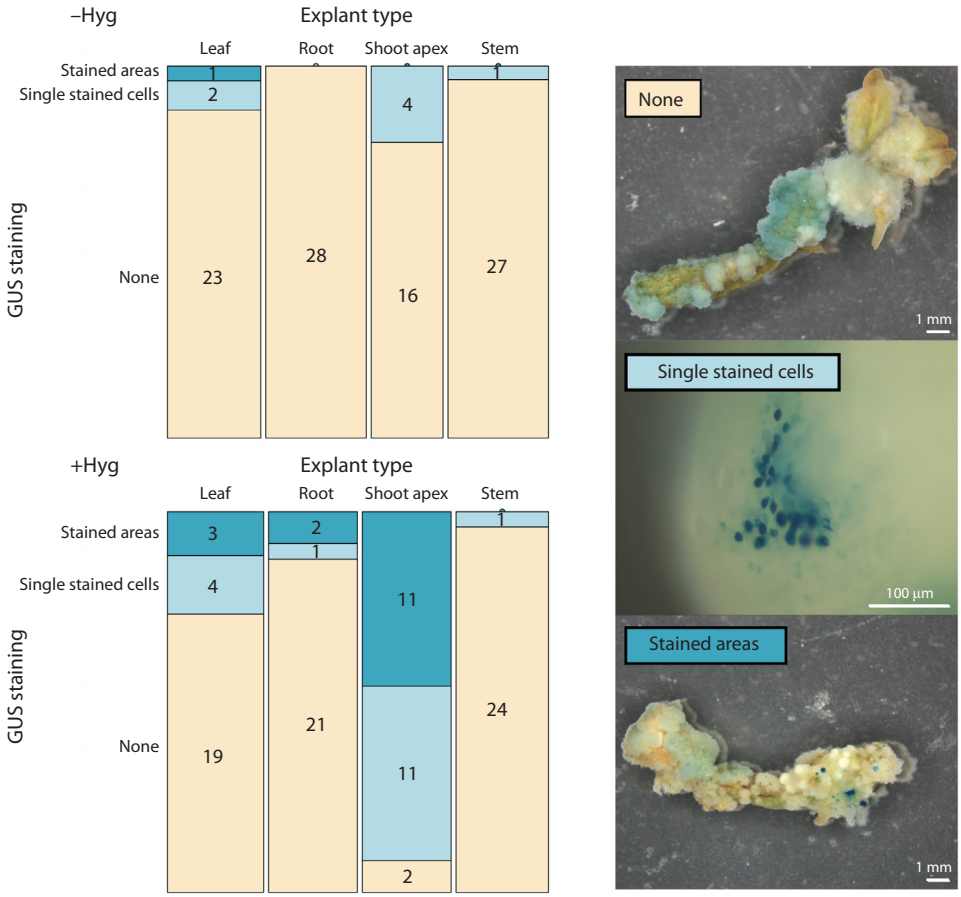


Fig. 2. Mosaic plots demonstrating the number of stained and non-stained calli developed from different types of explants after transformation with pMDC32_GUS plasmid, on the medium without hygromycin (above) or on the medium with addition of 10 mg/l hygromycin (below). The legend on the right demonstrates calli belonging to different categories relative to GUS staining: none – no detectable GUS staining apart from the light-blue color characteristic of endogenous GUS-like activity; single stained cells – small spots of GUS staining visible only with high (about 80×) magnification, stained areas – patches of GUS staining visible with naked eye or with low magnification.

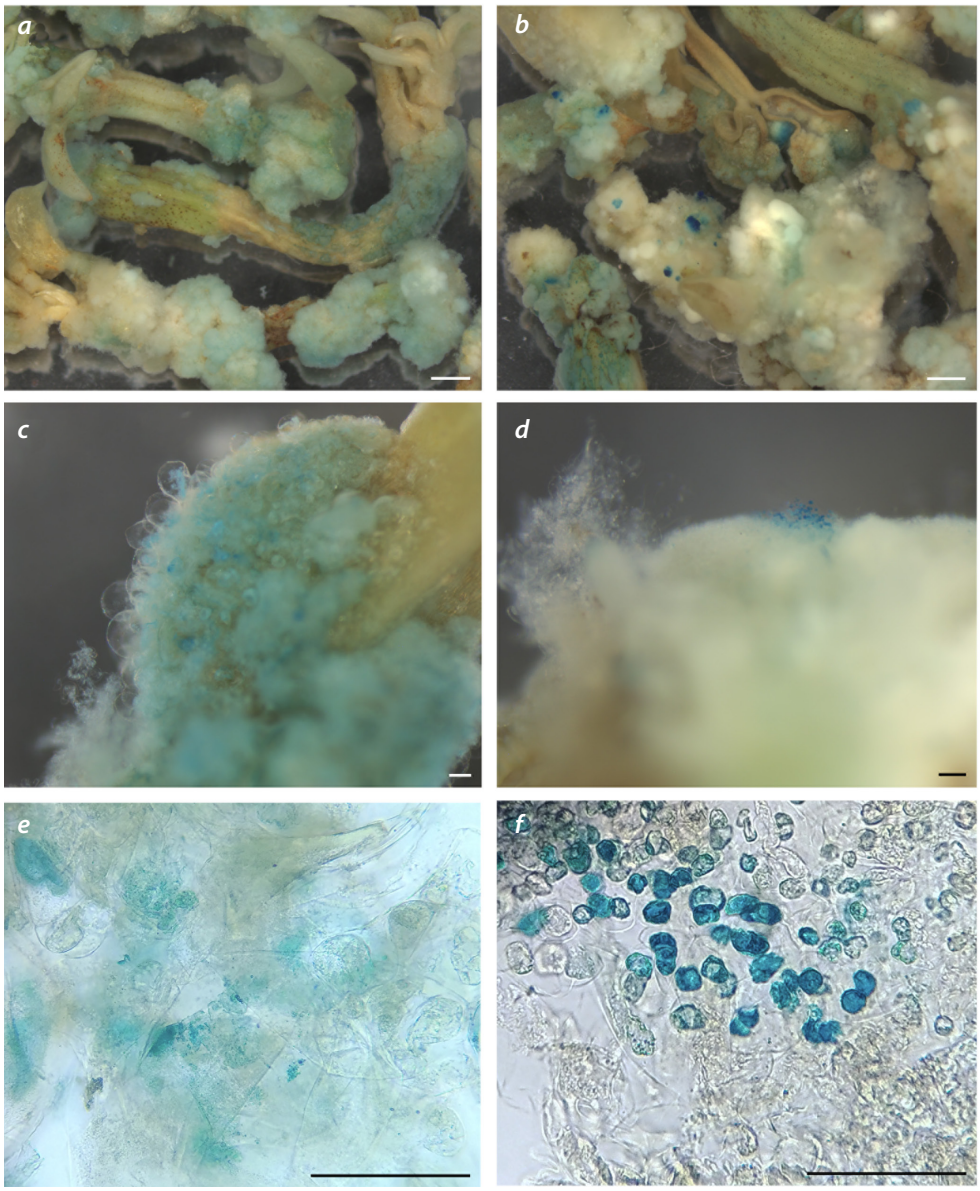


Fig. 3. GUS staining of control calli obtained from explants transformed with the pMDC32_eGFPer plasmid (*a*, *c*, *e*) or calli obtained from explants transformed with the pMDC32_GUS plasmid (*b*, *d*, *f*).

Size bar is 1 mm for *a*, *b* and 100 μm for *c*–*f*.

p-values 6.243734e–05, 2.191140e–07 and 7.429123e–10, respectively). In contrast, all types of explants grown on the hygromycin-free media mostly didn't develop detectable transgenic tissues. The difference was statistically significant when compared to explants cultivated on hygromycin-containing medium (Fisher's exact test, *p*-value 5.842e–12), confirming our results obtained for the GFP marker (Fig. 1).

Evaluation of GUS false positive staining

During the experiment with GUS activity analysis described above, we also performed GUS staining of explants transformed with the pMDC32_eGFPer construction, which didn't carry the *GUS* gene. Unexpectedly, these control explants also demonstrated blue staining, although the staining pattern differed from the one provided by true *GUS* expression (Hu et al., 1990). In false positives, the blue coloration was more diffuse and didn't have distinct borders compared with true GUS staining (Fig. 3).

Microscopic analysis also revealed differences between false positive and true GUS staining. *GUS*-expressing cells had deep blue coloring (Fig. 3*f*), which was not the case for false positive GUS-like staining (Fig. 3*e*). Therefore, the GUS marker can be used to detect transgenic tissue in lentil, but the staining of negative controls is necessary.

Discussion

Transformation is a powerful tool for obtaining new varieties of different plant species (Garcia Ruiz et al., 2018; Noack et al., 2024), which is why improving its effectiveness for lentil is important. Nevertheless, current *L. culinaris* transforma-

tion protocols have rather low efficiency, probably due to the low regeneration rate. One of the possible solutions for this problem is the usage of morphogenic regulators, i. e. genes that can be added to transformation vectors and that are able to induce proliferation of transgenic cells and plant regeneration from transgenic tissue. Still, the search for such morphogenic regulators, suitable for lentils, can be hampered with low initial transformation rate, i. e. low effectiveness of transgenic callus formation. Therefore, the development of effective protocols for maximizing the transgenic tissue amount can be very useful for further optimization of obtaining transformed and edited plants.

The protocols for obtaining transgenic lentil plants usually involve cotyledons and/or cotyledonary nodes, parts of embryos or embryo axes (Gulati et al., 2002; Cellikol Akcay et al., 2009; Khatib et al., 2011; Das et al., 2012, 2019; Polowick, Yan, 2023) or whole seeds (Chopra et al., 2012). In our study, we tried to perform *Agrobacterium*-mediated transformation using different types of explants obtained from 12-day old seedlings. According to our results, the transformation protocol, developed for *M. truncatula* leaf explants, can also be used for obtaining transgenic tissue from *L. culinaris* explants. Still, although the leaf explants are used in this protocol for *M. truncatula*, a different type of explant, shoot apex, gave better results for *L. culinaris*. It can possibly be related to a relatively high proportion of stem cells and, in general, proliferating cells in the shoot apex in comparison with other explant types. It would be interesting to evaluate the efficiency of usage of seedling shoot apices for *M. truncatula* transformation. The usage of hygromycin at the concentration 10 mg/l allowed us to obtain a significant amount of transgenic tissue from shoot apex explants. Nevertheless, the optimal hygromycin concentration has to be defined in the future.

There are also several studies aiming to obtain transgenic tissues in lentil from diverse explant types, including different parts of germinated seedlings (Warkentin, McHughen, 1992), cotyledonary nodes (Mahmoudian et al., 2002) and embryo axes (Lurquin et al., 1998), although in most cases, no selection agent and no 2,4-D-mediated callus induction was applied, in contrast to our research. As far as we know, the current study is also the first in which optimization of lentil transformation was performed with the usage of two different transgenic tissue markers.

In a study by Warkentin and McHughen (1992) *Agrobacterium*-mediated transformation of lentil shoot apex explants yielded a high percentage of explants, which contained transgenic tissue, without selection for transgenic cells at early stages of cultivation. In our study, cultivation on the medium without a selection agent was not effective, whereas hygromycin application made it possible to obtain a significant amount of transgenic tissue. The discrepancy between our results and results obtained previously may be due to differences in genotypes, *in vitro* cultivation and transformation conditions.

In experiments on evaluation of *in vitro* regeneration capacity in lentil, explants from shoot tips were shown to give callus on the medium with 2,4-D, as well as regenerating shoots on other types of media, with the addition of BAP, IAA or NAA (Polanco et al., 1988). Shoot regeneration from shoot apex explants was also observed after *Agrobacterium*-mediated transformation, although shoots regenerated during kanamycin selection were non-transgenic (Warkentin, McHughen, 1992). Therefore, the development of a protocol for producing transgenic lentil plants from shoot tip explants may be feasible upon identification of suitable selection and regeneration conditions. Based on our data, hygromycin selection appears to be a promising approach.

During our analysis, we noticed the GUS-like staining in calli that didn't contain plasmids for *GUS* gene expression. This can be explained with so-called intrinsic GUS-like activity, which was described in previous research (Hu et al., 1990). Indeed, beta-glucuronidase (GUS) genes or proteins were described in several species, including *Scutellaria baicalensis* (Zuo et al., 2025) and *Zea mays* (Muhitch, 1998), whereas endogenous GUS activity was detected in *Brassica napus* (Abdollahi et al., 2011), *Oryza sativa*, *A. thaliana*, *Nicotiana tabacum* and other plants (Sudan et al., 2006). Still, in other studies involving GUS staining of lentil explants, no false positive GUS staining was detected on control explants, which may be due to different staining buffer composition or different tissue features (Warkentin, McHughen, 1992; Lurquin et al., 1998; Mahmoudian et al., 2002).

Conclusion

Our results show that *Agrobacterium*-mediated transformation of lentil shoot apices according to the protocol developed for *M. truncatula*, can be successfully used for obtaining transgenic callus tissues. These data can be used for further development of effective and universal lentil transformation and editing protocols.

References

Abdollahi M.R., Memari H.R., van Wijnen A.J. Factor affecting the endogenous β-glucuronidase activity in rapeseed haploid cells: how to avoid interference with the GUS transgene in transformation studies. *Gene.* 2011;487(1):96-102. doi 10.1016/j.gene.2011.07.007

Bakulin S.D., Monakhos S.G., Bruskin S.A. Morphogenetic factors as a tool for enhancing plant regeneration capacity during *in vitro* transformation. *Int J Mol Sci*. 2025;26(17):8583. doi 10.3390/ijms26178583

Cao Z., Socquet-Juglard D., Daba K., Vandenberg A., Bett K.E. Understanding genome structure facilitates the use of wild lentil germplasm for breeding: a case study with shattering loci. *Plant Genome*. 2024;17(2):e20455. doi 10.1002/tpg2.20455

Celikkol Akcay U., Mahmoudian M., Kamci H., Yucel M., Oktem H.A. *Agrobacterium tumefaciens*-mediated genetic transformation of a recalcitrant grain legume, lentil (*Lens culinaris* Medik). *Plant Cell Rep.* 2009;28(3):407-417. doi 10.1007/s00299-008-0652-4

Chopra R., Aparna, Saini R. Use of sonication and vacuum infiltration for *Agrobacterium*-mediated transformation of an Indian lentil (*Lens culinaris* Medik.) cultivar. *Sci Hortic*. 2012;143:127-134. doi 10.1016/j.scienta.2012.06.019

Cosson V., Durand P., d'Erfurth I., Kondorosi A., Ratet P. *Medicago truncatula* transformation using leaf explants. *Methods Mol Biol*. 2006;343:115-127. doi 10.1385/1-59745-130-4:115

Curtis M.D., Grossniklaus U. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol*. 2003;133(2):462-469. doi 10.1104/pp.103.027979

Das S.K., Shethi K.J., Hoque M.I., Sarker R.H. *Agrobacterium*-mediated genetic transformation in lentil (*Lens culinaris* Medik.) followed by *in vitro* flowering and seed formation. *Plant Tissue Cult Biotechnol*. 2012;22(1):13-26. doi 10.3329/ptcb.v22i1.11243

Das S.K., Shethi K.J., Hoque M.I., Sarker R.H. *Agrobacterium*-mediated genetic transformation of lentil (*Lens culinaris* Medik.) with chitinase gene followed by *in vitro* flower and pod formation. *Plant Tissue Cult Biotechnol*. 2019;29(1):99-109. doi 10.3329/ptcb.v29i1.41982

Erskine W., Muehlbauer F., Sarker A., Sharma B. (Eds) The Lentil: Botany, Production and Uses. Wallingford: CABI, 2009

Fåhraeus G. The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *J Gen Microbiol*. 1957;16(2):374-381. doi 10.1099/00221287-16-2-374

Garcia Ruiz M.T., Knapp A.N., Garcia-Ruiz H. Profile of genetically modified plants authorized in Mexico. *GM Crops Food*. 2018;9(3):152-168. doi 10.1080/21645698.2018.1507601

Gulati A., Schryer P., McHughen A. Production of fertile transgenic lentil (*Lens culinaris* Medik) plants using particle bombardment. *In Vitro Cell Dev Biol Plant*. 2002;38(4):316-324. doi 10.1079/IVP2002303

Hoffmann B., Trinh T.H., Leung J., Kondorosi A., Kondorosi E. A new *Medicago truncatula* line with superior *in vitro* regeneration, transformation, and symbiotic properties isolated through cell culture selection. *Mol Plant-Microbe Interact*. 1997;10(3):307-315. doi 10.1094/MPMI.1997.10.3.307

Hu C.Y., Chee P.P., Chesney R.H., Zhou J.H., Miller P.D., O'Brien W.T. Intrinsic GUS-like activities in seed plants. *Plant Cell Rep*. 1990;9(1):1-5. doi 10.1007/BF00232123

Joshi M., Timilsena Y., Adhikari B. Global production, processing and utilization of lentil: a review. *J Integr Agric*. 2017;16(12):2898-2913. doi 10.1016/S2095-3119(17)61793-3

Kaale L.D., Siddiq M., Hooper S. Lentil (*Lens culinaris* Medik) as nutrient-rich and versatile food legume: a review. *Legume Sci*. 2023;5(2):e169. doi 10.1002/leg3.169

Karimi M., Inzé D., Depicker A. GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci*. 2002;7(5):193-195. doi 10.1016/S1360-1385(02)02251-3

Khatib F., Makris A., Yamaguchi-Shinozaki K., Kumar S., Sarker A., Erskine W., Baum M. Expression of the *DREB1A* gene in lentil (*Lens culinaris* Medik. subsp. *culinaris*) transformed with the Agrobacterium system. *Crop Pasture Sci*. 2011;62(6):488-495. doi 10.1071/CP10351

Kumar J., Gela T.S., Gupta D.S., Chandra A., Khazaei H. Recent advances in lentil genetics, genomics, and molecular breeding. In: Lentils: Production, Processing Technologies, Products, and Nutritional Profile. Hoboken: John Wiley & Sons Ltd., 2023;25-43. doi 10.1002/9781119866923.ch2

Li G., Liu R., Xu R., Varshney R.K., Ding H., Li M., Yan X., … Luo Y., Gao S., Wei P., Zong X., Yang T. Development of an *Agrobacterium*-mediated CRISPR/Cas9 system in pea (*Pisum sativum* L.). *Crop J*. 2023;11(1):132-139. doi 10.1016/j.cj.2022.04.011

Liber M., Duarte I., Maia A.T., Oliveira H.R. The history of lentil (*Lens culinaris* subsp. *culinaris*) domestication and spread as revealed by genotyping-by-sequencing of wild and landrace accessions. *Front Plant Sci*. 2021;12:628439. doi 10.3389/fpls.2021.628439

Lurquin P.F., Cai Z., Stiff C.M., Fuerst E.P. Half-embryo cocultivation technique for estimating the susceptibility of pea (*Pisum sativum* L.) and lentil (*Lens culinaris* Medik.) cultivars to *Agrobacterium tumefaciens*. *Mol Biotechnol*. 1998;9(2):175-179. doi 10.1007/BF02760819

Mahmoudian M., Yücel M., Öktem H.A. Transformation of lentil (*Lens culinaris* M.) cotyledonary nodes by vacuum infiltration of *Agrobacterium tumefaciens*. *Plant Mol Biol Rep*. 2002;20(3):251-257. doi 10.1007/BF02782460

Meyer D., Zeileis A., Hornik K. The strucplot framework: visualizing multi-way contingency tables with vcd. *J Statistical Software*. 2006;17(3):1-48. doi 10.18637/jss.v017.i03

Meyer D., Zeileis A., Hornik K., Friendly M. vcd: visualizing categorical data. R package version 1.4-12. 2024. URL: https://CRAN.R-project.org/package=vcd

Muhitch M.J. Characterization of pedicel β-glucuronidase activity in developing maize (*Zea mays*) kernels. *Physiol Plant*. 1998;104(3):423-430. doi 10.1034/j.1399-3054.1998.1040318.x

Noack F., Engist D., Gantois J., Gaur V., Hyjazie B.F., Larsen A., M'Gonigle L.K., Missirian A., Qaim M., Sargent R.D., Souza-Rodrigues E., Kremen C. Environmental impacts of genetically modified crops. *Science*. 2024;385(6712):eado9340. doi 10.1126/science.ado9340

Olhoft P.M., Flagel L.E., Donovan C.M., Somers D.A. Efficient soybean transformation using hygromycin B selection in the cotyledonary-node method. *Planta*. 2003;216(5):723-735. doi 10.1007/s00425-002-0922-2

Polanco M.C., Peláez M.I., Ruiz M.L. Factors affecting callus and shoot formation from *in vitro* cultures of *Lens culinaris* Medik. *Plant Cell Tissue Organ Cult*. 1988;15(2):175-182. doi 10.1007/BF00035759

Polowick P.L., Yan W. A protocol for *Agrobacterium*-mediated genetic transformation of *Lens culinaris* Medik (lentil). *Plant Cell Tissue Organ Cult*. 2023;152(3):605-618. doi 10.1007/s11240-022-02434-x

Potsenkovskaia E., Tvorogova V., Yakovleva D., Zlydneva N., Lutova L. Novel *NF-Y* genes expressed during somatic embryogenesis in *Medicago truncatula*. *Plant Gene*. 2022;31:100364. doi 10.1016/j.plgene.2022.100364

Runte M., Guth J.N., Ammann J. Consumers' perception of plant-based alternatives and changes over time. A linguistic analysis across three countries and ten years. *Food Qual Preference*. 2024;113:105057. doi 10.1016/j.foodqual.2023.105057

Sudan C., Prakash S., Bhomkar P., Jain S., Bhalla-Sarin N. Ubiquitous presence of β-glucuronidase (GUS) in plants and its regulation in some model plants. *Planta*. 2006;224(4):853-864. doi 10.1007/s00425-006-0276-2

Tisseyre P., Cartieaux F., Chabrillange N., Gully D., Hocher V., Svistoonoff S., Gherbi H. Setting up *Agrobacterium tumefaciens*-mediated transformation of the tropical legume *Aeschynomene evenia*, a powerful tool for studying gene function in nod factor-independent symbiosis. *PLoS One*. 2024;19(4):e0297547. doi 10.1371/journal.pone.0297547

Tvorogova V.E., Fedorova Y.A., Potsenkovskaya E.A., Kudriashov A.A., Efremova E.P., Kvitkovskaya V.A., Wolabu T.W., Zhang F., Tadege M., Lutova L.A. The WUSCHEL-related homeobox transcription factor MtWOX9-1 stimulates somatic embryogenesis in *Medicago truncatula*. *Plant Cell Tissue Organ Cult.* 2019;138(3):517-527. doi 10.1007/s11240-019-01648-w

Warkentin T.D., McHughen A. *Agrobacterium tumefaciens*-mediated beta-glucuronidase (GUS) gene expression in lentil (*Lens culinaris* Medik.) tissues. *Plant Cell Rep*. 1992;11(5-6):274-278. doi 10.1007/BF00235081

Wickham H., François R., Henry L., Müller K., Vaughan D. dplyr: a grammar of data manipulation. R package version 1.1.4. 2025. URL: https://CRAN.R-project.org/package=dplyr

Wu H.-Y., Liu K.-H., Wang Y.-C., Wu J.-F., Chiu W.-L., Chen C.-Y., Wu S.-H., Sheen J., Lai E.-M. AGROBEST: an efficient *Agrobacterium*-mediated transient expression method for versatile gene function analyses in *Arabidopsis* seedlings. *Plant Methods*. 2014;10(1):19. doi 10.1186/1746-4811-10-19

Yakovleva D., Efremova E., Smirnov K., Simonova V., Konstantinov Z., Tvorogova V., Lutova L. The *WOX* genes from the intermediate clade: influence on the somatic embryogenesis in *Medicago truncatula*. *Plants*. 2024;13(2):223. doi 10.3390/plants13020223

Zuo X., Li P., Ren G., Bai Z., Jiang D., Liu C. Functional characterization of β-glucuronidase genes involved in baicalein biosynthesis from *Scutellaria baicalensis* based on transcriptome analysis. *Int J Mol Sci*. 2025;26(5):4410. doi 10.3390/ijms26051793

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

Received November 4, 2025. Revised January 12, 2026. Accepted January 13, 2026.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):241-249

doi 10.18699/vjgb-26-27

Получение и характеристика двойной нокаутной линии *Arabidopsis thaliana*, лишенной экспрессии *AOX1a* и *VTC2*

Е.В. Гармаш, Е.С. Белых, К.В. Ядрихинский, Р.В. Малышев, И.О. Велегжанинов

Институт биологии Коми научного центра Уральского отделения Российской академии наук, Сыктывкар, Россия
garmash@ib.komisc.ru

Аннотация. У высших растений L-галактозный путь является основным путем биосинтеза витамина С (аскорбата, Аск), последний этап которого связан с функционированием электрон-транспортной цепи митохондрий. Помимо основного цитохромного пути, электрон-транспортная цепь растений содержит альтернативный путь через альтернативную терминальную оксидазу (АОХ). Вовлечение АОХ способствует синтезу Аск, поэтому предполагается, что подавление альтернативного пути в условиях дефицита Аск может привести к снижению жизнеспособности растений. В связи с этим была поставлена цель – рассмотреть последствия нокаутирования в растениях *Arabidopsis thaliana* одновременно двух генов: *AOX1a*, наиболее стресс-индуцибельного гена АОХ, и *VTC2*, гена ключевого фермента L-галактозного пути синтеза Аск. Для этого проведено скрещивание двух линий *A. thaliana* с Т-ДНК инсерцией в целевых генах и получены гибридные линии. Изучены характеристики семян первого (F_1) и второго (F_2) поколений. Семена F_1 отличались более крупными размерами по сравнению с родительскими линиями, что могло быть следствием гетерозиса. В поколении F_2 в результате самоопыления растений F_1 сформировались семена, значительно варьирующие по размерам, в том числе группа мелких семян, имеющих дегенеративные морфологические отклонения. Большинство мелких семян не прорастало или погибало на стадии появления проростка. Генотипирование этих семян с помощью ПЦР установило отсутствие нормальных копий *AOX1a* и *VTC2* в геноме, что указывает на появление летальной мутации при одновременном нокауте обоих генов. Обсуждаются причины летального исхода двойного нокаута. Одной из генетических причин, по-видимому, стало взаимодействие мутаций (неаллельных генов). На физиологическом уровне, возможно, возникли необратимые нарушения дыхания, в том числе из-за влияния в линии *vtc2* криптической мутации. Для подтверждения данных гипотез требуются дальнейшие исследования. В целом полученные результаты свидетельствуют о жизненно важном совместном функционировании альтернативного и L-галактозного путей биосинтеза Аск и могут быть полезны для развития генно-инженерных приемов управления синтезом витамина С в растениях.

Ключевые слова: *Arabidopsis thaliana*; альтернативная оксидаза; аскорбат; ГДФ-L-галактозофосфорилаза; линии с нокаутом *AOX1a* и *VTC2*; скрещивание; генотипирование; двойная нокаутная линия; летальная мутация

Для цитирования: Гармаш Е.В., Белых Е.С., Ядрихинский К.В., Малышев Р.В., Велегжанинов И.О. Получение и характеристика двойной нокаутной линии *Arabidopsis thaliana*, лишенной экспрессии *AOX1a* и *VTC2*. *Вавиловский журнал генетики и селекции*. 2026;30(2):241-249. doi 10.18699/vjgb-26-27

Финансирование. Работа выполнена в рамках темы госбюджетных НИОКТР «Физиологические и молекулярные механизмы интеграции клеточных процессов и целостности растительного организма: фотосинтез и дыхание» (регистрационный номер 125020301262-2).

Благодарности. Исследование проведено с использованием оборудования ЦКП «Молекулярная биология» Института биологии Коми научного центра Уральского отделения Российской академии наук. Авторы выражают благодарность научному сотруднику института к.б.н. М.А. Шелякину за помощь при скрещивании линий.

Generation and characterization of a double-knockout *Arabidopsis thaliana* line lacking expression of *AOX1a* and *VTC2*

E.V. Garmash, E.S. Belykh, K.V. Yadrikhinskiy, R.V. Malyshev, I.O. Velegzhaninov

Institute of Biology of Komi Science Centre of the Ural Branch of the Russian Academy of Sciences, Syktyvkar, Russia
garmash@ib.komisc.ru

Abstract. In higher plants, the L-galactose pathway is the main pathway for the biosynthesis of vitamin C (ascorbate, Asc), the final step of which is connected with the functioning of the mitochondrial electron transport chain (ETC). In addition to the main cytochrome pathway, plant ETC includes an alternative pathway (AP) via alternative terminal oxidase (AOX). The engagement of AOX promotes Asc synthesis, and it is hypothesized that AOX suppression under conditions of Asc deficiency may reduce plant viability. The aim of this work was to examine the consequences of simultaneously knocking

© Гармаш Е.В., Белых Е.С., Ядрихинский К.В., Малышев Р.В., Велегжанинов И.О., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

out two genes in *Arabidopsis thaliana*: *AOX1a*, the most stress-inducible AOX gene, and *VTC2*, encoding a key enzyme of the L-galactose pathway of Asc synthesis. Two lines of *A. thaliana* with T-DNA insertions in the target genes were crossed to generate hybrid lines. Seed characteristics of the first (F_1) and second (F_2) generations were analyzed. F_1 seeds were larger than those of parent lines, possibly due to heterosis. In the F_2 generation, self-pollination of F_1 plants resulted in seeds with significant size variation, including a group of small seeds with degenerative morphological abnormalities. Most of small seeds failed to germinate or died at the seedling stage. PCR genotyping of these seeds revealed the absence of native *AOX1a* and *VTC2* indicating a lethal mutation caused by simultaneous knockout of both genes. One likely genetic cause is the interaction of mutations in non-allelic genes. At the physiological level, irreversible respiratory damage may occur, possibly including the impact of a cryptic mutation in the *vtc2* line. Further studies are necessary to confirm these hypotheses. In general, the results obtained indicate the vital co-functioning of the AP and the L-galactose pathway of Asc biosynthesis and may be useful for the development of genetically engineered techniques for the control of vitamin C synthesis in plants.

Key words: *Arabidopsis thaliana*; alternative oxidase; ascorbate; GDP-L-galactose phosphorylase; *AOX1a* and *VTC2* knockout lines; crossing; genotyping; double mutants; lethal mutation

For citation: Garmash E.V., Belykh E.S., Yadrikhinskiy K.V., Malyshev R.V., Velegzhaninov I.O. Generation and characterization of a double-knockout *Arabidopsis thaliana* line lacking expression of *AOX1a* and *VTC2*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):241-249. doi 10.18699/vjgb-26-27

Введение

Аскорбат (витамин С) – важное многофункциональное антиоксидантное соединение, участвующее в обеспечении устойчивости к стрессу и редокс-сигналинге (Smirnoff, 2018; Foyer et al., 2020; Matos et al., 2022). В растениях существует несколько путей пополнения его пула – биосинтетические процессы и регенерация аскорбата (Аск) в аскорбат-глутатионовом цикле. Среди всех возможных путей синтеза Аск L-галактозный путь считается доминирующим (Dowdle et al., 2007; Wheeler et al., 2015). В L-галактозном пути ГДФ-D-манноза последовательно превращается в ГДФ-L-галактозу, L-галактозо-1-фосфат, L-галактозу, L-галактоно-1,4-лактон и, наконец, в L-аскорбат (рис. 1). Ключевой стадией этого пути является реакция, катализируемая ГДФ-L-галактозофосфорилазой (ГГФ) (Yoshimura et al., 2014; Matos et al., 2022). ГГФ кодируется генами-паралогами *VTC2* и *VTC5* (*VITAMIN C*), проявляющими светозависимый характер экспрессии (Yoshimura et al., 2014; Smirnoff, 2018). Из них *VTC2* наиболее важен для биосинтеза Аск. Нокаут *VTC2*, в отличие от его паралога *VTC5*, вызывал резкое снижение уровня Аск (на 80 % и более) и роста растений (Dowdle et al., 2007; Lim et al., 2016).

Заключительный этап L-галактозного пути связан с электрон-транспортной цепью (ЭТЦ) митохондрий, где L-галактоно-1,4-лактондегидрогеназа (ГЛДГ) окисляет L-галактоно-1,4-лактон до Аск, передавая электроны на лабильный переносчик электронов между комплексами III и IV – цитохром *c* (см. рис. 1). ЭТЦ растений отличается от ЭТЦ животных по составу переносчиков. Помимо основного энергогенерирующего цитохромного пути (ЦП), цепь содержит альтернативный путь (АП) через альтернативную терминальную цианидустойчивую оксидазу (АОХ) (Гармаш, 2022) (см. рис. 1). АП минует два участка генерации мембранного потенциала (комплексы III и IV) и поэтому является энергетически малоэффективным (Vanlerberghe et al., 2020). Главные функции АП связаны с его участием в поддержании редокс-баланса в ЭТЦ за счет более быстрого окисления НАДН и, тем самым, в предотвращении образования избыточного количества активных форм кислорода (АФК). Это влияет на метаболизм не только митохондрий, но и клетки в целом (Del-Saz et al., 2018; Vanlerberghe et al., 2020; Garmash, 2023). Кроме того, АП способствует синтезу Аск, принимая электроны с убихинона и поддерживая пул цитохрома *c* в более окисленном состоянии (Bartoli et al., 2006; Matos et al., 2022).

Исследования мутантных линий с подавлением *VTC2* показали наличие в них многочисленных дефектов и нарушений метаболизма, особенно в стрессовых условиях (Lim et al., 2016; Matos et al., 2022). Полагают, что дефицит аскорбата и фенотип низкорослости мутантных растений могут быть результатом длительного воздействия мутации на дыхательную активность, функционирование ЭТЦ митохондрий и, в частности, транспорта электронов через АОХ (Matos et al., 2022).

Ядерный геном *A. thaliana* содержит пять генов АОХ, среди которых яркую экспрессию в ответ на различные виды стресса проявляет ген *AOX1a* (Del-Saz et al., 2018; Гармаш, 2022). При стрессе (повышенной освещенности) растения *A. thaliana* со сверхэкспрессией *AOX1a* характеризовались усилением синтеза Аск (Garmash et al., 2022; Sweetman et al., 2022), а линии с подавлением *AOX1a* демонстрировали снижение содержания данного метаболита (Vishwakarma et al., 2015; Garmash et al., 2022). При этом у растений *A. thaliana* линии *vtc2* выявлена активация АОХ по сравнению с диким типом, а обработка мутантных растений специфическим ингибитором АП вызывала значительное (на 40 %) снижение концентрации АТФ (Garmash et al., 2024). Другими авторами на основе изучения реакции растений *A. thaliana* с нокаутом гена ГДФ-маннозофосфорилазы, катализирующей образование ГДФ-D-маннозы (линия *vtc1*), на ингибирование ЭТЦ митохондрий и хлоропластов сделано предположение, что Аск и АОХ могут взаимодополнять друг друга для предотвращения избыточного накопления АФК и обеспечения защиты фотосинтеза от фотоокисления (Talla et al., 2011). Эти данные подтверждают участие АОХ в поддержании синтеза Аск, оптимизации энергетического баланса и указывают на то, что подавление АП в условиях дефицита Аск может привести к снижению жизнеспособности растений.

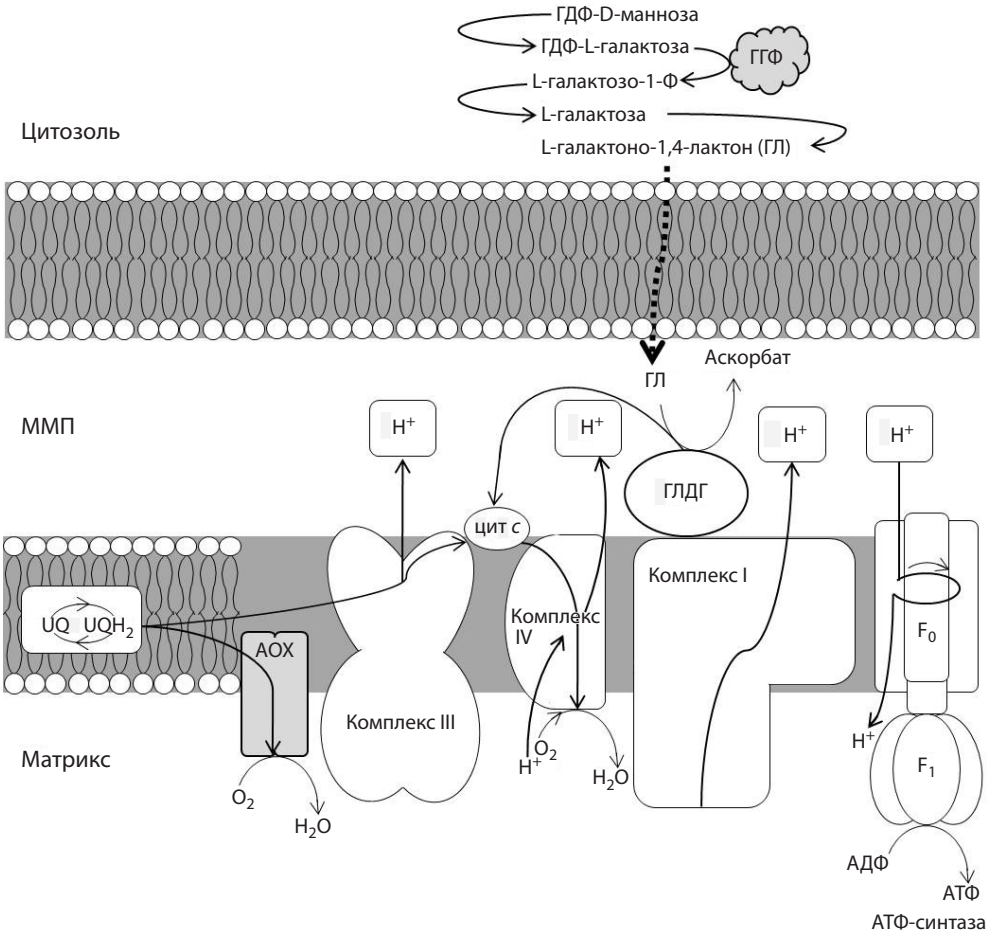


Рис. 1. Схема связи синтеза аскорбата с электрон-транспортной цепью (ЭТЦ) митохондрий растений.

Показан участок ЭТЦ во внутренней мембране митохондрий от пула убихинона (UQ/UQH_2), отдающего электроны по альтернативному пути через альтернативную оксидазу (AOX) и по цитохромному пути через комплекс III, цитохром *c* и комплекс IV. ГГФ – ГДФ-L-галактозофосфорилаза. Образованный по L-галактозному пути синтеза аскорбата L-галактоно-1,4-лактон (ГЛ), попадая из цитозоля в межмембранное пространство митохондрий (ММП), окисляется до аскорбата L-галактоно-1,4-лактондегидрогеназой (ГЛДГ), передающей электроны на цитохром *c*. Перенос электронов по ЭТЦ сопровождается прокачкой протонов в ММП и их «возвращением» в матрикс через АТФ-синтазу, генерирующую АТФ.

Поэтому возник вопрос о ключевой роли *AOX1a* и *VTC2* в энергообеспечении и жизнедеятельности растительного организма.

Нами была поставлена цель – рассмотреть последствия нокаутирования в растениях *A. thaliana* одновременно двух генов, *AOX1a* и *VTC2*. Для этого применена техника скрещивания мутантных линий и изучены характеристики семян первого (F_1) и второго (F_2) поколений, а также проведена идентификация двойного нокаутирования путем установления локализации вставок Т-ДНК с помощью подобранных праймеров к генам интереса.

Материал и методы

Для гибридизации использовали мутантные линии SALK_084897 со вставкой Т-ДНК (инсерцией) в *AOX1a* и SALK_146824C с инсерцией в *VTC2* (рис. 2).

Семена мутантных линий получены из Arabidopsis Biological Resources Centre (ABRC, Университет Огайо, США). Линия с нокаутом *AOX1a* (*aox1a*) имеет вставку Т-ДНК только в одной копии *AOX1a*, что делает линию *aox1a* гетерозиготой. Линия *aox1a* имеет также вставку в гене *EFS*, кодирующем гистонлизин-N-метилтрансферазу – ингибитор цветения. Линия с нокаутом *VTC2* (*vtc2*) имеет вставку в двух копиях гена, т. е. является гомозиготной. Для контроля показателей относительного содержания транскриптов генов использовали линию дикого типа Columbia-0 (Col-0). Семена линии Col-0 получены из Nottingham Arabidopsis Stock Centre (NASC, Великобритания) и любезно предоставлены О.И. Грабельных и В.И. Тарасенко (СИФИБР СО РАН).

Семена высаживали, стратифицировали и проращивали в емкостях объемом 200 $см^3$ в смеси перлит : вермикулит : грунт (1 : 1 : 2). Затем пересаживали в контролируемые условия при температуре 22 °C, уровне освещения 90 мкмоль квантов/($м^2$·с) и фотопериоде 16 ч до фазы бутонизации (стадия развития 5.1; Boyes et al., 2001). Свет обеспечивали люминесцентными лампами (TL-D 30W и TL-D 30WAquarelle, Philips, Нидерланды).

Скрещивание проводили путем переноса пыльцы с растения линии SALK_146824C (*vtc2*) на линию

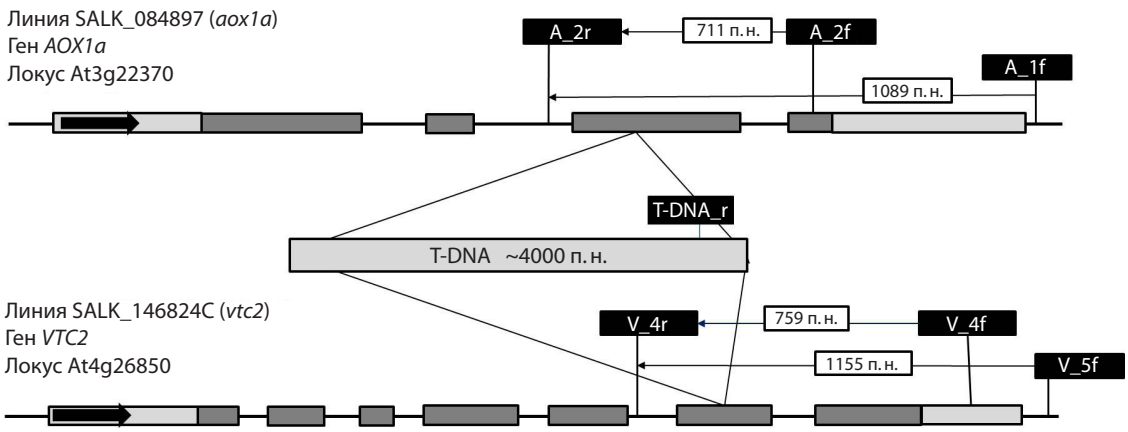


Рис. 2. Схема расположения праймеров и длины ПЦР-продуктов и локализация инсерций в гене *AOX1a* линии SALK_084897 (*aox1a*) и гене *VTC2* линии SALK_146824C (*vtc2*).

Пары праймеров V_4f–V_4r и V_5f–V_4r, а также A_2f–A_2r и A_1f–A_2r позволяют идентифицировать ген линии «дикого типа» *VTC2* или *AOX1a* соответственно, а пары праймеров V_4f–T-DNA_r и V_5f–T-DNA_r, а также A_2f–T-DNA_r и A_1f–T-DNA_r – соответствующий мутантный ген.

Таблица 1. Праймеры для ПЦР в конечной точке, использованные для анализа Т-ДНК вставок в ген *VTC2* (V) в линии SALK_146824C (*vtc2*) и *AOX1a* (A) в линии SALK_084897 (*aox1a*)

Название	Последовательность
T-DNA_r	CTTCCTTTCTCGCCACGTTC
V_4f	ACTGAAGGACAAGGCACTCG
V_4r	ACCTTTTCACCCTTGCTCATCT
V_5f	TCATAACCAGAGCTTGTGGTGG
A_1f	GATCCACATCCATTGGGTCCT
A_2r	CACTTCGTTAGTGCTTTGGATTCTTGG
A_2f	GCTTCCTTTAGTTCACGACCTTGG

SALK_084897 (*aox1a*), контролируя процесс под микроскопом. Для этого на материнском растении нераспустившиеся (за один день до распускания) бутоны аккуратно вскрывали, затем пинцетом тщательно удаляли все тычинки с пыльцой. Со второго (отцовского) бутона пыльцу переносили на рыльце пестика материнского растения. Оплодотворенный цветок помечали, оплодотворение повторяли три дня. Все цветки вокруг постоянно удаляли на фазе их бутонизации. Растение изолировали пластиковым прозрачным цилиндром до созревания семян.

В поколениях F_1 и F_2 определяли размеры семян (n = 111 и 496 для каждого поколения соответственно). Под микроскопом измеряли длину (большая полуось эллипса, a, мм) и ширину (меньшая полуось эллипса, b, мм) семени. Площадь эллипса (S, мм^2) рассчитывали по формуле $S = a \cdot b \cdot \pi$. Затем семена поколения F_2 разделяли по группам согласно их размерам – большие, средние, мелкие (n = 49, 365, 82 для каждой группы соответственно).

Для анализа наличия зародыша проводили под микроскопом препарирование набухших после размачивания в воде семян – нескольких представителей каждой группы.

Генотипирование и ОТ-ПЦР анализ на содержание относительного количества транскриптов генов осуществляли в образцах растений, выращенных до вегетативной фазы роста 1.14 (Boyes et al., 2001), и семенах.

Для генотипирования родительских линий *aox1a*, *vtc2* и гибридных линий потомков использовали ДНК соответствующих образцов и разработанные праймеры (табл. 1, см. рис. 2). ДНК выделяли из воздушно-сухой навески (50–100 мг) розеточных листьев или сухих семян с помощью набора Сорб-ГМО-Б («Синтол», Россия) согласно инструкции производителя. Для ПЦР в конечной точке использовали 4 мкл готовой смеси Screen Mix («Евроген», Россия), 4 мкл каждого из пары соответствующих праймеров (0.3 мкМ) («Синтол», Россия), 7 мкл ddH_2O («Панэко», Россия) и 1 мкл ДНК-матрицы. Амплификация включала предварительную денатурацию в течение 5 мин при температуре 95 °C и далее 35 циклов: денатурация (15 с) при температуре 95 °C, отжиг праймеров (30 с) при температуре 55 °C и элонгация (60 с) при температуре 72 °C; с окончательной элонгацией (2 мин) при температуре 72 °C. Продукты ПЦР разделяли методом электрофореза в 1 % агарозном геле в Трис-ацетатном буферном растворе. Фрагменты ДНК окрашивали Safe Green Stain (APGENA, Россия). Фотодокументирование электрофореграмм выполняли с помощью SuperRay Maxi (APGENA, Россия). Маркером длины фрагментов ДНК служил 100+ bp Ladder DNA marker (100–1500 bp) («Евроген», Россия).

Относительное содержание транскриптов генов в образцах анализировали при помощи ПЦР в реальном времени на термоциклере/амплификаторе CFX96 (Bio-Rad, США). Розеточные листья растений фиксировали в жидком азоте и до начала анализа хранили при температуре –80 °C. Выделение РНК проводили набором HiPure Plant RNA Kit (Magen, Китай) согласно инструкции производителя. Концентрацию выделенной РНК определяли на флюориметре Qubit (Thermo Fisher Scientific, США) с использованием набора Qubit™ RNA HS Assay Kit (Thermo Fisher

Таблица 2. Список праймеров для проведения ОТ-ПЦР *Arabidopsis thaliana*

Ген	Номер локуса	Название по NCBI	Прямой праймер (5′–3′) Обратный праймер (5′–3′)	Длина ПЦР продукта, п.н.
AOX1a	At3g22370	Альтернативная оксидаза 1A	ATGATAACTCGCGGTGGAGC TCGTCGGAGCTCTAGTCCAT	166
VTC2	At4g26850	ГДФ-L-галактозофосфорилаза 1	GGACCTCCATGGGTTCTGTT GAGGTTACTGCTCTCGCCTT	131
*AT2G28390**	At2g28390	Белок семейства SAND	AACTCTATGCAGCATTTGATCCACT TGATTGCATATCTTTATCGCCATC	61
*AT4G34270**	At4g34270	Белок семейства TIP41-подобных белков	GTGAAAACTGTTGGAGAGAAGCAA TCAACTGGATACCCTTTCGCA	61

Примечание. Разработка праймеров выполнена с помощью Primer-BLAST (Ye et al., 2012) таким образом, чтобы они одновременно амплифицировали фрагменты всех известных сплайс-вариантов гена интереса. * Референсные гены, по (Czechowski et al., 2005).

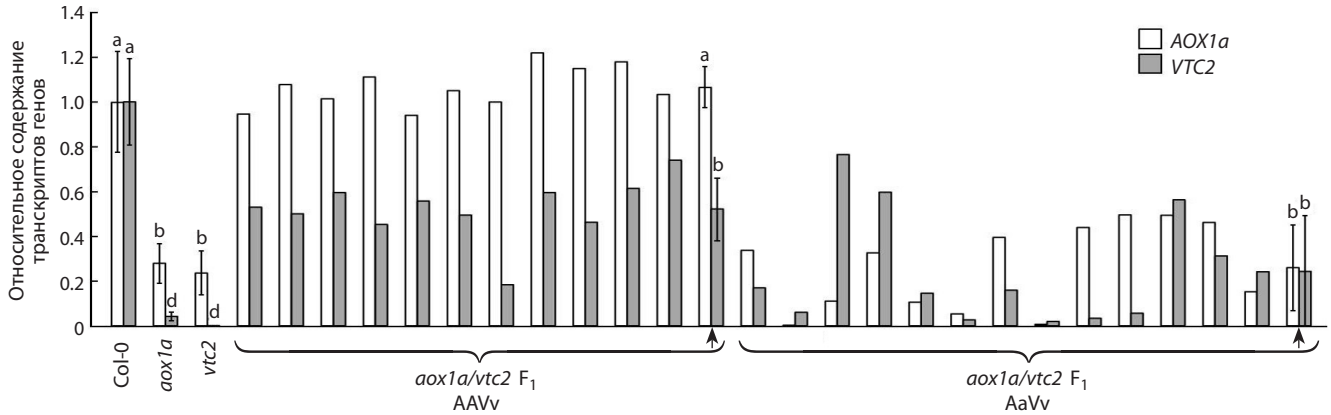


Рис. 3. Относительное количество транскриптов *AOX1a* и *VTC2* в листьях растений *A. thaliana* линий дикого типа (Col-0), родительских линий (*aox1a* и *vtc2*) и 24 растений *A. thaliana* гибридных линий (*aox1a/vtc2*) поколения F_1.

Стрелками показаны среднеарифметические величины и их стандартные ошибки параметра у предполагаемых генотипов (AAVv и AaVv). Достоверность изменений параметра отдельно для каждого гена обозначена разными строчными латинскими буквами (ANOVA, тест Дункана, $p \leq 0.05$).

Scientific, США). Для синтеза кДНК с одноцепочечной РНК-матрицы для ОТ-ПЦР применяли MMLV RT kit («Евроген», Россия) с олиго-dT праймером. ПЦР осуществляли с готовой смесью qPCRmix-HS SYBR («Евроген», Россия) и разработанными праймерами (табл. 2). Экспрессию генов нормализовали на геометрическую среднюю экспрессии генов *AT2G28390* и *AT4G34270* (Czechowski et al., 2005). Значения относительного уровня транскриптов генов рассчитывали методом $2^{-\Delta\Delta C_T}$ согласно (Livak, Schmittgen, 2001). Анализ проводили в трех биологических образцах, включающих листья одного гибридного растения или трех-пяти растений оригинальных линий, в двух технических повторностях для каждого образца.

Результаты

Гены *AOX1a* (At3g22370), *VTC2* (At4g26850) и *EFS* (At1g77300) расположены на третьей, четвертой и первой хромосомах *A. thaliana* соответственно. Следовательно, потомство гетерозигот в результате скрещивания линий SALK_084897 и SALK_146824C согласно несцепленному наследованию и третьему закону Менделя позволяет получить в поколении F_1 комбинирование признаков во всех возможных сочетаниях, а в поколении F_2 – потомство, гомозиготное по обеим мутациям.

ОТ-ПЦР анализ показал, что относительный уровень транскриптов *AOX1a* в линии *aox1a* был значительно ниже, чем в линии Col-0, а уровень мРНК *VTC2* в линии *vtc2* почти не детектировался (рис. 3). При этом семена линии *aox1a* заметно варьировали по размеру в отличие от линии *vtc2* (рис. 4). Эти данные косвенно свидетельствовали о гетерозиготности линии *aox1a* и гомозиготности линии *vtc2* по мутантному варианту гена интереса, что также было подтверждено наличием вставок Т-ДНК в соответствующих линиях (рис. 5). На электрофореграмме линии *vtc2* отсутствовали полосы ПЦР-продуктов пар праймеров V_4f–V_4r и V_5f–V_4r, характерных для «нормального» варианта гена, и присутствовали полосы V_4f–T-DNA_r и V5f–T-DNA_r, свидетельствующие о наличии вставки Т-ДНК. Ранее нами выявлена точная локализация вставки Т-ДНК в линии *vtc2* в положении 13499579 в экзоне 6 гена At4g26850 (Belykh et al., 2024). Полоса ПЦР-продукта пары праймеров A_2f–A_2r и ее отсутствие для пары

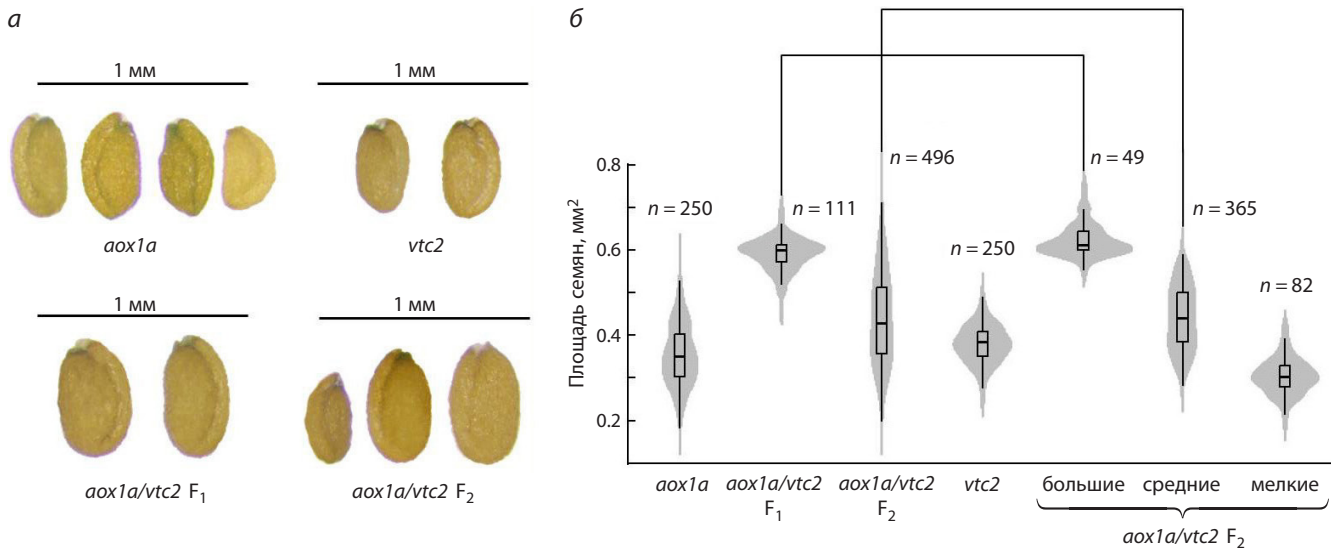


Рис. 4. Внешний вид (*а*) и площадь семян (*б*) родительских и гибридных линий (*aox1a/vtc2*) *A. thaliana* поколений F_1 и F_2.

Графики (*б*) представлены в виде виолоплотов с боксоплотами согласно среднеарифметическим значениям и их стандартным ошибкам; *n* – размер выборки. Отличия между всеми парными параметрами, не отмеченными квадратной скобкой, значимы при $p < 0.01$ (тест Данна с поправкой Бонферрони).

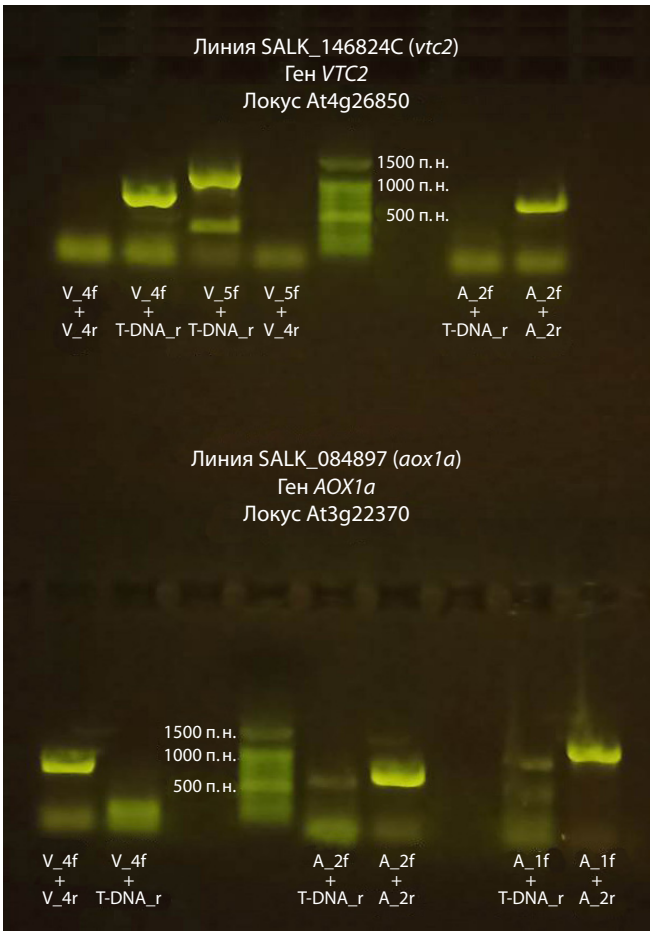


Рис. 5. Электрофореграмма ампликонов ДНК растений *A. thaliana* линий *vtc2* и *aox1a*, показывающая наличие Т-ДНК-вставок в ген *VTC2* и ген *AOX1a* соответственно.

Описание праймеров см. в табл. 1.

A_2f–T-DNA_r указывали на «нормальный» вариант гена *AOX1a*.

Линия *aox1a*, напротив, характеризовалась присутствием ПЦР-продукта пары праймеров V_4f–V_4r и отсутствием V_4f-T–DNA_r, что указывало на «нормальную» копию гена *VTC2* (см. рис. 5). Наличие ПЦР-продуктов всех пар праймеров для *AOX1a*, включая последовательность Т-ДНК (A_2f–A_2r и A_2f–T-DNA_r; A_1f–A_2r и A_1f–T-DNA_r), свидетельствовало как о «нормальной», так и о поврежденной вставкой Т-ДНК копиях гена, что свойственно для гетерозиготной линии.

Для предсказания расщепления в поколении F_1 составлена схема на основе решетки Пеннета, позволяющей выяснить возможные варианты генотипов потомства по генотипам родителей и их гамет путем деления числа ячеек с данным генотипом на общее число ячеек. Согласно полученной схеме, расщепление по генотипу составило 1:1:1:1 (четыре фенотипа в равных количествах) (табл. 3). Другими словами, 25 % семян F_1 принадлежало мутантной линии AaVvEE с дефицитом генов *VTC2* и *AOX1a*, 25 % семян приходилось на генотип AaVvEe с дефицитом всех генов генов *AOX1a*, *VTC2*, *EFS.* Остальная часть представлена генотипами с дефицитом *VTC2* и с дефицитом генов *VTC2*, *EFS.*

Семена F_1, полученные при скрещивании, статистически достоверно отличались более крупными размерами по сравнению с родительскими линиями (см. рис. 4). Это могло быть следствием гетерозиса, проявляющегося в увеличении жизнеспособности гибридов из-за унаследования определенного набора аллелей различных генов от своих разнородных родителей.

Из семян F_1 были выращены 24 растения, которые характеризовались снижением количества транскриптов *VTC2* относительно Col-0 (см. рис. 3). Из них более поло-

Таблица 3. Решетка Пеннета для поколения F_1, полученного в результате скрещивания линий SALK_146824C (*vtc2*) и SALK_084897 (*aox1a*)

Генотип линии	SALK_084897 (*aox1a*) AaVVEe				
SALK_146824C (*vtc2*) AAvvEE		AVE	aVE	AVe	aVe
	AvE	AAVvEE	AaVvEE	AAVvEe	AaVvEe
	AvE	AAVvEE	AaVvEE	AAVvEe	AaVvEe
	AvE	AAVvEE	AaVvEE	AAVvEe	AaVvEe
	AvE	AAVvEE	AaVvEE	AAVvEe	AaVvEe
		Потомство с дефицитом *VTC2*	Потомство с дефицитом *VTC2* и *AOX1a*	Потомство с дефицитом *VTC2* и *EFS*	Потомство с дефицитом *VTC2*, *AOX1a*, *EFS*

Примечание. A – нормальный (как в Col-0) ген *AOX1a*; a – инсерция в *AOX1a*; V – нормальный (как в Col-0) ген *VTC2*; v – инсерция в *VTC2*; E – нормальный (как в Col-0) ген *EFS*; e – инсерция в *EFS*.

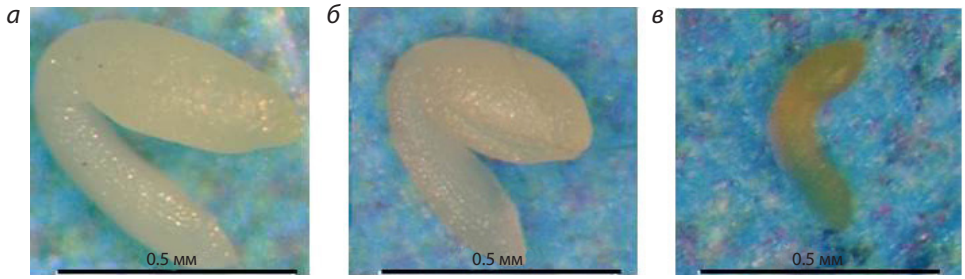


Рис. 6. Препарированные из больших (*а*), средних (*б*) и мелких семян (*в*) зародыши растений *A. thaliana* гибридных линий *aox1a/vtc2* поколения F_2.

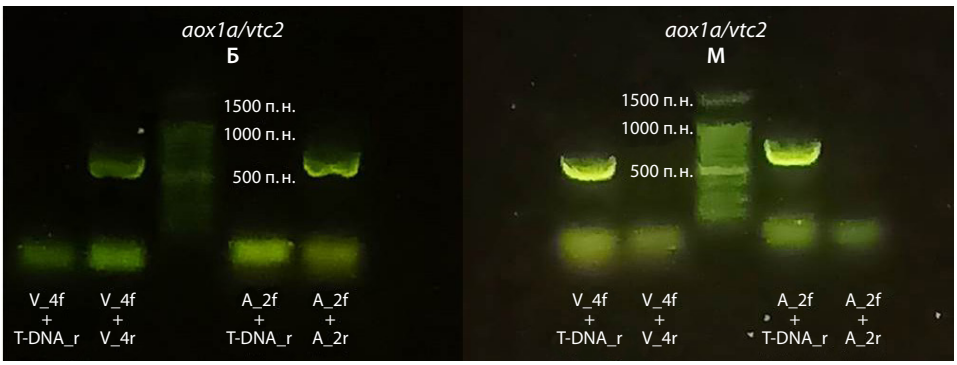


Рис. 7. Электрофореграмма ампликонов ДНК больших (Б) и мелких (М) семян *A. thaliana* гибридных линий *aox1a*/*vtc2* поколения F_2, показывающая наличие/отсутствие Т-ДНК-вставок в гены *AOX1a* и *VTC2*.

ДНК мелких семян показывает отсутствие полос ПЦР-продуктов пар праймеров V_4f–V_4r и A_2f–A_2r, характерных для «нормального» варианта генов *VTC2* и *AOX1a* соответственно. Описание праймеров см. в табл. 1.

вины образцов (54 %) показывали снижение транскриптов *AOX1a*. Геном остальных 46 % растений, предположительно, включал обе копии гена *AOX1a*, как в диком типе. Поэтому можно полагать, что гибридизация прошла успешно в предсказанном расщеплении.

На растениях F_1 в результате самоопыления сформировались плоды – стручки, содержащие семена F_2. Семена значительно варьировали по размерам (см. рис. 4). Для классификации по размерам семена были разделены на три достоверно различающиеся между собой группы. Интересно, что группа больших по размеру семян не отличалась от семян поколения F_1, что позволяет предположить для них генотип AAVv или AaVv, как в поколении F_1. Мелкие семена имели дегенеративные морфологические отклонения: при препарировании детектировали недоразвитый зародыш (рис. 6). Большинство мелких семян не прорастало или погибало на стадии появления проростка.

Генотипирование больших семян поколения F_2 показало наличие нормальных копий *AOX1a* и *VTC2* в геноме (рис. 7). На электрофореграмме ДНК, выделенной из мелких семян, отсутствовали полосы, соответствующие «нормальным» генам. Это свидетельствовало о том, что генотип образцов мелких семян соответствовал двойному нокаут-мутанту *aox1a/vtc2*.

Обсуждение

Исследование линий с дефицитом Аск, в первую очередь *vtc2*, показало важную роль L-галактозного пути биосинтеза витамина С и, в частности, ключевого фермента ГГФ в пополнении основного пула этого метаболита в растении. Вместе с тем полагают, что снижение роста и жизнеспособности мутантов связано не столько с пониженным уровнем Аск, сколько с нарушениями метаболизма и дисбалансом дыхательной ЭТЦ (Lim et al., 2016; Matos et al., 2022).

Такие нарушения могут быть вызваны возникающей в мутантной линии криптической мутацией (Lim et al., 2016). Однако эти же авторы показали, что дефекты роста мутантов *vtc2* сегрегируются независимо от мутации в гене *VTC2*, что указывает на локализацию потенциальной криптической мутации в данной линии на другой хромосоме и поэтому позволяет нам анализировать связь фенотипа потомства именно со статусом гена *VTC2*.

Причину нарушений метаболизма и роста выявляли в ряде работ с использованием мутантных комбинаций с *vtc2*. Фенотип снижения роста *vtc2* описан в сочетании с мутациями, вызывающими хронический фотоокислительный стресс (Müller-Moulé et al., 2004), с нокаутом хлоропластной аскорбатпероксидазы (Giacomelli et al., 2007), супрессией аутоиммунного ответа (Zhu et al., 2013). В результате скрещивания линии *A. thaliana vtc2-1* (loss-of-function мутант с сильно сниженной транскрипцией *VTC2*) с мутантной линией, нечувствительной к абсцизовой кислоте (*abi4*), получен двойной мутант *vtc2-1/abi4*, который демонстрировал рост и габитус, как у дикого типа (Kerchev et al., 2011). Авторы предположили, что АБК поддерживало фенотип низкорослости *vtc2-1*. Вместе с тем все описанные двойные нокауты сохраняли способность к росту. Летальным оказался лишь мутант с нокаутом обоих генов ГГФ – *VTC2* и *VTC5* (Dowdle et al., 2007).

Ранее первые мутанты с дефицитом Аск, *vtc1* (Conklin et al., 1996) и *vtc2* (Jander et al., 2002), были выделены по их чувствительности к озону. Известно, что чувствительность к озону может быть также результатом дисбаланса в дыхательных путях и повышенной генерации АФК на фоне низкой скорости синтеза Аск. При увеличении уровня озона ингибирование ЦП сопровождалось резким повышением активности АОХ (Ederli et al., 2006; Pasqualini et al., 2007). Это дало основание полагать, что усиление АП в линии *vtc2* предотвращает развитие окислительного стресса и возможную гибель растений.

Данное предположение подтвердилось в нашем исследовании, в котором при одновременном нокауте *AOX1a* и *VTC2* поколение F_2 было нежизнеспособным. Возможно, генетической причиной образования летальной мутации стало взаимодействие мутаций (неаллельных генов), в том числе криптической. Однако в пользу роли взаимодействия именно мутаций *AOX1a* и *VTC2*, а не *AOX1a* и криптической мутации свидетельствует тот факт, что все семена F_2, не способные к прорастанию, содержали мутацию в *VTC2*, которая, по всей видимости, с предполагаемой криптической не сцеплена (Lim et al., 2016). Для окончательного выяснения вопроса требуются более детальные молекулярно-генетические исследования, включающие секвенирование геномной ДНК, анализ РНК для выявления мутаций, влияющих на сплайсинг пре-мРНК и т. д.

Заключение

Таким образом, при гибридизации линий SALK_084897 и SALK_146824C были получены семена F_2 разного размера. Группа мелких семян охарактеризована как аберрантные семена, имеющие дегенеративные морфологические отклонения и не способные к прорастанию. Генотипирование этих семян свидетельствовало об отсутствии *AOX1a* и *VTC2* в геноме. Полученные данные указывали на появление летальной мутации при одновременном нокауте обоих генов и жизненно важном совместном функционировании двух систем – альтернативного пути через АОХ и L-галактозного пути биосинтеза Аск, связанного с ЭТЦ митохондрий. Дальнейшее изучение аспектов взаимосвязи и координации фукционирования АП и L-галактозного пути синтеза Аск может быть направлено на получение двойных мутантов с дефектами в генах, связанных с синтезом Аск, компонентов других дыхательных электрон-транспортных путей, участников гормонального и редокс-сигналинга. Такие исследования могут быть полезны для разработки генно-инженерного инструментария для улучшения синтеза Аск в целях повышения стрессоустойчивости растений, имеющих важное продовольственное значение.

Список литературы / References

Гармаш Е.В. Сигнальные пути регуляции экспрессии генов альтернативной оксидазы растений. *Физиология растений.* 2022; 69(1):3-19. doi 10.31857/S0015330322010055
[Garmash E.V. Signaling pathways of alternative oxidase genes expression in plants. *Russ J Plant Physiol.* 2022;69(1):1. doi 10.1134/S1021443722010058]

Bartoli C.G., Yu J., Gómez F., Fernández L., McIntosh L., Foyer C.H. Inter-relationships between light and respiration in the control of ascorbic acid synthesis and accumulation in *Arabidopsis thaliana* leaves. *J Exp Bot.* 2006;57(8):1621-1631. doi 10.1093/jxb/erl005

Belykh E., Yadrikhinskiy K., Golubev M., Belykh N., Velegzhaninov I., Garmash E. Localization data of the T-DNA insertion site in Arabidopsis line SALK_146824C. *Data Brief.* 2024;54:110309. doi 10.1016/j.dib.2024.110309

Boyes D.C., Zayed A.M., Ascenzi R., McCaskill A.J., Hoffman N.E., Davis K.R., Görlach J. Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell.* 2001;13(7):1499-1510. doi 10.2307/3871382

Conklin P.L., Williams E.H., Last R.L. Environmental stress sensitivity of an ascorbic acid-deficient Arabidopsis mutant. *Proc Natl Acad Sci USA.* 1996;93(18):9970-9974. doi 10.1073/pnas.93.18.9970

Czechowski T., Stitt M., Altmann T., Udvardi M.K., Scheible W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* 2005; 139(1):5-17. doi 10.1104/pp.105.063743

Del-Saz N.F., Ribas-Carbo M., McDonald A.E., Lambers H., Fernie A.R., Florez-Sarasa I. An *in vivo* perspective of the role(s) of the alternative oxidase pathway. *Trends Plant Sci.* 2018;23(3):206-219. doi 10.1016/j.tplants.2017.11.006

Dowdle J., Ishikawa T., Gatzek S., Rolinski S., Smirnoff N. Two genes in *Arabidopsis thaliana* encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. *Plant J.* 2007;52(4):673-689. doi 10.1111/j.1365-313X.2007.03266.x

Ederli L., Morettini R., Borgogni A., Wasternack C., Miersch O., Reale L., Ferranti F., Tosti N., Pasqualini S. Interaction between nitric oxide and ethylene in the induction of alternative oxidase in ozone-treated tobacco plants. *Plant Physiol.* 2006;142(2):595-608. doi 10.1104/pp.106.085472

Foyer C.H., Kyndt T., Hancock R.D. Vitamin C in plants: novel concepts, new perspectives, and outstanding issues. *Antioxid Redox Signal.* 2020;32(7):463-485. doi 10.1089/ars.2019.7819

Garmash E.V. Suppression of mitochondrial alternative oxidase can result in upregulation of the ROS scavenging network: some possible mechanisms underlying the compensation effect. *Plant Biol.* 2023; 25(1):43-53. doi 10.1111/plb.13477

Garmash E.V., Dymova O.V., Silina E.V., Malyshev R.V., Belykh E.S., Shelyakin M.A., Velegzhaninov I.O. *AOX1a* expression in *Arabidopsis thaliana* affects the state of chloroplast photoprotective systems under moderately high light conditions. *Plants* (*Basel*). 2022; 11(22):3030. doi 10.3390/plants11223030

Garmash E.V., Silina E.V., Belykh E.S., Shelyakin M.A., Malyshev R.V. Alternative oxidase affects ascorbate metabolism in *Arabidopsis thaliana* plants under high-light conditions: possible links between respiratory electron transport pathways in mitochondria. *J Biosci.* 2024;49:94. doi 10.1007/s12038-024-00471-6

Giacomelli L., Masi A., Ripoll D.R., Lee M.J., van Wijk K.J. *Arabidopsis thaliana* deficient in two chloroplast ascorbate peroxidases shows accelerated light-induced necrosis when levels of cellular ascorbate are low. *Plant Mol Biol*. 2007;65:627-644. doi 10.1007/s11103-007-9227-y

Jander G., Norris S.R., Rounsley S.D., Bush D.F., Levin I.M., Last R.L. Arabidopsis map-based cloning in the post-genome era. *Plant Physiol.* 2002;129(2):440-450. doi 10.1104/pp.003533

Kerchev P.I., Pellny T.K., Vivancos P.D., Kiddle G., Hedden P., Driscoll S., Vanacker H., Verrier P., Hancock R.D., Foyer C.H. The transcription factor ABI4 is required for the ascorbic acid-dependent regulation of growth and regulation of jasmonate-dependent defense signaling pathways in *Arabidopsis*. *Plant Cell.* 2011;23(9):3319-3334. doi 10.1105/tpc.111.090100

Lim B., Smirnoff N., Cobbett C.S., Golz J.F. Ascorbate-deficient *vtc2* mutants in *Arabidopsis* do not exhibit decreased growth. *Front Plant Sci.* 2016;7:1025. doi 10.3389/fpls.2016.01025

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

Matos I.F., Morales L.M.M., Santana D.B., Silva G.M.C., Gomes M.M.A., Ayub R.A., Costa J.H., de Oliveira J.G. Ascorbate synthesis as an alternative electron source for mitochondrial respiration: possible implications for the plant performance. *Front Plant Sci*. 2022;13: 987077. doi 10.3389/fpls.2022.987077

Müller-Moulé P., Golan T., Niyogi K.K. Ascorbate-deficient mutants of Arabidopsis grow in high light despite chronic photooxidative stress. *Plant Physiol.* 2004;134(3):1163-1172. doi 10.1104/pp.103.032375

Pasqualini S., Paolocci F., Borgogni A., Morettini R., Ederli L. The overexpression of an alternative oxidase gene triggers ozone sensitivity in tobacco plants. *Plant Cell Environ*. 2007;30(12):1545-1556. doi 10.1111/j.1365-3040.2007.01730.x

Smirnoff N. Ascorbic acid metabolism and functions: a comparison of plants and mammals. *Free Radic Biol Med.* 2018;122:116-129. doi 10.1016/j.freeradbiomed.2018.03.033

Sweetman C., Waterman C.D., Wong D.C.J., Day D.A., Jenkins C.L.D., Soole K.L. Altering the balance between AOX1A and NDB2 expression affects a common set of transcripts in Arabidopsis. *Front Plant Sci*. 2022;13:876843. doi 10.3389/fpls.2022.876843

Talla S., Riazunnisa K., Padmavathi L., Sunil B., Rajsheel P., Raghavendra A.S. Ascorbic acid is a key participant during the interactions between chloroplasts and mitochondria to optimize photosynthesis and protect against photoinhibition. *J Biosci.* 2011;36:163-173. doi 10.1007/s12038-011-9000-x

Van Aken O. Mitochondrial redox systems as central hubs in plant metabolism and signaling. *Plant Physiol.* 2021;186(1):36-52. doi 10.1093/plphys/kiab101

Vanlerberghe G.C., Dahal K., Alber N.A., Chadee A. Photosynthesis, respiration and growth: a carbon and energy balancing act for alternative oxidase. *Mitochondrion.* 2020;52:197-211. doi 10.1016/j.mito.2020.04.001

Vishwakarma A., Tetali S.D., Selinski J., Scheibe R., Padmasree K. Importance of the alternative oxidase (AOX) pathway in regulating cellular redox and ROS homeostasis to optimize photosynthesis during restriction of the cytochrome oxidase pathway in *Arabidopsis thaliana*. *Ann Bot.* 2015:116(4):555-569. doi 10.1093/aob/mcv122

Wheeler G., Ishikawa T., Pornsaksit V., Smirnoff N. Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife*. 2015:13(4):e06369. doi 10.7554/eLife.06369

Ye J., Coulouris G., Zaretskaya I., Cutcutache I., Rozen S., Madden T.L. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics.* 2012;13:134. doi 10.1186/1471-2105-13-134

Yoshimura K., Nakane T., Kume S., Shiomi Y., Maruta T., Ishikawa T., Shigeoka S. Transient expression analysis revealed the importance of *VTC2* expression level in light/dark regulation of ascorbate biosynthesis in Arabidopsis. *Biosci Biotechnol Biochem.* 2014;78(1): 60-66. doi 10.1080/09168451.2014.877831

Zhu Y., Du B., Qian J., Zou B., Hua J. Disease resistance gene-induced growth inhibition is enhanced by *rcd1* independent of defence activation in Arabidopsis. *Plant Physiol.* 2013;161(4):2005-2013. doi 10.1104/pp.112.213363

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 18.07.2025. После доработки 13.11.2025. Принята к публикации 14.11.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):250-258

doi 10.18699/vjgb-26-28

Генетические корреляции между болезнями человека и плейотропные гены

И.В. Зоркольцева, Н.М. Белоногова, А.В. Кириченко, Я.А. Цепилов, Т.И. Аксенович

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

Аннотация. Характеристикой глобальной генетической общности признаков человека считается их генетическая корреляция. Ее основной механизм – плейотропия, проявляющаяся на различных уровнях. Наиболее интересна плейотропия на уровне генов, поскольку именно они являются фундаментальными функциональными единицами генома. Используя результаты полногеномного анализа ассоциаций 324 болезней, находящиеся в открытом доступе, мы отобрали группу из 45 болезней, в которой каждая пара показала значимую генетическую корреляцию. Эти болезни принадлежали 10 нозологическим категориям. Поиск генов с плейотропными эффектами осуществляли с помощью трех подходов: полногеномного анализа ассоциаций на уровне гена; выбора однонуклеотидных полиморфизмов (SNP) внутри кодирующей части гена, значимо ассоциированных хотя бы с двумя болезнями; и метаанализа сигналов ассоциации SNP со всеми болезнями с последующей идентификацией независимых локусов и приоритизации генов в них. Для всех отобранных таким образом генов мы провели биоинформатический анализ. Всего было идентифицировано 167 плейотропных генов, вовлеченных в контроль 39 болезней. Наиболее плейотропными в нашем исследовании были гены *LPA*, *TCF7L2*, *SLC22A3*, *FES*, *CDKN2B* и *APOE*, каждый из которых контролировал семь-девять болезней. Мы провели биоинформатический анализ всех генов и показали, что найденные нами для 39 болезней плейотропные гены участвуют в контроле еще 501 заболевания. Полученные данные указывают на высокую плейотропную способность этих генов, которая обеспечивается их участием в различных биологических процессах, таких как поддержание гомеостаза, межклеточная сигнализация, регуляция клеточной пролиферации, транспорт веществ и каталитическая активность, а также выполнением ими различных молекулярных функций, в частности связывания с сигнальными рецепторами. Таким образом, мы показали, что 87 % болезней, представляющих полносвязную группу, имеют общие гены с хотя бы еще одной болезнью. Это свидетельствует о том, что генетические корреляции между болезнями в значительной степени обусловлены плейотропными эффектами генов.

Ключевые слова: генетическая корреляция; распространенные болезни; плейотропные гены; анализ ассоциаций на уровне гена; метаанализ; биоинформатический анализ

Для цитирования: Зоркольцева И.В., Белоногова Н.М., Кириченко А.В., Цепилов Я.А., Аксенович Т.И. Генетические корреляции между болезнями человека и плейотропные гены. *Вавиловский журнал генетики и селекции*. 2026; 30(2):250-258. doi 10.18699/vjgb-26-28

Финансирование. Работа выполнена при финансовой поддержке бюджетного проекта № FWNR-2026-0023.

Pleiotropic genes underlying genetic correlations across human diseases

I.V. Zorkoltseva, N.M. Belonogova, A.V. Kirichenko, Y.A. Tsepilov, T.I. Axenovich

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

Abstract. Genetic correlation is a key characteristic of the global genetic similarity of human traits. Its primary underlying mechanism is pleiotropy, which operates at various biological levels. Gene-level pleiotropy is of particular interest, as genes are the fundamental functional units of the genome. Using publicly available results from genome-wide association studies for 324 diseases, we selected a set of 45 diseases in which every pair exhibited a significant genetic correlation. These diseases belonged to 10 nosological categories. The search for genes with pleiotropic effects was carried out using three approaches: (1) gene-based association analysis, (2) selection of single nucleotide polymorphisms (SNP) within gene coding regions significantly associated with at least two diseases, and (3) a cross-trait meta-analysis of SNP association signals followed by the identification of independent loci and gene prioritization within those loci. A comprehensive bioinformatic analysis was performed on all genes identified through these methods. We identified 167 pleiotropic genes implicated in 39 diseases. The most pleiotropic genes in our study were *LPA*, *TCF7L2*, *SLC22A3*, *FES*, *CDKN2B*, and *APOE*, which were associated with 7 to 9 diseases each. Bioinformatic analysis revealed that the pleiotropic genes identified for these 39 diseases are also involved in the genetic architecture of 501 other diseases and traits. This indicates a high degree of pleiotropy, facilitated by the involvement of these genes in diverse biological processes – including homeostasis, cell-cell signaling, regulation of cell proliferation, transport, and catalytic activity – and various

© Зоркольцева И.В., Белоногова Н.М., Кириченко А.В., Цепилов Я.А., Аксенович Т.И., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

molecular functions, such as signaling receptor binding. Thus, we demonstrated that 87 % of diseases within a fully connected correlation network share associated genes with at least one other disease. This finding strongly suggests that genetic correlations between human diseases are largely driven by the pleiotropic effects of shared genes.
Key words: genetic correlation; common diseases; pleiotropic genes; gene-based association analysis; cross-trait meta-analysis; functional enrichment analysis

For citation: Zorkoltseva I.V., Belonogova N.M., Kirichenko A.V., Tsepilov Y.A., Axenovich T.I. Pleiotropic genes underlying genetic correlations across human diseases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2026;30(2): 250-258. doi 10.18699/vjgb-26-28

Введение

Традиционно каждое заболевание изучается отдельно, независимо от других. Но пациенты часто страдают несколькими хроническими заболеваниями. Это серьезная проблема в медицине, поскольку стратегии лечения и его прогноз во многом зависят от наличия сопутствующих патологий. Некоторые группы заболеваний встречаются у пациентов чаще, чем можно ожидать, если они развиваются независимо друг от друга. Это явление, называемое коморбидностью, во многом определяется генетическим сходством заболеваний (Rzhetsky et al., 2007; Wang et al., 2017; Jia et al., 2023). Было показано, что генетическая общность болезней объясняет 46 % коморбидности (Dong et al., 2021).

С развитием техники массового генотипирования и увеличением доступности результатов полногеномного анализа ассоциаций (GWAS) для огромного числа болезней человека стало возможным изучение их генетической общности. Создание метода оценки генетических корреляций между признаками позволило оценить глобальную попарную генетическую общность признаков и болезней человека и составить атлас генетических корреляций (Bulik-Sullivan et al., 2015). Оказалось, что практически каждая болезнь генетически коррелирована хотя бы с одной болезнью или признаком человека.

Однако знание одних только генетических корреляций между болезнями не позволяет понять механизмы этой общности, ключевым механизмом которой является плейотропия, проявляющаяся на различных уровнях. Для признаков человека недавно была показана широкая распространенность плейотропии на уровне генетических вариантов (Mackay, Anholt, 2024). Анализ результатов GWAS для 116 комплексных признаков выявил 2293 независимых локуса и обнаружил, что почти во всех этих локусах топовые варианты, полученные при анализе одного признака, воспроизводятся хотя бы еще для одного признака (Qi et al., 2024).

Наиболее интересна плейотропия на уровне генов, поскольку именно гены являются основными функциональными единицами генома. Авторы подавляющего числа работ изучают плейотропные эффекта генов на ограниченный набор патологий, принадлежащих одной или двум нозологическим категориям. Так, была изучена плейотропия для сердечно-сосудистых (Song J. et al., 2024), психиатрических (Song Q. et al., 2025), респираторных (Chen et al., 2024) болезней. Значительную группу исследований представляют сравнения болезни из одной категории с набором болезней из другой. Например, диабета и сердечно-сосудистых (Adebekun et al., 2024) или желудочно-кишечных (Adewuyi et al., 2024) заболеваний, посттравматического шока и набора сердечно-сосудистых болезней (Shen et al., 2025). В некоторых работах сравниваются два набора болезней из разных категорий, например желудочно-кишечные и психиатрические (Gong et al., 2023).

Хотя результаты этих исследований важны для специалистов, занимающихся определенными патологиями, они не позволяют оценить вклад плейотропии в генетическую общность множества болезней.

В данной работе мы оцениваем вклад плейотропных генов в генетическую общность обширного спектра болезней из разных нозологических категорий. Для этого мы используем находящиеся в открытом доступе результаты GWAS, полученные для 324 болезней на выборке UK Biobank. Среди них мы отбираем группу болезней, в которой каждая пара болезней имеет значимую генетическую корреляцию. Далее будем называть эту группу полносвязной.

Материалы и методы

Схема исследования представлена на рис. 1.

Для 324 болезней мы оценили попарные генетические корреляции между ними и отобрали для дальнейшего анализа полносвязную группу болезней. Анализ проводили с помощью трех подходов, основанных на результатах GWAS этих болезней. Первый из них использует метод полногеномного анализа ассоциаций на уровне гена для идентификации генов, участвующих в контроле хотя бы двух болезней. Второй подход выбирает однонуклеотидные полиморфизмы (SNP), значимо ассоциированные хотя бы с двумя болезнями и лежащие внутри кодирующей части гена. Третий подход основан на метаанализе сигналов ассоциации каждого SNP со всеми болезнями, идентификации независимых локусов и приоритизации генов в них. Для всех отобранных таким образом генов мы провели биоинформатический анализ.

Материалом для исследования служили суммарные статистики SAIGE для фенотипов UK Biobank, полученные на выборке около 400 тыс. человек европейского происхождения (Zhou et al., 2018). Этим фенотипам соответствовали коды и нозологические категории Международной классификации болезней PheCodes (табл. S1 Приложения)[1].

Из 1403 нозологических кодов болезней мы выбрали те, для которых имелось более 1000 больных в выборке, и исключили сильно перекрывающиеся фенотипы. В итоге был сформирован список из 324 наиболее распространенных заболеваний. Суммарные статистики для каждого из них содержали данные о примерно 28 млн SNP.

[1] Табл. S1–S11 Приложения см. по адресу:
https://vavilovj-icg.ru/download/pict-2026-30/appx18.xlsx

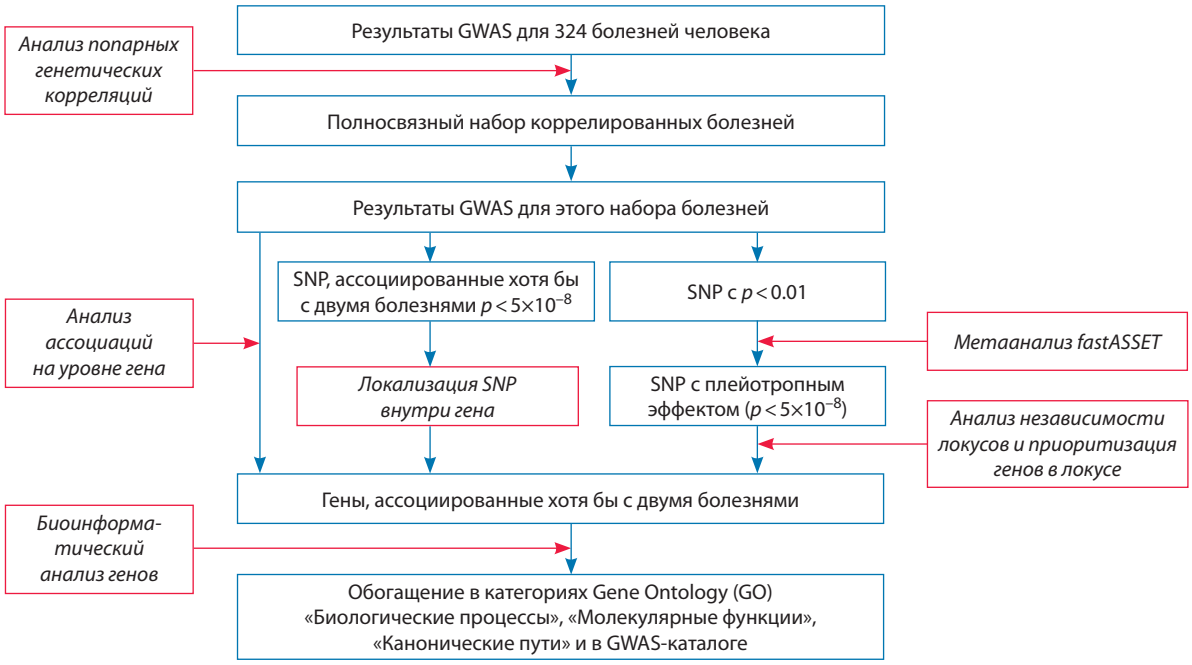


Рис. 1. Схема исследования.

Синими рамками показаны данные и полученные результаты, красными рамками и курсивом – методы анализа.

В анализ включали аутосомные SNP с частотой минорного аллеля (MAF) > 1×10^{-5}. Исключали из анализа регион MHC (Major Histocompatibility Complex) на хромосоме 6, позиции от 24 до 35 млн п. н.

Для анализа ассоциаций на уровне гена использовали матрицы корреляций между генотипами SNP внутри гена, рассчитанные нами ранее (Belonogova et al., 2022).

Анализ генетических корреляций. Для 324 отобранных болезней с помощью программы LDSC были рассчитаны глобальные генетические корреляции (Bulik-Sullivan et al., 2015). Выделена полносвязная группа болезней, в которых все парные генетические корреляции были номинально значимыми ($p \le 0.05$).

Анализ плейотропии на уровне гена. Гены с плейотропным эффектом были идентифицированы с помощью анализа ассоциаций на уровне гена отдельно для каждой болезни, с использованием суммарной статистики (z-score и размер эффекта) для каждого SNP и матрицы корреляций между генотипами всех SNP внутри гена. В анализ включали все SNP, лежащие между стартом и завершением транскрипции.

Анализ проводили с помощью платформы sumSTAAR (Belonogova et al., 2022). Применяли два метода, основанных на суммарных статистиках: SKAT-O (Liu D.J. et al., 2014) и PCA (Wang, Abbott, 2008). Эти методы были реализованы в пакете sumFREGAT (Svishcheva et al., 2019). Мы использовали весовую функцию, зависящую от MAF и определенную бета-распределением с параметрами (1, 1). При PCA учитывали первые главные компоненты, описывающие не менее 0.85 общей дисперсии признака.

Объединение результатов проводили методом ACAT-O (Liu Y. et al., 2019). Рассматривали только белок-кодирующие гены, имеющие хотя бы два SNP с известными суммарными статистиками. Уровень значимости был задан как 2.5×10^{-6} (0.05/20000).

Некоторые гены находились на расстоянии менее 500 тыс. п. н. друг от друга. Если они имели одинаковый паттерн сигналов ассоциации (т. е. были ассоциированы с одними и теми же болезнями), мы не могли их приоритизировать. Однако мы могли исключить ложноположительные сигналы ассоциации в этой группе генов, возникающие из-за неравновесия по сцеплению. Для этого мы использовали базу данных Open Targets Platform (Buniello et al., 2025). Гены, которые не показали ассоциацию ни с одной болезнью, общей для каждой группы генов (association score < 0.3), из исследования исключались.

Анализ плейотропии на уровне SNP. На первом этапе мы отбирали статистически значимые ассоциации по данным GWAS, соответствующие полногеномному уровню значимости (p-value < 5×10^{-8}). В дальнейшем рассматривали только те SNP, которые ассоциировались как минимум с двумя различными заболеваниями. Функциональную роль значимых SNP определяли с помощью аннотации Variant Effect Predictor (VEP, Ensembl) (McLaren et al., 2016). Если вариант локализовался в кодирующей области гена (являлся stop gained, missense variant или synonymous variant), соответствующий ген считался ассоциированным с заболеваниями, для которых данный SNP показал значимую ассоциацию. Такие гены включались в список потенциально плейотропных.

Анализ плейотропии на уровне локуса включает два этапа. На первом этапе мы проводили метаанализ ассоциации каждого SNP с полным набором болезней, применяя метод fastASSET (Fast Association Tests for Multiple Traits), реализованный в пакете ASSET v.1.0.0 (R/Bioconductor). Помимо значимости плейотропных эффектов каж-

дого SNP, этот метод идентифицирует набор признаков, дающих минимальное значение *p*-value метаанализа. Для уменьшения времени счета в метаанализ включают только признаки с *p*-value меньше заданного порога и применяют разбиение всех признаков на относительно независимые группы. Методологическая основа подхода детально описана в работах (Bhattacharjee et al., 2012; Qi et al., 2024).

Для проведения метаанализа каждого SNP мы отбирали болезни с *p*-value < 0.1. Плейотропный эффект считали значимым на полногеномном уровне при *p*-value < 5×10^{-8}. В качестве метрики плейотропии для каждого SNP мы рассматривали число признаков, дающих значимые *p*-value при метаанализе этого SNP.

На втором этапе мы формировали независимые локусы. С этой целью сначала проводили клампинг (clumping) в программе PLINK v1.9 (www.cog-genomics.org/plink/1.9) с параметрами: '--clump-p1 5e-8 --clump-p2 1e-5 --clump-r2 0.1 --clump-kb 1000'. В анализе использовались данные о неравновесии по сцеплению (LD) из проекта 1000 Genomes (фаза 3, европейская популяция) и координаты референсного генома GRCh37/hg19.

Затем мы искали независимые генетические локусы, используя следующий алгоритм обработки результатов клампинга. Если два лидирующих SNP были расположены ближе 500 тыс. п. н. друг от друга, то регионы объединяли и выбирали в качестве лидирующего SNP с меньшим *p*-value. При сопоставимой значимости (разница *p*-value меньше одного порядка) приоритет отдавали SNP, расположенному в гене или имеющему большее число ассоциированных признаков. Для SNP в пределах одного гена применялось правило наименьшего *p*-value.

Каждому лидирующему SNP в независимых локусах мы приписывали ген, в котором он локализован, или ближайший к нему ген, находящийся в пределах локуса. Этот ген рассматривался нами как плейотропный.

Биоинформатический анализ. Анализ обогащения наборов генов был выполнен с помощью модуля GENE2FUNC платформы FUMA (Watanabe et al., 2017) (http://fuma.ctglab.nl/). На вход подавали гены, идентифицированные в ходе проведенных анализов.

Для анализа функционального обогащения мы использовали наборы генов, аннотированные в базах данных Gene Ontology (GO) (биологические процессы и молекулярные функции), канонических путей и GWAS-каталога.

Во всех анализах применялись параметры по умолчанию. Поправка на множественное тестирование выполнялась методом Бенджамини–Хохберга. Статистически значимыми считались результаты со скорректированным *p*-value < 0.05.

Результаты

Анализ генетических корреляций

Среди 324 болезней 220 были генетически коррелированы хотя бы с одной болезнью. В полносвязную группу болезней вошли только 45 из них (рис. 2).

Эти болезни представляли 10 нозологических категорий, основанных на классификации PheCodes (рис. 3). Список болезней и категорий приведен в табл. S1. Наиболее представленными были категории желудочно-кишечных и костно-мышечных болезней.

Анализ плейотропии на уровне гена

Значимые сигналы ассоциации на уровне генов были обнаружены для 26 из 45 болезней (табл. S2). Наибольшее число ассоциированных генов присутствовало в категориях сердечно-сосудистых патологий (400) и метаболических нарушений (165). Всего выявлено 680 генов, ассоциированных хотя бы с одной болезнью. Среди них 129 генов (19 %) продемонстрировали ассоциацию не менее чем с двумя болезнями.

Мы исключили из рассмотрения 8 генов, которые были тесно сцеплены с другими генами и не прошли контроль на основе Open Targets Platform (см. Материалы и методы). Таким образом, в дальнейший анализ был включен 121 ген с плейотропными эффектами. Эти гены участвовали в контроле от двух до шести болезней. Общее число болезней, контролируемых хотя бы одним из 121 гена, составило 19. Они входили в шесть нозологических категорий (табл. S3). Наибольшее число генов было общим для болезней сердечно-сосудистой категории и метаболических нарушений, а также болезней внутри этих категорий.

Анализ плейотропных эффектов SNP

Из результатов GWAS мы отобрали 1389 SNP, значимо ассоциированных (*p*-value < 5×10^{-8}) хотя бы с двумя признаками. Для каждого из этих SNP была определена его локализация относительно генов. 36 SNP находились в белок-кодирующих областях 23 генов. Далее мы рассматривали их как гены с возможным плейотропным эффектом. Общие гены были найдены для 11 болезней из трех нозологических категорий (табл. S4).

Анализ плейотропных эффектов локусов

Метаанализ выявил 3210 SNP, значимо ассоциированных ($p < 5\times10^{-8}$) с двумя и более признаками. Полный список ассоциаций представлен в табл. S5. После проведения клампинга и фильтрации был идентифицирован 71 независимый локус (табл. S6). Наибольшее количество независимых локусов наблюдалось на хромосомах 2 и 10: 9 и 7 локусов соответственно. В 11 локусах не обнаружено генов, к которым можно было отнести лидирующий SNP. То есть из 71 лидирующего SNP только 60 маркировали гены. Общие гены были найдены для 36 болезней из 9 нозологических категорий.

Сравнение результатов различных анализов

На рис. 4, *а* представлена диаграмма, иллюстрирующая число плейотропных генов, обнаруженных каждым методом, и их перекрывание. Максимальное число плейотропных генов выявлено при анализе на уровне гена, а минимальное – на уровне SNP.

Распределение числа заболеваний, для которых был идентифицирован хотя бы один общий ген с другими болезнями в зависимости от используемого метода, показано на рис. 4, *б*. Наибольшее количество таких заболеваний выявлено анализом на уровне локуса, в то время как наименьшее – на уровне SNP.

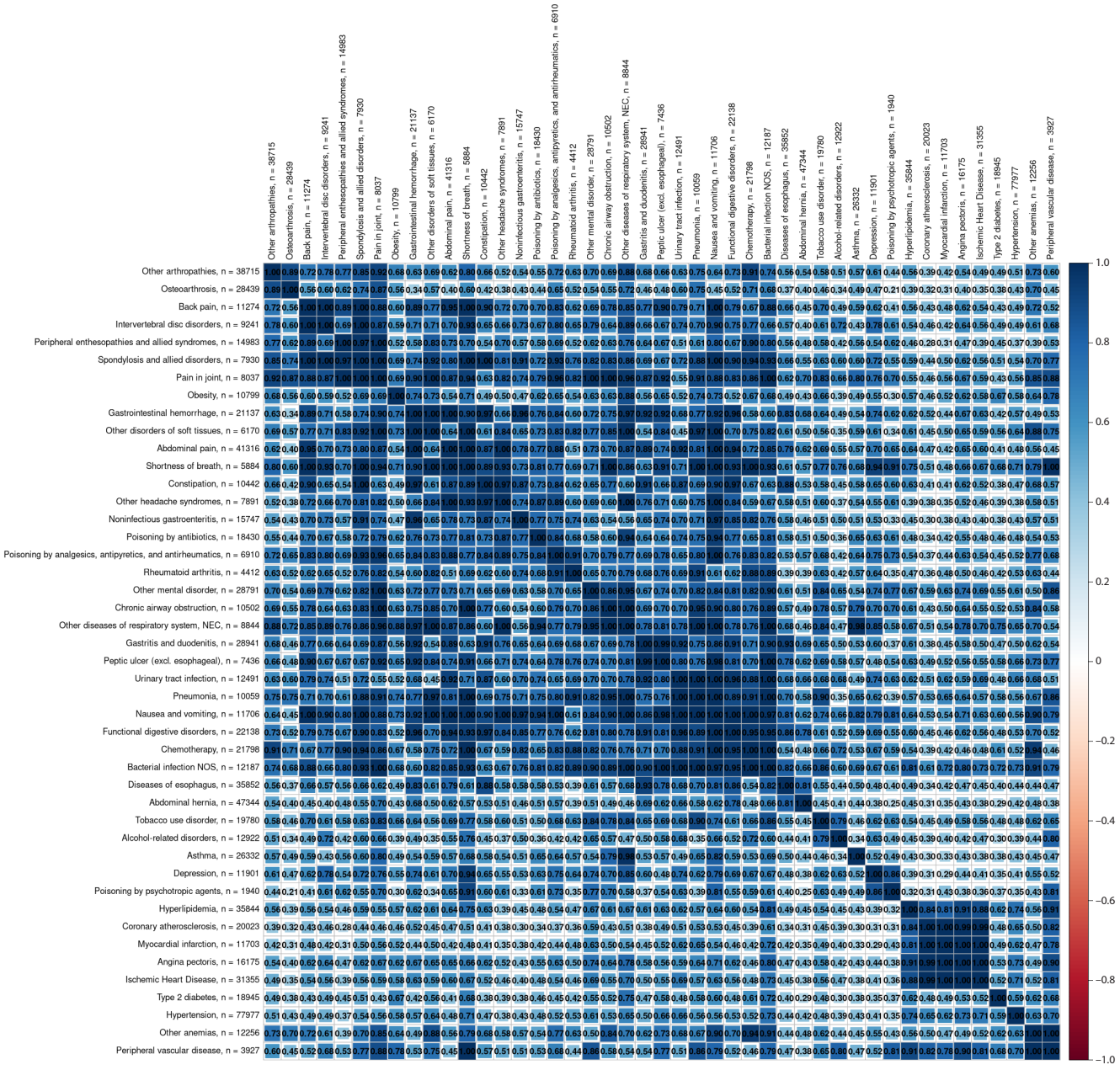


Рис. 2. Тепловая карта генетических корреляций между 45 полносвязными признаками.
n – число больных для каждого признака. Метод 'hclust' был использован для кластеризации.

Анализ суммарных результатов, полученных всеми методами

Мы провели объединение результатов, полученных всеми методами. Оказалось, что шесть болезней – отравление психотропными агентами (poisoning by agents), болезни периферических сосудов (peripheral vascular disease), ревматоидный артрит (rheumatoid arthritis), одышка (shortness of breath), запор (constipation) и химиотерапия (chemotherapy) – не имели общих генов ни с одной другой болезнью. Первые четыре болезни отличались самой низкой распространенностью (от 0.005 до 0.015) среди выбранных 45 болезней (см. табл. S1). Распространенность измеряли долей больных в большой популяционной выборке UK Biobank.

Распределение 39 болезней по нозологическим категориям было близко к распределению исходных 45 болезней, представленному на рис. 3.

Общее число плейотропных генов, найденных всеми тремя методами, составило 167 (табл. S7). Распределение этих генов по геному приведено на рис. 5. Как видно, они локализованы на всех аутосомах, но с разной плотностью. Число болезней, ассоциированных с каждым геном, варьирует от 2 до 9.

Далее мы построили сеть взаимодействий болезней, опосредованных плейотропными генами (рис. 6). Наибольшее число общих генов обнаружилось у болезней сердечно-сосудистой системы и гиперлипидемии. Все болезни, кроме двух (инфекции мочевыводящих путей



Рис. 3. Распределение анализируемых болезней по нозологическим категориям.

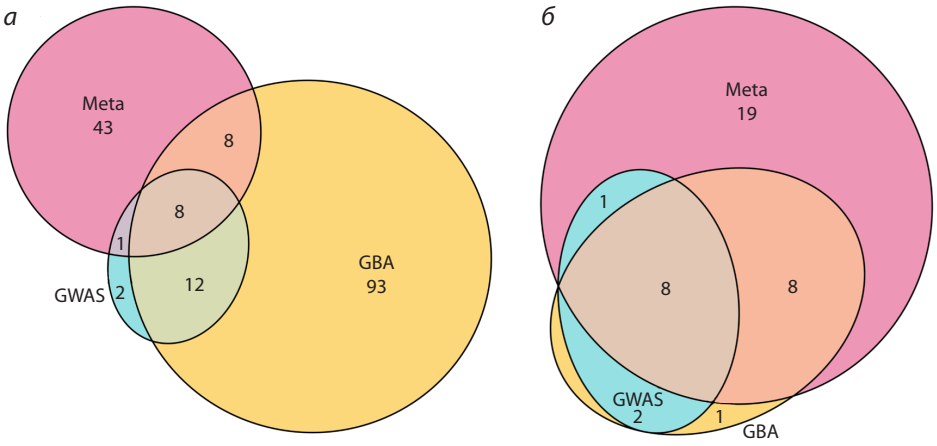


Рис. 4. Диаграмма результатов анализа трех методов: на уровне гена (GBA), на уровне локуса (Meta) и на уровне SNP (GWAS).

Кругами показано: *а* – число генов, идентифицированных каждым методом; *б* – число болезней, имеющих хотя бы один общий ген с другими болезнями для каждого метода.

и дегенерация межпозвонковых дисков), имеют общие гены хотя бы с двумя другими болезнями. Болезни, принадлежащие одной категории, в большинстве случаев соединены непосредственно или через одну болезнь другой категории. Большая часть болезней характеризуется высокой степенью связанности с другими болезнями той же или иной категории.

Плейотропные гены часто разделяют на те, которые являются общими для одной нозологической категории болезней, и те, плейотропный эффект которых распространяется на болезни разных категорий. Из 167 плейотропных генов 44 были общими только для болезней одной категории, а остальные объединяли от двух до шести категорий. Наибольшее число ассоциаций обнаружили гены *LPA*, *TCF7L2*, *SLC22A3*, *FES*, *CDKN2B* и *APOE*, общие для семи-девяти болезней каждый.

Биоинформатический анализ

Была выявлена 261 статистически значимо обогащенная категория Gene Ontology: 252 категории в разделе «Биологические процессы» (Biological processes) и 9 категорий в разделе «Молекулярные функции» (Molecular functions). Среди обогащенных биологических процессов наибольшая представленность наблюдалась для генов, участвующих в регуляции гомеостаза, реакции на эндогенные стимулы, межклеточной сигнализации, метаболизме малых молекул, регуляции клеточной пролиферации, транспорте веществ и каталитической активности. В категории молекулярных функций лидирующую позицию заняли гены, аннотированные термином «связывание с сигнальными рецепторами». Основная доля генов, вовлеченных в канонические пути, пришлась на процессы связывания с сигнальными рецепторами и транспорта малых молекул. При этом наиболее значимое обогащение было достигнуто для путей метаболизма липидов. К ним относятся: ингибирование синтеза холестерина статинами, метаболизм холестерина, сборка, ремоделирование и клиренс липопротеинов плазмы, метаболические пути ЛПНП, ЛПВП и триглицеридов (включая ассоциированные заболевания), ремоделирование липопротеинов плазмы, а также путь, связанный с семейной гиперлипидемией 2-го типа. Полные списки генов для обогащенных категорий биологических процессов, молекулярных функций и канонических путей представлены в табл. S8–S10.

Анализ обогащения с использованием наборов генов из GWAS-каталога выявил 501 значимо обогащенный при-

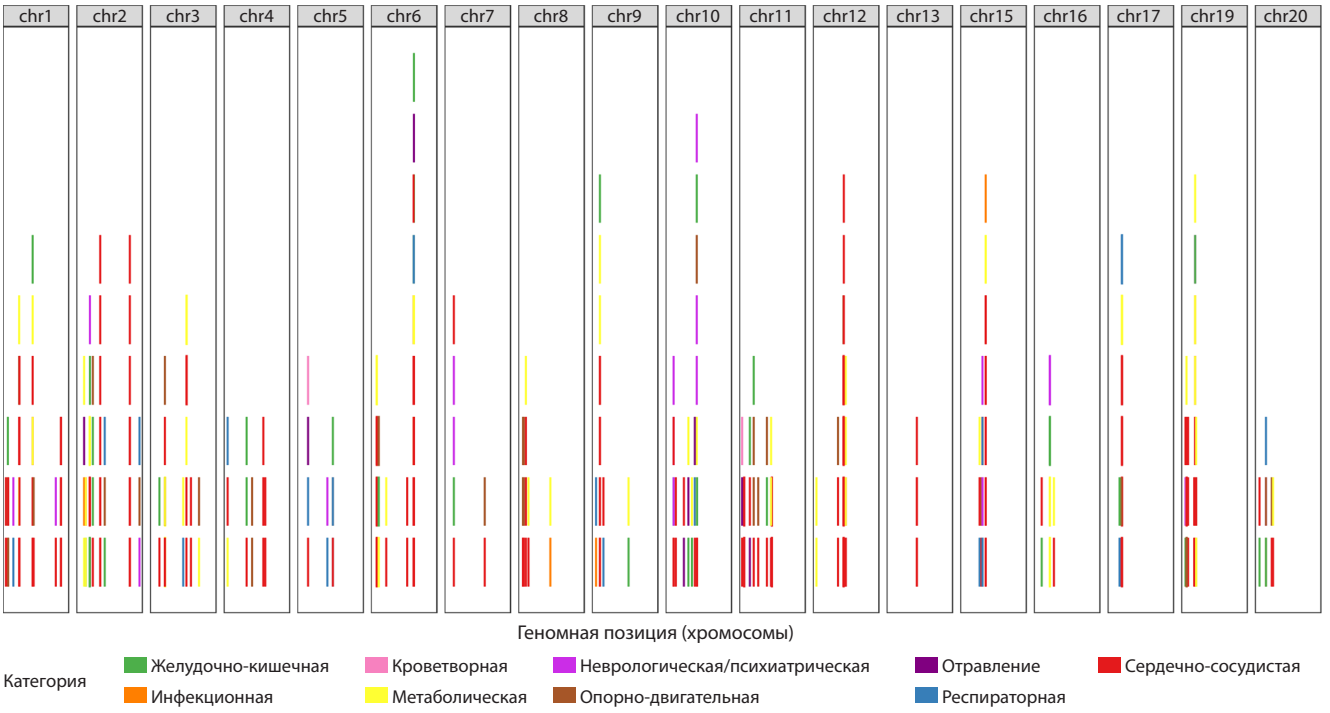


Рис. 5. Геномная карта ассоциаций ген-болезнь.

Каждый вертикальный отрезок соответствует одной болезни, а его цвет – нозологической категории болезней.

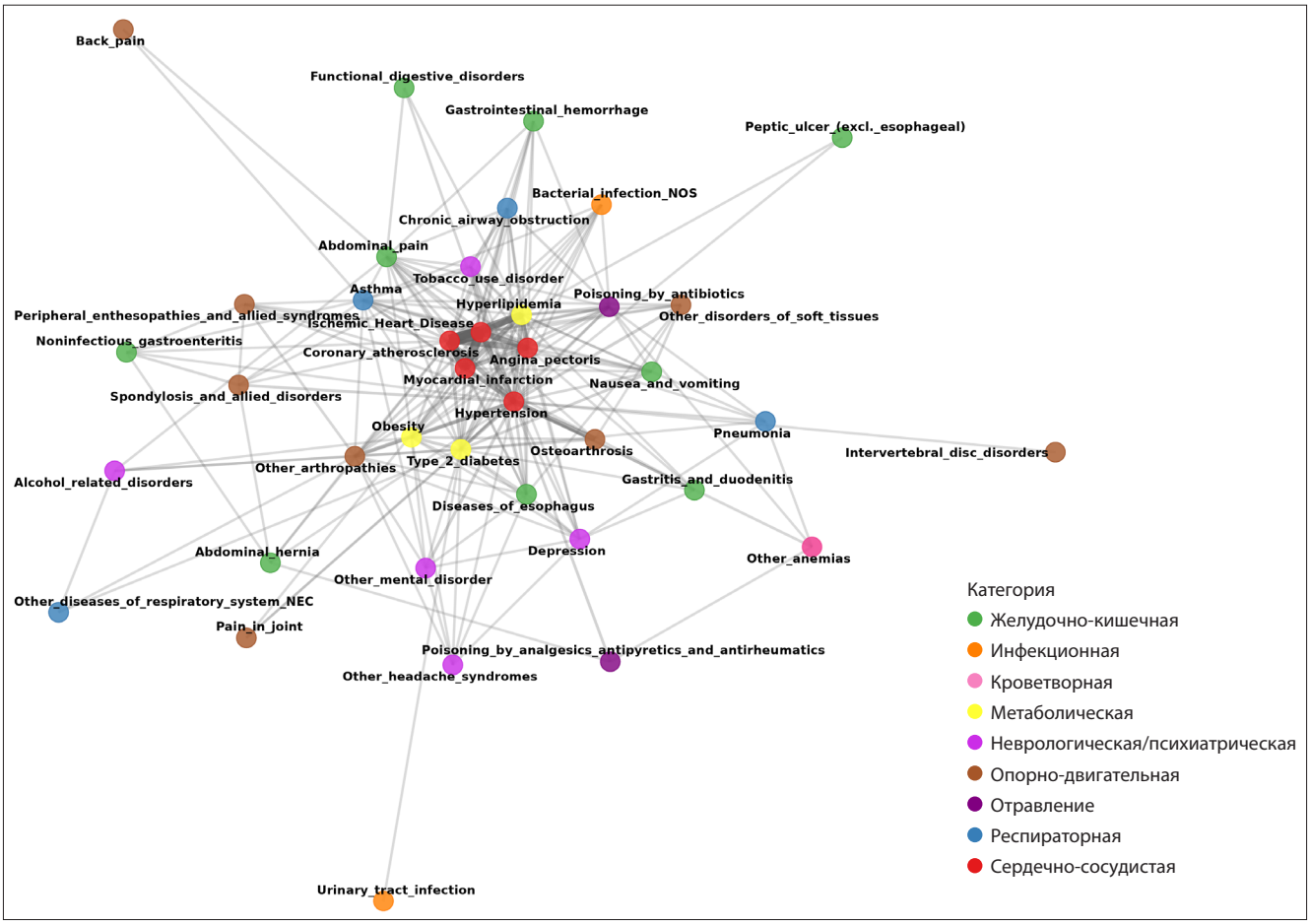


Рис. 6. Сеть болезней, связанных общими генами.

Цвет кружков соответствует нозологической категории болезни, линии – общим для пары болезней генам, толщина линий – числу общих генов.

знак. Самые сильные ассоциации наблюдались с сердечно-сосудистыми заболеваниями (ишемическая болезнь сердца, инфаркт миокарда), показателями артериального давления, фенотипами липидного обмена и индексом массы тела (табл. S11). Это хорошо соотносится с тем, что наиболее значимо обогащенные канонические пути связаны с метаболизмом липидов.

Обсуждение

В данной работе мы исследовали вклад плейотропных генов в генетическую общность распространенных болезней человека. Объектом анализа стала полносвязная группа, состоящая из 45 болезней. В анализ вошли не только классические мультифакториальные заболевания, но и состояния, развитие которых в значительной степени связано с внешними причинами. Однако сам факт наличия у них значимых генетических корреляций с другими болезнями свидетельствует о существовании общей генетической компоненты. Мы показали, что 39 из них (87 %) имеют общие гены хотя бы с одной другой болезнью и могут быть представлены в виде связного графа (см. рис. 6). Для остальных шести болезней мы не нашли общих генов из-за недостаточной информативности болезней, обусловленной низкой распространенностью, или из-за проявления плейотропии на других уровнях, а не на уровне гена. В пользу этого свидетельствует тот факт, что анализ ассоциаций на уровне гена не обнаружил ни одного значимого гена для пяти из этих болезней.

Для выявления плейотропных генов были применены три разных подхода, основанных на анализе SNP, локусов и генов. Всего с помощью этих методов мы идентифицировали 167 генов с плейотропным эффектом. Сравнение рассмотренных подходов показало, что наибольшее число плейотропных генов было идентифицировано с помощью анализа ассоциаций на уровне гена. Данный метод одновременно учитывает информацию о всех SNP внутри гена при тестировании ассоциаций и считается наиболее эффективным методом нахождения ассоциированных генов, особенно если в анализ вовлечены редкие SNP (Lee et al., 2012). Наименьшей эффективностью обладал подход, основанный на анализе SNP для поиска плейотропных генов. Однако это, вероятно, связано не столько с особенностью самого подхода, сколько с жесткими критериями аннотации SNP к генам, установленными нами для минимизации ложноположительных результатов.

Найденные нами гены различались по степени плейотропии: число ассоциированных с ними болезней варьировало от двух до девяти. Самые плейотропные гены – *LPA*, *TCF7L2*, *SLC22A3*, *FES*, *CDKN2B* и *APOE* – были ассоциированы с семью-девятью заболеваниями. Известно, что эти гены задействованы в фундаментальных клеточных и метаболических процессах, что объясняет их широкое влияние на различные фенотипы.

Так, ген *TCF7L2* кодирует транскрипционный фактор, который играет ключевую роль в Wnt-сигнальном пути. Этот путь контролирует рост, движение, выживание и специализацию клеток в развивающемся организме (Jin, Liu 2008; Facchinello et al., 2017). Ген *CDKN2B* кодирует белок, который является клеточным регулятором роста и контролирует прохождение G1-фазы клеточного цикла (Xia et al., 2021). *FES* кодирует тирозинкиназу, участвующую в регуляции актинового цитоскелета, сборке микротрубочек, а также в процессах клеточной адгезии и миграции (Laurent et al., 2004). Нарушение столь фундаментальных процессов закономерно приводит к повышенному риску развития широкого спектра заболеваний.

Мы провели биоинформатический анализ всех 167 плейотропных генов, оценив их представленность в различных функциональных аннотациях. Выявлено значительное обогащение множества категорий, что свидетельствует о широкой функциональной вовлеченности данных генов. Идентифицированные плейотропные гены значимо обогащены ассоциациями с сотнями (501) признаков и заболеваний из GWAS-каталога, что подтверждает их высокий плейотропный потенциал. Эта универсальность обусловлена вовлеченностью данных генов в ключевые биологические процессы, включая поддержание гомеостаза, межклеточную сигнализацию, регуляцию клеточной пролиферации, транспорт веществ и каталитическую активность, а также выполнением ими различных молекулярных функций, таких как связывание с сигнальными рецепторами.

Плейотропные эффекты генов во многом объясняют феномен коморбидности – одновременного развития нескольких заболеваний у одного пациента, что существенно осложняет диагностику и лечение (Gratten, Visscher, 2016). Выявление плейотропных генов позволяет определить новые терапевтические мишени и создавать препараты, воздействующие на несколько патологий одновременно (Bao et al., 2024). Плейотропия также открывает возможности для перепрофилирования существующих лекарств на новые заболевания (Pushpakom et al., 2019). В то же время плейотропность генов усложняет разработку таргетных препаратов, требуя оценки потенциальных побочных воздействий на другие заболевания (Nguyen et al., 2019).

Заключение

Таким образом, мы показали, что 87 % болезней, представляющих полносвязную группу, имеют общие гены с хотя бы еще одной болезнью. Более того, все эти болезни оказались включенными в одну связную группу. Найденные нами гены контролируют болезни, принадлежащие как одной, так и нескольким нозологическим категориям. Все это свидетельствует о том, что генетические корреляции между болезнями в значительной степени обусловлены плейотропными эффектами генов.

Список литературы / References

Adebekun J., Nadig A., Saarah P., Asgari S., Kachuri L., Alagpulinsa D.A. Genetic relations between type 1 diabetes, coronary artery disease and leukocyte counts. *Diabetologia.* 2024;67(11):2518-2529. doi 10.1007/s00125-024-06247-9

Adewuyi E.O., Porter T., O'Brien E.K., Olaniru O., Verdile G., Laws S.M. Genome-wide cross-disease analyses highlight causality and shared biological pathways of type 2 diabetes with gastrointestinal disorders. *Commun Biol.* 2024;7(1):643. doi 10.1038/s42003-024-06333-z

Bao C., Tan T., Wang S., Gao C., Lu C., Yang S., Diao Y., Jiang L., Jing D., Chen L., Lv H., Fang H. A cross-disease, pleiotropy-driven approach for therapeutic target prioritization and evaluation. *Cell Rep Methods*. 2024;4(4):100757. doi 10.1016/j.crmeth.2024.100757

Belonogova N.M., Svishcheva G.R., Kirichenko A.V., Zorkoltseva I.V., Tsepilov Y.A., Axenovich T.I. sumSTAAR: a flexible framework for gene-based association studies using GWAS summary statistics. *PLoS Comput Biol.* 2022;18(6):e1010172. doi 10.1371/journal.pcbi.1010172

Bhattacharjee S., Rajaraman P., Jacobs K.B., Wheeler W.A., Melin B.S., Hartge P.; GliomaScan Consortium; Yeager M., Chung C.C., Chanock S.J., Chatterjee N. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am J Hum Genet.* 2012;90(5):821-835. doi 10.1016/j.ajhg.2012.03.015

Bulik-Sullivan B., Finucane H.K., Anttila V., Gusev A., Day F.R., Loh P.R.; ReproGen Consortium; … Patterson N., Robinson E.B., Daly M.J., Price A.L., Neale B.M. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-1241. doi 10.1038/ng.3406

Buniello A., Suveges D., Cruz-Castillo C., Llinares M.B., Cornu H., Lopez I., Tsukanov K., … Ghoussaini M., Dunham I., Hulcoop D.G., McDonagh E.M., Ochoa D. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res.* 2025;53(D1):D1467-D1475. doi 10.1093/nar/gkae1128

Chen J., Zhang X., Sun G. Causal relationship between type 2 diabetes and common respiratory system diseases: a two-sample Mendelian randomization analysis. *Front Med* (*Lausanne*). 2024;11:1332664. doi 10.3389/fmed.2024.1332664

Dong G., Feng J., Sun F., Chen J., Zhao X.M. A global overview of genetically interpretable multimorbidities among common diseases in the UK Biobank. *Genome Med*. 2021;13(1):110. doi 10.1186/s13073-021-00927-6

Facchinello N., Tarifeno-Saldivia E., Grisan E., Schiavone M., Peron M., Mongera A., Ek O., Schmitner N., Meyer D., Peers B., Tiso N., Argenton F. Tcf7l2 plays pleiotropic roles in the control of glucose homeostasis, pancreas morphology, vascularization and regeneration. *Sci Rep.* 2017;7(1):9605. doi 10.1038/s41598-017-09867-x

Gong W., Guo P., Li Y., Liu L., Yan R., Liu S., Wang S., Xue F., Zhou X., Yuan Z. Role of the gut-brain axis in the shared genetic etiology between gastrointestinal tract diseases and psychiatric disorders: a genome-wide pleiotropic analysis. *JAMA Psychiatry.* 2023;80(4):360-370. doi 10.1001/jamapsychiatry.2022.4974

Gratten J., Visscher P.M. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med*. 2016;8(1):78. doi 10.1186/s13073-016-0332-x

Jia G., Li Y., Zhong X., Wang K., Pividori M., Alomairy R., Esposito A., … Kubo M., Cox N.J., Evans J., Gao X., Rzhetsky A. The high-dimensional space of human diseases built from diagnosis records and mapped to genetic loci. *Nat Comput Sci.* 2023;3(5):403-417. doi 10.1038/s43588-023-00453-y

Jin T., Liu L. The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus. *Mol Endocrinol*. 2008;22(11):2383-2392. doi 10.1210/me.2008-0135

Laurent C.E., Delfino F.J., Cheng H.Y., Smithgall T.E. The human c-Fes tyrosine kinase binds tubulin and microtubules through separate domains and promotes microtubule assembly. *Mol Cell Biol.* 2004;24(21):9351-9358. doi 10.1128/MCB.24.21.9351-9358.2004

Lee S., Wu M.C., Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics.* 2012;13(4):762-775. doi 10.1093/biostatistics/kxs014

Liu D.J., Peloso G.M., Zhan X., Holmen O.L., Zawistowski M., Feng S., Nikpay M., … Willer C.J., Hveem K., Melander O., Kathiresan S., Abecasis G.R. Meta-analysis of gene-level tests for rare variant association. *Nat Genet.* 2014;46(2):200-204. doi 10.1038/ng.2852

Liu Y., Chen S., Li Z., Morrison A.C., Boerwinkle E., Lin X. ACAT: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am J Hum Genet*. 2019;104(3):410-421. doi 10.1016/j.ajhg.2019.01.002

Mackay T.F.C., Anholt R.R.H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat Rev Genet.* 2024;25(9):639-657. doi 10.1038/s41576-024-00711-3

McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R., Thormann A., Flicek P., Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122. doi 10.1186/s13059-016-0974-4

Nguyen P.A., Born D.A., Deaton A.M., Nioi P., Ward L.D. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat Commun*. 2019;10(1):1579. doi 10.1038/s41467-019-09407-3

Pushpakom S., Iorio F., Eyers P.A., Escott K.J., Hopper S., Wells A., Doig A., Guilliams T., Latimer J., McNamee C., Norris A., Sanseau P., Cavalla D., Pirmohamed M. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019;18(1):41-58. doi 10.1038/nrd.2018.168

Qi G., Chhetri S.B., Ray D., Dutta D., Battle A., Bhattacharjee S., Chatterjee N. Genome-wide large-scale multi-trait analysis characterizes global patterns of pleiotropy and unique trait-specific variants. *Nat Commun.* 2024;15(1):6985. doi 10.1038/s41467-024-51075-5

Rzhetsky A., Wajngurt D., Park N., Zheng T. Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci USA.* 2007;104(28):11694-11699. doi 10.1073/pnas.0704820104

Shen J., Valentim W., Friligkou E., Overstreet C., Choi K.W., Koller D., O'Donnell C.J., … Lv H., Sun L., Falcone G.J., Polimanti R., Pathak G.A. Shared genetic architecture of posttraumatic stress disorder with cardiovascular imaging, risk, and diagnoses. *Nat Commun.* 2025;16(1):5631. doi 10.1038/s41467-025-60487-w

Song J., Gao N., Chen Z., Xu G., Kong M., Wei D., Sun Q., Dong A. Shared genetic etiology of vessel diseases: a genome-wide multi-traits association analysis. *Thromb Res.* 2024;241:109102. doi 10.1016/j.thromres.2024.109102

Song Q., Zhang C., Wang W., Wang C., Yi C. Exploring the genetic landscape of the brain-heart axis: a comprehensive analysis of pleiotropic effects between heart disease and psychiatric disorders. *Prog Neuropsychopharmacol Biol Psychiatry.* 2025;136:111172. doi 10.1016/j.pnpbp.2024.111172

Svishcheva G.R., Belonogova N.M., Zorkoltseva I.V., Kirichenko A.V., Axenovich T.I. Gene-based association tests using GWAS summary statistics. *Bioinformatics.* 2019;35(19):3701-3708. doi 10.1093/bioinformatics/btz172

Wang K., Abbott D. A principal components regression approach to multilocus genetic association studies. *Genet Epidemiol.* 2008;32(2):108-118. doi 10.1002/gepi.20266

Wang K., Gaitsch H., Poon H., Cox N.J., Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat Genet.* 2017;49(9):1319-1325. doi 10.1038/ng.3931

Watanabe K., Taskesen E., van Bochoven A., Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1826. doi 10.1038/s41467-017-01261-5

Xia Y., Liu Y., Yang C., Simeone D.M., Sun T.T., DeGraff D.J., Tang M.S., Zhang Y., Wu X.R. Dominant role of *CDKN2B/p15INK4B* of 9p21.3 tumor suppressor hub in inhibition of cell-cycle and glycolysis. *Nat Commun.* 2021;12(1):2047. doi 10.1038/s41467-021-22327-5

Zhou W., Nielsen J.B., Fritsche L.G., Dey R., Gabrielsen M.E., Wolford B.N., LeFaive J., … Hveem K., Kang H.M., Abecasis G.R., Willer C.J., Lee S. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet.* 2018;50(9):1335-1341. doi 10.1038/s41588-018-0184-y

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 04.09.2025. После доработки 01.11.2025. Принята к публикации 01.12.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):259-266

doi 10.18699/vjgb-26-29

Novel regulatory SNPs that can be activated due to metformin treatment may orchestrate liver gluconeogenesis and add to the variability in AMPK-dependent mechanisms of metformin response

E.E. Korbolina, I.S. Damarov, T.I. Merkulova

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

Abstract. Metformin is a first-line therapy for type 2 diabetes, yet individual response varies significantly, with over 30 % of patients failing to achieve optimal glycemic control. The specific regulatory mechanisms of this phenomenon remain poorly understood and genetic variants involved are mainly undiscovered. There are multiple lines of evidence that the leading role in determining the variance in phenotypic outcome belongs to regulatory SNPs (rSNPs) as they directly modify gene expression. Therefore, the genome-wide search for such functional variants and deciphering associated phenotypes stands as a fundamental challenge. Previously, based on the results of bioinformatics analysis of allele-specific expression and binding landscape in human peripheral blood mononuclear cells, we have established an original panel of 14 796 rSNPs within promotors of 5132 genes. Aiming to pinpoint functional variants most likely linked to metformin hepatic response and impacts on liver gluconeogenesis, we analyzed the relevant open-access data as well as rSNPs from our panel and the corresponding genes. 1196 genes reported to be regulated by metformin in human hepatocytes and 115 genes involved in gluconeogenesis and/or its regulation via Gene Ontology annotations were intersected. Free R software and STRING v.11 tools were used for functional annotation. A number of genes harboring rSNPs within promotor regions were found to be particularly implicated in the mechanisms of metformin's action. Functional enrichment analyses revealed enrichment in critical pathways including FoxO, TNF-α and TGF-β signaling, also implicated in diabetes complications. Among these, six genes (*ARPP19*, *ATF4*, *NR3C1*, *PFKFB3*, *TCF7L2*, and *WDR5*) were strongly associated with regulation of gluconeogenesis, and may be modulated by metformin in the liver. We conclude that metformin therapy response may be influenced by the newly identified functional SNPs including rSNPs within the promotors of genes for gluconeogenic enzymes and transcription regulators.

Key words: metformin; individual drug response; TD2M; regulatory SNPs; gluconeogenesis; AMPK-dependent mechanisms; signal transduction pathways (FoxO, TNF-α, TGF-β)

For citation: Korbolina E.E., Damarov I.S., Merkulova T.I. Novel regulatory SNPs that can be activated due to metformin treatment may orchestrate liver gluconeogenesis and add to the variability in AMPK-dependent mechanisms of metformin response. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):259-266. doi 10.18699/vjgb-26-29

Funding. This study was supported by the RSF, Russian Science Foundation (No. 23-15-00113).

Новые регуляторные однонуклеотидные полиморфизмы (rSNPs) потенциально участвуют в регуляции глюконеогенеза в печени и вносят вклад в вариабельность АМФК-зависимых механизмов индивидуального ответа на метформин

Е.Е. Корболина, И.С. Дамаров, Т.И. Меркулова

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

Аннотация. Метформин является препаратом первого выбора для лечения сахарного диабета второго типа, однако более чем у 30 % пациентов не достигается оптимальный гликемический контроль. Лежащие в основе этого феномена регуляторные механизмы и ассоциированные генетические варианты остаются неизученными. Накопленные данные свидетельствуют о том, что ключевую роль в определении фенотипической вариабельности играют функциональные генетические варианты, в особенности регуляторные однонуклеотидные полиморфизмы (rSNP), непосредственно влияющие на уровень экспрессии генов. Полногеномный поиск таких функциональных вариантов и выявление их фенотипических последствий – актуальная научная задача. Ранее на основе результатов биоинформатического анализа аллель-специфической экспрессии и профиля распределения активирующих гистоновых модификаций в мононуклеарных клетках периферической крови человека нами была сформирована

© Korbolina E.E., Damarov I.S., Merkulova T.I., 2026
This work is licensed under a Creative Commons Attribution 4.0 License

оригинальная панель из 14 796 rSNP, локализованных в промоторах 5132 генов. Целью настоящей работы был поиск в нашей панели rSNP вариантов, потенциально связанных с регуляцией процессов глюконеогенеза в печени, в том числе под действием монотерапии метформином, с помощью анализа независимых релевантных полногеномных данных. Анализировали 1196 генов, экспрессия которых в гепатоцитах человека изменялась под влиянием метформина АМФК-зависимым образом, а также 115 генов, вовлеченных в глюконеогенез и/или его регуляцию, согласно данным генных онтологий. Для функциональной аннотации использовали инструменты свободного программного обеспечения R и базы данных STRING (v.11). В результате анализа был выделен ряд генов, которые, вероятно, играют особую роль в механизмах действия метформина и индивидуальной чувствительности к препарату, в промоторных районах которых нами были картированы rSNP. Функциональный анализ обогащения выявил вовлеченность этих генов в ключевые сигнальные пути, ассоциированные также с осложнениями сахарного диабета второго типа. Для шести генов (*ARPP19*, *ATF4*, *NR3C1*, *PFKFB3*, *TCF7L2* и *WDR5*) экспрессия не только тесно связана с регуляцией глюконеогенеза, но также может модулироваться в гепатоцитах под действием метформина. Мы предполагаем, что rSNP в промоторах генов ферментов, участвующих в глюконеогенезе, и транскрипционных факторов могут играть существенную роль в механизмах формирования индивидуального ответа на терапию метформином.

Ключевые слова: метформин; индивидуальный ответ на препарат; сахарный диабет второго типа; регуляторные полиморфизмы (rSNPs); глюконеогенез; АМФК-зависимая киназа; FoxO, TNF-α, TGF-β-сигнальные пути

Introduction

Metformin, an oral biguanide agent, is the first-line medication to reduce blood glucose levels in patients with T2DM. Although its mechanism of action is still not fully understood, metformin has been shown to decrease liver gluconeogenesis and enhance insulin sensitivity in muscle and adipose tissue (Rena et al., 2017; LaMoia, Shulman, 2021). Moreover, its beneficial pleiotropic effects beyond glycemic control are well known, which include lowering the incidence of cardiovascular events (Bu et al., 2022), correcting energy metabolism in neurologic disorders (Loan et al., 2024), and regulating inflammatory markers (Karbalaee-Hasani et al., 2021).

At the molecular level, metformin can inhibit complex I of the mitochondrial electron transport chain in hepatocytes (El-Mir et al., 2000), leading to decreases in ATP and increases in AMP levels and in turn activating AMP-activated protein kinase (AMPK), a master controller of metabolic homeostasis (Smiles et al., 2024). The activation of AMPK results in suppression of gluconeogenic gene expression and reduction in glucose output (LaMoia, Shulman, 2021). Moreover, there is evidence that the metformin glucose-lowering effects in hepatocytes can occur via AMPK-independent mechanisms. In particular, increased AMP levels can directly suppress gluconeogenesis via allosteric inhibition of fructose-1,6-bisphosphatase 1, FBP1 (Hunter et al., 2018). There is also an increasing interest in examining the effects of AMPK activity modulation in chronic metabolic diseases and cancer (Strang et al., 2025), including using the activators and inhibitors with effects independent of AMPK, i. e. compound C, an ATP-competitive AMPK inhibitor (which was reported to disrupt various biological events).

However, metformin's response varies among individuals. As is shown, metformin is ineffective in over 30 % of cases because of various factors, including genetic ones (Cook et al., 2007; St-Amour et al., 2025). Currently, the search for genetic variants (mainly SNPs) associated with drug response efficacy is mainly focused on the genes coding for phase I and II drug metabolizing enzymes, drug transporters, and a number of upstream transcription regulators (Gaedigk et al., 2020; Rykova et al., 2022). Metformin is not metabolized in the human body (Gong et al., 2012); correspondingly, most studies aimed to identify the SNPs associated with the response to metformin were focused on the genes coding for organic cation transporters (OCTs) and the multidrug and toxin extrusion (MATE) proteins. This allowed for detection of a number of SNPs in the genes of these groups (Zhou et al., 2015; Xiao et al., 2016). The use of unbiased genome-wide approach (GWAS) succeeded in finding several SNPs associated with the response to metformin in genes, the products of which are involved in other functions. These SNPs have been discovered in regulatory regions of the *ATM* gene, the product of which is involved in the maintenance of redox homeostasis in the cell (GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group et al., 2011); the *PRPF31* gene, coding for pre-mRNA processing factor 31; the *CPA6* (carboxypeptidase A6) gene, associated with the regulation of the Akt/mTOR signaling pathway; and the *STAT3* gene, coding for eponymous transcription factor, which is a well-known regulator of metabolic and immune processes (Rotroff et al., 2018).

Nevertheless, GWAS application in pharmacogenomics yet failed to provide a considerable progress, mainly because of the difficulties (and more often, impossibility) in forming large cohorts (tens and hundreds of thousands) of patients to determine the effects of each drug. Modern methods of functional genomics – eQTL analysis and detection of allele-specific events in omics data – do not suffer from this limitation (Degtyareva et al., 2021). However, the studies utilizing these approaches are few and have even absent at all for antidiabetic drugs, including metformin. Earlier, we measured the allelic imbalance at the level of gene expression and the profiles of active chromatin marks in the paired RNA-seq and ChIP-seq data for peripheral blood mononuclear cells (PBMCs) of nine healthy donors and constructed a panel of 14 796 regulatory SNPs (rSNPs) potentially able to influence the expression levels of the genes harboring these SNPs in their promoters (Damarov et al., 2024). Then, using the RNA-seq data (GSE153315, GEO DataSets) on the blood cells of T2DM non-responders to metformin ($N = 10$) and T2DM responders ($N = 10$) (Vohra et al., 2022), we found 367 rSNPs from our panel in the promoters of 131 corresponding differentially expressed genes (DEGs). Primary analysis of this gene set revealed a number of transcriptional regulators both known

to be involved in T2DM pathogenesis and still poorly studied (Damarov et al., 2024). Here, we performed further analysis of the obtained rSNP list using independent transcriptomic data on metformin hepatic response (Luizon et al., 2016) and available data on metformin impacts on liver gluconeogenesis.

Materials and methods

Gene sets analyzed. Three original gene sets were compared, and first was the set of 5132 genes harboring the rSNPs in promotor regions based on our previous analysis of allele-asymmetric events in omics data (RNAseq and ChIP-seq) for human PBMCs of nine healthy donors (Damarov et al., 2024). Each rSNP in the reported panel was located within ± 1000 bp from known TSS (transcription start site) and associated with allele-asymmetric binding of activating histone modifications (H3K4me3 and/or H3K27ac, ChIP-seq). Moreover, when analyzing RNA-seq data obtained for the same PBMCs samples, allele-specific expression events were observed within the transcribed sequences of the corresponding genes. Based on the association of the identified regulatory variants with two types of allele-specific events, we assumed that our rSNPs were most likely affecting the expression of certain human genes.

The second gene set comprised 1196 AMPK-dependent genes that were identified through clustering analysis on all DEGs (padj < 0.05) identified from RNA-seq data across the three conditions (treated with metformin, metformin + an AMPK inhibitor (compound C) or no treatment for 8 hours in triplicate) of cultured primary human hepatocytes from a sole male donor (Luizon et al., 2016). The DEGs were termed AMPK-dependent when found differently expressed between hepatocytes with metformin treatment and controls; however, their expression was restored to the normal level (or similarly altered) in hepatocytes with the combined treatment of compound C and metformin. And the third gene set consisted of 115 genes that were classified as involved only in gluconeogenesis, only in regulation of gluconeogenesis, or shared both Gene ontology terms (Ashburner et al., 2000). The analyzed human datasets were under ethical consent agreements as stated in authorized submissions.

Functional annotation. To predict the biological processes and metabolic pathways associated with the analyzed gene lists, a search for associated terms from GO and KEGG gene ontologies and metabolic pathways (Kanehisa, 2000) was performed using the clusterProfiler package (v4.15.1) (Yu et al., 2012) when taking into account the Benjamini–Hochberg correction (p.adj < 0.1).

A functional protein association network (PPI) was built and analyzed by STRING v.11 tools (Szklarczyk et al., 2023). The k-means clustering was applied in STRING on the basis of evidence score and connection cutoff at 0.40; the enrichments of the KEGG and Wiki pathways were analyzed with a *p*-value cutoff of < 0.05 and minimum count in the network setting of 2.

Results and discussion

Matching well the generally accepted model for metformin action, our panel of rSNPs located within the promotors of DEGs associated with metformin response (Damarov et al., 2024) included rSNPs within promotors of genes for key enzymes of mitochondrial electron transfer chain (NDUFA11 and NDUFB1) and upstream/downstream AMPK regulators (Fig. 1). So, AK5 adenylate kinase increases the phosphorylation of AMPK (Dzeja, Terzic, 2009); and N-myristoyltransferase (NMT1) controls the N-myristoylation and lysosomal localization of the AMPK protein (Chen et al., 2020). Activated AMPK inhibits the mTOR (mammalian target of rapamycin) pathway, a pivotal regulator of cell metabolism and growth (Gwinn, Shaw, 2010; Smiles et al., 2024).

To gain a better understanding of the possible involvement of our rSNPs in the mechanisms underlying the antihyperglycemyc effects of metformin treatment, we compared and analyzed three gene sets (Fig. 2). The ‘AMPKdep’ set of 1196 AMPK-dependent genes associated with metformin hepatic response was obtained from (Luizon et al., 2016), the ‘rSNPs’ set of 5132 genes with the regulatory variants originally identified within promoter regions was previously reported in (Damarov et al., 2024). Then, the ‘Glu’ set consisted of genes related to gluconeogenesis in GO terms (see the Materials and Methods section for details).

Combining the results of our previous analysis with the AMPKdep gene set allowed to identify 402 genes, the expression of which in hepatocytes may be influenced by our rSNPs. The results of functional enrichment analysis using R signified that Protein processing in endoplasmic reticulum, Apoptosis and FoxO signaling KEGG pathways were enriched in a given set (Fig. 3*A*). Using the enrichment analysis in STRING, we observed the enrichment of the same gene list belonging to several Wiki metabolic pathways: similarly, apoptosis and, in addition, adipogenesis and signaling pathways including TNF-alpha-, TGF-beta- and angiogenesis-related VEGFA\VEGFR2-signaling (Fig. 3*B*).

It is worth to mention that the signaling pathways above were widely reported to be affected by metformin (Yi et al., 2016; Kristófi, Eriksson, 2021; Pan et al., 2024). For example, some mice data indicate the Foxo1 protein to be one of the key targets of metformin in regulating blood glucose and hepatic glucose production under therapeutic concentrations (Guo et al., 2021). Moreover, the same pathways are commonly known as mediators that orchestrate and propagate inflammatory responses and thereby play a crucial role in pathological signaling mechanisms shared among common TD2M complications, including diabetic retinopathy (Behl et al., 2022) and diabetic kidney disease (Ansari et al., 2025; Hou et al., 2025).

When the genes annotated to the ‘Glu’ set were put together with our rSNP panel, we discovered 28 genes in the overlap. The encoded products include important enzymes, i. e. MDH2, malate dehydrogenase 2, which catalyzes a reversible NAD-dependent dehydrogenase reaction and glycolysis enzymes (i. e. ENO1, enolase 1, and PGP, phosphoglycolate phosphatase), as well as a number of regulatory proteins (i. e. FoxO1, USP7, SIRT6). Thus, FoxO1 has a critical role in the insulin-mediated regulation of de novo glucose synthesis affecting the transcription of rate-limiting enzymes, G6Pc and Pck1 (Hall et al., 2014). Notably, multiple studies support that sirtuin 6, SIRT6, an NAD-dependent deacetylase/deacylase/mono-ADP ribosyltransferase, is involved in inhibition of FoxO1 leading

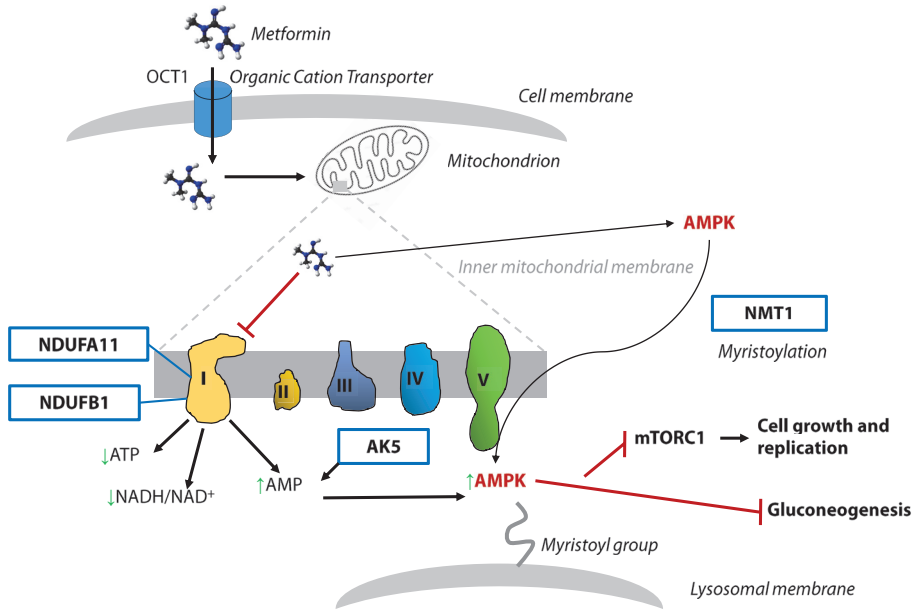


Fig. 1. Antihyperglycemic effects of metformin are implemented through inhibition of complex I of mitochondrial electron transport chain and activation of AMP kinase. The proteins associated with the AMPK pathway are denoted with blue frames. Green vertical arrows indicate a decrease or an increase in the amount of metabolite or enzyme activity; blunt red arrows indicate inhibition; black arrows indicate membrane transport or a functional connectivity; the complexes of the electron transport chain embedded in the inner mitochondrial membrane are labeled with roman numerals (I through V); NDUFA11, NADH:ubiquinone oxidoreductase subunit A11; NDUFB1, NADH:ubiquinone oxidoreductase subunit B1; AK5, adenylate kinase 5; AMPK, AMP-activated protein kinase; NMT1, N-myristoyltransferase 1; and OCT1, organic cation/carnitine transporter 1.

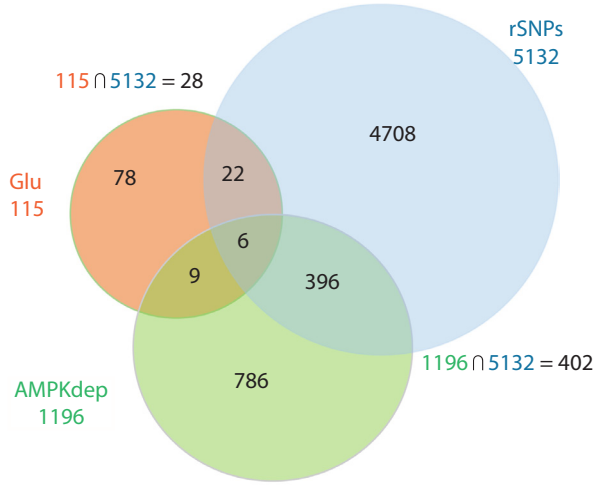


Fig. 2. Venn diagram representing the intersection between the 'Glu' (orange), 'rSNP' (blue) and 'AMPKdep' (green) gene sets. ∩ indicates overlapping; the numerical values in the intersections show the counts for corresponding gene groups.

to suppression of gluconeogenesis (Liu et al., 2021; Wang et al., 2023). Moreover, it was shown that the activation of deubiquitinating enzyme USP7 similarly results in a decrease in FoxO1 activity (Jiang et al., 2017).

Using STRING, we constructed and analyzed a PPI network for this subset of 28 genes (Fig. 4). Three clusters were identified by the KMeans algorithm: nine protein products of cluster 1 were found to be involved in NADH metabolism and gluconeogenesis; seven from cluster 2, in regulation of gluconeogenesis, circadian rhythms and AMP-activated protein kinase activity (PRKAG2 and PRKACA); and cluster 3 was comprised of two subunits of protein phosphatase 4 complex: PPP4R3A and PPP4R3B (Fig. 4*A*). It is also to be noted that the protein products of four genes were functionally related to KEGG AMPK-signaling pathway (Fig. 4*B*). The results of functional enrichment analysis based on Gene Ontology are visualized in Figure 4*C*.

For a number of these genes, the functional links with regulation of gluconeogenesis have been well described as given on the schematic representation in Figure 5. From this background we hypothesized that our rSNPs located within the corresponding promotors may be involved in regulatory signatures.

Adding the data on AMPK-dependent metformin hepatic response allows us to highlight six genes (*ARPP19*, *ATF4*, *NR3C1*, *PFKFB3*, *TCF7L2*, and *WDR5*) from a list of 28. These genes belong to the intersection of all three compared gene sets, which means the corresponding promotors harbor rSNPs; the genes were shown to be expressed in AMPK-dependent manner in hepatocytes; and the resulting proteins were involved in gluconeogenesis.

Thus, NR3C1, better known as the glucocorticoid receptor (GR), has been functionally linked to stress response, immune regulation, inflammation and control of hepatic gluconeogenesis (Zhang et al., 2019). The encoded protein functions as a transcription factor enhancing the transcription of glu-

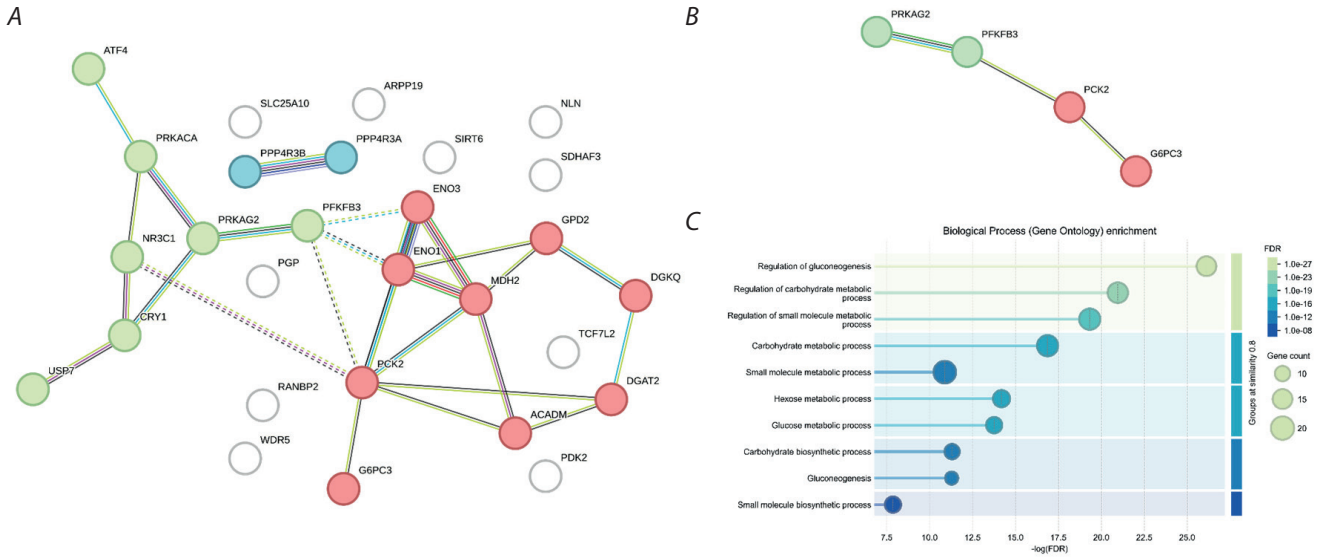


Fig. 3. Interpreting a list of 402 genes with pathway enrichment analysis. (*A*) The most relevant enriched KEGG terms obtained using R (Materials and Methods). (*B*) Wiki pathway enrichment determined and visualized by STRING. ID, gene or pathway identifier from the database; FE, Fold Enrichment; p.Adjust, a *p*-value as corrected with the Benjamini–Hochberg procedure; count, the number of genes presented in the list; and ER, endoplasmic reticulum (*A*). Different colors indicate top enriched terms with corresponding *p*-values (FDR) according to the color panel (*B*).

A

KEGG ID	Description	FE	*p*-value	p.Adjust	Count	geneID_symbols
hsa04141	Protein processing in ER	4.13	7.045*e-6	0.0012	14	*ATF4/NFE2L2/UBXN4/SEC62/SEC24B/HSPA1B/HSPA1A/PDIA4/RAD23B/UBQLN1/HYOU1/AMFR/PPP1R15A/DNAJC5*
hsa04068	FoxO signaling pathway	3.82	2.86*e-4	0.019	10	*BCL2L11/PLK2/SGK1/SOD2/CDKN1A/FBXO32/MDM2/GABARAPL1/IRS2/CREBBP*
hsa04210	Apoptosis	3.71	3.63*e-4	0.019	10	*ATF4/MCL1/CTSS/CFLAR/BCL2L11/DAXX/FAS/FADD/TUBA1B/NFKBIA*

B

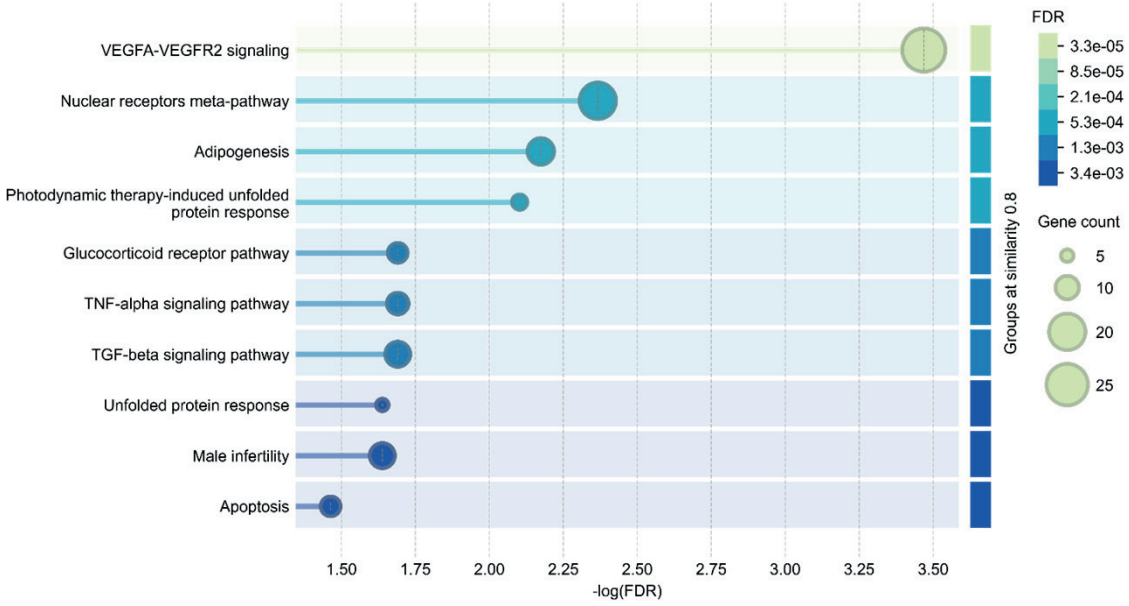


Fig. 4. PPI network for the intersection of 28 genes as by STRING. (*A*) The clusters identified with K-means cluster analysis. (*B*) Four closely connected PPI members belonging to KEGG AMPK-signaling pathway (hsa04152). (*C*) A visualization of the most relevant enriched Biological Process Gene Ontology terms. Nodes of different color indicate the members of different clusters (*A* and *B*): red – cluster 1, green – cluster 2, and blue – cluster 3. The solid and the dotted lines (*A*) indicate connections within the same and with a different cluster respectively. Edges of different color indicate both functional and physical protein interactions (*A* and *B*): cyan – from curated databases; pink – experimentally determined; blue – gene co-occurrence; khaki – from text mining; black – coexpression; light blue – protein homology. Different enriched terms with corresponding *p*-values (FDR) are highlighted in different colors according to the color panel (*C*).

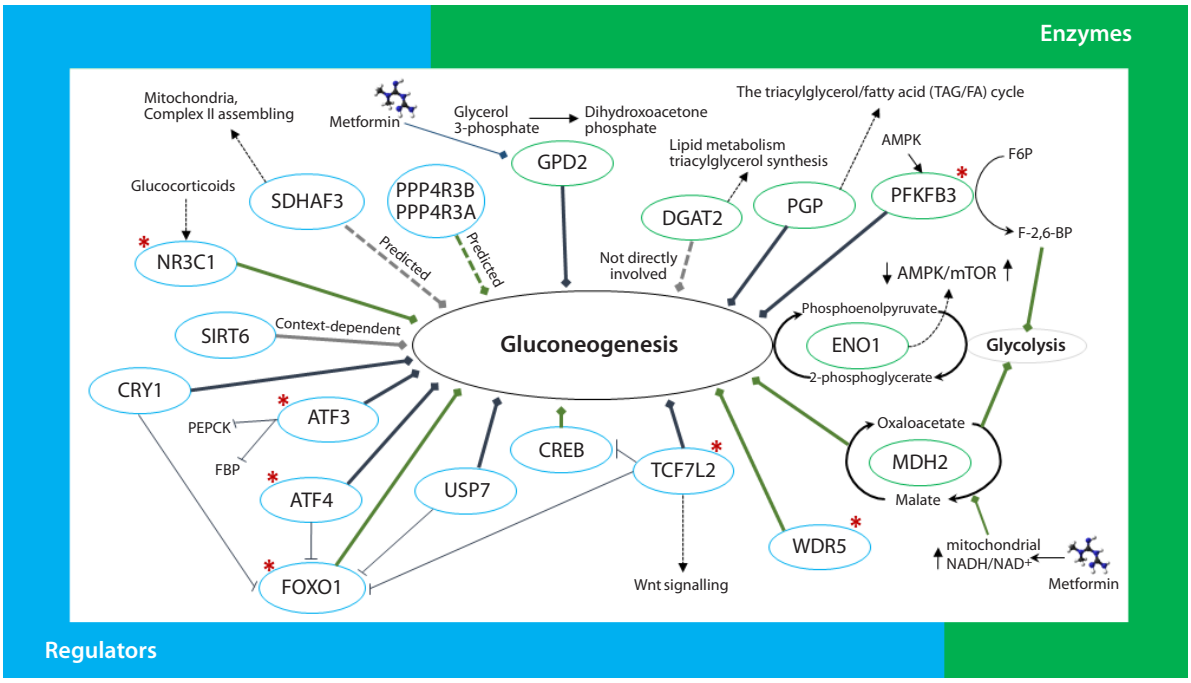


Fig. 5. Multiple potential rSNP targets, including enzymes and regulatory proteins, known to be involved in regulation of gluconeogenesis. Blunt orange arrows indicate the activation of gluconeogenesis and deep-blue ones stand for the suppression of the process; T-shaped blue arrows indicate the suppression of transcription/translation or protein activity; red stars indicate genes from the AMPKdep gene set; AMPK, 5′ adenosine monophosphate-activated protein kinase; mTOR, mammalian target of rapamycin; Wnt, Wingless-related integration site; PEPCK, phosphoenolpyruvate carboxykinase; FBP, fructose-1,6-bisphosphatase; F6P, fructose 6-phosphate; F-2,6-BP, fructose-2,6-bisphosphate; NADH, nicotinamide adenine dinucleotide; and NAD+, the oxidized form of NAD.

coneogenic genes such as PEPCK and G6Pase under fasting conditions (Beaupere et al., 2021).

ATF4, activating transcription factor 4, is a key protein in metabolic regulation involved in insulin secretion and sensitivity control and regulation of gluconeogenesis by affecting the transcriptional activity of FoxO1 (Li et al., 2022). Interestingly, a number of findings demonstrate epigenetic regulatory mechanisms targeting ATF4. For example, it was reported that microRNA (miR)-214 upregulates ATF4 expression, leading to elevated expression of FoxO1 and thus to suppression of gluconeogenesis (Zhu et al., 2018; Wang et al., 2024).

WRD5 performs multiple scaffolding functions in the context of chromatin and was reported to modulate gluconeogenic gene expression by potentiating the activity of KAT2B histone lysine acetyltransferase. Although these two genes seem to regulate a relatively small subset of gene targets in hepatocytes, both KAT2B and WDR5 are required for the expression of a large fraction of glucagon-inducible genes, indicating that they represent the important cofactors for the CREB pathway (Ravnskjaer et al., 2013).

The TCF7L2 protein is a transcription factor that plays an important role in the Wnt signaling pathway and has been implicated in blood glucose homeostasis (Ip et al., 2012) through actions even beyond pancreatic beta cells (Bailey et al., 2015).

Finally, the protein encoded by *PFKFB3* is a bifunctional enzyme involved in both the synthesis and degradation of fructose-2,6-bisphosphate (F-2,6-BP), and is a potent activator of 6-phosphofructokinase-1 (PFK-1) – a trigger of aerobic oxidation for glucose metabolism. Recent studies have reported the pivotal role of PFKFB3 in the regulation of insulin resistance (Yang et al., 2023).

Conclusion

This work identifies several novel genes and gene regulatory variants that can be activated due to metformin treatment and thus provides candidates in the human genome where nucleotide variation can lead to differences in metformin response. We conclude that metformin therapy response may depend on the identified functional SNPs within the promotors of genes for gluconeogenic enzymes and regulatory proteins.

Список литературы / References

Ansari Z., Chaurasia A., Neha, Sharma N., Bachheti R.K., Gupta P.C. Exploring inflammatory and fibrotic mechanisms driving diabetic nephropathy progression. *Cytokine Growth Factor Rev.* 2025;84: 120-134. doi 10.1016/j.cytogfr.2025.05.007

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., … Matese J.C., Richardson J.E., Ringwald M., Rubin G.M., Sherlock G. Gene Ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25-29. doi 10.1038/75556

Bailey K.A., Savic D., Zielinski M., Park S.-Y., Wang L., Witkowski P., Brady M., Hara M., Bell G.I., Nobrega M.A. Evidence of non-pancreatic beta cell-dependent roles of Tcf7l2 in the regulation of glucose metabolism in mice. *Hum Mol Genet.* 2015;24(6):1646-1654. doi 10.1093/hmg/ddu577

Beaupere C., Liboz A., Fève B., Blondeau B., Guillemain G. Molecular mechanisms of glucocorticoid-induced insulin resistance. *Int J Mol Sci.* 2021;22(2):623. doi 10.3390/ijms22020623

Behl T., Wadhwa M., Sehgal A., Singh S., Sharma N., Bhatia S., Al-Harrasi A., Aleya L., Bungau S. Mechanistic insights into the role of FOXO in diabetic retinopathy. *Am J Transl Res.* 2022;14(6):3584-3602

Bu Y., Peng M., Tang X., Xu X., Wu Y., Chen A.F., Yang X. Protective effects of metformin in various cardiovascular diseases: Clinical evidence and AMPK-dependent mechanisms. *J Cell Mol Med.* 2022;26(19):4886-4903. doi 10.1111/jcmm.17519

Chen Y.-C., Navarrete M.S., Wang Y., McClintock N.C., Sakurai R., Wang F., Chen K.T., Chou T.-F., Rehan V.K., Lee D.J., Diaz B. N-myristoyltransferase-1 is necessary for lysosomal degradation and mTORC1 activation in cancer cells. *Sci Rep.* 2020;10(1):11952. doi 10.1038/s41598-020-68615-w

Cook M.N., Girman C.J., Stein P.P., Alexander C.M. Initial monotherapy with either metformin or sulphonylureas often fails to achieve or maintain current glycaemic goals in patients with Type 2 diabetes in UK primary care. *Diabet Med.* 2007;24(4):350-358. doi 10.1111/j.1464-5491.2007.02078.x

Damarov I.S., Korbolina E.E., Rykova E.Y., Merkulova T.I. Multi-omics analysis revealed the rSNPs potentially involved in T2DM pathogenic mechanism and metformin response. *Int J Mol Sci.* 2024;25(17):9297. doi 10.3390/ijms25179297

Degtyareva A.O., Antontseva E.V, Merkulova T.I. Regulatory SNPs: altered transcription factor binding sites implicated in complex traits and diseases. *Int J Mol Sci.* 2021;22(12):6454. doi 10.3390/ijms22126454

Dzeja P., Terzic A. Adenylate kinase and AMP signaling networks: metabolic monitoring, signal communication and body energy sensing. *Int J Mol Sci.* 2009;10(4):1729-1772. doi 10.3390/ijms10041729

El-Mir M.-Y., Nogueira V., Fontaine E., Avéret N., Rigoulet M., Leverve X. Dimethylbiguanide inhibits cell respiration via an indirect effect targeted on the respiratory chain complex I. *J Biol Chem.* 2000;275(1):223-228. doi 10.1074/jbc.275.1.223

Gaedigk A., Whirl-Carrillo M., Pratt V.M., Miller N.A., Klein T.E. PharmVar and the landscape of pharmacogenetic resources. *Clin Pharmacol Ther.* 2020;107(1):43-46. doi 10.1002/cpt.1654

GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group; Wellcome Trust Case Control Consortium 2; Zhou K., Bellenguez C., Spencer C.C., Bennett A.J., Coleman R.L., … McCarthy M.I., Holman R.R., Palmer C.N., Donnelly P., Pearson E.R. Common variants near *ATM* are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet*. 2011;43(2):117-120. doi 10.1038/ng.735

Gong L., Goswami S., Giacomini K.M., Altman R.B., Klein T.E. Metformin pathways. *Pharmacogenet Genomics*. 2012;22(11):820-827. doi 10.1097/FPC.0b013e3283559b22

Guo X., Li X., Yang W., Liao W., Shen J.Z., Ai W., Pan Q., Sun Y., Zhang K., Zhang R., Qiu Y., Dai Q., Zheng H., Guo S. Metformin targets Foxo1 to control glucose homeostasis. *Biomolecules*. 2021;11(6):873. doi 10.3390/biom11060873

Gwinn D.M., Shaw R.J. AMPK Control of mTOR signaling and growth. In: Tamanoi F., Hall M.N. (Eds). The Enzymes. Vol. 28. Academic Press, 2010;49-75. doi 10.1016/S1874-6047(10)28003-4

Hall J.A., Tabata M., Rodgers J.T., Puigserver P. USP7 attenuates hepatic gluconeogenesis through modulation of FoxO1 gene promoter occupancy. *Mol Endocrinol.* 2014;28(6):912-924. doi 10.1210/me.2013-1420

Hou G., Dong Y., Jiang Y., Zhao W., Zhou L., Cao S., Li W. Immune inflammation and metabolic interactions in the pathogenesis of diabetic nephropathy. *Front Endocrinol.* (*Lausanne*). 2025;16:1602594. doi 10.3389/fendo.2025.1602594

Hunter R.W., Hughey C.C., Lantier L., Sundelin E.I., Peggie M., Zeqiraj E., Sicheri F., Jessen N., Wasserman D.H., Sakamoto K. Metformin reduces liver glucose production by inhibition of fructose-1-6-bisphosphatase. *Nat Med.* 2018;24(9):1395-1406. doi 10.1038/s41591-018-0159-7

Ip W., Shao W., Chiang Y.A., Jin T. The Wnt signaling pathway effector TCF7L2 is upregulated by insulin and represses hepatic gluconeogenesis. *Am J Physiol Metab.* 2012;303(9):E1166-E1176. doi 10.1152/ajpendo.00249.2012

Jiang L., Xiong J., Zhan J., Yuan F., Tang M., Zhang C., Cao Z., … Wang Hui, Wang L., Wang J., Zhu W.-G., Wang Haiying. Ubiquitin-specific peptidase 7 (USP7)-mediated deubiquitination of the histone deacetylase SIRT7 regulates gluconeogenesis. *J Biol Chem.* 2017;292(32):13296-13311. doi 10.1074/jbc.M117.780130

Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):2730. doi 10.1093/nar/28.1.27

Karbalaee-Hasani A., Khadive T., Eskandari M., Shahidi S., Mosavi M., Nejadebrahimi Z., Khalkhali L., Sangdari A., Mohammadi D., Soltani A., Khodabandehloo H., Hosseini H., Koushki M. Effect of metformin on circulating levels of inflammatory markers in patients with type 2 diabetes: A systematic review and meta-analysis of randomized controlled trials. *Ann Pharmacother.* 2021;55(9):1096-1109. doi 10.1177/1060028020985303

Kristófi R., Eriksson J.W. Metformin as an anti-inflammatory agent: a short review. *J Endocrinol.* 2021;251(2):R11-R22. doi 10.1530/JOE-21-0194

LaMoia T.E., Shulman G.I. Cellular and molecular mechanisms of metformin action. *Endocr Rev.* 2021;42(1):77-96. doi 10.1210/endrev/bnaa023

Li J., Yan H., Xiang R., Yang W., Ye J., Yin R., Yang J., Chi Y. ATP secretion and metabolism in regulating pancreatic beta cell functions and hepatic glycolipid metabolism. *Front Physiol.* 2022;13:918042. doi 10.3389/fphys.2022.918042

Liu G., Chen H., Liu H., Zhang W., Zhou J. Emerging roles of SIRT6 in human diseases and its modulators. *Med Res Rev.* 2021;41(2):1089-1137. doi 10.1002/med.21753

Loan A., Syal C., Lui M., He L., Wang J. Promising use of metformin in treating neurological disorders: biomarker-guided therapies. *Neural Regen Res.* 2024;19(5):1045-1055. doi 10.4103/1673-5374.385286

Luizon M.R., Eckalbar W.L., Wang Y., Jones S.L., Smith R.P., Laurance M., Lin L., … Molony C., Innocenti F., Yee S.W., Giacomini K.M., Ahituv N. Genomic characterization of metformin hepatic response. *PLoS Genet.* 2016;12(11):e1006449. doi 10.1371/journal.pgen.1006449

Pan Q., Ai W., Guo S. TGF-β1 signaling impairs metformin action on glycemic control. *Int J Mol Sci.* 2024;25(4):2424. doi 10.3390/ijms25042424

Ravnskjaer K., Hogan M.F., Lackey D., Tora L., Dent S.Y.R., Olefsky J., Montminy M. Glucagon regulates gluconeogenesis through KAT2B- and WDR5-mediated epigenetic effects. *J Clin Invest.* 2013;123(10):4318-4328. doi 10.1172/JCI69035

Rena G., Hardie D.G., Pearson E.R. The mechanisms of action of metformin. *Diabetologia*. 2017;60(9):1577-1585. doi 10.1007/s00125-017-4342-z

Rotroff D.M., Yee S.W., Zhou K., Marvel S.W., Shah H.S., Jack J.R., Havener T.M., … Giacomini K.M., Pearson E.R., Wagner M.J., Buse J.B., Motsinger-Reif A.A. Genetic variants in *CPA6* and *PRPF31* are associated with variation in response to metformin in individuals with type 2 diabetes. *Diabetes*. 2018;67(7):1428-1440. doi 10.2337/db17-1164

Rykova E., Ershov N., Damarov I., Merkulova T. SNPs in 3′UTR miRNA target sequences associated with individual drug susceptibility. *Int J Mol Sci.* 2022;23(22):13725. doi 10.3390/ijms232213725

Smiles W.J., Ovens A.J., Kemp B.E., Galic S., Petersen J., Oakhill J.S. New developments in AMPK and mTORC1 cross-talk. *Essays Biochem.* 2024;68(3):321-336. doi 10.1042/EBC20240007

St-Amour S., Tessier L., Harnois J., Allard C., Lavoie A., Caron P., Bouchard L., Perron P., Tremblay K. PCK1 and SLC22A2 gene variants associated with response to metformin treatment in type 2 diabetes. *PLoS One*. 2025;20(2):e0305511. doi 10.1371/journal.pone.0305511

Strang J., Astridge D., Nguyen V., Reigan P. Small molecule modulators of AMP-Activated Protein Kinase (AMPK) activity and their potential in cancer therapy. *J Med Chem.* 2025;68(3):2238-2254. doi 10.1021/acs.jmedchem.4c02354

Szklarczyk D., Kirsch R., Koutrouli M., Nastou K., Mehryary F., Hachilif R., Gable A.L., Fang T., Doncheva N.T., Pyysalo S., Bork P., Jensen L.J., von Mering C. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 2023;51(D1): D638-D646. doi 10.1093/nar/gkac1000

Vohra M., Sharma A.R., Mallya S., Prabhu N.B., Jayaram P., Nagri S.K., Umakanth S., Rai P.S. Implications of genetic variations, differential gene expression, and allele-specific expression on metformin response in drug-naïve type 2 diabetes. *J Endocrinol Invest.* 2022;46(6):1205-1218. doi 10.1007/s40618-022-01989-y

Wang Y., Liu T., Cai Y., Liu W., Guo J. SIRT6's function in controlling the metabolism of lipids and glucose in diabetic nephropathy. *Front Endocrinol* (*Lausanne*). 2023;14:1244705. doi 10.3389/fendo.2023.1244705

Wang Y., Wang X., Du C., Wang Z., Wang J., Zhou N., Wang B., Tan K., Fan Y., Cao P. Glycolysis and beyond in glucose metabolism: exploring pulmonary fibrosis at the metabolic crossroads. *Front Endocrinol* (*Lausanne*). 2024;15:1379521. doi 10.3389/fendo.2024.1379521

Xiao D., Guo Y., Li X., Yin J.-Y., Zheng W., Qiu X.-W., Xiao L., Liu R.-R., Wang S.-Y., Gong W.-J., Zhou H.-H., Liu Z.-Q. The impacts of *SLC22A1* rs594709 and *SLC47A1* rs2289669 polymorphisms on metformin therapeutic efficacy in chinese type 2 diabetes patients. *Int J Endocrinol.* 2016;2016:4350712. doi 10.1155/2016/4350712

Yang Q., Huo E., Cai Y., Zhang Z., Dong C., Asara J.M., Wei Q. PFKFB3-mediated glycolysis boosts fibroblast activation and subsequent kidney fibrosis. *Cells*. 2023;12(16):2081. doi 10.3390/cells12162081

Yi Q.-Y., Deng G., Chen N., Bai Z.-S., Yuan J.-S., Wu G.-H., Wang Y.-W., Wu S.-J. Metformin inhibits development of diabetic retinopathy through inducing alternative splicing of VEGF-A. *Am J Transl Res.* 2016;8(9):3947-3954

Yu G., Wang L.-G., Han Y., He Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi A J Integr Biol.* 2012;16(5):284-287. doi 10.1089/omi.2011.0118

Zhang X., Yang S., Chen J., Su Z. Unraveling the regulation of hepatic gluconeogenesis. *Front Endocrinol* (*Lausanne*). 2019;9:802. doi 10.3389/fendo.2018.00802

Zhou Y., Ye W., Wang Y., Jiang Z., Meng X., Xiao Q., Zhao Q., Yan J. Genetic variants of OCT1 influence glycemic response to metformin in Han Chinese patients with type-2 diabetes mellitus in Shanghai. *Int J Clin Exp Pathol.* 2015;8(8):9533-9542

Zhu X., Li H., Wu Y., Zhou J., Yang G., Wang W. lncRNA MEG3 promotes hepatic insulin resistance by serving as a competing endogenous RNA of miR-214 to regulate ATF4 expression. *Int J Mol Med.* 2018;43(1):345-357. doi 10.3892/ijmm.2018.3975

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

Received October 17, 2025. Revised November 21, 2025. Accepted December 1, 2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):267-273

doi 10.18699/vjgb-26-30

Особенности коннективности дефолтной сети мозга в зависимости от полиморфизма гена транспортера серотонина (5-HTTLPR)

А.В. Бочаров, А.Н. Савостьянов, С.С. Таможников, А.Е. Сапрыгин, Д.А. Лебедкин, Е.А. Меркулова, Г.Г. Князев

Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия
bocharovav@neuronm.ru

Аннотация. Полиморфизм гена транспортера серотонина имеет большое значение в регуляции серотонинергической системы, влияющей на настроение и регуляцию эмоций и поведения. В исследовании была проведена запись 128-канальной электроэнцефалограммы и взяты образцы буккального эпителия у 53 добровольцев (32 женщины). Аллели *La*, *Lg* и *S* были определены методом полимеразной цепной реакции. Целью работы было изучение особенностей коннективности дефолтной сети мозга, измеренной с помощью электрофизиологических данных состояния покоя, в зависимости от полиморфизма гена транспортера серотонина. Локализация источников биоэлектрической активности коры мозга осуществлялась методом формирователя пучка. Сравнение групп носителей *LaLa* генотипа и носителей одного из аллелей *S* или *Lg* было выполнено с помощью Т-контраста показателей коннективности, рассчитанных между узлами дефолтной сети мозга и остальным мозгом. Обнаружено, что для носителей *S* или *Lg* аллеля характерна повышенная коннективность дефолтной сети мозга со зрительной ассоциативной корой и со структурами, образующими задний узел дефолтной сети мозга, а также повышенная коннективность заднего узла дефолтной сети мозга с правой парагиппокампальной извилиной, что может предрасполагать к возникновению и/или поддержанию навязчивых мыслей. У носителей *LaLa* генотипа по сравнению с носителями *S* или *Lg* аллеля была выше коннективность переднего узла дефолтной сети мозга с правой вентромедиальной лобной корой, с медиальной лобной извилиной и с задней поясной корой, являющейся структурой заднего узла дефолтной сети мозга. Также у носителей *LaLa* генотипа по сравнению с носителями *S* или *Lg* аллеля была выше коннективность заднего узла дефолтной сети мозга с кластером, охватывающим правую дорсолатеральную префронтальную кору. Можно предположить, что увеличение коннективности дефолтной сети мозга со структурами мозга (дорсолатеральная и вентромедиальная префронтальная кора), участвующими в процессах когнитивной регуляции, способствует регуляции процессов дефолтной сети мозга, связанных с автобиографическими воспоминаниями.

Ключевые слова: ЭЭГ; 5-HTTLPR; полиморфизм гена транспортера серотонина; дефолтная сеть мозга; дефолт-система мозга; коннективность; сети покоя

Для цитирования: Бочаров А.В., Савостьянов А.Н., Таможников С.С., Сапрыгин А.Е., Лебедкин Д.А., Меркулова Е.А., Князев Г.Г. Особенности коннективности дефолтной сети мозга в зависимости от полиморфизма гена транспортера серотонина (5-HTTLPR). *Вавиловский журнал генетики и селекции*. 2026;30(2):267-273. doi 10.18699/vjgb-26-30

Финансирование. Работа выполнена за счет средств федерального бюджета на проведение фундаментальных научных исследований (тема № 126020316369-9, 2026–2028).

Features of connectivity of default mode network depending on polymorphism of serotonin transporter gene (5-HTTLPR)

A.V. Bocharov, A.N. Savostyanov, S.S. Tamozhnikov, A.E. Saprygin, D.A. Lebedkin, E.A. Merkulova, G.G. Knyazev

Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia
bocharovav@neuronm.ru

Abstract. Serotonin transporter gene polymorphism is important in the regulation of the serotoninergic system that affects mood and the regulation of emotions and behavior. In this study, 128 channel electroencephalogram recordings were performed, and buccal epithelium samples were obtained from 53 volunteers (32 females). *La*, *Lg*, and *S* alleles were identified by polymerase chain reaction. The aim of the study was to investigate the connectivity of the default mode network, measured using resting state electrophysiologic data, depending on the serotonin transporter gene polymorphism. Localization of the sources of bioelectrical activity of the cerebral cortex was performed by the beamformer method. Comparisons of *LaLa* genotype carriers and *S* or *Lg* allele carriers were performed using T-contrast of connectivity indices calculated between the nodes of the default mode network and the rest of the brain. It was found that carriers of the *S* allele were characterized by increased connectivity of the default mode network with the visual association cortex and

© Бочаров А.В., Савостьянов А.Н., Таможников С.С., Сапрыгин А.Е., Лебедкин Д.А., Меркулова Е.А., Князев Г.Г., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

with structures forming the posterior node of the default mode network, as well as increased connectivity of the posterior node of the default mode network with the right parahippocampal gyrus, and this pattern of connectivity may predispose to the onset and/or maintenance of intrusive thoughts. Whereas carriers of the *LaLa* genotype had higher connectivity of the anterior node of the default mode network with the right ventromedial prefrontal cortex, with the medial frontal gyrus, and with the posterior cingulate cortex, which is the structure of the posterior node of the default mode network, compared to carriers of the *S* or *Lg* allele. Also, carriers of the *LaLa* genotype had higher connectivity of the posterior node of the default mode network with the cluster involving the right dorsolateral prefrontal cortex compared to carriers of the *S* or *Lg* allele. It could be hypothesized that increased connectivity of the default mode network with brain structures (i.e., dorsolateral and ventromedial prefrontal cortex) involved in cognitive regulation processes may contribute to the regulation of the processes of the default mode network associated with autobiographical memory.

Keywords: EEG; 5-HTTLPR; serotonin transporter polymorphism; default mode network; resting-state networks; connectivity

For citation: Bocharov A.V., Savostyanov A.N., Tamozhnikov S.S., Saprygin A.E., Lebedkin D.A., Merkulova E.A., Knyazev G.G. Features of connectivity of default mode network depending on polymorphism of serotonin transporter gene (5-HTTLPR). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):267-273. doi 10.18699/vjgb-26-30

Введение

Многочисленные исследования, проведенные с использованием функциональной магнитно-резонансной томографии (фМРТ), в состоянии покоя обнаружили характерный паттерн активации структур головного мозга, образующих сеть. Эта сеть получила название «дефолтная сеть мозга» (ДСМ). Среди структур, входящих в ее состав, – медиальная префронтальная кора, задняя поясная кора, предклинье, медиальная, латеральная и нижняя теменные доли (Raichle et al., 2001). Показано, что при выполнении большинства внешне ориентированных задач, требующих концентрации внимания на внешних стимулах, ДСМ демонстрирует устойчивое снижение активности/коннективности. Однако в состоянии покоя, когда внимание не сосредоточено на внешнем мире, наблюдается увеличение активности/коннективности в структурах ДСМ (Raichle, 2015). Дальнейшие исследования значительно расширили понимание функций ДСМ. Так, были выявлены паттерны увеличения активности/коннективности ДСМ во время самореферентной обработки информации и извлечения автобиографических воспоминаний, при размышлениях о себе и об отношениях с близкими людьми, при планировании будущего и при переживании внутренних эмоциональных состояний (Raichle et al., 2001; Knyazev, 2013; Raichle, 2015; Buckner, DiNicola, 2019; Hutchinson-Wong et al., 2024; Si et al., 2025). Кроме того, обнаружено, что нарушения в работе ДСМ связаны с рядом психических расстройств (Hutchinson-Wong et al., 2024). Так, у людей с выраженностью симптомов депрессии и у пациентов с большим депрессивным расстройством (БДР) наблюдается увеличение в активности/коннективности ДСМ, что может быть связано с особенностями процессов саморефлексии и эмоциональной регуляции (Berman et al., 2011; Hamilton et al., 2011, 2013; Knyazev et al., 2016).

Внимание ученых привлекает изучение полиморфизма гена транспортера серотонина, который имеет большое значение в регуляции серотонинергической системы, влияющей на настроение, эмоциональную регуляцию и поведение (Cools et al., 2008). Транспортер серотонина (5-HTT) выполняет важную функцию – обратный захват серотонина, тем самым контролируя уровень серотонина в синаптической щели и определяя интенсивность и продолжительность его действия. Полиморфизм гена 5-HTT представляет собой вариацию в промоторной области гена и имеет два основных аллеля: длинный (*L*) и короткий (*S*). Различие в длине промоторной области влияет на эффективность транскрипции гена, т.е. на количество синтезируемого белка-транспортера. Аллель *S* ассоциируется со сниженной экспрессией гена 5-HTT, что приводит к меньшему количеству транспортера серотонина в синапсе (Lesch et al., 1996). Длинный *L* аллель содержит a/g единичный нуклеотидный полиморфизм. Согласно (Hu et al., 2006), аллель *Lg* функционально схож с аллелем *S*. Генотип *La/La* классифицируется как *LL* (высокий уровень транскрипционной эффективности), генотипы *La/S* и *La/Lg* – как *LS* (промежуточный уровень транскрипционной эффективности), а генотипы *Lg/S*, *Lg/Lg* и *S/S* – как *SS* (низкий уровень транскрипционной эффективности) (Hu et al., 2006).

В исследованиях была установлена связь между полиморфизмом гена 5-HTT и предрасположенностью к различным психическим расстройствам, включая БДР, посттравматическое стрессовое расстройство (ПТСР) и суицидальное поведение. Обнаружено, что люди, имеющие хотя бы один *S* аллель, характеризовались повышенной чувствительностью к стрессовым ситуациям и негативным эмоциям и при наличии неблагоприятных факторов у них в большей мере были выражены симптомы БДР (Caspi et al., 2003) и ПТСР (Koenen et al., 2009). Следует отметить, что в других работах не были подтверждены связи между психическими расстройствами и полиморфизмом гена 5-HTT (Risch et al., 2009). Показано также, что носители *S* аллеля более склонны к проявлению импульсивного поведения (Walderhaug et al., 2010), имеют более высокую выраженность личностной черты «нейротизм», обратным полюсом которой является эмоциональная стабильность (Greenberg et al., 2000), а также обладают сниженной способностью к когнитивному контролю в задании стоп-сигнал (Landro et al., 2015).

фМРТ эксперименты установили, что носители *S* аллеля в ответ на негативные стимулы демонстрировали повышенную активацию миндалевидного тела – структуры мозга, ответственной за детекцию угроз и переживание эмоций страха и тревоги (Adolphs, 2008). Совместно с этим у носителей *S* аллеля наблюдалось снижение активности в вентромедиальной префронтальной коре – структуре мозга, отвечающей за процессы регуляции эмоций (Rao et

al., 2007). В другом исследовании, в котором участвовали 23 женщины, объем латеральной префронтальной коры – структуры мозга, которая принимает участие в регуляции эмоций и когнитивном контроле, был отрицательно связан с вниманием к эмоциональным стимулам (как негативным, так и позитивным) только у носителей *S* аллеля, но не обнаружено связей полиморфизма с объемом миндалевидного тела и медиальной префронтальной коры (Beevers et al., 2010). В исследовании детей и подростков с использованием метода функциональной коннективности была установлена более низкая коннективность заднего узла ДСМ с верхней медиальной префронтальной корой у носителей *SS* генотипа по сравнению с носителями *L* аллеля (Wiggins et al., 2012). A. Meyer-Lindenberg (2009) обнаружил, что показатели функциональной коннективности в мозге являются более надежными предикторами влияния полиморфизма гена 5-HTT, чем просто показатели активности отдельных областей мозга. J. Wiggins с коллегами (2012) изучали с помощью фМРТ выраженность коннективности между передним и задним узлами ДСМ у детей и подростков и ее связь с возрастом в группах в зависимости от полиморфизма гена транспортера серотонина.

В нашей работе мы не ограничимся изучением коннективности между узлами ДСМ, а проанализируем связи ДСМ в пределах коры головного мозга в состоянии покоя в зависимости от полиморфизма гена транспортера серотонина у взрослых участников исследования. Целью исследования стало изучение особенностей коннективности ДСМ, измеренной на основе ЭЭГ данных в состоянии покоя, в зависимости от полиморфизма гена транспортера серотонина.

Материалы и методы

Участниками исследования были 53 добровольца (32 женщины и 21 мужчина) – правши с нормальным или скорректированным до нормы зрением (средний возраст 28 лет, стандартное отклонение = 10). Предварительно все участники заполнили анкету, в которой ответили на вопросы о состоянии своего психического и физического здоровья, о самочувствии и употреблении психоактивных веществ накануне исследования. Женщины дополнительно отвечали на вопросы о текущей фазе менструального цикла и приеме гормональных контрацептивов. Критерием исключения было наличие психических заболеваний, травм головы, употребление наркотических, психоактивных и гормональных веществ. Участники были проинформированы о методах исследования, после чего подписали информированное согласие на участие в эксперименте. Исследование было организовано в соответствии с этическими нормами, установленными Хельсинкской декларацией, и одобрено локальным биоэтическим комитетом НИИ нейронаук и медицины (Новосибирск).

Генотипирование. ДНК была выделена из клеток буккального эпителия с использованием набора для выделения «Биосилика» (Россия). Аллели *La*, *Lg* и *S* полиморфизма 5-HTT определяли в образцах ДНК с помощью полимеразной цепной реакции с праймерами: 5′-gagggactg agctggacaacccac-3′ и 5′-gcgttgccgctctgaattgc-3′ (Lesch et al., 1996). Полученные продукты разделяли путем агарозно-гелевого электрофореза. Размеры аллелей *L* и *S* для гена 5-HTT составили 529 и 489 п. н. Для выявления аллелей *La* и *Lg* проводили гидролиз продуктов амплификации в течение трех часов эндонуклеазой MspI. После этого размеры продуктов составили 340, 127 и 62 п. н. для аллеля *La* и 174, 166, 127 и 62 п. н. для аллеля *Lg*. Выборка включала 17 носителей *LL*, 28 *LS* и 8 *SS*.

Запись ЭЭГ осуществляли с использованием многоканального усилителя биопотенциалов Brain Products (Германия), который был оснащен шлемом, содержащим 128 электродов, один из электродов был применен для записи окулограммы. Электроды были расположены в соответствии с международной системой 10–5. Полоса пропускания была в диапазоне от 0.1 до 100 Гц, частота дискретизации составила 1000 Гц. Электрод Cz был задействован в качестве референта.

Анализ данных. Артефакты удаляли с помощью анализа независимых компонент под зрительным контролем. Согласно данным предыдущих исследований осцилляторных сетей покоя (Knyazev et al., 2016; Bocharov et al., 2021, 2022), анализ ограничился дельта частотным диапазоном. После удаления артефактов ЭЭГ данные были отфильтрованы в дельта частотном диапазоне (1–4 Гц) с использованием фильтра Butterworth и функции filtfilt. В качестве модели головы была выбрана модель граничных элементов (Fuchs et al., 2001). Корковая сетка содержала 5124 вершины и была создана с помощью шаблона, основанного на модели мозга Монреальского неврологического института (MNI). Локализация корковых источников биоэлектрической активности производилась методом формирователя пучка (beamformer) (Van Veen et al., 1997) посредством набора инструментов программы SPM-12 (https://www.fil.ion.ucl.ac.uk/spm/). Ковариационные матрицы для анализа рассчитывались на данных ЭЭГ состоянии покоя продолжительностью 5 мин. Была проведена коррекция «утечки сигнала», которая может возникать из-за недостаточного пространственного разрешения метода локализации источников, для этого использовали метод ортогонализации (Brookes et al., 2011; Hipp et al., 2012).

После ортогонализации применяли преобразование Гильберта и выделяли огибающую сигнала (envelope), представляющую собой кривую амплитуд колебаний сигнала во времени. Затем рассчитывали карты коннективности между временным рядом активности выбранных областей интереса (ОИ) и временными рядами активности остальных вокселей мозга. Выбор ОИ был основан на данных предыдущих исследований. Для ДСМ были выбраны медиальная префронтальная кора (–1, 49, –2), задняя поясная кора (–5, –53, 41) и левая (–45, –71, 35) и правая (45, –71, 35) латеральные теменные доли (Gusnard, Raichle, 2001).

К коэффициентам корреляции Пирсона между временной активностью ОИ и остальным мозгом было применено преобразование Фишера. Полученные карты коннективности были пространственно сглажены (FWHM 8 мм) и конвертированы в формат NIFTI. Статистический анализ

второго уровня проводили с использованием Т-контраста, что позволило выявлять различия между группами (носители *S* аллеля *vs* носители генотипа *LL*). Для оценки статистической значимости обнаруженных эффектов вводили двойной порог: на уровне вокселя ($p < 0.005$) и на уровне кластера (размер кластера больше 100 вокселей).

Результаты

Статистический анализ (Т-контраст) коннективности ДСМ в двух группах (носители *S* аллеля *vs* носители генотипа *LL*) был выполнен в программе SPM 12.

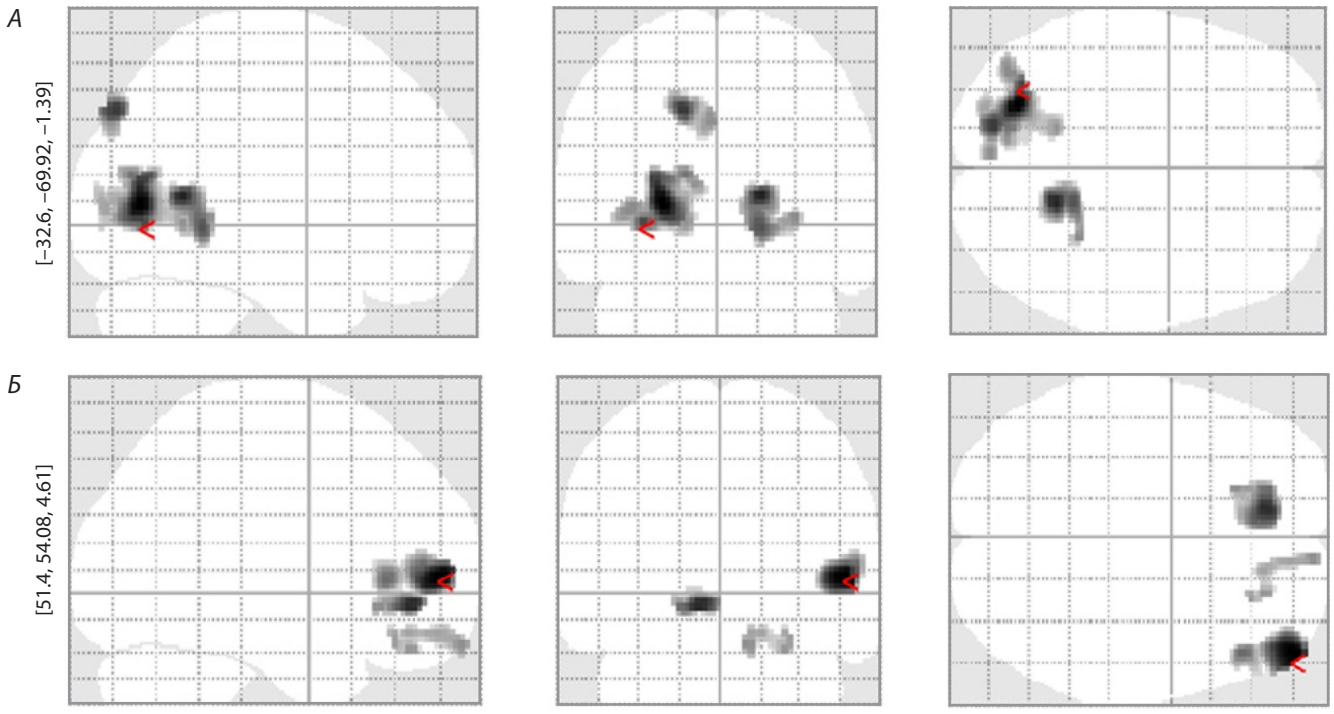
Результаты сравнения коннективности ДСМ у носителей *S* аллеля и носителей *LL* генотипа показаны на рисунке, *А*. Выявлено, что у носителей *S* аллеля по сравнению с носителями генотипа *LL* была больше коннективность ДСМ в левом полушарии с предклиньем (ОБ (область Бродманна) 19, $x = –17$, $y = –82$, $z = 41$, размер кластера = 166, T = 3.10, $p = 0.001$) и с кластером в левом полушарии, охватывающим заднюю поясную кору (ОБ 30, $x = –29$, $y = –70$, $z = 15$, размер кластера = 943, T = 3.08, $p = 0.001$), область клина (ОБ 30, $x = –25$, $y = –72$, $z = 7$, размер кластера = 943, T = 3.32, $p < 0.001$) и среднюю затылочную извилину (ОБ 19, $x = –29$, $y = –70$, $z = 15$, размер кластера = 943, T = 3.05, $p = 0.001$), и с кластером в правом полушарии, охватывающим заднюю поясную кору (ОБ 30, $x = 17$, $y = –54$, $z = 9$, размер кластера = 341, T = 3.19, $p = 0.001$) и парагиппокампальную извилину (ОБ 19, $x = 17$, $y = –46$, $z = –3$, размер кластера = 341, T = 3.03, $p = 0.001$) (см. рисунок, *А*).

Результаты Т-контраста коннективности ДСМ у носителей *LL* генотипа больше, чем у носителей *S* аллеля (см. рисунок, *Б*). У носителей *LL* генотипа была больше коннективность ДСМ с медиальной лобной извилиной (ОБ 10, $x = –11$, $y = 40$, $z = –7$, размер кластера = 251, T = 2.97, $p = 0.002$), с верхней лобной извилиной (ОБ 11, $x = 9$, $y = 62$, $z = –21$, размер кластера = 110, T = 2.76, $p = 0.003$) и с кластером, расположенным в правой дорсолатеральной префронтальной коре (ОБ 10, $x = 51$, $y = 54$, $z = 5$, размер кластера = 473, T = 3.06, $p = 0.001$) (см. рисунок, *Б*).

Согласно (Raichle, 2015), передний и задний узлы ДСМ могут иметь разное функциональное значение. Для того чтобы понять вклад переднего и заднего узлов ДСМ в выявленные эффекты, мы провели сравнения для переднего и заднего узлов ДСМ отдельно в группах носителей генотипа *LL* и носителей *S* аллеля.

Коннективность переднего узла ДСМ у носителей *S* аллеля была больше, чем у носителей *LL* генотипа. Сравнение не выявило статистически значимых эффектов. Тогда как у носителей *LL* генотипа по сравнению с носителями *S* аллеля была выше коннективность переднего узла ДСМ с задней поясной корой (ОБ 23, $x = 2$, $y = –44$, $z = 29$, размер кластера = 1800, T = 3.77, $p < 0.001$), с правой верхней лобной извилиной (ОБ 10, $x = 23$, $y = 60$, $z = –3$, размер кластера = 581, T = 3.08, $p = 0.001$) и медиальной лобной извилиной (ОБ 6, $x = –11$, $y = –12$, $z = 57$, размер кластера = 292, T = 3.05, $p = 0.001$).

У носителей *S* аллеля по сравнению с носителями *LL* генотипа была больше коннективность заднего узла



Результаты статистического анализа (Т-контраст) коннективности ДСМ между группой носителей *S* аллеля и группой носителей генотипа *LL*.

А – Т-контраст коннективность ДСМ у носителей *S* аллеля больше, чем коннективность ДСМ у носителей *LL* генотипа; *Б* – Т-контраст коннективность ДСМ у носителей *LL* генотипа больше, чем коннективность ДСМ у носителей *S* аллеля.

ДСМ с кластером в правом полушарии, охватывающим заднюю поясную кору (ОБ 30, $x = 17$, $y = –56$, $z = 9$, размер кластера = 249, T = 2.95, $p = 0.002$) и парагиппокампальную извилину (ОБ 19, $x = 19$, $y = –46$, $z = –5$, размер кластера = 249, T = 2.79, $p = 0.002$), и с кластером в левом полушарии, охватывающим заднюю поясную кору (ОБ 30, $x = –29$, $y = –70$, $z = 15$, размер кластера = 782, T = 3.08, $p = 0.001$) и область клина (ОБ 17, $x = –11$, $y = –76$, $z = 13$, размер кластера = 782, T = 3.10, $p = 0.002$).

У носителей *LL* генотипа по сравнению с носителями *S* аллеля была больше коннективность заднего узла ДСМ с кластером вокселей, охватывающим правую дорсолатеральную префронтальную кору (ОБ 46, $x = 43$, $y = 30$, $z = 17$, размер кластера = 234, T = 2.77, $p = 0.003$).

Обсуждение

В нашем исследовании у носителей *S* аллеля по сравнению с представителями генотипа *LL* был выявлен паттерн коннективности ДСМ с предклиньем, с задней поясной корой, а также с теменной и затылочной корой. Известно, что предклинье и задняя поясная кора входят в нейронный субстрат самосознания и образуют задний узел ДСМ (Raichle et al., 2001; Utevsky et al., 2014). Неоднократно было показано, что ДСМ сопряжена с внутренне направленными мыслительными процессами и повышает свою активность во время самореферентной обработки, автобиографических воспоминаний, размышлений о будущем и построении планов (Knyazev, 2013; Menon, D'Esposito, 2022). Согласно (Raichle, 2015), существует функциональное разделение переднего и заднего узлов ДСМ. Задний узел ДСМ преимущественно связан с процессами воспоминаний о прошлых событиях (Raichle, 2015). Также в исследованиях было показано, что совместная активация задней поясной коры и предклинья наблюдалась во время извлечения из памяти и переживания травматических событий у пациентов с ПТСР (Ramage et al., 2013; Thome et al., 2020). А в условиях регуляции эмоций, по сравнению с пассивным наблюдением за эмоциональными стимулами, у пациентов с ПТСР и у группы здоровых испытуемых наблюдалось снижение активации задней поясной коры и предклинья (Nicholson et al., 2022).

Интересно, что для носителей *S* аллеля, помимо увеличенной коннективности ДСМ с предклиньем и с задней поясной корой, было характерно увеличение коннективности с теменной и затылочной корой, входящими в состав зрительной ассоциативной коры. Выявлено, что толщина зрительной ассоциативной коры коррелирует с выраженностью депрессии и тревоги (Peterson et al., 2009). Также было показано, что зрительная ассоциативная кора участвует в процессах руминации (Cooney et al., 2010).

У носителей *S* аллеля по сравнению с носителями генотипа *LL* обнаружено увеличение коннективности ДСМ с правой парагиппокампальной извилиной, являющейся частью височной коры. Известно, что ДСМ ассоциирована с интроспективной мыслительной деятельностью, а височная кора и парагиппокампальная извилина участвуют в процессах извлечения информации из памяти (Buckner, DiNicola, 2019). В более ранней работе мы выявили положительную корреляцию выраженности симптомов депрессии с паттерном коннективности ДСМ с правой височной корой (Bocharov et al., 2021). Позднее нам удалось обнаружить, что увеличение такого паттерна коннективности коррелировало с выраженностью навязчивых мыслей – руминаций и предпочтением неадаптивной стратегии регуляции эмоций – подавления мыслей (Bocharov et al., 2022). Надо отметить, что в исследовании (Rassin, 2003) было выявлено, что при использовании такой стратегии стремление подавить навязчивые мысли парадоксальным образом приводит к их усилению. В исследованиях показано, что носители *S* аллеля в большей степени подвержены развитию БДР и имеют трудности в регуляции эмоционального состояния (Caspi et al., 2003; Walderhaug et al., 2010). Можно предположить, что выявленный паттерн, выражающийся в более высокой коннективности ДСМ с задним узлом ДСМ, задней ассоциативной корой и правой височной корой (парагиппокампальная извилина), характерный для носителей *S* аллеля, может быть связан и/или предрасполагает к возникновению навязчивых мыслей и может лежать в основе предрасположенности к возникновению и поддержанию симптомов депрессии.

В отличие от носителей *S* аллеля, у которых была обнаружена повышенная коннективность заднего узла ДСМ, у носителей *LL* генотипа была выше коннективность ДСМ в переднем узле: с медиальной лобной извилиной и с верхней лобной извилиной, являющейся частью вентромедиальной лобной коры. Структуры переднего узла ДСМ могут иметь разное функциональное значение (Raichle, 2015): так, дорсомедиальная префронтальная кора принимает участие в процессах самореференции, а вентромедиальная часть лобной коры участвует в процессах социального поведения, мотивации, настроения и эмоциональной обработки. В исследовании (Rao et al., 2007) была обнаружена сниженная активация вентромедиальной лобной коры и повышенная активация миндалевидного тела в ответ на негативные стимулы у носителей *S* аллеля по сравнению с носителями генотипа *LL*. Считается, что дисфункция вентромедиальной лобной коры играет ключевую роль в патогенезе аффективных и тревожных расстройств. Вероятно, недостаточная активность вентромедиальной лобной коры может приводить к нарушению ингибирования активности миндалевидного тела и, как следствие, к патологически повышенному уровню негативного аффекта (Motzkin et al., 2015).

Отметим, что для носителей *LL* генотипа было характерно увеличение коннективности переднего узла ДСМ с задней поясной корой, входящей в состав заднего узла ДСМ, т. е. была повышена коннективность между передним и задним узлами ДСМ. В некоторой степени этот результат согласуется с данными работы (Cha et al., 2018). Авторы обнаружили, что у носителей генотипов с низкой транскрипционной эффективностью понижена коннективность между верхней лобной извилиной, входящей в состав переднего узла ДМН, и задним узлом ДМН. Снижение коннективности между этими структурами опосредовало повышенную импульсивность у носителей *SS* генотипа (Cha et al., 2018).

Кроме того, в нашем исследовании было выявлено, что для носителей *LL* генотипа характерно увеличение коннективности заднего узла ДСМ, участвующего в процессах автобиографической памяти, с правой дорсолатеральной префронтальной корой, ответственной за когнитивный контроль и регулирующие функции (Scult et al., 2017). Так, в работах было показано, что правая дорсолатеральная префронтальная кора часто используется как целевая область для применения транскраниальной магнитной стимуляции (ТМС) при лечении ПТСР, обессивно-компульсивного расстройства и зависимостей. В эксперименте, проведенном B.D. Greenberg, U. Ziemann (1998), воздействие с помощью ТМС на область правой дорсолатеральной префронтальной коры у пациентов с обессивно-компульсивным расстройством способствовало снижению компульсивных симптомов и улучшению настроения. H. Kober с коллегами (2010) выявили, что у курильщиков снижение тяги к курению связано с повышением активности дорсолатеральной префронтальной коры. В исследовании (Cohen et al., 2004) продемонстрированы эффективность ТМС в снижении симптомов ПТСР и значимость этой области мозга в регуляции стресса и эмоциональных реакций. Согласно данным (Wu et al., 2020), стимуляция правой дорсолатеральной префронтальной коры мозга может улучшить способности человека к саморегуляции, особенно в контексте управления своими желаниями и влечениями, а также при регуляции негативных эмоций. Можно предположить, что у носителей *LL* генотипа, которые, как известно, более устойчивы к стрессу и развитию депрессии (Caspi et al., 2003; Koenen et al., 2009) и проявлению импульсивного поведения (Walderhaug et al., 2010), увеличение коннективности ДСМ со структурами головного мозга, участвующими в процессах когнитивной регуляции (дорсолатеральная и вентромедиальная префронтальная кора), может способствовать более эффективной регуляции процессов ДСМ, связанных с автобиографическими воспоминаниями.

Заключение

Выявленные особенности коннективности ДСМ в состоянии покоя у носителей *LL* генотипа и носителей *S* аллеля вносят вклад в понимание связей полиморфизма гена транспортера серотонина с предрасположенностью к аффективным расстройствам. У носителей *S* аллеля по сравнению с носителями *LL* генотипа был выявлен паттерн более высокой коннективности заднего узла ДСМ, участвующего в процессах автобиографической памяти, с предклиньем, задней поясной и зрительной ассоциативной корой, а также с правой парагиппокампальной извилиной, что может предрасполагать к возникновению и поддержанию навязчивых мыслей. Тогда как у носителей *LL* генотипа была выше коннективность ДСМ в переднем узле ДСМ: с медиальной лобной извилиной, вентромедиальной лобной корой и задней поясной извилиной, входящей в состав заднего узла ДСМ. Также у носителей *LL* генотипа по сравнению с носителями *S* аллеля в состоянии покоя была выше коннективность заднего узла ДСМ с кластером, охватывающим правую дорсолатеральную кору. Можно предположить, что увеличение коннективности ДСМ со структурами головного мозга, участвующими в процессах когнитивной регуляции (дорсолатеральная и вентромедиальная префронтальная кора), может способствовать регуляции процессов ДСМ, связанных с автобиографическими воспоминаниями.

Список литературы / References

Adolphs R. Fear, faces, and the human amygdala. *Curr Opin Neurobiol.* 2008;18:166-172. doi 10.1016/j.conb.2008.06.006

Beevers C.G., Pacheco J., Clasen P., McGeary J.E., Schnyer D. Prefrontal morphology, 5-HTTLPR polymorphism and biased attention for emotional stimuli. *Genes Brain Behav*. 2010;9(2):224-233. doi 10.1111/j.1601-183X.2009.00550.x

Berman M.G., Peltier S., Nee D.E., Kross E., Deldin P.J., Jonides J. Depression, rumination and the default network. *Soc Cogn Affect Neurosci*. 2011;6:548-555. doi 10.1093/scan/nsq080

Bocharov A.V., Knyazev G.G., Savostyanov A.N., Saprygin A.E., Proshina E.A., Tamozhnikov S.S. Relationship of depression, anxiety and rumination scores with EEG connectivity of resting state networks. *Hum Physiol*. 2021;47(2):123-127. doi 10.1134/S0362119721010023

Bocharov A.V., Savostyanov A.N., Tamozhnikov S.S., Proshina E.A., Knyazev G.G. The association between emotion regulation strategy and oscillation balance of resting state networks. *Hum Physiol*. 2022;48(1):30-36. doi 10.1134/S0362119722010030

Brookes M.J., Woolric M., Luckhoo H., Price D., Hale J.R., Stephenson M.C., Barnes G.R., Smith S.M., Morris P.G. Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proc Natl Acad Sci USA*. 2011;108(40):16783-16788. doi 10.1073/pnas.1112685108

Buckner R.L., DiNicola L.M. The brain's default network: updated anatomy, physiology and evolving insights. *Nat Rev Neurosci*. 2019;20(10):593-608. doi 10.1038/s41583-019-0212-7

Caspi A., Sugden K., Moffitt T.E., Taylor A., Craig I.W., Harrington H., McClay J., Mill J., Martin J., Braithwaite A., Poulton R. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*. 2003;301(5631):386-389. doi 10.1126/science.1083968

Cha J., Guffanti G., Gingrich J., Talat A., Wickramaratne P., Weissman M., Posner J. Effects of serotonin transporter gene variation on impulsivity mediated by default mode network: a family study of depression. *Cerebral Cortex*. 2018;28(6):1911-1921. doi 10.1093/cercor/bhx097

Cohen H., Kaplan Z., Kotler M., Kouperman I., Moisa R., Grisaru N. Repetitive transcranial magnetic stimulation of the right dorsolateral prefrontal cortex in posttraumatic stress disorder: a double-blind, placebo-controlled study. *Am J Psychiatry*. 2004;161(3):515-524. doi 10.1176/appi.ajp.161.3.515

Cools R., Roberts A.C., Robbins T.W. Serotoninergic regulation of emotional and behavioural control processes. *Trends Cogn Sci*. 2008;12:31-40. doi 10.1016/j.tics.2007.10.011

Cooney R.E., Joormann J., Eugène F., Dennis E.L., Gotlib I.H. Neural correlates of rumination in depression. *Cogn Affect Behav Neurosci*. 2010;10(4):470-478. doi 10.3758/CABN.10.4.470

Fuchs M., Wagner M., Kastner J. Boundary element method volume conductor models for EEG source reconstruction. *Clin Neurophysiol*. 2001;112:1400-1407. doi 10.1016/S1388-2457(01)00589-2

Greenberg B.D., Ziemann U. Decreased neuronal inhibition in cerebral cortex in obsessive-compulsive disorder on transcranial magnetic stimulation. *Lancet*. 1998;352(9131):881-882. doi 10.1016/S0140-6736(05)60009-8

Greenberg B.D., Li Q., Lucas F.R., Hu S., Sirota L.A., Benjamin J., Murphy D.L. Association between the serotonin transporter promoter polymorphism and personality traits in a primarily female population sample. *Am J Med Genet*. 2000;96:202-216. doi 10.1002/(SICI)1096-8628(20000403)96:2<202::AID-AJMG16>3.0.CO;2-J

Gusnard D.A., Raichle M.E. Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci*. 2001;2(10):685-694. doi 10.1038/35094500

Hamilton J.P., Furman D.J., Chang C., Thomason M.E., Dennis E., Gotlib I.H. Default-mode and task-positive network activity in major depressive disorder: implications for adaptive and maladaptive rumination. *Biol Psychiatry*. 2011;70(4):327-333. doi 10.1016/j.biopsych.2011.02.003

Hamilton J.P., Chen M.C., Gotlib I.H. Neural systems approaches to understanding major depressive disorder: an intrinsic functional organization perspective. *Neurobiol Dis*. 2013;52:4-11. doi 10.1016/j.nbd.2012.01.015

Hipp J.F., Hawellek D.J., Corbetta M., Siegel M., Engel A.K. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat Neurosci*. 2012;15(6):884-890. doi 10.1038/nn.3101

Hu X.Z., Lipsky R.H., Zhu G., Akhtar L.A., Taubman J., Greenberg B.D., Xu K., Arnold P.D., Richter M.A., Kennedy J.L., Murphy D.L., Goldman D. Serotonin transporter promoter gain-of-function genotypes are linked to obsessive-compulsive disorder. *Am J Hum Genet*. 2006;78:815-826. doi 10.1086/503850

Hutchinson-Wong N., Glue P., Adhia D., de Ridder D. How does depressive cognition develop? A state-dependent network model of predictive processing. *Psychol Rev.* 2025;132(2):442-469. doi 10.1037/rev0000512. Epub 2024 Nov 14. PMID: 39541510.

Knyazev G.G. EEG correlates of self-referential processing. *Front Hum Neurosci.* 2013;7:264. doi 10.3389/fnhum.2013.00264

Knyazev G.G., Savostyanov A.N., Bocharov A.V., Tamozhnikov S.S., Saprigyn A.E. Task-positive and task-negative networks and their relation to depression: EEG beamformer analysis. *Behav Brain Res*. 2016;306:160-169. doi 10.1016/j.bbr.2016.03.033

Kober H., Kross E.F., Mischel W., Hart C.L., Ochsner K.N. Regulation of craving by cognitive strategies in cigarette smokers. *Drug Alcohol Depend*. 2010;106(1):52-55. doi 10.1016/j.drugalcdep.2009.07.017

Koenen K.C., Aiello A.E., Bakshis E., Amstadter A.B., Ruggiero K.J., Acierno R., Galea S. Modification of the association between serotonin transporter genotype and risk of posttraumatic stress disorder in adults by county-level social environment. *Am J Epidemiol*. 2009;169(6):704-711. doi 10.1093/aje/kwn397

Landro N.I., Jonassen R., Clark L., Haug K.F., Aker M., Bo R., Berg J.P., Neumeister A., Stiles T.C. Serotonin transporter polymorphisms predict response inhibition in healthy volunteers. *Neurosci Lett*. 2015;584:109-112. doi 10.1016/j.neulet.2014.10.006

Lesch K.P., Bengel D., Heils A., Sabol S.Z., Greenberg B.D., Petri S., Murphy D.L. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*. 1996;274(5292):1527-1531. doi 10.1126/science.274.5292.1527

Menon V., D'Esposito M. The role of PFC networks in cognitive control and executive function. *Neuropsychopharmacol*. 2022;47(1):90-103. doi 10.1038/s41386-021-01152-w

Meyer-Lindenberg A. Neural connectivity as an intermediate phenotype: brain networks under genetic control. *Hum Brain Mapp*. 2009;30(7):1938-1946. doi 10.1002/hbm.20639

Motzkin J.C., Philippi C.L., Wolf R.C., Baskaya M.K., Koenigs M. Ventromedial prefrontal cortex is critical for the regulation of amygdala activity in humans. *Biol Psychiatry*. 2015;77(3):276-284. doi 10.1016/j.biopsych.2014.02.014

Nicholson A.A., Rabellino D., Densmore M., Frewen P.A., Steyrl D., Scharnowski F., Theberge J., Neufeld R.W.J., Schmahl C., Jetly R., Lanius R.A. Differential mechanisms of posterior cingulate cortex downregulation and symptom decreases in posttraumatic stress disorder and healthy individuals using real-time fMRI neurofeedback. *Brain Behav*. 2022;12(1):e2441. doi 10.1002/brb3.2441

Peterson B.S., Warner V., Bansal R., Zhu H., Hao X., Liu J., Weissman M.M. Cortical thinning in persons at increased familial risk for major depression. *Proc Natl Acad Sci USA*. 2009;106(15):6273-6278. doi 10.1073/pnas.0805311106

Raichle M.E. The brain's default mode network. *Annu Rev Neurosci*. 2015;38:433-447. doi 10.1146/annurev-neuro-071013-014030

Raichle M.E., MacLeod A.M., Snyder A.Z., Powers W.J., Gusnard D.A., Shulman G.L. A default mode of brain function. *Proc Natl Acad Sci USA*. 2001;98(2):676-682. doi 10.1073/pnas.98.2.676

Ramage A.E., Laird A.R., Eickhoff S.B., Acheson A., Peterson A.L., Williamson D.E., Telch M.J., Fox P.T. A coordinate-based meta-analytic model of trauma processing in posttraumatic stress disorder. *Hum Brain Mapp*. 2013;34(12):3392-3399. doi 10.1002/hbm.22155

Rao H.Y., Gillihan S.J., Wang J.J., Korczykowski M., Sankoorikal G.M.V., Kaercher K.A., Brodkin E.S., Detre J.A., Farah M.J. Genetic variation in serotonin transporter alters resting brain function in healthy individuals. *Biol Psychiatry*. 2007;62(6):600-606. doi 10.1016/j.biopsych.2006.11.028

Rassin E. The white bear suppression inventory (WBSI) focuses on failing suppression attempts. *Eur J Pers*. 2003;17(4):285-298. doi 10.1002/per.478

Risch N., Herrell R., Lehner T., Liang K.Y., Eaves L., Hoh J., Griem A., Kovacs M., Jurg J., Merikangas K.R. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *Jama*. 2009;301(23):2462-2471. doi 10.1001/jama.2009.878

Scult M.A., Knodt A.R., Swartz J.R., Brigidi B.D., Hariri A.R. Thinking and feeling: individual differences in habitual emotion regulation and stress-related mood are associated with prefrontal executive control. *Clin Psychol Sci*. 2017;5(1):150-157. doi 10.1177/2167702616654688

Si Q., Tian J., Savostyanov V.A., Lebedkin D.A., Bocharov A.V., Savostyanov A.N. Comparison of brain activity indexes in the Chinese and Russian students under recognition of self- and other-related information. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;28(8):982-992. doi 10.18699/vjgb-24-105

Thome J., Terpou B.A., McKinnon M.C., Lanius R.A. The neural correlates of trauma-related autobiographical memory in posttraumatic stress disorder: a meta-analysis. *Depress Anxiety*. 2020;37(4):321-345. doi 10.1002/da.22977

Utevsky A.V., Smith D.V., Huettel S.A. Precuneus is a functional core of the default-mode network. *J Neurosci*. 2014;34(3):932-940. doi 10.1523/JNEUROSCI.4227-13.2014

Van Veen B.D., van Drongelen W., Yuchtman M., Suzuki A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng*. 1997;44:867-880. doi 10.1109/10.623056

Walderhaug E., Herman A.I., Magnusson A., Morgan M.J., Landro N.I. The short (S) allele of the serotonin transporter polymorphism and acute tryptophan depletion both increase impulsivity in men. *Neurosci Lett*. 2010;473:208-211. doi 10.1016/j.neulet.2010.02.048

Wiggins J.L., Bedoyan J.K., Peltier S.J., Ashinoff S., Carrasco M., Weng S.J., Welsh R.C., Martin D.M., Monk C.S. The impact of serotonin transporter (5-HTTLPR) genotype on the development of resting-state functional connectivity in children and adolescents: a preliminary report. *Neuroimage*. 2012;59(3):2760-2770. doi 10.1016/j.neuroimage.2011.10.030

Wu L.L., Potenza M.N., Zhou N., Kober H., Shi X.H., Yip S.W., Xu J., Zhu L., Wang R., Liu G., Zhang J.T. A role for the right dorsolateral prefrontal cortex in enhancing regulation of both craving and negative emotions in internet gaming disorder: a randomized trial. *Eur Neuropsychopharmacol*. 2020;36:29-37. doi 10.1016/j.euroneuro.2020.04.003

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 21.02.2025. После доработки 17.04.2025. Принята к публикации 21.04.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):274-283

doi 10.18699/vjgb-26-31

Phylogeography and taxonomic status of the *Formica picea* complex (Hymenoptera: Formicidae)

Z.A. Zhigulskaya [1], S.V. Shekhovtsov [1, 2] ✉, S.V. Chesnokova [3], A.P. Burnasheva [4], A.A. Gurina [3], R.Yu. Dudko [3], T.V. Poluboyarova [2], S.V. Reshetnikov [3], Yu.N. Sundukov [5], D.I. Berman [1]

[1] Institute of Biological Problems of the North of the Far Eastern Branch of the Russian Academy of Sciences, Magadan, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibisk, Russia
[3] Institute of Systematics and Ecology of Animals of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[4] Institute of Biological Problems of Cryolithozone of the Siberian Branch of the Russian Academy of Sciences, Yakutsk, Russia
[5] Federal Scientific Center of the East Asia Terrestrial Biodiversity, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, Russia

✉ shekhovtsov@bionet.nsc.ru

Abstract. The black bog ant *Formica picea* complex is widespread from the Atlantic to the Pacific coasts of Eurasia. This complex was earlier believed to consist of one or two species (*F. picea* and *F. candida*). However, molecular analysis suggested that it includes three cryptic species. One is *F. picea* from Europe, another, *F. candida*, is currently known exclusively from Kyrgyzstan, while the third one, temporarily designated here as *Formica* sp., inhabits the easternmost part of Eurasia from China to Kamchatka. It is unknown how *F. picea* and *Formica* sp. are distributed in Siberia and whether their ranges intersect. Here we studied a sample of this complex from Siberia using mtDNA and found that their ranges overlap. The distribution of *Formica* sp. extends from the south of West Siberia, including Altai, to China, and the Russian Far East. No phylogeographic structure was detected, suggesting their recent dispersal from a single source. *F. picea* was found as far as East Siberia, but was relatively rare. While the European and West Siberian populations were genetically closely related, the specimens from Zabaykalsky Krai differed, suggesting a putative East Siberian refugium. We also determined that ecologically *F. picea* inhabits peat bogs in lowland areas and grassy communities above the tree line in the European mountains; in Altai, it is found in mountain steppes, while in Transbaikalia, in waterlogged areas along riverbanks. *Formica* sp. thrives in dry steppes and low riverbanks, but avoids bogs. Thus, *F. picea* and *Formica* sp. differ genetically, and have different distribution ranges, as well as habitat preferences. This supports the opinion that *Formica* sp. should be recognized as a distinct species.

Key words: *Formica picea*; *Formica candida*; cryptic species; Northern Eurasia; Siberia

For citation: Zhigulskaya Z.A., Shekhovtsov S.V., Chesnokova S.V., Burnasheva A.P., Gurina A.A., Dudko R.Yu., Poluboyarova T.V., Reshetnikov S.V., Sundukov Yu.N., Berman D.I. Phylogeography and taxonomic status of the *Formica picea* complex (Hymenoptera: Formicidae). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):274-283. doi 10.18699/vjgb-26-31

Funding. The work was funded solely through the budgets of Federal State Budgetary Scientific Institutions: the Institute of Biological Problems of the North FEB RAS (No. 122041900011-9), the Institute of Cytology and Genetics SB RAS (FWNR-2026-0031), and the Institute of Systematics and Ecology of Animals SB RAS (FWGS-2026-0008). Sanger sequencing was performed in the SB RAS Genomics Core Facility (ICBFM SB RAS, Novosibirsk, Russia).

Acknowledgements. The authors are grateful to T.A. Novgorodova for critical comments on the manuscript, and to O.V. Vaulin, T.M. Krugova, and P.V. Golyakov for providing us with the ant samples. The photo of the Kansk Steppe (Central Altai) was kindly provided by R.M. Shershnev.

Филогеография и таксономический статус видового комплекса *Formica picea* (Hymenoptera: Formicidae)

З.А. Жигульская [1], С.В. Шеховцов [1, 2] ✉, С.В. Чеснокова [3], А.П. Бурнашева [4], А.А. Гурина [3], Р.Ю. Дудко [3], Т.В. Полубоярова [2], С.В. Решетников [3], Ю.Н. Сундуков [5], Д.И. Берман [1]

[1] Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Институт систематики и экологии животных Сибирского отделения Российской академии наук, Новосибирск, Россия
[4] Институт биологических проблем криолитозоны Сибирского отделения Российской академии наук, Якутск, Россия
[5] Федеральный научный центр биоразнообразия наземной биоты Восточной Азии Дальневосточного отделения Российской академии наук, Владивосток, Россия

✉ shekhovtsov@bionet.nsc.ru

Аннотация. Комплекс *Formica picea*, или черный болотный муравей, широко распространен по всему северу Евразии, от Испании и Ирландии до побережья Тихого океана и Тибета, примерно между 35 и 66.5° северной широты. Ранее считалось, что он состоит из одного или двух видов (*F. picea* и *F. candida*). Однако молекулярный

© Zhigulskaya Z.A., Shekhovtsov S.V., Chesnokova S.V., Burnasheva A.P., Gurina A.A., Dudko R.Yu., Poluboyarova T.V., Reshetnikov S.V., Sundukov Yu.N., Berman D.I., 2026
This work is licensed under a Creative Commons Attribution 4.0 License

анализ показал, что в него входят три криптических вида. Один из них – *F. picea* из Европы, другой – *F. candida*, известный исключительно из Киргизии, а третий, временно обозначенный как *Formica* sp., обитает на востоке Евразии: от Китая до Камчатки. До сих пор было неизвестно, каково распределение *F. picea* и *Formica* sp. в Сибири, пересекаются ли их ареалы. Мы изучили образцы этого комплекса из Сибири с использованием митохондриальной ДНК и обнаружили, что их ареалы перекрываются. Ареал *Formica* sp. охватывает юг Западной Сибири, включая Алтай, на восток – до Тихоокеанского побережья. Филогеографической структуры не обнаружено, что указывает на недавнее расселение из одного источника. Мы нашли мало образцов *F. picea* в Сибири, это свидетельствует, что здесь этот вид встречается редко. Самая восточная находка вида была сделана в Забайкальском крае. *F. picea* обитает на торфяных болотах в пределах равнин и травяных сообществах выше линии леса в европейских горах; на Алтае – в горных степях, а в Забайкалье – на влажных участках у речных берегов. *Formica* sp. предпочитает сухие степи и низкие берега рек, но избегает болот. Таким образом, оказалось, что *F. picea* и *Formica* sp. отличаются генетически, имеют разные ареалы и биотопические предпочтения по местообитаниям. Это свидетельствует в пользу того, что *Formica* sp. можно считать отдельным видом.

Ключевые слова: *Formica picea*; *Formica candida*; виды-двойники; Северная Евразия; Сибирь

Introduction

The *Formica picea* Nylander, 1846 species complex, i. e. black bog ant, is widespread in the Palearctic, from Spain and Ireland to the Pacific coast and Tibet, and between approximately 35 and 66.5° N latitude (Ruzsky, 1905; Pleshanov, 1966; Dlussky, 1967; Dmitrienko, Petrenko, 1976; Kupyanskaya, 1990; Berman et al., 2007; Antonov, Pleshanov, 2008; Radchenko, 2016; Seifert, 2018) (Fig. 1). For a long time, *F. picea* was considered a single species with high geographic variation (Dlussky, 1967). Bolton (1995) proposed that the name *F. candida* Smith, 1878 takes precedence over *F. picea*. Not all specialists adopted this suggestion, leading to confusing parallel usage of both names in the literature.

Seifert (2004) suggested that *F. picea* and *F. candida* are two distinct species, based on their pubescence and certain morphometric parameters. Seifert (2004) also designated a neotype of *F. candida* collected from Kyrgyzstan due to the loss of the original type specimen. However, Zakharov et al. (2019) disagreed with Seifert's conclusions, pointing out the smooth gradient of morphological variation from Europe to the Far East.

There were several molecular studies that included members of the *F. picea*/*F. candida* complex (Goropashnaya, 2003; Goropashnaya et al., 2012; Antonov, Bukin, 2016; Chen, Zhou, 2017; Schär et al., 2018). However, none of these aimed to explore the differences between the two taxa, instead simply using one or the other name. Moreover, some authors used the mitochondrial *cytb* gene, and others, *cox1*/*COI*, so their results could not be directly compared.

A recent study on the *F. picea* complex showed that it included one more species, in addition to *F. picea* and *F. candida* (Zhigulskaya et al., 2022). While *F. picea* turned out to be restricted to Europe, as well as to one location from Tibet, ants from all east Eurasian locations (lowland China and the Far

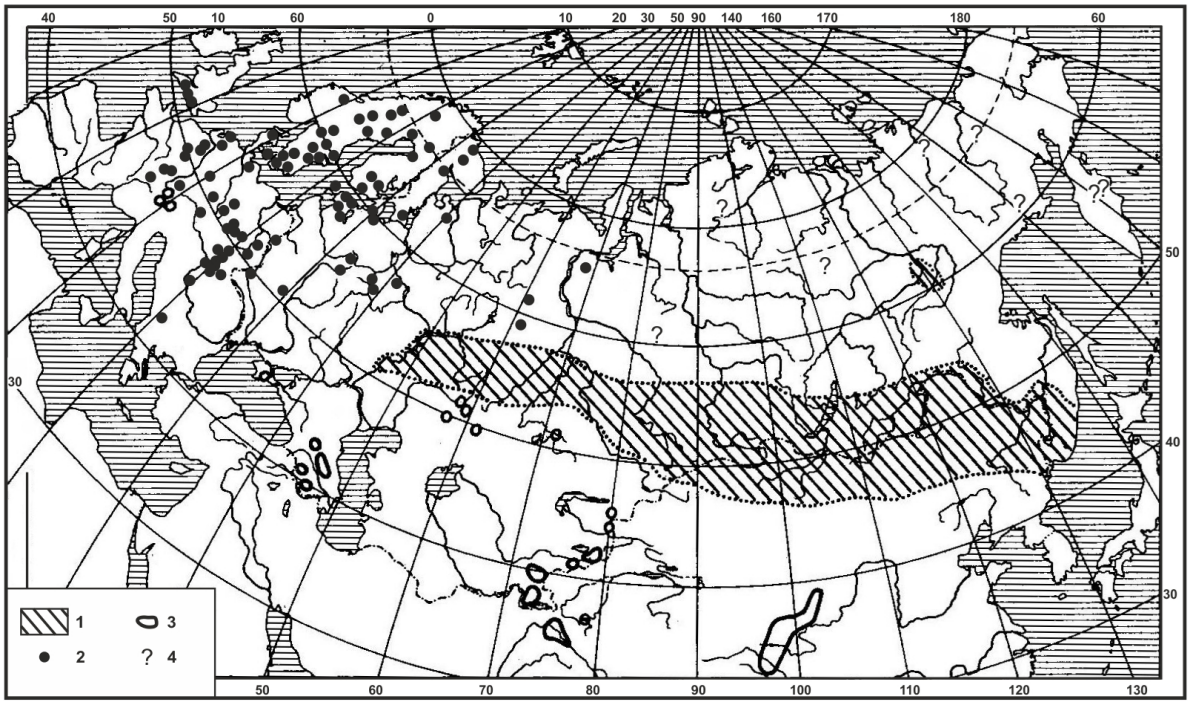


Fig. 1. Geographical distribution and ecology of the *F. picea* complex.

Characteristic habitats: 1 – diverse habitats; 2 – peat bogs; 3 – mountain meadows; 4 – no details provided (scheme from (Dlussky, 1967)).

East of Russia) were highly genetically divergent from other known *Formica* species. Geographic and genetic differences suggest that it should be regarded as a distinct species, although the formal description was not yet performed. Thus, this new taxon is hereafter referred to as *Formica* sp. There seem to be no obvious morphological distinctions for *Formica* sp. that could clearly differentiate it from *F. picea* and *F. candida*, supporting the observations of Zakharov and Dlussky (Zakharov et al., 2019) on high levels of intrapopulation and geographic variation for this complex.

Siberia currently remains a blank spot in this respect: it is unknown how far *F. picea* goes eastwards from the Urals, and how far west can *Formica* sp. be found. In this study we collected an extensive sample from different parts of Siberia in order to determine the ranges of both taxa, as well as to infer phylogeographic patterns. In order to search for putative ecological differences between *F. picea* and *Formica* sp., we also documented habitat preferences, which had been previously well described for European *F. picea* (Dlussky, 1967; Zakharov, 2015; Radchenko, 2016; Seifert, 2018; Zakharov et al., 2019).

Materials and methods

We surveyed a wide range of habitat types, from steppes to bogs, in both lowland and mountainous landscapes of Siberia and the Russian Far East (see the Table, Figs 2, 3). Ants were collected into 96 % ethanol in a series of 6–10 specimens from nests, as well as from forage plants. The collection material is stored in the Institute of Systematics and Ecology of Animals, Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia).

Ants of the *F. picea* complex were taken for genetic analysis from 85 locations (Fig. 2): for 47, obtained by us in this study, for the rest, they were taken from our previous study (Zhigulskaya et al., 2022) and from GenBank.

DNA was extracted from individual worker ants preserved in ethanol using commercial silica columns (BioSilica, Novosibirsk, Russia), following the method outlined in (Shekhovtsov et al., 2020). We amplified a segment of the mitochondrial cytochrome c oxidase subunit 1 (*cox1*) gene with the universal primers LCO1490m (5′-TACTC-AACAA-ATCAC-AAAGA-TATTG-G-3′; modified from (Folmer et al., 1994)) and HCO2198 (5′-TAAAC-TTCAG-GGTGA-CCAAA-AAATC-A-3′) (Folmer et al., 1994). Additionally, we amplified a portion of the cytochrome b (*cytb*) gene using primers Fcbl-F (5′-ACCCT-CACCT-GTAAA-TATTT-CTT-3′) and Fcbl-R (5′-GGAAT-AGATC-GTAAA-ATTGC-AT-3′) (Zhigulskaya et al., 2022). PCR was conducted with the Biomaster HS-Taq PCR Mix (Biolabmix, Novosibirsk, Russia).

The DNA fragments obtained were visualized through agarose gel electrophoresis. Unincorporated primers and nucleotide phosphates were eliminated using shrimp alkaline phosphatase and *Escherichia coli* exonuclease I mixture (New England Biolabs, Ipswich, MA, USA). Sanger sequencing was conducted on a 3130xl DNA Analyzer (Applied Biosystems, Framingham, MA, USA) at the SB RAS Genomics Core Facility (ICBFM SB RAS, Novosibirsk, Russia) using both forward and reverse primers. The resulting sequences were submitted to GenBank with accession numbers PQ807734–PQ807780 (*cox1*) and PQ810811–PQ810857 (*cytb*). Sequences from other researchers were also incorporated into this study (refer to the Table for GenBank accessions). This dataset includes the



Fig. 2. *F. picea* complex specimens used in this study.
Location numbers refer to the Table. Rectangles: *F. picea*; circles: *Formica* sp.; triangles: *F. candida*.

Representatives of the *F. picea* complex with available *cox1/cytb* sequences

No.	Loc.	Coord.	Altitude	*cox1*	*cytb*
		Formica picea			
1	United Kingdom: Cors Goch Nature Reserve	N53.308 E–4.253	78	–	AY786145 (as *F. candida*)
2	Denmark	N55.830 E12.560	30	LT977411	–
3	Sweden: Varmland	N59.426 E13.271	50	–	JX170886
4	Sweden: Trajmossen, Brattforsheden	N59.695 E13.889	170	–	AY786144 (as *F. candida*)
5	Finland: Nylandia, Raasepori, Harpar Stortraesket	N59.975 E23.436	4	MZ607401*	–
6	Finland: Helsinki, Vihti	N60.417 E24.317	39	–	AY786146* (as *F. candida*)
7	Estonia: Torma	N58.813 E26.738	86	–	AY786147 (as *F. candida*)
8	Armenia: Aragats Mountains	N40.432 E44.236	2728	ON220876	ON228281
9	Russia: Novgorod oblast	N58.517 E31.267	36	–	AY786148 (as *F. candida*)
10	Russia: Moscow oblast, Odintsovo district	N55.733 E36.850	184	ON220886-87	ON228278-79
11	Russia: Moscow oblast	N55.751 E37.617	148	–	AY786149 (as *F. candida*)
12	Russia: Novgorod oblast, Kerzhenskiy Natural Reserve, swamp	N56.467 E44.500	ca. 120	ON220875	ON228280
13	Russia: Perm Krai, Kachkanar	N58.700 E59.483	307	–	AY786152 (as *F. candida*)
14	Russia: Revda, Sverdlovsk oblast	N56.800 E59.917	260	–	AY786151 (as *F. candida*)
15	Russia: Yekaterinburg	N56.833 E60.583	369	–	AY786150 (as *F. candida*)
16	**Russia: Tomsk oblast, Melnikovo village, swamp**	**N56.550 E84.067**	**90**	**PQ810826**	**PQ807749**
17	**Russia: Altai Republic, Abay village, swamp**	**N50.419 E85.066**	**1068**	**PQ810820**	**PQ807743**
18	**Russia: Altai Republic, Amur village, residual hills, stony steppe**	**N50.402 E85.107**	**1100**	**PQ810821**	**PQ807744**
19	**Russia: Altai Republic, Amur village, residual hills, stony steppe**	**N50.402 E85.107**	**1100**	**PQ810835**	**PQ807758**
20	**Russia: Altai Republic, Kosh-Agach village, steppe**	**N49.992 E88.664**	**2275**	**PQ810842**	**PQ807765**
21	**Russia: Zabaikalsky Krai, Gazimursky Zavod village, Kotikha River floodplain, swamp with rare trees**	**N51.545 E118.342**	**770**	**PQ810811**	**PQ807734**
22	China: Quinghai, Tibet	N35.000 E96.000	ca. 4500	–	AY786157 (as *F. candida*)
		***Formica* sp.**			
23	**Russia: Novosibirsk oblast, Kabinetnoye village, halophytic meadow**	**N55.117 E81.233**	**140**	**PQ810851**	**PQ807774**
24	**Russia: Novosibirsk oblast, Ordynsky district, steppe meadow**	**N54.320 E81.850**	**116**	**PQ810852**	**PQ807775**
25	Russia: Novosibirsk oblast, Iskitim district, steppe	N54.450 E83.470	230	ON220885	–
26	**Russia: Republic of Khakassia, Lake Shira, steppe**	**N54.534 E90.199**	**350**	**PQ810857**	**PQ807780**
27	**Russia: Altai Krai, Tigirek village, meadow with shrubs**	**N51.146 E83.010**	**500**	**PQ810831**	**PQ807754**
28	**Russia: Altai Republic, Ilyinka village, meadow by the river**	**N51.517 E85.033**	**690**	**PQ810843**	**PQ807766**
29	**Russia: Altai Republic, Yabogan village, steppe**	**N50.940 E85.021**	**1090**	**PQ810834**	**PQ807757**
30	Russia: Altai Republic, Ongudai village, stony steppe	N50.733 E86.133	810	ON220880	ON228273
31	**Russia: Altai Republic, Abay village, steppe near the swamp**	**N50.419 E85.066**	**1070**	**PQ810830**	**PQ807753**
32	**Russia: Altai Republic, Ust-Koksa village, stony steppe**	**N50.270 E85.615**	**1080**	**PQ810841**	**PQ807764**
33	**Russia: Altai Republic, Kurai village, steppe**	**N50.224 E87.932**	**1500**	**PQ810822**	**PQ807745**
34	**Russia: Altai Republic, Kurai village, steppe**	**N50.224 E87.932**	**1500**	**PQ810823**	**PQ807746**
35	**Russia: Altai Republic, Amur village, meadow near the lake**	**N50.223 E84.898**	**1980**	**PQ810840**	**PQ807763**
36	**Russia: Altai Republic, Kosh-Agach village, solonchak**	**N49.992 E88.664**	**1800**	**PQ810824**	**PQ807747**
37	**Russia: Altai Republic, Kosh-Agach village, solonchak**	**N49.992 E88.664**	**1800**	**PQ810825**	**PQ807748**

Table (continued)

No.	Loc.	Coord.	Altitude	*cox1*	*cytb*
38	**Russia: Republic of Tuva, Lake Dus-Khol, desertified steppe**	**N50.384 E94.870**	**990**	**PQ810836**	**PQ807759**
39	Russia: Buryatia Republic, Kizhinga village, steppe	N51.846 E109.913	700	ON220883-84	ON228270, ON228274
40	**Russia: Zabaikalsky Krai, Gazimur River, mixed-grass meadow**	**N51.764 E118.425**	**655**	**PQ810832**	**PQ807755**
41	**Russia: Zabaikalsky Krai, Gazimur River, mixed-grass meadow**	**N51.764 E118.425**	**655**	**PQ810833**	**PQ807756**
42	**Russia: Zabaikalsky Krai, Mount Kozlikha, meadow on a slope near a stream**	**N51.760 E118.713**	**730**	**PQ810837**	**PQ807760**
43	**Russia: Zabaikalsky Krai, Talovka River, swampy meadow**	**N51.450 E118.791**	**815**	**PQ810853**	**PQ807776**
44	**Russia: Zabaikalsky Krai, Uktycha River, steppe slope**	**N51.545 E118.342**	**750**	**PQ810854**	**PQ807777**
45	**Russia: Zabaikalsky Krai, Solonechnaya River, steppe meadow**	**N51.440 E118.907**	**750**	**PQ810855**	**PQ807778**
46	**Russia: Republic of Yakutia, Yakutsk, hygromesophytic meadow**	**N62.087 E129.289**	**260**	**PQ810845**	**PQ807768**
47	**Russia: Republic of Yakutia, Yakutsk, xerophytic meadow**	**N62.055 E129.644**	**100**	**PQ810846**	**PQ807769**
48	**Russia: Republic of Yakutia, Yakutsk, solonchak meadow**	**N62.021 E129.627**	**250**	**PQ810847**	**PQ807770**
49	**Russia: Republic of Yakutia, Yakutsk, lake Ytyk-Kyuel, meadow near the shore**	**N62.017 E129.618**	**100**	**PQ810848**	**PQ807771**
50	**Russia: Republic of Yakutia, Yakutsk, xerophytic meadow**	**N62.017 E129.618**	**100**	**PQ810849**	**PQ807772**
51	**Russia: Republic of Yakutia, Khayirdakh River, bank with *Chosenia***	**N66.263 E144.268**	**606**	**PQ810850**	**PQ807773**
52	Russia: Magadan oblast, Yagodnoye village, bank of Sokhatina River	N62.517 E149.617	465	ON220878	ON228277
53	Russia: Magadan oblast, Yagodnoye village, bank of Debin River	N62.750 E150.467	403	ON220879	ON228271
54	**Russia: Magadan oblast, Madaun village, gravel high floodplain**	**N60.615 E150.696**	**520**	**PQ810839**	**PQ807762**
55	**Russia: Magadan oblast, Klepka village, gravel high floodplain**	**N59.754 E151.491**	**78**	**PQ810838**	**PQ807761**
56	Russia: Magadan oblast, Arman village, gravel high floodplain	N59.667 E150.117	1	ON220872	ON228275
57	Russia: Kamchatka oblast	N53.190 E158.480	60	LT977377 (as *F. candida*)	–
58	**Russia: Sakhalin, Aniva, dry meadow**	**N46.019 E143.413**	**< than 10**	**PQ810812**	**PQ807735**
59	**Russia: Sakhalin, Aniva, dry meadow**	**N46.019 E143.413**	**< than 10**	**PQ810829**	**PQ807752**
60	**Russia: Sakhalin, Aniva, dry meadow**	**N46.019 E143.413**	**< than 10**	**PQ810819**	**PQ807742**
61	**Russia: Komsomolsk-on-Amur, Silinka River, floodplain with poplars**	**N50.550 E137.000**	**30**	**PQ810856**	**PQ807779**
62	**Russia: Khabarovsk Krai, sandy area on forest road**	**N66.263 E134.330**	**74**	**PQ810844**	**PQ807767**
63	Russia: Primorsky Krai, Lazo town, Lazovka river's bank	N43.380 E133.898	220	ON220881	ON228276
64	**Russia: Primorsky Krai, Kievka River, sandy seaside meadow**	**N42.855 E133.675**	**2–3**	**PQ810817**	**PQ807740**
65	**Russia: Primorsky Krai, Kievka River, sandy seaside meadow**	**N42.855 E133.675**	**2–3**	**PQ810818**	**PQ807741**
66	**Russia: Primorsky Krai, Kievka River, sandy seaside meadow**	**N42.855 E133.675**	**2–3**	**PQ810828**	**PQ807751**
67	**Russia: Primorsky Krai, Gryaznaya River, overgrown sand and gravel shoal**	**N43.361 E131.556**	**130–135**	**PQ810814**	**PQ807737**

Table (end)

No.	Loc.	Coord.	Altitude	*cox1*	*cytb*
68	**Russia: Primorsky Krai, Gryaznaya River, overgrown sand and gravel shoal**	**N43.361 E131.556**	**130–135**	**PQ810816**	**PQ807739**
69	Russia: Primorsky Krai, Sinii Utes town, gravel area	N43.062 E131.356	40–45	ON220882	ON228272
70	**Russia: Primorsky Krai, Andreyevka village, wasteland**	**N42.649 E131.140**	**7**	**PQ810813**	**PQ807736**
71	**Russia: Primorsky Krai, Andreyevka village, wasteland**	**N42.649 E131.140**	**7**	**PQ810815**	**PQ807738**
72	**Russia: Primorsky Krai, Andreyevka village, wasteland**	**N42.649 E131.140**	**7**	**PQ810827**	**PQ807750**
73	South Korea	–	–	MT800217 (as *F. candida*)	–
74	China: Yanbian, Jilin	N43.133 E129.183	750	KX665026-27 (as *F. candida*)	–
75	China: Harbin	N45.750 E126.633	140	KX665031-32 (as *F. candida*)	–
76	China: Xilin Gol League, Inner Mongolia	N43.930 E116.090	1040	KX665022, KX665024 (as *F. candida*)	–
77	China: Ulanqab, Inner Mongolia	N40.994 E113.132	1370	KX665023, KX665025 (as *F. candida*)	–
78	China: Helan Mountains, Inner Mongolia	N38.900 E105.967	3260	KX665019-21 (as *F. candida*)	–
79	China: Hebei, Qinglong	N40.407 E118.950	245	KX665028 (as *F. candida*)	–
80	China: Hebei, Xiaowutai Mountain	N39.941 E115.043	2840	KX665029-30 (as *F. candida*)	–
81	China: Hebei	–	–	HQ619704, HQ619710 (as *F. candida*)	HQ651081, HQ651074 (as *F. candida*)
82	China: Zhong-Tiao Mountains	N35.116 E111.772	520	KX665033 (as *F. candida*)	–
83	China: Liupan Mountains, Ningxia	N35.657 E106.205	1300	KX665034 (as *F. candida*)	–
		Formica candida			
84	Kyrgyzstan: Alai Valley	N39.703 E73.457	3200	–	AY786154*, AY78615, JX170887
85	Kyrgyzstan: Tian-Shan	–	–	–	AY786155–AY786156

Note. Loc. – location numbers referring to Fig. 2; altitude provided in m above sea level. Asterisks denote type of material/locations. Specimens sequenced in this study are indicated in boldface.

cytb sequence for the *F. candida* neotype (AY786154). The haplotype network was constructed using the median joining algorithm in Pop ART v. 1.7 (Leigh, Bryant, 2015).

Results

Out of the 47 ant specimens sequenced in this study for *cox1* and *cytb*, 6 belonged to *F. picea*, and 41, to *Formica* sp. We did not detect haplotypes of *F. candida* (Seifert, 2004) in any of the sampling points, despite the extensive survey area. *F. picea* and *Formica* sp. significantly differed at the DNA level (Fig. 4).

For *F. picea*, a notable geographic split was evident: samples from Europe and West Siberia formed a tight cluster, while samples from Armenia and Transbaikalia diverged from them (Fig. 4). For *Formica* sp., the larger sample size resulted in a more informative network. However, no regions with a specific set of haplotypes were identified. Genetic diversity was substantial, especially for three samples (from different regions) that differed markedly from other haplotypes.

F. picea was found in five locations east of the Urals: in the Altai mountains, in the lowlands of West Siberia, as well as in East Siberia. The West Siberian location (Melnikovo, Tomsk oblast) was in a fen with a tall birch. The nest (with no external structures) was located in a sedge hummock with a fern; nest entry was in moss.

In the Altai mountains, we found nests of *F. picea*: three in Central Altai near the Abay (locations No. 17 in the Table and Fig. 2) and Amur (locations No. 18–19 in the Table and



Fig. 3. Habitats of *Formica* sp. and *F. picea*.

a – Central Altai, near Abay village, lowland swamp (*F. picea*); *b* – the same area; meadows on the site of drained swamps (co-occurrence of *Formica* sp. and *F. picea*); *c* – the same area; stony steppes on rocky outcrops (*F. picea*); *d* – Central Altai, near Ust-Kan village, steppe (*Formica* sp); *e* – Magadan oblast, Naledny River (tributary of the Ola River) (*Formica* sp.); *f* – Zabaikalsky Krai, floodplain of the Talovka River (*Formica* sp.); *g* – Sakha Republic (Yakutia), Momsky District, Khairdakh River (*Formica* sp.); *h* – Magadan oblast, upper reaches of the Kolyma River, steppe (*Formica* sp.); *i* – Magadan oblast, an island in the upper reaches of the Kolyma River near the eastern end of the Big Ananchag Range (*Formica* sp.); *j* – same area; nest exits of *Formica* sp.; *k* – Zabaikalsky Krai, floodplain of the Kotikha River (*F. picea*); *l* – Novosibirsk oblast, Kabinetnoye village, halophytic meadow (*Formica* sp.); *m* – Southeast Altai, near Kosh-Agach village, Tabozhok town, stony steppe (*F. picea*). For more details, see the Table.

Fig. 2) villages (ca. 1100 m a. s. l.), and one in the Southeast Altai (locations No. 20 in the Table and Fig. 2), near the Kosh-Agach village (2275 m a. s. l.). Near the Abay village, a single nest was detected in a damp area on the edge of a bog, where it transitioned into a steppe meadow, 150–200 m away from a small lake. The nest had a ground part in the form of a small dome made of thin blades of grass, no more than 10 cm high and about 15 cm in diameter. Two other nests were found on southern steppe slopes, with no external structures and exits located under the rocks.

In East Siberia, *F. picea* was detected in Zabaykalsky Krai, east of the Gazimursky Zavod town (locations No. 21 in the Table and Fig. 2), at 770 m a. s. l., within the forest belt. The nest was found on a swampy floodplain of the Kotikha River (a third-order tributary of the Gazimur River), with saplings of alder, birch, and larch. The nest, without visible external structures, was located under a layer of litter composed of fallen leaves and dry grass.

Specimens of *Formica* sp. were detected in 61 locations (Fig. 2). Their habitats included floodplain (including long-term and deeply flooded sandy-pebble spits) and non-floodplain terraces of different heights on rivers of various sizes, meadows of diverse composition and projective cover, sparse forests, steppes (from mixed grass to desert), foothills and highlands of Altai. This ant was found to avoid forests with closed canopies, as well as wetlands.

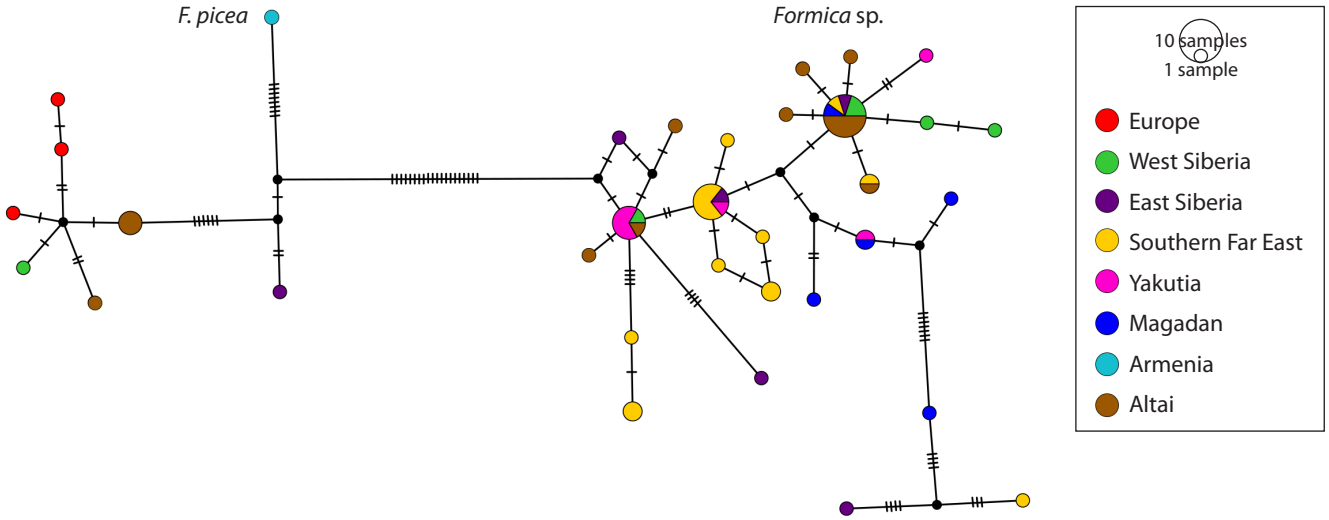


Fig. 4. Haplotype network for the combined *cox1+cytb* sequences.
Dashes represent the number of nucleotide substitutions.

Discussion

Genetic patterns

As in the previous study (Zhigulskaya et al., 2022), *F. picea* and *Formica* sp. were significantly different on mtDNA levels, with no intermediate haplotypes found (Fig. 4). Expectedly, we found no haplotypes of *F. candida* sensu Seifert (2004), which is probably limited to certain regions of Kyrgyzstan.

Throughout the distribution of *Formica* sp., from the Altai mountains to the Pacific coast and eastern China, there is a pretty high genetic variation, but with no obvious genetic structure (Fig. 4). This probably indicates recent colonization of this region without a bottleneck event and high dispersal abilities of the species. Unfortunately, there are no *cox1* data on *Formica* sp. from China, a region with a milder climate that could serve as a refugium/ancestral region of *Formica* sp.

Noteworthy, three haplotypes differed by 8–15 nucleotide substitutions from the main cluster (Fig. 4). However, all three were from different and geographically remote regions. This might imply that the history of the species is complex, with high dispersal rate and multiple colonization waves.

Although the data on the *cox1* gene in *F. picea* remain limited, several interesting insights can be gleaned from the haplotype network (Fig. 4). Most haplotypes form a tight cluster, indicating close genetic relationships. This cluster comprises specimens from the Altai and West Siberia, suggesting recent (likely Holocene) dispersal into these regions. The Armenian haplotype is distantly related to this group, as expected given the geographic distance and mountain barriers. Notably, the East Siberian haplotype is also distinct from the European cluster. Intriguingly, this pattern may indicate that East Siberian populations of *F. picea* belong to a different genetic lineage relative to the European/West Siberian group, and that an East Siberian refugium for the species may have existed. Of course, drawing firm conclusions based on a single specimen would be premature; thus, we intend to examine additional specimens from the region in future studies.

Geographic distribution

There are a lot of studies on the distribution and habitat preferences of the *F. picea/F. candida* complex. In various parts of its range, these ants are known to inhabit diverse habitats: fens, meadows, steppes, as well as sparse forests on plains, floodplains, and in highlands. In the south taiga zone of Siberia, black bog ants are found in swamps, in wet mixed forests, in pine woods (Omelchenko, Zhigulskaya, 1998; Dlussky, 2001; Gridina, 2003). They are ubiquitous in Central Yakutia and the Baikal Region, from boggy larch forests, dry pine-larch forests, and steppe-like and steppe biotopes to bare rock formations, with nest densities reaching up to 30 nests per 100 square meters (Pleshanov, 1966; Dmitrienko, Petrenko, 1976).

However, they are particularly ecologically versatile and abundant in the mountain-steppe landscapes of Tuva, Southeast and Central Altai, as well as in the steppes of Transbaikalia, where they dominate almost all habitats, from solonetzes and solonchaks to dry steppes on sandy and stony soils (Zhigulskaya, 1968, 2009, 2011; Chesnokova, Omelchenko, 2011).

In Northeast Asia, the complex is associated primarily with river valleys; high abundance is observed in sparse *Populus* and *Chosenia* groves and on gravel-sandy spits in floodplains. Much less frequently, and only as isolated nests, *Formica* sp. can be found in mesoxeromorphic sparse forests (in Yakutia, specifically in pine forests), in steppe habitats, or southern slopes with sphagnum-moss bogs (Berman et al., 2007).

Since the black bog ant turned out to include several cryptic taxa, this raises a question if they have different habitat preferences. Here we obtained certain data on this point, which are, of course, far from exhaustive. For *F. picea*, we currently know that it is found from West Europe to West Siberia. We also have single findings from Transbaikalia (this study) and Tibet (AY786157), implying the presence of this taxon between Siberia and Tibet. *F. picea* appears to be rare in Asia, since we found only six locations of this species compared to 41 for *Formica* sp. Of these, only two were not from the mountains: a location in Tomsk oblast and one from mid-altitude highlands

of Transbaikalia. Based on our findings, we can suggest that *F. picea* appears to be associated with fens, but not peatlands. The latter is probably due to the fact that peatlands in Siberia are still insufficiently studied by myrmecologists.

Formica sp., on the other hand, appears to be found even more widely than previously thought (Zhigulskaya et al., 2022). In this study it was recorded across a vast area, from West Siberia to the Pacific Ocean, and from 66.5°N in Yakutia (Khayyrdakh River) roughly to 35°N in the southern part of the Korean Peninsula (Fig. 2), as well as from sea level up to approximately 3500 m a. s. l., across a broad range of habitats (except for forests with closed canopy). Furthermore, unlike *F. picea*, *Formica* sp. has not been found in bogs in Europe, Siberia, or Altai. Conversely, only *Formica* sp. is known from river floodplains, and from these areas, it can spread to xeromorphic slopes and sections of low floodplain terraces.

The versatility of *Formica* sp. to moisture is striking: it can colonize both long-term flooded habitats on floodplains (that is, for about a month of the 4.5-month active period in the Subarctic) and dry zonal steppes. Notably, in the Altai Mountains, *Formica* sp. nests were found in both types of habitats.

In this study, we found that the distributions of *Formica* sp. and *F. picea* significantly overlap, probably from the Urals to Transbaikalia (Fig. 2), and we expect this overlap to expand when new data on bog ants are obtained. However, the ecological preferences of these two taxa seem to differ, which can be prospectively formulated as follows: *Formica* sp. avoids fens, while *F. picea* does not like low floodplains.

In this connection, the cases of syntopic occurrences of both taxa in the Central Altai are of particular note. Here, these two nests were found a few dozens of meters apart on the floodplain of the Abay River. The nest of *F. picea* was at the edge of a wet swampy depression about 1.5 m deep, while that of *Formica* sp. was on a dry steppe meadow that was formed in a dried fen. In the same location, about 3 km apart, we found a *F. picea* nest on a southern slope of the residual hills with rock outcrops, occupied by steppes.

We can suggest that in the Altai, *F. picea* inhabits meadows, and can move into both wet habitats and moderately dry steppes on southern slopes, as well as mountain steppes above the forest edge. However, it is unclear why this species can use both very wet and very dry habitats (Bönner, 1915; Skwarra, 1929; Stitz et al., 1939; Seifert, 2004, 2018).

To this, we can put forward the hypothesis that the generalist *F. picea* has to constrict itself to the habitats where it can compete with other ant species (Dlussky, 1967). "Apparently, while spreading westward, *F. picea* (a eurytopic forest-steppe species of Mongolian origin) was displaced by species of the *F. rufibarbis* group into bogs. In these bogs, the *F. rufibarbis* species are virtually absent, whereas *F. picea* can survive, owing to its eurytopy and plasticity in nest construction. In the mountains, this species inhabits higher altitudes than other species of the *F. rufibarbis* group (except for *F. subpilosa pamirica* and *litoralis* in Central Asia, which exclusively occupy river thalwegs, rocky screes, and sands). *F. picea* cannot compete with *F. fusca* and its closely related species, as it is virtually absent in forests; instead, it is associated with more or less open habitats (meadows, steppified areas, sparse forests, river floodplains, and bogs)" (translated from: Dlussky, 1967, p. 63).

If this idea is correct, *Formica* sp. outcompetes *F. picea* on the vast floodplains and xeromorphic habitats of Asian Russia. So, *F. picea* is outcompeted both in Siberia and in Europe by different local ant species.

Conclusions

We can summarize the following important points inferred from our study: only *F. picea* is found in Europe, whereas both *F. picea* and *Formica* sp. are found in North Asia. We failed to detect *F. candida* on this vast territory, and we can surmise that in the current understanding proposed by Seifert (2004), it is a Central Asian taxon. In the south of West and East Siberia, the ranges of *F. picea* and *Formica* sp. overlap, but they have different ecological preferences. *F. picea* is probably constrained to fens on the plains, and to subalpine meadows in the mountains. *Formica* sp. is broadly tolerant to humidity but is not found in fens. The cases of syntopic distribution of these ants, which we described in Altai, are apparently associated with transient successional processes. The relative rarity of *F. picea* in our collections possibly reflects the insufficient study of raised bogs in Asia, the typical habitats of the species in Europe. Differences in the distribution and habitat preference of *F. picea* and *Formica* sp. are probably the result of competition, according to G.M. Dlussky (1967).

F. picea and *Formica* sp. differ genetically and have different distribution ranges, as well as habitat preferences. This supports the opinion that *Formica* sp. should be recognized as a distinct species.

References

Antonov I.A., Pleshanov A.S. Landscape ecological complexes of ants in Baikalian Siberia. *Sibirskij Ekologicheskij Zhurnal* (*Siberian Journal of Ecology*). 2008;15(1):53-57 (in Russian)

Antonov I.A., Bukin Y.S. Molecular phylogenetic analysis of the ant genus *Formica* L. (Hymenoptera: Formicidae) from Palearctic region. *Russ J Genet.* 2016;52(8):810-820. doi 10.1134/S1022795416080020

Berman D.I., Alfimov A.V., Zhigulskaya Z.A., Leirikh A.N. Wintering and Cold Hardiness of Ants in the North-East of Asia. Moscow, 2007 (in Russian)

Bolton B. A New General Catalogue of the Ants of the World. Cambridge, Mass.: Harvard University Press, 1995

Bönner W. Die Überwinterung von *Formica picea* und andere biologische Beobachtungen. *Biol Zent Bl.* 1915;35:65-77 (in German)

Chen Y., Zhou S. Phylogenetic relationships based on DNA barcoding among 16 species of the ant genus *Formica* (Hymenoptera: Formicidae) from China. *J Insect Sci.* 2017;17(6):117. doi 10.1093/jisesa/iex092

Chesnokova S.V., Omelchenko L.V. Ants of Central Altai: Spatial-typological structure and classification of communities. *Entomol Rev.* 2011;91(2):253-263. doi 10.1134/S0013873811020138

Dlussky G.M. Ants of the Genus *Formica* (Hymenoptera: Formicidae g. *Formica*). Moscow, 1967 (in Russian)

Dlussky G.M. Seasonal dynamics of *Formica candida* colonies in an isolated bog population. In: Ants and Protection of the Forest. Proceedings of the XI All-Russian Myrmecological Symposium. Perm, 2001;69-71 (in Russian)

Dmitrienko V.K., Petrenko E.S. Ants of Taiga Biocenoses of Siberia. Novosibirsk, 1976 (in Russian)

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from

diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*. 1994; 3(5):294-299

Goropashnaya A. Phylogeographic Structure and Genetic Variation in *Formica* Ants. Upsala, 2003

Goropashnaya A.V., Fedorov V.B., Seifert B., Pamilo P. Phylogenetic relationships of palaearctic *Formica* species (Hymenoptera, Formicidae) based on mitochondrial cytochrome *b* sequences. *PLoS One*. 2012;7(7):e41697. doi 10.1371/journal.pone.0041697

Gridina T.I. Ants of the Urals and their geographical distribution. *Uspekhi Sovremennoi Biologii = Advances in Current Biology.* 2003; 123(3):289-298 (in Russian)

Kupyanskaya A.N. Ants (Hymenoptera, Formicidae) of the Far East. Vladivostok, 1990 (in Russian)

Leigh J.W., Bryant D. POPART: full-feature software for haplotype network construction. *Methods Ecol Evol.* 2015;6(9):1110-1116. doi 10.1111/2041-210X.12410

Omelchenko L.V., Zhigulskaya Z.A. Ecological-faunistic analysis of ants in the southern taiga of western and central Siberia. *Uspekhi Sovremennoi Biologii = Advances in Current Biology.* 1998;118(3): 299-305 (in Russian)

Pleshanov A.S. Family Formicidae (Ants). In: Pests of Siberian Larch. Moscow, 1966;194-302 (in Russian)

Radchenko A.G. Ants (Hymenoptera, Formicidae) of Ukraine. Kiev, 2016 (in Russian)

Ruzsky M.D. Ants of Russia. Part 1. *Trudy Obshchestva Yestestvoispytatelej pri Imperatorskom Kazanskom Universitete = Proceedings of the Naturalist Society through the Kazan Imperial University.* 1905; 38(5-7):3-798 (in Russian)

Schär S., Talavera G., Espadaler X., Rana J.D., Andersen A.A., Cover S.P., Vila R. Do Holarctic ant species exist? Trans-Beringian dispersal and homoplasy in the Formicidae. *J Biogeogr.* 2018;45(8): 1917-1928. doi 10.1111/jbi.13380

Shekhovtsov S.V., Derzhinsky Y.A., Poluboyarova T.V., Golovanova E.V., Peltek S.E. Phylogeography and genetic lineages of *Aporrectodea rosea* (Lumbricidae, Annelida). *Eur J Soil Biol*. 2020;99: 103191. doi 10.1016/j.ejsobi.2020.103191

Seifert B. The "Black Bog Ant" *Formica picea* Nylander, 1846 – a species different from *Formica candida* Smith, 1878 (Hymenoptera: Formicidae). *Myrmecologische Nachrichten*. 2004;6:29-38

Seifert B. The Ants of Central and North Europe. Lutra Verlags – und Vertriebsgesellschaft. Tauer, Germany, 2018

Skwarra E. *Formica fusca-picea* Nyl. als Moorameise Biologische. *Anzeiger.* 1929;82:46-55 (in German)

Stitz H. Hautflügler oder Hymenoptera: Ameisen oder Formicidae. In: Die Tierwelt Deutschlands. Jena, 1939 (in German)

Zakharov A.A. Ants of Forest Communities, their Life and Role in the Forest. Moscow, 2015 (in Russian)

Zakharov A.A., Dlussky G.M., Goryunov D.N., Gilyov A.V., Zryanin V.A., Fedoseeva E.B., Gorokhovskaya E.A., Radchenko A.G. Monitoring of the *Formica* Ants. Moscow, 2019 (in Russian)

Zhigulskaya Z.A. Ant (Formicidae) population of steppe Tuva landscapes. In: Animal Population of Soils in Treeless Biogeocenoses of the Altai-Sayan Mountain System. Novosibirsk, 1968;115-139 (in Russian)

Zhigulskaya Z.A. The ants of the Chuya depression and the Yustyd River basin in Southeastern Altai. *Contemp Probl Ecol.* 2009;2(3): 210-215. doi 10.1134/S1995425509030084

Zhigulskaya Z.A. Ants (Hymenoptera, Formicidae) of the Kuraiskaya depression of the Southeastern Altai. *Tomsk State University Journal.* 2011;350:189-195 (in Russian)

Zhigulskaya Z.A., Shekhovtsov S.V., Poluboyarova T.V., Berman D.I. *Formica picea* and *F. candida* (Hymenoptera: Formicidae): synonyms or two species? *Diversity*. 2022;14(8):613. doi 10.3390/d14080613

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

Received August 29, 2025. Revised September 23, 2025. Accepted October 6, 2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):284-292

doi 10.18699/vjgb-26-32

Функциональная симметрия и воспроизводимость эволюционного процесса

С.И. Барцев [1, 2]

[1] Институт биофизики Сибирского отделения Российской академии наук – обособленное подразделение Федерального исследовательского центра «Красноярский научный центр СО РАН», Красноярск, Россия
[2] Сибирский федеральный университет, Красноярск, Россия
bartsev@yandex.ru

Аннотация. Вопрос о воспроизводимости эволюционных процессов имеет в первую очередь фундаментальное значение, однако с развитием методов моделирования эволюционных процессов на компьютерных многоуровневых моделях ответ на этот вопрос необходим для прояснения статуса получаемых прогнозов. Экспериментальное получение ансамблей эволюционных исходов для последующей статистической обработки на реальных биологических системах представляется неосуществимым. В то же время прогнозы, сгенерированные многоуровневыми компьютерными моделями, вследствие их сложности и зависимости результатов моделирования от множества параметров с трудом поддаются интерпретации. Данная работа направлена на выявление общих свойств эволюционирующих систем с помощью простой эвристической модели, построенной на прозрачных общих принципах и представлениях о ключевых свойствах биологических систем, значимых для эволюционного процесса. Агенты, претерпевающие эволюционные изменения, являются рекуррентными нейронными сетями с четко определенной структурой, заданной функцией и определенным правилом модификации структуры в направлении максимальной приспособленности. Отдельный экземпляр нейронной сети, формируемой в ходе эволюционного процесса, назван нейросетевым модельным объектом (НМО). В работе проведены вычислительные эксперименты по генерации ансамблей структур НМО, выполняющих заданную функцию, и проанализированы закономерности распределения НМО в структурном пространстве. Этот анализ подтверждает наличие функциональной симметрии структуры НМО, выполняющих одну и ту же функцию. Оценены устойчивость и воспроизводимость индивидуальных эволюционных траекторий. Показано, что при определенных ограничениях, приводящих к редукции сложности структуры НМО (аналог – узкая экологическая специализация), финальные структуры НМО могут быть близки, но не идентичны. Это позволяет говорить о неточном воспроизведении эволюции структуры на фоне функциональной эквифинальности. Тем не менее можно утверждать, что в общем случае сама способность к эволюционным изменениям реализуется при избыточности потенциальной сложности структуры над функциональной сложностью и автоматически влечет за собой множественность эволюционных исходов, основанную на том, что одна и та же функция может реализовываться различными, но функционально инвариантными структурами.

Ключевые слова: воспроизводимость эволюционного процесса; эквифинальность эволюционных исходов; функциональная симметрия; эвристическая нейросетевая модель; функциональная сложность

Для цитирования: Барцев С.И. Функциональная симметрия и воспроизводимость эволюционного процесса. *Вавиловский журнал генетики и селекции*. 2026;30(2):284-292. doi 10.18699/vjgb-26-32

Финансирование. Исследование выполнено в рамках государственного задания Министерства науки и высшего образования РФ, проект № 0287-2021-0018.

Functional symmetry and reproducibility of the evolutionary process

S.I. Bartsev [1, 2]

[1] Institute of Biophysics of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia
[2] Siberian Federal University, Krasnoyarsk, Russia
bartsev@yandex.ru

Abstract. The question on the reproducibility of evolutionary processes is primarily of fundamental importance; however, with the development of methods for modeling evolutionary processes on computer multilevel models, an answer to this question is necessary to clarify the status of the predictions obtained. Experimental obtaining of ensembles of evolutionary outcomes for subsequent statistical processing on real biological systems seems to be impracticable. At the same time, the results obtained on multilevel computer models are difficult to interpret due to their complexity and the dependence of modeling results on a variety of parameters. This work is aimed at identifying common properties of evolving systems using a simple heuristic model based on transparent general principles and ideas about the key properties of biological systems

© Барцев С.И., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

that are important for the evolutionary process. Agents undergoing evolutionary changes are recurrent neural networks with a well-defined structure, a given function, and a specific rule for modifying the structure in the direction of maximum fitness. A separate instance of a neural network formed during the evolutionary process is called neural network model object (NNMO). Computational experiments have been carried out to generate ensembles of NNMO structures performing a given function, and the patterns of NNMO distribution in the structural space have been analyzed. This analysis confirms the presence of functional symmetry in the structure of NNMOs performing the same function. An assessment of the stability and reproducibility of individual evolutionary trajectories has been carried out. It is shown that under certain constraints leading to a reduction of the complexity of the NNMO structure (analogous to a narrow environmental specialization), the final NNMO structures may be close, but not identical. This suggests an inaccurate reproduction of the evolution of the structure with functional equivalence. Nevertheless, it can be argued that in the general case, the very ability for evolutionary change is possible with the redundancy of the potential complexity of the structure over the functional complexity and automatically entails a multiplicity of evolutionary outcomes based on the fact that the same function can be implemented by different, but functionally invariant structures.

Key words: reproducibility of the evolutionary process; equifinality of evolutionary outcomes; functional symmetry; heuristic neural network model; functional complexity

For citation: Bartsev S.I. Functional symmetry and reproducibility of the evolutionary process. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):284-292. doi 10.18699/vjgb-26-32

Истинно ищущий не знает, что он ищет,
истинно странствующий не знает, куда он идет.
Приписывается Лао Цзы

Введение

Степень нашего непонимания живого видна по обилию связанных с ним проблем. Одной из ключевых научных проблем биологии является проблема предсказуемости и/или возможной эквифинальности биологической эволюции.

По мнению Эйгена, «каждая отдельная система, возникшая в результате мутаций и отбора, непредсказуема в отношении своей структуры; тем не менее неизбежным результатом всегда является процесс эволюции – это закон. …Оптимизирующий процесс эволюции в принципе неизбежен, хотя выбор конкретного пути не детерминирован» (Эйген, 1973, с. 207).

В то же время известен определенный параллелизм в эволюции различных видов и родов, нашедший свое выражение в законе гомологических рядов в наследственной изменчивости Н.И. Вавилова (Мейен и др., 1977). Если существует некоторая структурная общность у систем с одной и той же функцией, но сформированных на разных эволюционных траекториях, то можно говорить об эквифинальности (Мейен, 1974; Мейен и др., 1977) эволюции в каком-то определенном смысле.

Вопрос о том, идет ли эволюция по детерминированному пути или может идти по альтернативным траекториям, привлекает многих исследователей (Povolotskaya, Kondrashov, 2010; Lobkovsky et al., 2011; Orgogozo, 2015; Xue et al., 2017). В настоящее время существуют противоречия между некоторыми теоретическими концепциями и экспериментальными данными по вопросу воспроизводимости эволюционных траекторий.

Например, в статье (Lobkovsky, Koonin, 2012) отмечается, что при высокой интенсивности мутаций эволюция может идти по разным траекториям на ландшафте приспособленности, но в конечном счете все они сходятся на единственном пике, соответствующем наиболее приспособленной структуре эволюционирующей биологической системы. А в работе (Orgogozo, 2015) рассматривается возможность по фенотипу прогнозировать генотип, т. е. автор допускает существование взаимно-однозначного отображения генотипа в фенотип, что означает единственность эволюционного исхода, т. е. одинаковость генотипа при одинаковом фенотипе.

Есть эксперименты в пользу единственности эволюционных исходов (Weinreigh et al., 2006; Dickins, Nekrutenko, 2009; Meyer et al., 2012), но есть данные, демонстрирующие существование нескольких конечных точек эволюции (Poelwijk et al., 2007; Dunham et al., 2009; Kvitek, Sherlock, 2011; Podgornaia, Laub, 2015; Starr et al., 2017).

Сложность исследований в этой области связана с тем, что прямое изучение эволюционных процессов биологических систем в природе натыкается на три основных препятствия: 1) единственность эволюционных исходов, которая не позволяет применять сравнительный анализ биологических структур; 2) характерное время эволюционных изменений, которое обычно многократно превышает длительность жизни ученого; 3) чрезвычайную сложность реальных биологических объектов.

Возможный подход к преодолению заявленных препятствий напрашивается сразу. Это математическое моделирование и его особая разновидность – компьютерное моделирование.

Компьютерное моделирование позволяет преодолеть первое и второе из указанных препятствий – исследователь может получить ансамбли модельных эволюционных траекторий, к которым можно применять статистическую обработку данных для выявления общих закономерностей эволюционных процессов. Но статус результатов компьютерного моделирования зависит от того, насколько удалось справиться со сложностью биологической системы

и обеспечить адекватность ее описания. Можно выделить два основных препятствия.

Во-первых, при моделировании реальных событий нужны модели, максимально приближенные по сложности к реальным системам, но здесь мы сталкиваемся с «проклятием размерности» – количество подгоночных параметров в условиях неопределенности точного вида функциональных зависимостей делает невозможным их определение вследствие дефицита эволюционных данных.

Во-вторых, существуют сомнения (Beckage et al., 2011; Garte et al., 2025) в возможности построения адекватных математических моделей биологических систем вообще и эволюционных моделей в частности. Авторы (Garte et al., 2025) отмечают, что «биологические концепции не поддаются четким определениям, поддающимся математической обработке: жизнь, приспособленность, наследственность, вид, ген, интеллект, чувственность», и на этом основании утверждают, что очередной «парадигмальный сдвиг потребует не только новых математических инструментов, но и новой научной ЭПИСТЕМОЛОГИИ», т. е. способов представления знаний. При этом, по их мнению, «выбор того, на какие данные обратить внимание, гораздо ценнее, чем сбор больших объемов данных».

Один из путей разрешения проблемы эквифинальности состоит в тщательном анализе имеющихся в печати теоретических аргументов и экспериментальных данных в пользу той или иной позиции.

Другой путь заключается в получении ожидаемых и неожиданных ответов на основе развития начальных принципов и общих представлений об эволюции живого. Наше исследование опирается на ранее опубликованные результаты (Барцев, Барцева, 2002, 2005, 2010; Bartsev, Baturina, 2019).

Методы и материалы

Поскольку подход к изучению свойств биологической эволюции, используемый в данной статье, достаточно необычен с точки зрения традиционной молекулярной биологии, то представляется необходимым остановиться на методологических основаниях предлагаемого подхода.

Модели и модельные объекты являются инструментами научного исследования начиная с Галилея, который, технически не имея возможности непосредственно изучать свободное падение тел, вывел формулы кинематики равноускоренного движения, используя физическую модель – скатывание шара с наклонного желоба. В дальнейшем нам пригодится следующее определение: «Если между двумя объектами может быть установлено сходство хотя бы в каком-либо одном определенном смысле, то между этими объектами существуют отношения оригинала и модели» (Лернер, 1967, с. 40).

В работе будут использованы эвристические модели (фон Нейман, 1971), которые, в отличие от традиционных моделей, направлены не на моделирование (описание) конкретных реальных систем, а представляют собой абстрактные модели, предназначенные для выявления удобных понятий, широко приложимых принципов и построения общей теории.

Полезность перехода к эвристическим модельным объектам фон Нейман обосновал так: «Поскольку у нас нет достаточно ясного представления о том, как функционируют живые организмы, то обращение к органике большой пользы нам не принесет. Мы займемся поэтому автоматами, которые мы в совершенстве знаем, ибо мы их сделали» (фон Нейман, 1971, с. 98).

Некоторые способы постановки задачи при таком подходе можно выделить из рассуждений фон Неймана: «В теории автоматов устанавливаются общие принципы организации, структуры, языка, информации и управления. Многие из них можно применить как к естественным, так и к искусственным системам, и поэтому сравнительное изучение этих двух типов автоматов может служить хорошей отправной точкой: необходимо описать и объяснить черты их сходства и различия, развить математические методы, применимые к обоим типам автоматов. …Тогда можно надеяться получить ответ или хотя бы попытаться исследовать такие вопросы: по каким принципам составляются из этих элементарных частей функционирующие организмы, каковы черты таких организмов и каковы их основные количественные характеристики?» (фон Нейман, 1971, с. 99).

Требования к эвристической модели эволюционирующей биологической системы давно сформулировал Дж. Бернал: «Биология методологически отличается от других естественных наук тем, что в фокусе внимания находятся, прежде всего, ***функционирование*** и эволюция систем. Структура здесь имеет значение только в связи с функцией и происхождением…» (Бернал, 1968, с. 112).

Выделение Берналом функционирования как особой характеристики живого согласуется с подходом Н. Рашевского и Р. Розена, которые рассматривали организм как «набор или систему функциональных отображений» (Rosen, 1958; Рашевский, 1968, с. 63).

Практически любое исследование живых объектов представляет собой сначала выявление функциональных закономерностей типа: «стимул–реакция» или «воздействие–отклик», а затем определение «механизмов», обеспечивающих реализацию данной функции (катализа, наследования, локомоции, распознавания и т. п.). Механизм функционирования системы считается раскрытым, если дано его описание в терминах взаимодействия частей, составляющих эту систему.

Формально выявлению механизма соответствует разложение (декомпозиция) исходной функции $F(x)$ на более простые, базовые функции $\{g_i\}$, реализуемые соответствующими подсистемами $\{S_i\}$ – элементами структуры. Коэффициенты связи $\{\alpha_{ij}\}$ описывают взаимодействие (отношения) между элементами структуры и собственно задают структуру системы.

Важно подчеркнуть, что процедура декомпозиции соответствует сетевому описанию, где $\{S_i\}$ – узлы сети, а $\{\alpha_{ij}\}$ – коэффициенты связи между узлами. Сетевое описание уже очень давно широко используется в различных областях биологии: от описания корреляционного соотношения структура-свойство химического соединения (Голованов и др., 1998) до метаболических путей, белок-белковых

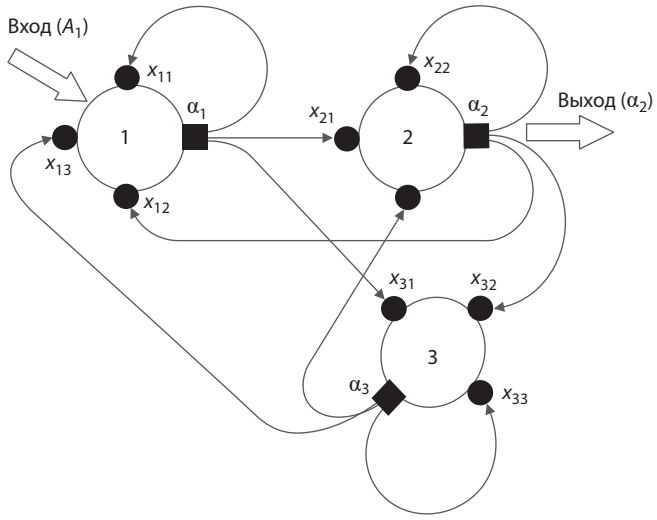


Рис. 1. Полносвязная нейронная сеть из трех нейронов: 1 – входной нейрон; 2 – выходной; 3 – ассоциативный (скрытый) нейрон, который не связан непосредственно со входом и выходом.

взаимодействий, генетических сетей, нервной системы, пищевых цепей экосистем и др. (Albert et al., 2000; Amaral et al., 2000; Edelman, Gally, 2001; Strogatz, 2001; Dunne et al., 2002; Sole, 2002; Stumpf et al., 2008; De Las Rivas, Fontanillo, 2010).

Для описания биологической системы сетевая модель должна обладать способностью к эволюции. Под эволюционным процессом будет пониматься изменение структуры биологической системы в соответствии с некоторым функциональным критерием оптимальности (Эйген, 1973; Forst et al., 1995; Schuster, 1996).

Анализ существующих формальных моделей показывает, что наиболее адекватными для целей данного исследования являются сети из формальных нейронов, так называемые нейросети (НС). Достоинство НС как эвристической модели эволюции состоит в том, что для нейронных сетей легко определить ФУНКЦИЮ, выделить СТРУКТУРУ и запустить процесс обучения, формальное описание которого по ключевым признакам не отличается от процесса биологической ЭВОЛЮЦИИ.

В настоящей работе использованы полносвязные рекуррентные НС, которые могут функционировать в потоке дискретных событий. В качестве иллюстрации на рис. 1 приведен орграф простейшей НС из трех нейронов, описывающий пути поступления и обмена сигналами.

Формулы, описывающие функционирование НС, имеют вид:

$$\alpha_i^{k+1} = \frac{\rho_i^k}{a + |\rho_i^k|}, \quad \rho_i^k = \sum_j x_{ij}\,\alpha_j^k + A_j^k,$$

где α_i^k – выходной сигнал i-го нейрона в k-й момент времени; ρ_i^k – взвешенная сумма выходных сигналов нейронов, поступивших на i-й нейрон в k-й момент времени, плюс внешний входной сигнал A_i^k, поступивший на i-й входной нейрон в k-й момент времени; x_{ij} – матрица весовых коэффициентов (синапсов); a – коэффициент, задающий крутизну функции активации.

Из формулы можно видеть, что выходной сигнал каждого нейрона (уровень его возбуждения) в каждый момент времени зависит от состояния НС (и внешнего сигнала для входных нейронов) в предыдущий момент времени.

Модификация структуры НС осуществлялась с помощью алгоритма случайного поиска, который реализовывался следующим образом: 1) каждому синапсу текущей НС добавлялось случайное приращение, причем сразу создавались две модифицированные НС, различающиеся тем, что одинаковые по модулю случайные приращения прибавлялись к соответствующим синапсам с разными знаками; 2) все три НС – исходная и два ее модифицированных варианта – запускались на функционирование в потоке событий длительностью 100 символов, и для каждой НС оценивалась суммарная по этому фрагменту целевая функция (функция потерь) – сумма квадратов разностей между выходами НС и требуемым откликом; 3) по результатам сравнения отбиралась НС с минимальной целевой функцией, которая сравнивалась с требуемым уровнем точности, и в случае превышения осуществлялся переход к пункту 1. Формально эта процедура соответствует случайным мутационным заменам.

Чтобы НС могла выступать в качестве модели эволюционного процесса, необходимо четко прописать, в чем имеется сходство между ней и биологическим видом в процессе эволюции. Ключевыми компонентами или составляющими эволюционного процесса являются наследственность, изменчивость и естественный отбор, т. е. выживание наиболее приспособленных. Меру приспособленности для биологических систем сложно вывести (Garte et al., 2025), но очевидно, что в этой мере должны учитываться способность к размножению и успешность конкуренции за ресурсы.

В ходе естественного отбора выживают наиболее приспособленные, поэтому для моделирования отбора необходима мера приспособленности того или иного варианта структуры НС. Для нейронных сетей в качестве меры приспособленности выступает целевая функция или функция потерь, которая и определяет, какой вариант структуры НС будет продолжать эволюцию (обучение). То есть целевая функция в интегральном виде включает в себя оценку успешности в размножении и конкуренции за ресурсы. Носителем наследственности в нашем случае является структура межнейронных связей, которая наследуется следующим поколением НС. Изменчивость структуры обеспечивается случайными приращениями синапсов, что приводит к вариации качества функционирования НС, т. е. изменению ее приспособленности к среде, в которой она обучается. Мы рассматриваем эволюционный процесс в принципе, а особенности реализации в материале – это нюансы.

При обучении (эволюции) и в процессе функционирования на вход НС подавалась непрерывная, квазислучайная последовательность, состоящая из сигналов: «пауза» – (00), A – (10), B – (01) и C – (11), где числа в скобках обозначают наличие или отсутствие сигнала на соответ-

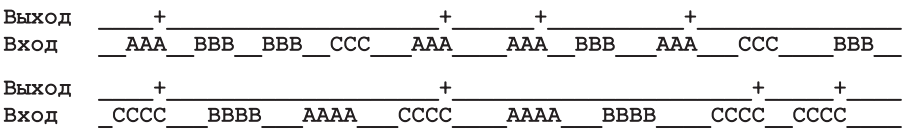


Рис. 2. Фрагменты входной и выходной последовательностей для функций А3 и С4.
Знак «+» в выходной последовательности обозначает «правильное» срабатывание НМО.

ствующих входных нейронах. Случайными параметрами последовательности были: вид сигнала, который поступит на вход, и длительность паузы между сигналами. Согласно обучающему паттерну, нейросеть должна выдавать на выходе единицу, если на вход поступает «правильный» сигнал, и ноль – в противном случае. Последовательности могут различаться количеством тактов времени, предоставляемых нейросети для обработки входного сигнала (3 или 4), что указывается в обозначении функции (рис. 2).

Несмотря на простоту, подобные последовательности подходят для имитации широкого класса процессов в биологических системах.

Биологическая система существует в непрерывном потоке времени, однако можно выделить дискретные состояния, в которых она в конкретном интервале времени пребывает: фермент может быть в свободном или в связанном с субстратом состоянии; гликолитический путь может находиться в режиме либо гликолиза, либо глюконеогенеза, что, в частности, определяется состоянием ключевого фермента фосфофруктокиназы/фруктозо-бифосфатазы; животное может находиться в состоянии сна, питания, преследования и т. п. У. Эшби (1959, с. 46–47) обратил внимание, что смена типов (паттернов) поведения животного может быть описана как смена состояний конечного автомата. В таком случае алгебра регулярных событий (например, Хопкрофт и др., 2008, гл. 3) может формально описывать среду (и ее сложность), в которой «живет» автомат и «эволюционирует» нейронная сеть.

Поскольку вычислительные эксперименты были направлены на выявление общих закономерностей эволюционного процесса, то в ходе экспериментов создавались ансамбли нейронных сетей, выполняющих одну и ту же функцию. Для краткости отдельный экземпляр обученной нейронной сети будем называть нейросетевым модельным объектом (НМО).

Начальная структура НМО (матрица весовых коэффициентов) задавалась генератором случайных чисел, по модулю близких к 0. Обучение заканчивалось при достижении заданного уровня целевой функции, соответствующего полному отсутствию ошибок распознавания стимулов во входном потоке. То есть все экземпляры, прошедшие обучение, реализовывали требуемую функцию одинаково хорошо. Генерация и редукция сложности НМО осуществлялись с помощью оригинальной программы в среде Lazarus.

Для визуализации и статистического анализа структуру НМО можно представить точкой в пространстве весовых коэффициентов, а евклидово расстояние между точками, соответствующими разным НМО, принять за меру близости их структур. Кроме распределения структур НМО в структурном пространстве, в работе оценивалась расходимость траекторий при старте из одного начального состояния. Визуализация осуществлялась с помощью оригинальной программы в среде Scilab 6.1.1.

Результаты

Сразу отметим, что сетевое описание предполагает наличие операций сложения и умножения для описания взаимодействия элементов, их взаимовлияния друг на друга. Но конкретные значения суммы или произведения могут быть получены бесконечным числом способов, если мы используем не целые числа. Отсюда сразу можно предположить, а точнее с уверенностью сказать, что в общем случае максимальные значения функции приспособленности могут достигаться разными способами, т. е. разными функционально-инвариантными структурами. Результаты вычислительных экспериментов подтвердили это предположение.

Отметим, что минимальное количество нейронов в НМО рассмотренной конфигурации, обеспечивающее требуемое качество функционирования, равно 6, причем два нейрона являются входными, два – выходными, а два нейрона ассоциативные или скрытые. Попытки использования НС с меньшим количеством нейронов, предпринимаемые в этой и более ранних работах (Барцев, Барцева, 2010), оказались безуспешными. Следовательно, можно принять, что минимальное пространство структур НМО, выполняющих заданные функции, является 36-мерным. Визуализировалась локализация структур НМО путем проецирования 36-мерного пространства в трехмерное, причем в качестве координат проекционного пространства выбирались различные комбинации номеров весовых коэффициентов.

Ранее было показано, что структуры НМО в пространстве распределены не случайным (диффузным) образом, а образуют скопления – кластеры (рис. 3). Поскольку точки соответствует структурам НМО, выполняющим одну и ту же функцию, то можно говорить о наличии функциональной симметрии, т. е. сохранения инварианта (функции) при определенного рода преобразованиях (перемещениях в пространстве структур НМО).

Отметим, что перемещения внутри кластеров соответствуют локальной симметрии, а перескоки из кластера в кластер – дискретным преобразованиям перестановки и смены знака (Барцев, Барцева, 2005, 2010). Количество кластеров N для НМО с нечетной функцией активации определяется формулой $N = 2^K K!$, где K – количество ассоциативных нейронов. Степень двойки описывает количество комбинаций смены знака, а факториал – коли-

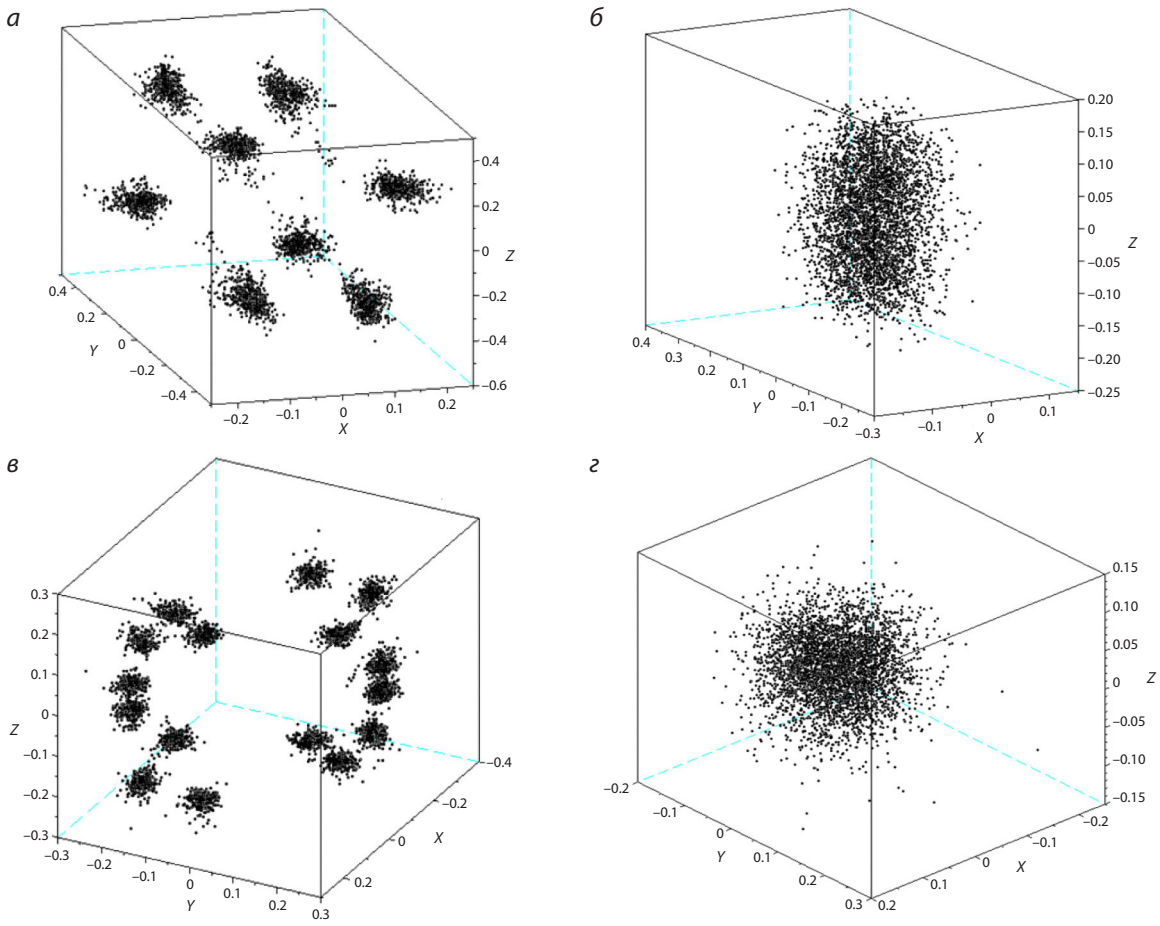


Рис. 3. Расположение структур НМО с функцией А3 (*а*, *б*) и С4 (*в*, *г*).
Координаты *XYZ* для (*а*) и (*в*) соответствуют номерам синапсов 25/26/31, для (*б*) и (*г*) – 9/12/22.

чество перестановок ассоциативных нейронов в матрице весовых коэффициентов. В случае функции С4 кластеров в два раза больше, так как сигналы входных нейронов идентичны и возможна перестановка этих нейронов в матрице синапсов.

Каждый кластер на рис. 3 соответствует вершине со срезанной верхушкой на ландшафте приспособленности. Подчеркнем, что качество функционирования (уровень приспособленности) приведенных на рисунке структур идентично, что следует из природы функционально-инвариантных преобразований и подтверждается в вычислительном эксперименте.

Из сравнения левых и правых 3D изображений можно сделать полезный методологический вывод, что представления о свойствах ландшафта приспособленности критическим образом зависят от того, какие элементы структуры принимаются во внимание. В данном случае 8 и 16 структурных кластеров проявляются только при определенном подборе параметров (весовых коэффициентов) НМО. Один из индикаторов наличия кластерной структуры – би- и полимодельность частотных распределений попарных декартовых расстояний между структурами НМО (Барцев, Барцева, 2010).

Как уже отмечалось, начальные весовые коэффициенты НМО задавались генератором случайных чисел и все НМО стартовали из разных начальных точек пространства структуры. Тогда представленный результат говорит о возможности разных исходов эволюции, но ничего не говорит об устойчивости и воспроизводимости эволюционной траектории.

Воспроизводимость траектории более простой НС и на других задачах исследовалась ранее (Bartsev, Baturina, 2019). Было показано, что на протяжении всего процесса обучения каждая точка траектории может породить веер альтернативных траекторий, направленных, тем не менее, в сторону притяжения кластера, соответствующего максимальной приспособленности. В данной работе для демонстрации в общем-то ожидаемых результатов было получено несколько реализаций обучающих траекторий из одного и того же начального состояния (рис. 4). Легко видеть быструю расходимость эволюционных траекторий, которая обусловлена, во-первых, случайностью возмущений (мутаций) весовых коэффициентов, а во-вторых, невоспроизводимостью потока событий (см. рис. 2) (окружающий мир), в котором обучались НМО.

Выше уже говорилось, что приспособиться (обучиться данной задаче) к заданной среде могут только НМО, состоящие из 6 нейронов и более. Для оценки минимального количества синапсов, обеспечивающего выполнение заданной функции (в данном случае – А3 и С4), была

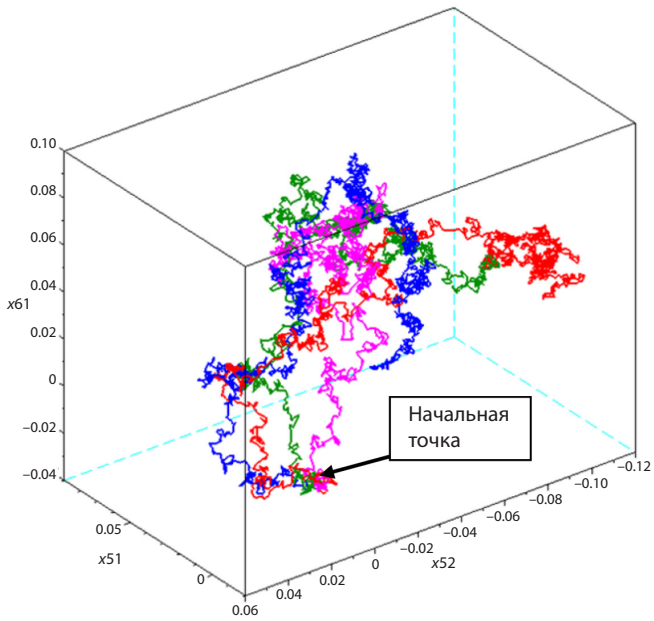


Рис. 4. Пример эволюционных траекторий НМО для функции A3 в координатах 25/26/31.

использована процедура редукции сложности НМО. Эта процедура состояла в последовательном удалении минимальных по модулю синапсов и дальнейшем дообучении до заданного качества функционирования.

Было показано, что предельное количество весовых коэффициентов, при которых НМО сохраняет то же самое качество функционирования, равно 11 для последовательности A3 (рис. 5) и 13 для последовательности C4. Такие различия в минимальном количестве синапсов ожидаемы, поскольку функция C4 сложнее A3 из-за того, что нейросети нужно дольше хранить память о стимуле, прежде чем выдать ответ.

Надо отметить еще один важный момент: если взять структуру редуцированного НМО и заменить оставшиеся ненулевые синапсы на случайные числа (даже с сохранением знака), то обучения НМО не происходит. Из этого можно сделать вывод, что для обучения (эволюционного процесса) необходимо наличие степеней свободы, т. е. весовых коэффициентов, не задействованных полностью на реализацию функции и позволяющих обходить препятствия на ландшафте функции приспособленности по другим измерениям.

При этом необходимо отметить, что минимальное количество синапсов зависит не от размера исходной нейросети, а только от выполняемой функции (рис. 6). Этот факт позволяет утверждать, что сложность (при заданной переходной характеристике нейрона) является инвариантным количественно выражаемым свойством самой функции.

Процедуру редукции НМО можно интерпретировать как наличие цены за ненулевой весовой коэффициент. В этом случае формирование редуцированной структуры соответствует жесткой специализации организма, не дающей впоследствии вырваться из эволюционного тупика.

Минимальное количество синапсов, необходимое для функционирования НМО, можно принять за оценку функциональной сложности. Тогда общее количество синапсов НМО можно рассматривать как потенциальную сложность, которая аналогична информационной емкости класса (Эйген, 1973).

Из вычислительных экспериментов следует, что для того, чтобы эволюционный процесс мог осуществляться, необходима избыточность потенциальной сложности над функциональной, а эта избыточность ведет, в свою очередь, к множественности эволюционных исходов, что возможно, только когда одна и та же функция реализуется разными структурами.

$$\begin{pmatrix} -0.005 & 0.024 & 0.020 & -0.032 & 0.031 & 0.057 \\ 0.002 & 0.036 & -0.006 & -0.033 & -0.035 & 0.079 \\ -0.002 & -0.006 & 0.026 & -0.024 & -0.129 & 0.118 \\ -0.001 & -0.003 & 0.062 & -0.064 & -0.124 & 0.115 \\ -0.202 & -0.008 & 0.022 & -0.025 & 0.093 & 0.048 \\ -0.293 & 0.212 & -0.120 & -0.123 & 0.303 & -0.105 \end{pmatrix} \Rightarrow \begin{pmatrix} 0.000 & 0.206 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.213 \\ 0.000 & 0.000 & 0.000 & 0.000 & -0.077 & 0.061 \\ 0.000 & 0.000 & 0.000 & 0.000 & -0.077 & 0.061 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.412 & 0.035 \\ -0.710 & 0.589 & 0.000 & 0.000 & 0.775 & 0.000 \end{pmatrix}$$

Рис. 5. Пример редукции сложности 6-нейронного НМО с функцией A3.

$$\begin{pmatrix} 0.000 & 0.246 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.386 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.062 & 0.000 & 0.000 & 0.000 & 0.077 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.062 & 0.000 & 0.000 & 0.000 & 0.077 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.623 & 0.563 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.742 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.042 & 0.000 & 0.000 & 0.000 & 0.450 \end{pmatrix} \equiv \begin{pmatrix} 0.000 & 0.246 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.386 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.062 & 0.077 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.062 & 0.077 \\ -0.623 & 0.563 & 0.000 & 0.000 & 0.000 & -0.0742 \\ 0.000 & 0.000 & 0.000 & 0.000 & -0.042 & 0.450 \end{pmatrix}$$

Рис. 6. Перевод уже редуцированной структуры 10-нейронного НМО (функция A3) в 6-нейронную конфигурацию путем удаления нулевых строк (и соответствующих столбцов) матрицы синапсов.

Обсуждение

Результаты исследования эвристической модели эволюции показали, что, с одной стороны, при значительной избыточности потенциальной сложности над сложностью функциональной финальную структуру НМО предсказать невозможно. С другой стороны, введение ограничений, например требование минимального количества синапсов, приводит к тому, что количество возможных финальных вариантов структуры уменьшается и возможно их почти полное совпадение, даже при различии размеров НМО. Например, сравнивая правые матрицы весовых коэффициентов на рис. 5 и 6, легко заметить, что матрицы редуцированных 6- и 10-нейронных НМО очень похожи: достаточно поменять местами две нижние строки и изменить знаки синапсов в строках, т. е. произвести дискретные функционально-инвариантные преобразования структуры.

Выводы эвристической модели (от 2005 г.) о потенциальной множественности эволюционных исходов нашли свою иллюстрацию в работах (Podgornaia, Laub, 2015; Starr et al., 2017), где показано, что помимо природного набора из четырех аминокислот в обоих рассмотренных белках их функцию могут выполнять (так же качественно) еще около 1 % комбинаций из общего набора в 160 тыс. При этом в пространстве последовательностей изученная область распадается на подобласти (кластеры), внутри которых эволюционные перемещения могут происходить относительно свободно, тогда как переходы между кластерами затруднены, что хорошо соответствует эвристической нейросетевой модели.

В то же время при наличии ограничений и при нахождении в окрестности максимума функции приспособленности эволюционные траектории могут воспроизводиться (Weinreigh et al., 2006; Meyer et al., 2012; Orgogozo, 2015), что тоже находит соответствие в эвристической модели на примере редуцированных НМО.

Отметим, что полученные результаты не являются чисто эвристическими и вычислительно-экспериментальными. Возможность функционально-инвариантных преобразований приводит к тому, что могут возникать «мутации» структуры с идентичным или очень близким качеством функционирования (равной или близкой приспособленностью), что соответствует нейтральным мутациям Кимуры (1985). Вследствие этого на очередном шаге эволюционная траектория может пойти по нескольким эквивалентным с точки зрения функции приспособленности направлениям.

Резюмируя, можно согласиться с претензией к математическим моделям (Garte et al., 2025), что они в общем случае не в состоянии предсказать конкретную эволюционную траекторию. По-видимому, требование спрогнозировать конкретную траекторию представляет собой в большинстве случаев неправильный запрос.

Вместо этого модель, работая с ансамблями траекторий, позволяет рассчитать вероятности и условия того или иного эволюционного исхода, что обеспечит понимание общих свойств эволюционирующих систем. Так, например, вычислительное построение «гомологической карты» Агекяна–Аносовой (Lehto et al., 2008) показало, что зависимость времени существования системы трех тел от начального состояния имеет фрактальную структуру, и становится понятно, что от этой системы можно в принципе ожидать.

Заключение

Абстрактная эвристическая модель эволюции, построенная на основе общих предположений о природе живого, позволяет сформулировать предположения (гипотезы) об общих свойствах эволюционирующих систем. В частности, сама возможность *эволюционных изменений* зависит от *избыточности структуры* над сложностью приспособительной функции, что в общем случае приводит к потенциальной *множественности эволюционных исходов*, обусловленной *неоднозначностью структурно-функциональных отношений*. То есть в общем случае конкретный путь и конечная структура (и само ее достижение) эволюционирующей системы не детерминированы при функциональной эквифинальности эволюционных исходов.

Список литературы / References

Барцев С.И., Барцева О.Д. Симметрии структуры и эквифинальность эволюционных исходов в простых нейросетевых моделях. *Доклады Академии наук.* 2002;386(1):114-117
[Bartsev S.I., Bartseva O.D. Symmetric structures and equifinality of evolution outcomes in simple neural network models. *Dokl Biochem Biophys.* 2002;386(1-6):235-238. doi 10.1023/A:1020747225202]

Барцев С.И., Барцева О.Д. Функционально-инвариантный подход к проблеме уникальности биологических систем: простая нейросетевая модель. *Доклады Академии наук.* 2005;406(3):394-397
[Bartsev S.I., Bartseva O.D. Functional invariant approach to the biological system uniqueness: a simple neuronet model. *Doklady Biochemistry and Biophysics.* 2006;406(1):15-18. doi 10.1134/S1607672906010054]

Барцев С.И., Барцева О.Д. Эвристические нейросетевые модели в биофизике: приложение к проблеме структурно-функционального соответствия. Красноярск: СФУ, 2010
[Bartsev S.I., Bartseva O.D. Heuristic Neural Network Models in Biophysics: an application to the problem of structural and functional relations. Krasnoyarsk: Siberian Federal University, 2010 (in Russian)]

Бернал Дж.Д. Молекулярная структура, биохимическая функция и эволюция. В: Теоретическая и математическая биология. М.: Мир, 1968;110-151
[Bernal J.D. Molecular structure, biochemical physics and evolution. In: Theoretical and Mathematical Biology. Moscow: Mir Publ., 1968;110-151 (in Russian)]

Голованов И.Б., Иваницкий Г.Р., Цыганкова И.Б. Простая форма корреляционного соотношения и физико-химический смысл входящих в него параметров. *Доклады Академии наук.* 1998;359(2):258-262
[Golovanov I.B., Ivanitsky G.R., Tsygankova I.B. The simple form of the correlation relation and the physico-chemical meaning of its parameters. *Doklady Akademii Nauk.* 1998;359(2):258-262 (in Russian)]

Кимура М. Молекулярная эволюция: теория нейтральности. М.: Мир, 1985
[Kimura M. The Neutral Theory of Molecular Evolution. *Sci American.* 1979;241(5):98-129. http://www.jstor.org/stable/24965339]

Лернер А.Я. Начала кибернетики. М.: Наука, 1967
[Lerner A.Y. Fundamentals of Cybernetics. Springer, 1972]

Мейен С.В. О соотношении номогенетического и тихогенетического аспекта эволюции. *Журнал общей биологии.* 1974;35(3):353-364

[Meyen S.V. The relationship between the nomogenetic and tychogenetic aspects of evolution. *J Gener Biol*. 1974;35(3):353-364 (in Russian)]

Мейен С.В., Соколов Б.С., Шрейдер Ю.А. Классическая и неклассическая биология. Феномен Любищева. *Вестник АН СССР*. 1977;10:112-124
[Meyen S.V., Sokolov B.S., Schrader Yu.A. Classical and nonclassical biology: the Lyubishchev phenomenon. *Vestnik Akademii Nauk USSR = Herald of the Academy of Sciences (USSR).* 1977;10:112-124 (in Russian)]

Рашевский Н. Модели и математические принципы в биологии. В: Теоретическая и математическая биология. М.: Мир, 1968; 48-66
[Rashevsky N. Models and mathematical principles in biology. In: Theoretical and Mathematical Biology. Moscow: Mir Publ., 1968;48-66 (in Russian)]

фон Нейман Дж. Теория самовоспроизводящихся автоматов. М.: Мир, 1971
[Von Neumann J. Theory of Self-Reproducing Automata. University of Illinois Press, USA, 1966]

Хопкрофт Д., Мотвани Р., Ульман Д. Введение в теорию автоматов, языков и вычислений. М.: Изд. дом «Вильямс», 2008
[Hopcroft J.E., Motwani R., Ullman J.D. Introduction to Automata Theory, Languages, and Computations. 3rd ed. Pearson, Addison-Wesley, USA, 2007]

Эйген М. Самоорганизация материи и эволюция биологических макромолекул. М.: Мир, 1973
[Eigen M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften.* 1971;58(10):465-523. doi 10.1007/BF00623322]

Эшби У.Р. Введение в кибернетику. М., 1959
[Ashby W.R. An Introduction to Cybernetics. New York, J. Wiley & Sons Inc, 1956]

Albert R., Jeong H., Barabasi A. Error and attack tolerance of complex networks. *Nature*. 2000;406:378-382. doi 10.1038/35019019

Amaral L.A., Scala A., Barthelemy M., Stanley H.E. Classes of small-world networks. *Proc Natl Acad Sci USA.* 2000;97(21):11149-11152. doi 10.1073/pnas.200327197

Bartsev S., Baturina P. A heuristic neural network model in the research of properties of evolutionary trajectories. *IOP Conf Ser Mater Sci Eng*. 2019;537(4):042001. doi 10.1088/1757-899X/537/4/042001

Beckage B., Gross L.J., Kauffman S. The limits to prediction in ecological systems. *Ecosphere.* 2011;2(11):1-12. doi 10.1890/ES11-00211.1

De Las Rivas J., Fontanillo C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Computat Biol*. 2010;6(6):e1000807. doi 10.1371/journal.pcbi.1000807

Dickins B., Nekrutenko A. High-resolution mapping of evolutionary trajectories in a phage. *Genome Biol Evol.* 2009;1:294-307. doi 10.1093/gbe/evp029

Dunham E.J., Dugan V.G., Kaser E.K., Perkins S.E., Brown I.H., Holmes E.C., Taubenberger J.K. Different evolutionary trajectories of European avian-like and classical swine H1N1 influenza A viruses. *J Virol.* 2009;83(11):5485-5494. doi 10.1128/JVI.02565-08

Dunne J.A., Williams R.J., Martinez N.D. Food-web structure and network theory: the role of competence and size. *Proc Natl Acad Sci USA*. 2002;99(20):12917-12922. doi 10.1073/pnas.192407699

Edelman G.M., Gally J.A. Degeneracy and complexity in biological systems. *Proc Natl Acad Sci USA*. 2001;98(24):13763-13768. doi 10.1073/pnas.231499798

Forst C.V., Reidys C., Weber J. Evolutionary dynamics and optimization. In: Morán F., Moreno A., Merelo J.J., Chacón P. (Eds) Advances in Artificial Life. ECAL 1995. Lecture Notes in Computer Science. Vol. 929. Springer, 1995;128-147. doi 10.1007/3-540-59496-5_294

Garte S., Marshall P., Kauffman S. The reasonable ineffectiveness of mathematics in the biological sciences. *Entropy*. 2025;27(3):280. doi 10.3390/e27030280

Kvitek D.J., Sherlock G. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* 2011;7(4):e1002056. doi 10.1371/journal.pgen.1002056

Lehto H.J., Kotiranta S., Valtonen M.J., Heinämäki P., Mikkola S., Chernin A.D. Mapping the three-body system – decay time and reversibility. *Mon Not R Astron Soc.* 2008;388(3):965-970. doi 10.1111/j.1365-2966.2008.13450.x

Lobkovsky A.E., Koonin E.V. Replaying the tape of life: quantification of the predictability of evolution. *Front Genet.* 2012;3:246. doi 10.3389/fgene.2012.00246

Lobkovsky A.E., Wolf Y.I., Koonin E.V. Predictability of evolutionary trajectories in fitness landscapes. *PLoS Computat Biol.* 2011;7(12):e1002302. doi 10.1371/journal.pcbi.1002302

Meyer J.R., Dobias D.T., Weitz J.S., Barrick J.E., Quick R.T., Lenski R.E. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science.* 2012;335(6067):428-432. doi 10.1126/science.1214449

Orgogozo V. Replaying the tape of life in the twenty-first century. *Interface Focus*. 2015;5(6):20150057. doi 10.1098/rsfs.2015.0057

Podgornaia A.I., Laub M.T. Pervasive degeneracy and epistasis in a protein-protein interface. *Science*. 2015;347(6222):673-677. doi 10.1126/science.1257360

Poelwijk F.J., Kiviet D.J., Weinreich D.M., Tans S.J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature.* 2007;445(7126):383-386. doi 10.1038/nature05451

Povolotskaya I.S., Kondrashov F.A. Sequence space and the ongoing expansion of the protein universe. *Nature.* 2010;465(7300):922-926. doi 10.1038/nature09105

Rosen R. A relational theory of biological systems. *Bull Math Biophys*. 1958;20:245-260. doi 10.1007/BF02478302

Schuster P. How does complexity arise in evolution? In: Götschl J. (Ed.) Evolution and Progress in Democracies. Theory and Decision Library. Vol. 31. Springer, 1996;147-159. doi 10.1007/978-94-017-1504-1_8

Sole R.V. Modelling macroevolutionary patterns: an ecological perspective. In: Lässig M., Valleriani A. (Eds) Biological Evolution and Statistical Physics. Lecture Notes in Physics. Vol. 585. Springer, 2002;312-337. doi 10.1007/3-540-45692-9_18

Starr T.N., Picton L.K., Thornton J.W. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature.* 2017;549(7672):409-413. doi 10.1038/nature23902

Strogatz S.H. Exploring complex networks. *Nature*. 2001;410(6825):268-276. doi 10.1038/35065725

Stumpf M.P.H., Thorne T., de Silva E., Stewart R., An H.J., Lappe M., Wiuf C. Estimating the size of the human interactome. *Proc Natl Acad Sci USA*. 2008;105(19):6959-6964. doi 10.1073/pnas.0708078105

Weinreigh D.M., Delaney N.F., DePristo M.A., Hart D.L. Darwinian evolution can follow only few mutational paths to fitter proteins. *Science*. 2006;312(5770):111-114. doi 10.1126/science.1123539

Xue K.S., Stevens-Ayers T., Campbell A.P., Englund J.A., Pergam S.A., Boeckh M., Bloom J.D. Parallel evolution of influenza across multiple spatiotemporal scales. *eLife.* 2017;6:e26875. doi 10.7554/eLife.26875

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию 01.08.2025. После доработки 16.10.2025. Принята к публикации 27.10.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):293-298

doi 10.18699/vjgb-26-33

Alembic: от разрозненных биологических данных к структурированным ресурсам

И.В. Бездворных#, К.И. Юдыцкий#, Н.А. Черкасов, А.А. Самсонова, А.А. Канапин

Институт трансляционной биомедицины, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
a.kanapin@gmail.com

Аннотация. Развитие технологий высокопроизводительного секвенирования и методов анализа больших данных создает устойчивую потребность в повторном анализе накопленной в открытых репозиториях гетерогенной информации. Серьезной проблемой при этом остается преобладание свободного текстового описания биологических экспериментов, что затрудняет продуктивный поиск, систематизацию и дальнейшее использование соответствующих наборов данных. Прогресс в области искусственного интеллекта, особенно в развитии методов обработки естественного языка (natural language processing, NLP), обуславливает новые методологические возможности для эффективного решения этой задачи. Интегрированная система баз данных Entrez, поддерживаемая Национальным центром биотехнологической информации США (NCBI), предоставляет развитый и надежный доступ как к исходным данным секвенирования, так и к сопутствующей метаинформации, включающей детальное описание параметров экспериментов, через программный интерфейс (application programming interface, API). Это позволяет идентифицировать и загружать данные секвенирования и соответствующие им метаданные с описаниями экспериментов, используя поиск по ключевым словам и различным терминам, таким, например, как имена генов, в репозиториях; преобразовывать и систематизировать текстовые описания с применением современных NLP-методов и обеспечивать исследователям структурированную информацию для интеграции в локальные базы данных и форматированный перечень ссылок для загрузки исходных данных. Программный пакет Alembic предлагает комплексное решение для поиска и загрузки данных, автоматизируя все указанные этапы. Платформа использует клиент-серверную архитектуру и предназначена для локальной установки. Для анализа биомедицинских текстов, сопровождающих данные секвенирования, в Alembic интегрированы современные алгоритмы искусственного интеллекта на основе архитектуры трансформеров. В частности, используется имеющаяся в открытом доступе платформа AIONER, обученная на данных репозитория PubMed с помощью модели PubMedBERT. Такой подход обеспечивает эффективное распознавание именованных сущностей (named entity recognition, NER) биомедицинского характера (гены, заболевания и др.), предоставляя пользователю структурированные результаты поиска по ключевым словам. Формируемый пакетом список дает возможность исследователю анализировать результаты, отбирать наиболее релевантные наборы данных и получать всю необходимую информацию (включая исходные данные) для создания локального репозитория, ориентированного на конкретную исследовательскую задачу. В отличие от имеющихся аналогов, Alembic является универсальным решением для интеграции данных из репозиториев открытого доступа и работы с разнородными типами данных секвенирования.

Ключевые слова: обработка естественных языков; анализ биомедицинских текстов; семантическая аннотация; гармонизация данных; интеграция омиксных даных

Для цитирования: Бездворных И.В., Юдыцкий К.И., Черкасов Н.А., Самсонова А.А., Канапин А.А. Alembic: от разрозненных биологических данных к структурированным ресурсам. *Вавиловский журнал генетики и селекции*. 2026;30(2):293-298. doi 10.18699/vjgb-26-33

Финансирование. Работа поддержана грантом Российского научного фонда 23-14-00134.

Alembic: a framework for converting disparate biological data into structured resources

I.V. Bezdvornykh#, K.I. Yuditskiy#, N.A. Cherkasov, A.A. Samsonova, A.A. Kanapin

Institute for Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia
a.kanapin@gmail.com

Abstract. The imperative to re-analyze existing public sequencing data is central to modern biology, driven by new hypotheses and advanced analytical methods. However, this effort is critically hampered by the profound heterogeneity of repository data, particularly the non-standardized, free-text descriptions of biological experiments. This lack of structural and semantic homogeneity prevents systematic search, integration, and comparative analysis, effectively locking away the full potential of accumulated datasets. Advances in Natural Language Processing (NLP) offer a pivotal pathway to overcome this bottleneck by transforming unstructured text into computable, homogeneous information.

© Бездворных И.В., Юдыцкий К.И., Черкасов Н.А., Самсонова А.А., Канапин А.А., 2026

Авторы внесли равный вклад в исследование.

Контент доступен под лицензией Creative Commons Attribution 4.0

The integrated Entrez database system, maintained by the National Center for Biotechnology Information (NCBI), provides sophisticated programmatic access via an API to primary sequencing data and its associated metadata, including detailed experimental descriptions. This interface enables researchers to identify and retrieve relevant data through keyword searches, including those based on gene names, and to apply modern NLP techniques to transform textual metadata into structured information. The output is formatted data ready for integration into local databases, accompanied by a systematic list of links for downloading primary files. The Alembic software package offers a comprehensive and automated solution for the entire workflow. Designed as a locally deployable client-server system, Alembic incorporates state-of-the-art transformer-based AI algorithms for analyzing the biomedical text that accompanies sequencing data. Its core utilizes the openly available AIONER platform, which is built upon the PubMedBERT model trained on the PubMed repository, to ensure efficient and accurate recognition of biomedical named entities (e. g., genes, diseases). This provides users with structured and meaningful keyword search results. By delivering a curated list of datasets, Alembic streamlines the path from search to analysis. Researchers can efficiently identify high-value targets and obtain a complete package of metadata and primary data to construct a tailored local repository. This positions Alembic as a universal solution that overcomes the fragmented approach of existing tools, offering an integrated workflow for diverse public sequencing data.

Key words: natural language processing; biomedical text mining; semantic annotation; data harmonization; omics data integration

For citation: Bezdvornykh I.V., Yuditskiy K.I., Cherkasov N.A., Samsonova A.A., Kanapin A.A. Alembic: a framework for converting disparate biological data into structured resources. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):293-298. doi 10.18699/vjgb-26-33

Введение

Данные секвенирования различных модальностей (таких как WGSeq, RNASeq, BSSeq) и сопутствующие публикации накапливаются со все возрастающей скоростью, что напрямую связано с развитием технологий высокопроизводительного секвенирования. Обработка таких массивов информации требует создания специализированных программных систем для автоматического эффективного извлечения биологической информации. Перспективным направлением в этой области стало применение алгоритмов обработки естественного языка (natural language processing, NLP). Современные NLP-методы автоматизируют анализ научных текстов, существенно повышая эффективность работы с большими данными. Одним из примеров применения NLP в биологии выступает система MetaMap (Aronson, Lang, 2010), реализующая распознавание именованных сущностей (named entity recognition, NER) для идентификации концептов метатезауруса UMLS (Unified Medical Language System) в текстах. Дальнейшее развитие этого направления воплотилось в пакете BIONER для распознавания биомедицинских сущностей (Wang et al., 2019).

Появление нейросетевых архитектур и техники векторных представлений слов (word embeddings) существенно повысило точность распознавания биомедицинских сущностей. Значительным шагом стало внедрение контекстнозависимых моделей типа ELMo (embeddings from language model), формирующих представления слов с учетом их окружения. Современный этап развития связан с доминированием трансформерных архитектур, прежде всего BERT (Devlin et al., 2019), кардинально изменивших методы решения NLP-задач. Дальнейшая адаптация этих моделей к биомедицинской предметной области привела к созданию домен-специфичных решений (BioBERT, PubMedBERT и др.) (Lee et al., 2020) и специализированных инструментов преобразования данных, таких как scispaCy (Neumann et al., 2019).

Наиболее комплексной системой биомедицинских баз данных остается Entrez, разрабатываемая Национальным центром биотехнологической информации США (NCBI). Платформа объединяет 38 репозиториев, включая PubMed, PMC, базы нуклеотидных и белковых последовательностей. Пополнение системы осуществляется как автоматически (путем анализа научных публикаций), так и исследователями: депонирование данных секвенирования в открытых архивах (SRA, ENA и др.) требует обязательного предоставления метаинформации об экспериментах и биологических образцах. Ключевое преимущество Entrez – наличие программного интерфейса Entrez Programming Utilities (E-utilities) (Sayers, 2022), позволяющего осуществлять расширенный поиск и автоматизированное извлечение данных.

В задачи исследователя довольно часто входит поиск существующих наборов сырых данных, полученных в экспериментах по определенной тематике. Комбинация программного доступа к ресурсам Entrez с современными NLP-алгоритмами для анализа биомедицинских текстов позволяет создать унифицированную систему эффективного отбора интересующих данных (включая исходные данные секвенирования и сопроводительную метаинформацию). Существующие решения, например iSeq (Chao et al., 2024) и SampleExplorer (Chin, Lassmann, 2024), не обеспечивают универсального подхода и имеют ограниченную функциональность. В настоящей работе представлена система Alembic, предназначенная для анализа данных в открытых репозиториях NCBI и структурированного извлечения нужной информации. Название системы отражает сущность процесса обработки данных (alembic, аламбик – тип перегонного куба, предназначенного для извлечения полезных компонентов из больших объемов сырья).

Материалы и методы

В основе Alembic лежит клиент-серверная архитектура. Клиентский модуль управляет формированием запросов и отвечает за визуализацию результатов, тогда как серверная часть обрабатывает запросы и выполняет извлечение

именованных сущностей. Общая схема модулей системы и процессов обработки приведена на рис. 1.

Система Alembic использует данные из архива коротких прочтений NCBI (SRA), доступные посредством открытого API Entrez. Поиск поддерживает практически любые термины: названия генов, организмов, биологических процессов, метаболитов и других сущностей. Результаты запроса обрабатываются модулем AlembicDump следующим образом: извлекаются универсальные идентификаторы NCBI (UID), на их основе загружаются метаданные в XML-формате, которые затем стандартизируются и преобразуются в таблицу с заданной структурой столбцов. Метаданные представлены в стандарте Entrez SRA, включающем 21 поле, которые могут содержать как свободный неформатированный текст (например, Study Abstract), так и набор идентификаторов заданного формата (Sample Accession и др.).

На следующем этапе полученные структурированные данные обрабатываются алгоритмом AlembicNLP для извлечения именованных сущностей с использованием предобученных биомедицинских языковых моделей. Сначала входной текст последовательно проходит предобработку с помощью sciSpaCy (версия 2.0.18): преобразование в нижний регистр, очистку от специальных символов, пунктуации и изолированных цифр, фильтрацию стоп-слов (по словарю NLTK, www.ntlk.org, версия 3.8.1), лемматизацию и токенизацию. После чего трансформированный текст анализируется биомедицинской моделью AIONER (Luo et al., 2023) на основе архитектуры Bioformer, которая идентифицирует фрагменты текста, соответствующие биомедицинским сущностям. Исходный код AIONER, находящийся в открытом доступе, встроен в Alembic и не является зависимым элементом, требующим дополнительной установки.

В Alembic используются модели AIONER bioformer-cased-v1.0 и BiomedNLP-PubMedBERT-base-uncased-abstract, предобученные на данных PubMed и имеющиеся в открытом доступе (https://huggingface.co/lingbionlp/AIONER-0415/tree/main). Эти модели обучались на информации, имеющейся в описании статей (abstracts), депонированных в каталоге PubMed (Luo et al., 2023). После распознавания полученные элементы автоматически классифицируются по установленным категориям – гены (Gene), болезни (Disease), виды (Species), клеточные линии (Cell Line), генетические варианты (Variant) и химические соединения (Chemical). Финальный результат работы алгоритма AlembicNLP представляет собой аннотированный список сущностей с указанием их типа и позиционных координат в исходном тексте.

Клиентская часть приложения реализована на базе фреймворка Vite JS (версия 7.0), обеспечивающего быструю сборку и высокую производительность разработки. Приложение построено с помощью библиотеки React версии 19. Для создания интерфейса применяются компоненты из библиотеки Material UI (MUI) версии 7, а также специализированные компоненты для работы с таблицами данных (MUI X DataGrid). Взаимодействие с



Рис. 1. Схема пакета программ Alembic, включающего модули AlembicDump (предобработка результатов поиска) и AlembicNLP (извлечение именованных сущностей).

сервером осуществляется через асинхронные запросы с использованием библиотеки Axios (версия 1.6.7).

Функционал приложения реализован в виде двух взаимосвязанных модулей. Первый модуль предоставляет интерфейс поиска биомедицинскх терминов с последующим отображением результатов в виде агрегированной таблицы. Табличное представление результатов поиска по ключевым словам оптимизировано для работы с большими объемами данных.

Второй модуль отвечает за подготовку данных для ИИ-обработки. После выбора пользователем нужных ему полей система автоматически извлекает ключевые термины, классифицируя их по заданным категориям. Обработка выполняется серверной частью, тогда как клиентский интерфейс обеспечивает структурированное представление результатов в удобном формате. Такая архитектура минимизирует когнитивную нагрузку на пользователя и существенно упрощает анализ сложных текстовых описаний условий экспериментов.

Визуализация результатов группировки оснащена интерактивными инструментами: фильтрами терминов, быстрым поиском и динамической подсветкой релевантных элементов, что позволяет эффективно анализировать большие объемы данных, избавляя пользователя от необходимости изучения полных описаний экспериментов. Интерфейс адаптивен и оптимизирован для всех типов устройств – от десктопов до мобильных платформ.

Типичный сценарий работы с системой Alembic включает три стадии. В качестве примера рассмотрим поиск транскриптомных данных, полученных в ходе экспериментов по изучению болезни Альцгеймера. На первом этапе пользователь проводит поиск экспериментов в базе данных NCBI SRA с помощью набора ключевых слов, например названия генов, заболеваний, протоколов секвенирования и т. д., используя окно поиска (рис. 2). В данном случае это “Alzheimer”, “disease” и “RNASeq”. Система начинает поиск и показывает число найденных экспериментов (130 в приведенном примере), имеющихся в базе SRA. В случае когда число найденных экспериментов превы-

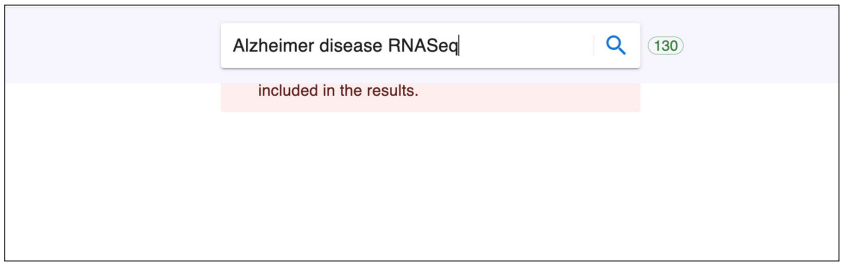


Рис. 2. Окно поиска системы Alembic.

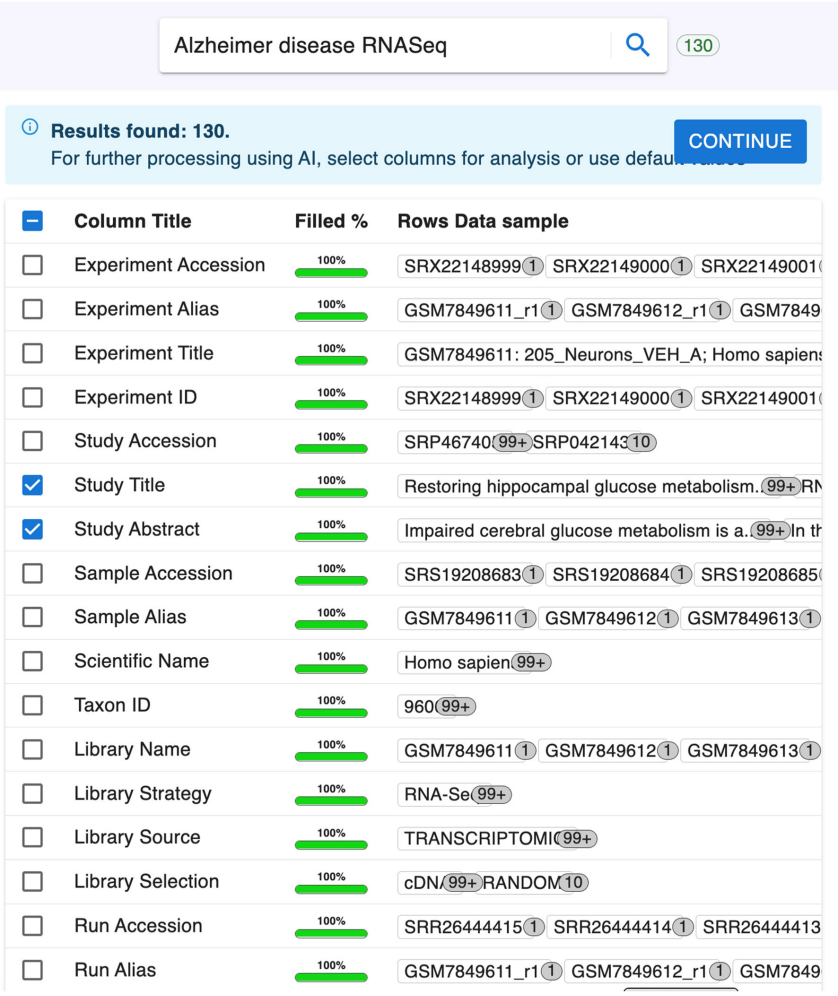


Рис. 3. Табличное представление результатов поиска по ключевым словам.

шает 1000 шт., выводится предупреждение об этом и исследователь может либо изменить термины, по которым ведется поиск, либо использовать первые 1000 из найденных экспериментов (сценарий по умолчанию).

На втором этапе Alembic структурирует полученные метаданные и представляет их в табличном виде (рис. 3). Формат включает в себя названия стандартных полей метаданных EntRez (Column Title), долю в процентах найденных экспериментов, имеющих в своих метаданных такое поле (Filled %) и примеры метаданных (Rows Data sample). Используя флажки в крайней левой колонке, пользователь может выбирать несколько типов метаданных для дальнейшего анализа, включая обработку полей данных средствами искусственного интеллекта.

На третьем этапе, нажав кнопку CONTINUE, пользователь запускает NLP-обработку и получает окончательный результат (рис. 4). В данном примере приведена визуализация результатов поиска сущностей в формате «облако слов», wordCloud. В правом верхнем углу окна приведены ссылки для загрузки файлов с результатами в текстовом формате, с полями, разделенными табуляцией (tab-separated text, tsv), как до обработки ИИ – RESULTS.TSV,



Рис. 4. Визуализация результатов обработки метаданных PubMedBERT.
Найденные именованные сущности представлены в виде облака слов (wordcloud).

так и после обработки – AI_PROCESSED.TSV. Полученные форматированные файлы могут быть напрямую импортированы в реляционные базы данных на основе sqLite, MySql или Postgress SQL. Наконец, файл DOWNLOADS.TXT содержит прямые ссылки для скачивания сырых данных секвенирования из репозитория SRA. В нижней части страницы приводятся список экспериментов и для каждого из них – список образцов, для которых получены данные секвенирования, доступные для скачивания.

Внешний интерфейс проекта реализован с использованием модулей JavaScript. Архитектура спроектирована с учетом масштабируемости; добавление новых функциональных модулей не требует структурных изменений. Исходный код проекта и полная документация к нему находятся в репозитории открытого доступа GitHub (лицензия MIT): https://github.com/shitohana/Alembic.

Для установки Alembic на компьютере пользователя необходимо наличие в системе языка python версии не ниже 3.12 и платформы docker версии не ниже 24.0.2. Остальные пакеты, необходимые для функционирования, в том числе библиотеки python TensorFlow (версия 2.3.0), Transformer (версия 4.18.0) и stanza (версия 1.4.0), необходимые для функционирования AIONER, устанавливаются автоматически. После установки пакета взаимодействие с системой осуществляется через локальный веб-интерфейс.

Результаты и обсуждение

Обработка больших объемов биомедицинских текстов потребовала создания многочисленных систем для распознавания именованных сущностей. Наиболее развитые решения в этой области – scispaCy и AIONER. scispaCy предлагает модели, предобученные на биомедицинских корпусах (GENIA, MedMentions), которые позволяют эффективно решать ключевые задачи NLP: токенизацию, POS-разметку (part-of-speech – разметка по частям речи) – технику в обработке естественного языка (NLP), присваивающую грамматическую категорию (например, существительное, глагол или прилагательное) каждому слову в тексте, синтаксический анализ и NER, учитывая при этом биомедицинскую специфику текста.

В отличие от этого, AIONER специализируется исключительно на BioNER, реализуя инновационный подход All-in-One (AIO), сущность которого заключается в одновременном распознавании нескольких терминов из различных наборов данных, применяемых для обучения. По сравнению с другими алгоритмами, где разметка текста по токенам происходит отдельно для каждой категории (например, категорий Gene, Disease), AIO обрабатывает все категории одновременно, используя специальные метки (tags), для того чтобы избежать ошибок при атрибутировании отдельных слов или их комбинаций. Данная архитектура повышает точность токенизации и, таким образом, решает проблему переобучения и слабой обобщающей способности моделей, вызванную нехваткой размеченных биомедицинских данных. В основе AIONER лежат современные предобученные языковые модели (PubMedBERT и аналоги), дополненные слоем условных случайных полей (conditional random fields, CRF) для точного определения границ сущностей. Благодаря этим преимуществам AIONER выбран базовым алгоритмом для пакета Alembic.

Растущий объем биомедицинских текстов стимулировал разработку многочисленных систем извлечения именованных сущностей (named entity recognition, NER), ведущими примерами которых являются scispaCy и AIONER. scispaCy предлагает модели, предварительно обученные на биомедицинских корпусах, таких как GENIA и MedMentions. Эти модели эффективно выполняют основные задачи обработки естественного языка, а именно

токенизацию, разметку частей речи (part-of-speech, POS), синтаксический анализ и NER, одновременно учитывая доменно-специфические лингвистические особенности. Под разметкой частей речи здесь понимается фундаментальный метод NLP, заключающийся в присвоении каждому слову в тексте грамматической категории (например, существительное, глагол, прилагательное).

Среди существующих решений для извлечения структурированных данных из открытых репозиториев выделяются iSeq и SampleExplorer, функциональность которых принципиально отличается от NLP-подхода Alembic. iSeq функционирует как инструмент командной строки для пакетной загрузки NGS-данных по известным идентификаторам из репозиториев GSA/SRA/ENA/DDBJ. Он автоматизирует скачивание с проверкой целостности, но полностью исключает семантический анализ метаданных, работая только с предопределенными списками идентификаторов. SampleExplorer, напротив, применяет языковые модели для поиска транскриптомных данных в архивах типа ARCHS4. Его алгоритм сочетает векторное представление текстовых метаданных с транскриптомным сходством, выявляя семантически близкие эксперименты по генам или описаниям. Однако инструмент остается ориентированным на поиск аналогичных образцов, а не на детальную аннотацию сущностей внутри описаний конкретных экспериментов.

В отличие от iSeq и SampleExplorer, Alembic обеспечивает принципиально иной уровень обработки данных: пакет выполняет глубинный NLP-анализ биомедицинских текстов, включая метаданные экспериментов, с применением передовых моделей – BERT, BioBERT, scispaCy и AIONER. Ключевое преимущество Alembic заключается в использовании PubMedBERT, дообученного на обширных биомедицинских корпусах (PubMed, PMC). Эта модель демонстрирует превосходство как над базовым BERT, так и над классическими подходами, достигая исключительной точности в распознавании сложной биомедицинской терминологии и контекстуальных зависимостей за счет платформы AIONER, составляющей ядро Alembic (Luo et al., 2023).

Заключение

Система Alembic реализует принципиально новый подход к анализу биомедицинских данных открытых репозиториев через глубинное NLP-структурирование информации. Alembic генерирует готовые к импорту выгрузки метаданных (совместимые с SQLite/PostgreSQL) и автоматизирует получение соответствующих сырых данных секвенирования. Главное преимущество решения заключается в создании специализированных, проектно-ориентированных, локальных баз данных. Используемая в Alembic платформа AIONER превосходит аналогичные инструменты благодаря комплексной обработке текстов на основе современных языковых моделей, обеспечивающей точную категоризацию биомедицинских сущностей. Интуитивный интерфейс Alembic с интерактивной визуализацией результатов является принципиальным упрощением по сравнению с инструментами поиска, использующими командную строку, облегчая рутинный поиск данных.

Таким образом, пакет Alembic существенно облегчает операции по извлечению информации из репозиториев открытого доступа, преобразуя неструктурированные метаданные в формализованное машинно-читаемое знание.

Список литературы / References

Aronson A.R., Lang F.M. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc*. 2010;17(3):229-236. doi 10.1136/jamia.2009.002733

Chao H., Li Z., Chen D., Chen M. iSeq: an integrated tool to fetch public sequencing data. *Bioinformatics*. 2024;40(11):btae641. doi 10.1093/bioinformatics/btae641

Chin W.L., Lassmann T. SampleExplorer: using language models to discover relevant transcriptome data. *Bioinformatics*. 2024;41(1):btae759. doi 10.1093/bioinformatics/btae759

Devlin J., Chang M.W., Lee K., Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv.* 2019. doi 10.48550/arXiv.1810.04805

Lee J., Yoon W., Kim S., Kim D., Kim S., So C.H., Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics.* 2020;36(4):1234-1240. doi 10.1093/bioinformatics/btz682

Luo L., Wei C.-H., Lai P.-T., Leaman R., Chen Q., Lu Z. AIONER: all-in-one scheme-based biomedical named entity recognition using deep learning. *Bioinformatics.* 2023;39(5):btad310. doi 10.1093/bioinformatics/btad310

Neumann M., King D., Beltagy I., Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. In: Proceedings of the 18th BioNLP Workshop and Shared Task. Association for Computational Linguistics, 2019;319-327. doi 10.18653/v1/W19-5034

Sayers E. The E-utilities in-depth: parameters, syntax and more. In: Entrez Programming Utilities Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US), 2022. Available at: https://www.ncbi.nlm.nih.gov/books/NBK25499/. Accessed: Jul. 30, 2025

Wang X., Zhang Y., Ren X., Zhang Y., Zitnik M., Shang J., Langlotz C., Han J. Cross-type biomedical named entity recognition with deep multi-task learning. *Bioinformatics*. 2019;35(10):1745-1752. doi 10.1093/bioinformatics/bty869

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 06.08.2025. После доработки 21.10.2025. Принята к публикации 06.11.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):299-310

doi 10.18699/vjgb-26-34

Разработка и валидация программы PipeSeq для анализа данных секвенирования РНК на модели *Chlamydomonas reinhardtii*

А.М. Нерезенко, П.А. Виролайнен, С.А. Тупицына, Е.М. Чекунова

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
alexnerezenko@gmail.com

Аннотация. Секвенирование РНК (РНК-сек) – высокочувствительный метод анализа транскриптома, позволяющий одновременно оценивать экспрессию тысяч генов и выявлять паттерны экспрессии в различных условиях. Существующее разнообразие форматов данных РНК-сек, методов нормализации и подходов к статистической обработке результатов затрудняет сопоставление данных разных исследований и снижает воспроизводимость анализа. В настоящей работе представлен автоматизированный пайплайн PipeSeq, объединяющий стандартные этапы обработки данных РНК-сек – от загрузки (SRA Toolkit), выравнивания прочтений на референсный геном (HISAT2) и сборки транскриптов (StringTie) до подсчета транскриптов (FeatureCounts) и статистического анализа дифференциальной экспрессии генов в различных экспериментальных условиях (DESeq2). Программа PipeSeq имеет простой визуальный интерфейс, поддерживает многопоточность и формирует готовые для анализа тепловые карты экспрессии генов и отчеты в форме таблиц и графиков. Функциональность пайплайна продемонстрирована на трех наборах пакетов сырых данных секвенирования РНК клеток зеленой водоросли *Chlamydomonas reinhardtii*, доступных в открытой базе данных NCBI SRA. Результаты этих экспериментов были использованы для анализа дифференциальной экспрессии генов *C. reinhardtii*, кодирующих факторы транскрипции семейства GATA, в различных световых условиях культивирования. Полученные методами *in silico* данные верифицированы методом полимеразной цепной реакции в реальном времени с обратной транскрипцией (ОТ-ПЦР-РВ) по 12 генам GATA, что позволило выдвинуть предположения об их функциях, а также оценить степень согласованности между массовым (РНК-сек) и таргетным (ОТ-ПЦР-РВ) подходами. Результаты нашего исследования показали, что методы секвенирования РНК и ОТ-ПЦР-РВ выявляют схожие направления изменения экспрессии генов, но демонстрируют различия по оценке степени размера эффекта и чувствительности, что подчеркивает необходимость совместного применения двух подходов. Таким образом, программа PipeSeq представляет собой инструмент для проведения полного цикла биоинформатического анализа данных РНК-сек, дает возможность обрабатывать данные ОТ-ПЦР-РВ и выполнять сравнительный статистический анализ полученных результатов.

Ключевые слова: секвенирование РНК; РНК-сек; ОТ-ПЦР-РВ; пайплайн; транскриптом; экспрессия генов; факторы транскрипции семейства GATA; ФТ GATA; *Chlamydomonas reinhardtii*

Для цитирования: Нерезенко А.М., Виролайнен П.А.,Тупицына С.А., Чекунова Е.М. Разработка и валидация программы PipeSeq для анализа данных секвенирования РНК на модели *Chlamydomonas reinhardtii*. *Вавиловский журнал генетики и селекции*. 2026; 30(2):299-310. doi 10.18699/vjgb-26-34

Финансирование. Работа выполнена при поддержке СПбГУ, шифр проекта 124032000041-1.

Development and validation of the PipeSeq program for RNA-seq data analysis in the *Chlamydomonas reinhardtii* as a model

A.M. Nerezenko, P.A. Virolainen, S.A. Tupitsyna, E.M. Chekunova

Saint-Petersburg University, St. Petersburg, Russia
alexnerezenko@gmail.com

Abstract. RNA sequencing (RNA-seq) is a highly sensitive method for transcriptome analysis that allows simultaneous assessment of expression of thousands of genes and identification of expression patterns under various conditions. The existing variety of RNA-seq data formats, normalization methods, and approaches to statistical processing of results complicates comparison of data from different studies and reduces reproducibility of the analysis. This study presents an automated pipeline PipeSeq that combines standard steps of RNA-seq data processing: loading (SRA Toolkit), read alignment to the reference genome (HISAT2), transcript assembly (StringTie), transcript counting (FeatureCounts) and statistical analysis of differential gene expression under various experimental conditions (DESeq2). PipeSeq has a simple visual interface, supports multithreading, and generates ready-to-analyze gene expression heat maps, tables and graphs. The functionality of the pipeline is demonstrated on three sets of raw RNA-seq data from the green alga *Chlamydomonas reinhardtii* cells available in the NCBI SRA database. The data from these experiments were used to

© Нерезенко А.М., Виролайнен П.А.,Тупицына С.А., Чекунова Е.М., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

analyze the differential expression of *C. reinhardtii* genes encoding the GATA family transcription factors under different light cultivation conditions. The data obtained by *in silico* methods were verified by real-time reverse transcription polymerase chain reaction (RT-qPCR) for 12 GATA genes, which allowed us to hypothesize their functions and evaluate the correlation between the bulk (RNA-seq) and targeted (RT-qPCR) approaches. Our results showed that RNA-seq and RT-qPCR methods reveal similar directions of gene expression changes, but demonstrate differences in the effect size and sensitivity, which emphasizes the need for a combined use of the two approaches. Thus, the PipeSeq program is a tool for conducting a full cycle of bioinformatic analysis of RNA-seq data, additionally providing the opportunity to process RT-qPCR data and perform a comparative statistical analysis of the results obtained.
Key words: RNA sequencing; RNA-seq; RT-qPCR; pipeline; transcriptome; gene expression; GATA family transcription factors; GATA TFs; *Chlamydomonas reinhardtii*

For citation: Nerezenko A.M., Virolainen P.A., Tupitsyna S.A., Chekunova E.M. Development and validation of the PipeSeq program for RNA-seq data analysis in the *Chlamydomonas reinhardtii* as a model. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):299-310. doi 10.18699/vjgb-26-34

Введение

В последние годы метод секвенирования РНК (РНК-сек) получил широкое распространение как надежный подход для массового количественного анализа экспрессии генов в различных биологических системах (Marioni et al., 2008; Mortazavi et al., 2008; Conesa et al., 2016; Li X., Wang, 2021). Благодаря «цифровой» природе данных метод РНК-сек позволяет напрямую подсчитывать количество прочтений для каждого транскрипта, обеспечивая широкий динамический диапазон измерений и высокую воспроизводимость результатов эксперимента. Кроме того, этот метод дает возможность обнаруживать ранее не аннотированные транскрипты и варианты альтернативного сплайсинга (Wang et al., 2009; Li X., Wang, 2021). Эти особенности делают секвенирование РНК мощным инструментом для системных исследований транскриптома, позволяя надежно измерять уровни транскриптов в различных условиях эксперимента и выявлять дифференциальную экспрессию генов объекта исследования (Wang et al., 2009).

Обработка, накопление и объединение этих данных могут способствовать получению новых знаний о функционировании живых систем. Для реализации подобного подхода необходимо решить проблему стандартизации процесса обработки сырых данных РНК-сек (Conesa et al., 2016; Li X., Wang, 2021). Для оценки экспрессии генов до сих пор применяются устаревшие методы нормализации, которые характеризуются низкой воспроизводимостью, не позволяют корректно нормализовывать композиционные смещения и проводить прямое сравнение результатов между независимыми исследованиями, – это RPKM (reads per kilobase per million mapped reads), FPKM (fragments per kilobase of exon model per million mapped fragments) и TPM (transcripts per million). В настоящее время стандартом для проведения анализа дифференциальной экспрессии генов считаются методы, основанные на отрицательном биномиальном распределении, – DESeq2 и edgeR. Нормализованные счетчики (DESeq2) продемонстрировали наименьший коэффициент вариации и максимальную воспроизводимость (Zhao S. et al., 2020; Zhao Y. et al., 2021; Elahimanesh, Najafi, 2024).

Существующее разнообразие методов обработки данных РНК-сек, форматов и способов их представления в различных базах данных затрудняет консолидацию сведений для проведения комплексного анализа. Для решения этой проблемы требуется разработка интегрированных инструментов, способных минимизировать ручной труд, обеспечить высокую степень воспроизводимости результатов и простоту использования для биологов без специальной подготовки в области информационных технологий (Conesa et al., 2016).

Проведение сравнительного анализа результатов обработки данных РНК-сек и молекулярно-биологических данных осложнено рядом методологических проблем. Хотя секвенирование РНК признано достоверным методом глобального профилирования экспрессии, полученные с его помощью результаты обычно представляют относительные изменения уровня транскриптов в масштабе всего генома. Исторически сложилась практика применения метода полимеразной цепной реакции в реальном времени с обратной транскрипцией (ОТ-ПЦР-РВ) как «золотого стандарта», позволяющего подтвердить данные, полученные по результатам транскриптомных исследований (Derveaux et al., 2010; Coenye, 2021). Однако методы на основе ПЦР-РВ и технология секвенирования РНК основаны на разных протоколах, что затрудняет их прямое количественное сравнение: на получаемые результаты могут оказывать влияние эффективность процессов обратной транскрипции и амплификации, методы нормализации данных, чувствительность методик.

В настоящее время метод РНК-сек используется для охвата всего транскриптома объекта исследования и выявления круга дифференциально экспрессируемых генов-кандидатов в ответ на определенное воздействие, тогда как метод ОТ-ПЦР-РВ применяется для точной количественной оценки этих изменений прицельно для ограниченного пула генов интереса (Shi, He, 2014; He et al., 2015; Coenye, 2021).

Для понимания механизмов регуляции метаболизма особое внимание уделяется изучению факторов транскрипции (ФТ) как ключевых регуляторов активности генов. ФТ представляют собой белки, способные связываться с определенными последовательностями ДНК в промоторных областях и тем самым усиливать или подавлять транскрипцию целевых генов. Эти регуляторы координируют работу генов в ответ на различные изменения среды и сигнальные воздействия, участвуя в глобальных процессах роста, развития и адаптации к стрессовым факторам (Riechmann et al., 2000).

У фотосинтезирующих организмов особый интерес представляют гены, кодирующие ФТ семейства GATA, – белки, несущие консервативный домен цинкового пальца типа IV (общая формула: $CX_2CX_{18–20}CX_2C$). Этот домен осуществляет связывание консенсусной последовательности (A/T)GATA(A/G) в промоторах целевых генов (Reyes et al., 2004). Факторы GATA растений участвуют в регуляции процессов фотоморфогенеза, метаболизма азота и углерода, гормонального контроля (Manfield et al., 2007; Naito et al., 2007; Luo et al., 2010; Schwechheimer et al., 2022; Schröder et al., 2023; Ren et al., 2025). В последние годы наблюдается существенный рост научного интереса к исследованию ФТ GATA у мохообразных и водорослей, поскольку их биологические функции и эволюционная роль остаются недостаточно изученными (Schwechheimer et al., 2022; Virolainen, Chekunova, 2024).

Целью настоящего исследования стала разработка интегрированного решения для унификации процессов обработки сырых данных секвенирования РНК и результатов ОТ-ПЦР-РВ.

Для реализации этой цели необходимо:

1. Разработать автоматизированный алгоритм (пайплайн) для проведения полного цикла анализа данных РНК-сек.
2. Апробировать пайплайн на сырых данных РНК-сек, доступных в открытой базе NCBI SRA, на модели генов, кодирующих ФТ семейства GATA у модельного объекта генетики фотосинтеза – зеленой водоросли *Chlamydomonas reinhardtii*, в ответ на изменение световых условий выращивания.
3. Провести анализ экспрессии генов GATA *C. reinhardtii* методом ОТ-ПЦР-РВ в ответ на изменение световых условий выращивания.
4. Провести сравнительный статистический анализ полученных данных РНК-сек и ОТ-ПЦР-РВ.

Материалы и методы

Количественный анализ экспрессии генов в программе PipeSeq. Для анализа экспрессионного профиля 12 генов, кодирующих ФТ GATA у *C. reinhardtii* в различных условиях культивирования методами *in silico*, были отобраны следующие доступные пакеты сырых данных РНК-сек из базы данных NCBI SRA (Wheeler et al., 2005): SRX8380269, SRX8380270, SRX8380271 (повышенная освещенность – 600 мкмоль/м^2/с, 1 ч), SRX7413406, SRX7413407, SRX7413412, SRX7413413, SRX7413414, SRX7413415 (акклиматизация к свету, 30 мин), SRX5120530, SRX5120531, SRX5120532, SRX5120533, SRX5120534, SRX5120535 (акклиматизация к темноте).

Выбранные три набора, PRJNA634446, PRJNA596622, PRJNA509798, содержат данные эксперимента (выращивание в измененных условиях) и данные контроля (выращивание в стандартных условиях) для штамма *C. reinhardtii* дикого типа CC-124 (*wt*, *mt*-), выполненные в трехкратной биологической и технической повторности. Для выравнивания прочтений с портала Phytozome (Goodstein et al., 2012) были загружены геном *C. reinhardtii* (v5.6) (Merchant et al., 2007) и его аннотация в формате GTF.

Загрузку и обработку данных осуществляли с помощью PipeSeq – автоматизированного программного комплекса из 17 скриптов и более 2000 строк кода, предназначенного для системного анализа транскриптомных данных, полученных методом РНК-сек, а также обработки данных ПЦР-РВ методом ΔΔCt (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008) и проведения сравнительного анализа результатов (с расчетом коэффициентов Пирсона, Спирмена, Кендалла). Программа написана на языке программирования Python (версия 3.9) с использованием библиотек Pandas (McKinney, 2011), Matplotlib (Hunter, 2007), PyQt6 (https://www.riverbankcomputing.com/software/pyqt/), PyDESeq2 (Muzellec et al., 2023) и инструментов биоинформатики SRA Toolkit, FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), MultiQC (Ewels et al., 2016), Cutadapt (Martin, 2011), HISAT2 (Kim et al., 2019), SAMtools (Li H. et al., 2009), FeatureCounts (Liao et al., 2014), StringTie (Pertea et al., 2015), DESeq2 (Love et al., 2014).

Штаммы и условия культивирования. Штамм дикого типа *C. reinhardtii* CC-124 (*wt*, *mt*-) из Петергофской генетической коллекции СПбГУ (Квитко и др., 1983) выращивали в чашках Петри на агаризованной среде TAP (1.5 %) (Harris, 1989) с добавлением аргинина (50 мг на литр среды) и дрожжевого автолизата (4 г на литр среды) при температуре 20–25 °C в цикле день (14 ч, освещенность 90 мкмоль/м^2/с) – ночь (10 ч) с пересевом каждые три дня. Для работы отбирали культуры в стандартных условиях освещенности (90 мкмоль/м^2/с) и темноты (контрольные условия), в условиях повышенной освещенности (215 мкмоль/м^2/с) в течение 30 мин и 2 ч (перенос культур из стандартных условий), в условиях стандартной освещенности в течение 30 мин и 2 ч (перенос культур из темноты), в условиях темноты в течение 30 мин и 2 ч (перенос культур из стандартных условий) (экспериментальные условия).

Фиксация культур и выделение РНК. Фиксацию выращенных в контрольных и экспериментальных условиях культур клеток *C. reinhardtii* и выделение суммарной РНК проводили с использованием реагента “ExtractRNA” («Евроген», Россия) в строгом соответствии с инструкцией производителя. Препараты РНК очищали от примеси геномной ДНК с помощью фермента ДНКаза I (Thermo Fisher Scientific, США) и переосаждали этанолом. Концентрацию суммарной РНК измеряли на спектрофотометре Eppendorf BioPhotometer plus (Eppendorf, Германия).

Дизайн праймеров. Ген-специфичные праймеры подбирали с помощью программ IDT PrimerQuest Tool (https://www.idtdna.com/pages/tools/primerquest) и NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) таким образом, чтобы место посадки одного из праймеров располагалось на стыке экзонов либо между сайтами отжига праймеров находилась последовательность интрона. В качестве референс-генов для нормализации получаемых значений задействовали конститутивно экспрессируемые гены *RPL19* (*ribosomal protein L19*) и *RPL32* (*ribosomal protein L32*) *C. reinhardtii* (Liu et al., 2012); для контроля

Последовательности праймеров для количественного анализа экспрессии генов GATA

Ген	Идентификатор	Праймер	Последовательность (5′–3′)	Литературный источник
GATA-1	Cre01.g025050	F	GTTGTTTGGCGACCTCTTTGTG	Настоящее исследование
		R	GATCAGCGGCGGCTATGTC	
GATA-2	Cre10.g435450	F	ACTACGACGAGCGGGAAGA	
		R	GCCTTCTTCGCCATGTACTCC	
GATA-3	Cre08.g378800	F	GAGCTGGACGGGAACGAAAC	
		R	GTGCGGTGCCGAGTAGTTT	
GATA-4	Cre11.g467581	F	GATCCTATCACCACCAAGGTTGC	
		R	CCATGCCGCCCATGTTCA	
GATA-5	Cre03.g160600	F	TCACGGGACGACGACATCA	
		R	CGGGTGAAGAATATGCCACAGG	
GATA-6	Cre03.g160700	F	GAAAAGGCAGGACAAGTCCAAG	
		R	TGTGAGGCGGGATGAAGAT	
GATA-7	Cre05.g242600	F	AGGAGCAGCAGCAGCAATC	
		R	CTGGTTAGTGCGGCGGTATC	
GATA-8	Cre06.g266850	F	TGTGCAACGCATGTGGGATA	
		R	CGGTCTTGGCTGACACATAGTT	
GATA-9	Cre06.g266950	F	ACATCAGCGGCTGCGATAAT	
		R	CGCCTGAGCCACTTTCGG	
GATA-10	Cre07.g319701	F	TCCGCTGCTGCGTAGAGT	
		R	GCAAAGACATCCTCGTCGGC	
GATA-11	Cre08.g358532	F	TCAGCAACAGCCCTCACTTC	
		R	CGCTCAAACCACTTGACCTCTAT	
GATA-12	Cre08.g358534	F	TGTCAAGTGTTTCCACGACAAGA	
		R	GCACCAGAACCACTCGCA	
RPL32	Cre06.g289550	F	CCCAACGGCTTCCTGAAGTA	
		R	AAGCGACGGTTGTGCATCAT	
RPL19	Cre02.g075700	F	CCTGAAGAAGTACCGCGACTC	Liu et al., 2012
		R	AACACGTTACCCTTGACCTTCA	
RBCS	Cre02.g120150	F	ACCCCGGTCAACAACAAGATG	Sanchez-Tarre, Kiparissides, 2021
		R	GTCGTAGTACAGGCAAGACACG	

изменения условий эксперимента использовали ген *RBCS* (*Ribulose bisphosphate carboxylase small subunit*) *C. reinhardtii* (Sanchez-Tarre, Kiparissides, 2021). Последовательности праймеров приведены в таблице.

Для оценки специфичности использовали анализ кривых плавления праймеров на серии контролей (контроль без матрицы, контроль без обратной транскрипции, положительный контроль), а также визуализацию ПЦР-продуктов методом гель-электрофореза (Derveaux et al., 2010). Эффективность ПЦР оценивали средствами программного обеспечения амплификатора QuantStudio 5 (Thermo Fisher Scientific).

Количественный анализ экспрессии генов методом ОТ-ПЦР-РВ. Реакции ПЦР-РВ проводили в один шаг с использованием компонентов набора реагентов “OneTube RT-PCR SYBR” («Евроген», Россия), строго следуя инструкции производителя, на амплификаторе QuantStudio 5 (Thermo Fisher Scientific) со считыванием флюоресценции на шаге элонгации и плавления по следующему протоколу: 55 °C – 15 мин, 95 °C – 1 мин, далее 50 циклов: 95 °C – 15 с, 62 °C – 20 с, 72 °C – 20 с, плавление: 55–95 °C с шагом 0.5 °C. Каждый образец анализировали в трехкратной усредненной биологической повторности (Derveaux et al., 2010). Обработку и визуализацию данных осуществляли методом ΔΔCt (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008) в разработанной программе PipeSeq.

Сравнительный анализ результатов РНК-сек и ОТ-ПЦР-РВ. Сравнительный статистический анализ полученных разными методами данных выполняли в программе PipeSeq.

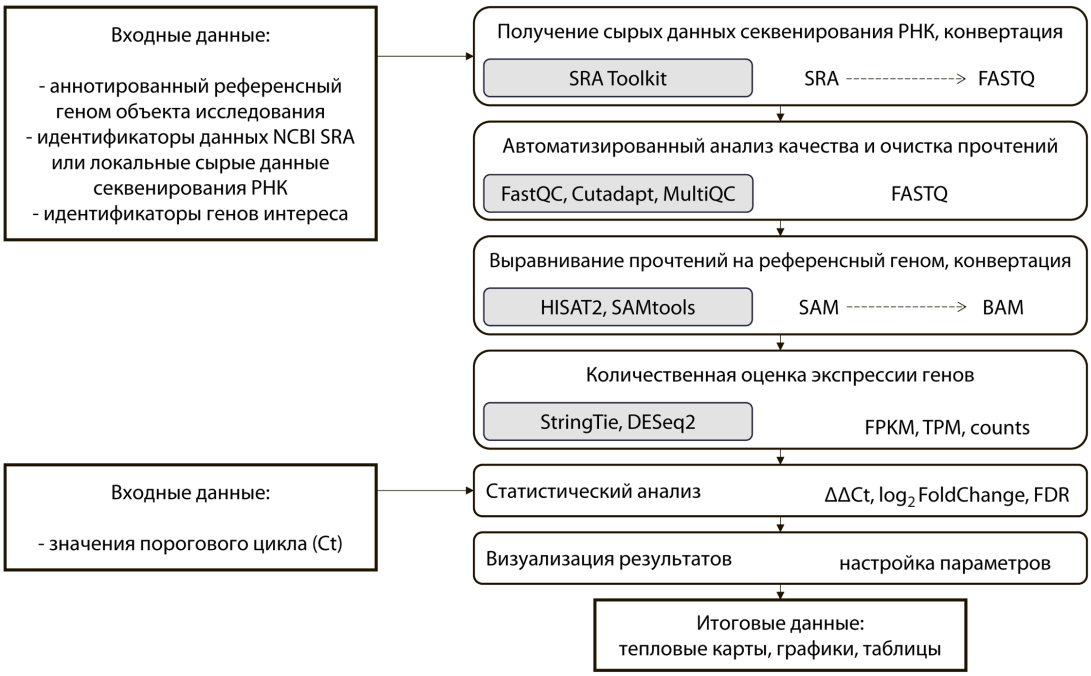


Рис. 1. Схема работы программы PipeSeq.

Результаты

Разработка пайплайна PipeSeq

Целью работы было создание такого инструмента (пайплайна), который позволил бы проводить полный цикл анализа данных РНК-сек через простой визуальный интерфейс с минимальной потребностью для пользователя в системном администрировании и был ориентирован на запуск в локальной среде Windows с использованием подсистемы WSL для выполнения команд Linux. Схема работы программы PipeSeq представлена на рис. 1.

Все шаги исполнения алгоритма полностью автоматизированы и позволяют анализировать большие объемы данных за счет оптимизации команд и параллельных вычислений. Входными данными являются файлы формата FASTQ, содержащие необработанные короткие прочтения секвенирования, получаемые после конвертации из формата SRA с помощью входящего в пакет программы инструмента SRA Toolkit. На начальном этапе выполняется автоматизированная подготовка прочтений, включающая контроль качества (FastQC), удаление последовательностей адаптеров, нуклеотидов с низким качеством, коротких прочтений (Cutadapt), агрегирование сводных отчетов (MultiQC). Далее проводится выравнивание прочтений на референсный геном с помощью инструмента HISAT2 (Kim et al., 2019), при этом автоматически создается индекс референсного генома в случае его отсутствия. Преимущества HISAT2 – учет сплайсинга, точное определение структуры транскриптов и быстрый анализ качества выравнивания. После выравнивания осуществляются конвертация файлов из формата SAM в формат BAM, сортировка и подготовка к дальнейшему анализу с использованием инструментов SAMtools (Li H. et al., 2009).

Разработанная программа PipeSeq интегрирует три подхода нормализации: DESeq2 (для анализа дифференциальной экспрессии), TPM (абсолютная квантификация) и FPKM (устаревший показатель, включен для обеспечения обратной совместимости). На данном этапе проводятся количественная оценка экспрессии генов методом подсчета числа прочтений, приходящихся на гены и транскрипты, с использованием пакета FeatureCounts (Liao et al., 2014), а также сборка транскриптов и расчет значений FPKM и TPM с помощью инструмента StringTie (Pertea et al., 2015), позволяющего анализировать выравнивания ридов, строить транскрипты и определять экзоны, интроны и сплайс-сайты. В программе PipeSeq реализована возможность отключать функцию построения новых транскриптов для минимизации вероятности получения ложных результатов. Данные этапы выполняются с учетом различных режимов, таких как строгая аннотация (опция «-e» в StringTie) и изменяемая чувствительность (параметр «-c»).

Для статистического анализа дифференциальной экспрессии генов используется пакет на основе отрицательного биномиального распределения DESeq2 (Love et al., 2014), реализованный в библиотеке PyDESeq2 (Muzellec et al., 2023). Пайплайн PipeSeq автоматически готовит входные данные для DESeq2, осуществляет статистический расчет и формирует итоговые таблицы с логарифмическими изменениями уровня экспрессии ($\log_2$FoldChange) и соответствующими значениями скорректированного p-значения – FDR (false discovery rate) по методу Бенджамини–Хохберга. Статистические гипотезы о дифференциальной экспрессии генов проверяются на основе обобщенной линейной модели отрицательного биномиального распределения. После нормализации для счета данных и

оценки дисперсии (с последующим усреднением оценок дисперсии по всем генам с помощью байесовского сглаживания) применяется тест Уолда.

В программе также реализована возможность обработки данных ПЦР-РВ методом ΔΔCt (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008) и проведения сравнительного статистического анализа с тремя используемыми методами нормализации данных РНК-сек (коэффициент линейной корреляции Пирсона, коэффициент ранговой корреляции Спирмена, коэффициент ранговой корреляции Кендалла).

На выходе программа PipeSeq формирует тепловые карты экспрессии, отражающие изменения в уровне экспрессии генов интереса при различных экспериментальных условиях, итоговые таблицы со значениями $\log_2$FoldChange и графики. Пользователю предоставлен широкий выбор настроек отображения данных. Программа доступна для скачивания по ссылке: https://github.com/MarvinMarss/PipeSeq.

Обработка доступных данных секвенирования РНК клеток C. *reinhardtii* с помощью пайплайна PipeSeq

С использованием разработанной программы мы проанализировали три доступных набора данных РНК-сек клеток *C. reinhardtii* штамма дикого типа CC-124 (*wt*, *mt-*) при изменении условий освещения – повышенная освещенность (600 мкмоль/м^2/с) в течение 1 ч, акклиматизация к свету в течение 30 мин, акклиматизация к темноте. Построенная тепловая карта демонстрирует сложную динамику экспрессии генов, кодирующих ФТ GATA, в ответ на изменение световых условий (рис. 2). Большинство полученных значений логарифма изменения уровня экспрессии генов оказались статистически незначимыми.

В условиях повышенной освещенности наблюдается значимое повышение уровня экспрессии генов *GATA-7*, *GATA-9*, *GATA-10*, *GATA-11*, в то время как транскрипция генов *GATA-3*, *GATA-5*, *GATA-8* значимо репрессируется. Вероятно, излишне высокая интенсивность освещения оказывает негативное влияние на жизнедеятельность клеток *C. reinhardtii*.

В ходе акклиматизации к свету в течение 30 мин происходит перенастройка метаболизма клеток, что маркирует значимое повышение уровня транскриптов генов *GATA-3*, *GATA-5*, *GATA-6*, *GATA-7*, *GATA-8*. Экспрессия генов *GATA-2*, *GATA-4*, *GATA-10* в этих условиях достоверно репрессируется.

В темноте достоверно происходят подавление экспрессии генов *GATA-2*, *GATA-3*, *GATA-7*, *GATA-8*, *GATA-11*, *GATA-12* и активная экспрессия гена *GATA-9*.

Доступные наборы данных характеризуются ограниченностью условий эксперимента и скудной охарактеризованностью, а также низким уровнем достоверности изменений. Мы приняли решение воспользоваться методом ОТ-ПЦР-РВ для получения полных данных по экспрессии генов GATA. Анализ данных РНК-сек позволил нам осуществить выбор референс-генов со стабильной экспрессией при изменении условий освещения для проведения собственных экспериментов – были обнаружены и затем подтверждены литературными данными (Liu et al., 2012) гены *RPL19* и *RPL32*, кодирующие рибосомные белки. Экспрессия гена *RBCS*, согласно литературным данным (Sanchez-Tarre, Kiparissides, 2021), изменяется в зависимости от спектра и интенсивности освещения, в связи с чем этот ген выбран нами в качестве контроля изменения условий эксперимента.

Анализ экспрессии генов GATA *C. reinhardtii* методом ОТ-ПЦР-РВ

Полученные методом ОТ-ПЦР-РВ результаты анализа экспрессии генов GATA *C. reinhardtii* представлены на рис. 3.

Вся группа генов GATA отвечает на изменение условий освещения быстро и согласованно. Самый сильный ответ наблюдается в первые 30 мин повышенной освещенности (215 мкмоль/м^2/с) после переноса культуры из стандартных условий (90 мкмоль/м^2/с): почти все гены демонстрируют кратное увеличение уровня экспрессии в ответ на стресс-воздействие. Примечательно, что только в этих условиях происходит значительная активация экспрессии гена *GATA-6*. Повышенная освещенность интенсифицирует метаболизм клеток *C. reinhardtii*, обеспечивая единое направление изменения профиля экспрессии изучаемых генов. Через 2 ч наблюдается стабилизация – экспрессия ряда генов (*GATA-1*, *GATA-6*, *GATA-8*, *GATA-9*, *GATA-10*, *GATA-11*, *GATA-12*) значительно снижается, а генов *GATA-5* и *GATA-7* – усиливается.

В первые 30 мин после переноса культуры из темноты на свет (90 мкмоль/м^2/с) происходит перенастройка метаболизма – наблюдается репрессия транскрипции генов *GATA-1* и *GATA-3* одновременно с активацией генов *GATA- 2*, *GATA-4*, *GATA-5*, *GATA-7*, *GATA-9* и *GATA-10*. К 2 ч экспозиции формируется два профиля, обеспечивающих поддержание процессов роста и развития в оптимальных световых условиях: 1) активно транскрибируемые гены (*GATA-1*, *GATA-4*, *GATA-5*, *GATA-8*, *GATA-9*, *GATA-11*, *GATA-12*); 2) репрессированные гены (*GATA-2*, *GATA-3*, *GATA-6*, *GATA-7*, *GATA-10*).

В первые 30 мин после переноса культуры из света (90 мкмоль/м^2/с) в темноту происходит подавление экспрессии всех генов GATA, за исключением гена *GATA-7*. По-видимому, его продукт участвует в процессах переключения метаболизма клеток *C. reinhardtii* при переходе свет/темнота и темнота/свет. Через 2 ч происходит стабилизация метаболизма клеток с выявлением трех активно транскрибируемых генов, по-видимому, задействованных в обеспечении адаптации к темноте, – *GATA-1*, *GATA-3*, *GATA-9*.

Выявленная динамика экспрессии генов GATA согласуется со стандартной моделью ответа на возникающий стресс: изменение световых условий запускает широкий аварийный каскад (быстрый ответ на раздражитель); длительная экспозиция сужает реакцию до специфических регуляторных модулей. В каждых условиях культивирования наблюдается уникальный профиль экспрессии,

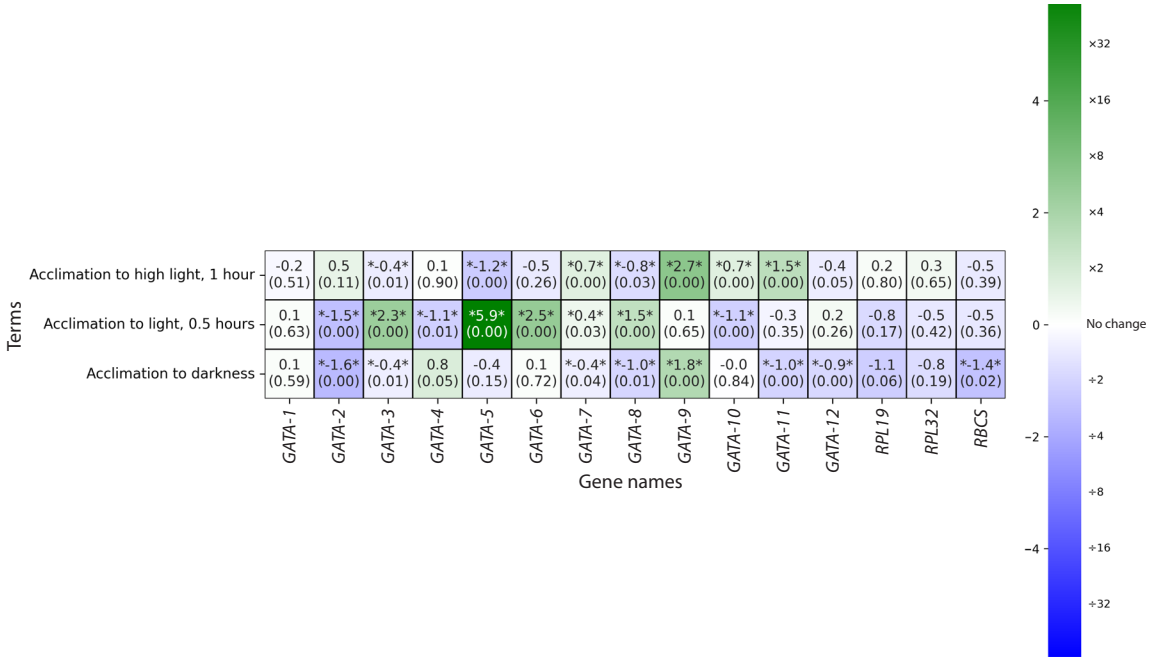


Рис. 2. Тепловая карта экспрессии (на уровне транскрипции) 12 генов, кодирующих факторы транскрипции семейства GATA и трех контрольных генов (*RPL19*, *RPL32*, *RBCS*) *C. reinhardtii* при различных условиях акклиматизации по данным секвенирования РНК.

Слева направо – гены; сверху вниз – экспериментальные условия: акклиматизация к высокому освещению в течение 1 ч (ряд 1), акклиматизация к свету в течение 30 мин (ряд 2), акклиматизация к темноте (ряд 3).
Здесь и на рис. 3: ячейки окрашены в соответствии со значениями $\log_2$FoldChange, где положительные значения (зеленая шкала) отражают индукцию экспрессии, отрицательные (синяя шкала) – репрессию, а нулевые значения (белый цвет) – сохранение уровня экспрессии. Статистическая значимость изменений оценивалась по скорректированным p-значениям с порогом 0.05. Контрольные транскрипты включены для верификации качества нормализации данных. Изображение создано в программе PipeSeq.

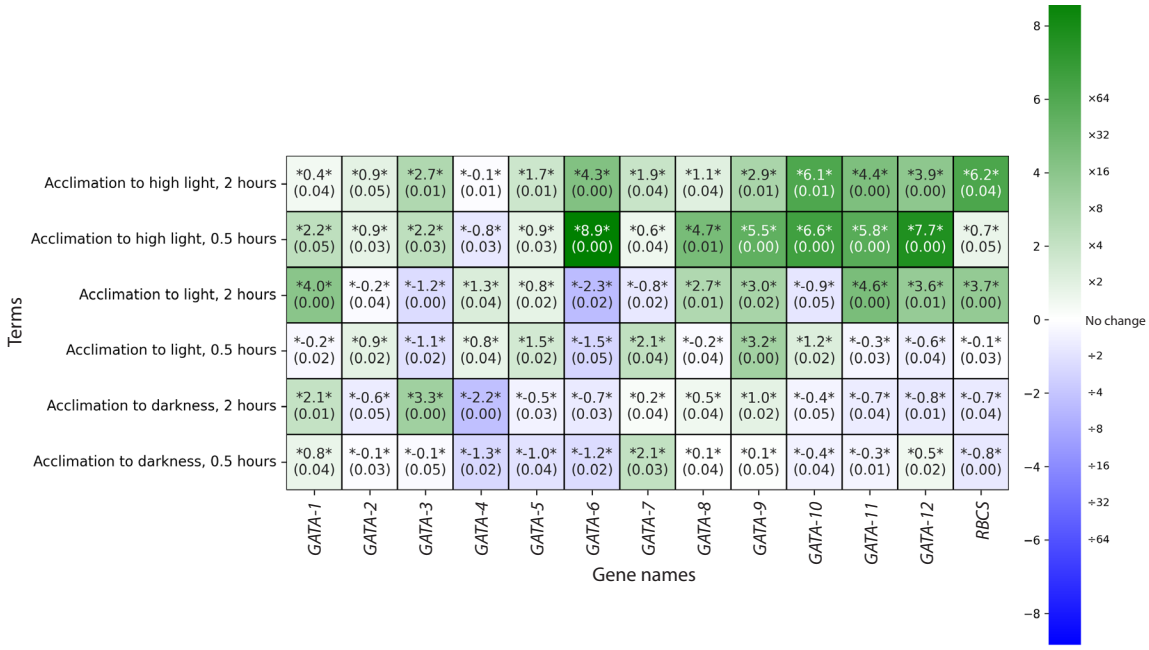


Рис. 3. Тепловая карта экспрессии (на уровне транскрипции) 12 генов, кодирующих факторы транскрипции семейства GATA и гена-контроля изменения условий (*RBCS*) *C. reinhardtii* при различных условиях акклиматизации по данным ОТ-ПЦР-РВ.

Данные нормированы на пару референс-генов *RPL19–RPL32*. Слева направо – гены; сверху вниз – экспериментальные условия: акклиматизация к высокому освещению в течение 2 ч (ряд 1), в течение 30 мин (ряд 2), акклиматизация к свету в течение 2 ч (ряд 3), в течение 30 мин (ряд 4), акклиматизация к темноте в течение 2 ч (ряд 5), в течение 30 мин (ряд 6).

причем часть генов GATA предположительно действует как «переключатель» между световой и темновой программами метаболизма.

Сравнительный анализ полученных результатов

Возможность проведения сравнительного анализа по результатам обработки данных РНК-сек с применением трех методов нормализации, FPKM, TPM, DESeq2, и обработки результатов ОТ-ПЦР-РВ методом ΔΔCt с построением корреляционной матрицы в программе PipeSeq позволяет оценить степень согласованности полученных разными методами и подходами данных (рис. 4).

Межплатформенная согласованность полученных результатов ограничена и зависит от контекста. Несмотря на методологические различия, воспроизводимые совпадения включают индукцию экспрессии генов *GATA-7*, *GATA-9*, *GATA-10* и *GATA-11* при повышенном освещении, репрессию генов *GATA-2*, *GATA-11*, *GATA-12* в темноте и усиление экспрессии гена *GATA-9* в темноте. Ранняя репрессия генов *GATA-3*, *GATA-5*, *GATA-8* при повышенном освещении через 1 ч, по данным РНК-сек, оказалась инвертированной, по данным ОТ-ПЦР-РВ, через 2 ч, что согласуется с потенциальным сдвигом фазы реакции, но для подтверждения этого требуется выборка, согласованная по времени. В темноте оба набора данных подтверждали общую репрессию связанных с фотосинтезом транскриптов генов *RBCS*, *GATA-2*, *GATA-11*, *GATA-12* и повышением *GATA-9*, что указывает на скоординированную программу адаптации к темноте, полученную обоими методами.

Корреляция между методами продемонстрировала сильную внутреннюю согласованность между показателями, полученными на основе данных РНК-сек (DESeq2, FPKM, TPM), и более слабую согласованность с ΔΔCt, что, вероятно, отражает различия во времени отбора проб, поведении референс-генов и разрешающей способности методик. Наибольшая корреляция наблюдается в паре DESeq2–ΔΔCt (см. рис. 4). Результаты сравнительного анализа показывают, что, хотя обе платформы фиксируют перекрывающиеся регуляторные тренды для некоторых генов GATA, каждая из них также обнаруживает специфические для условий и времени изменения в экспрессии. Это подчеркивает ценность интеграции обоих подходов.

Обсуждение

Преимущества разработанной программы PipeSeq

Программа PipeSeq разрабатывалась с учетом минимальной потребности в системном администрировании и ориентирована на запуск в локальной среде Windows. Она предоставляет удобный графический интерфейс, который снижает порог вхождения пользователя, объединяет современные методы анализа данных и подходит для работы с малыми и средними наборами данных на обычном персональном компьютере, что делает ее удобным и практичным решением для индивидуальных исследователей и небольших лабораторий, не имеющих значительных вычислительных мощностей и опыта системного администрирования.

На сегодняшний день существует ряд инструментов и платформ для транскриптомного анализа, таких как Galaxy (Afgan et al., 2018), Nextflow (Di Tommaso et al., 2017), Snakemake (Mölder et al., 2021), а также специализированные связки типа HISAT2-StringTie-Ballgown (Pertea et al., 2016) и Kallisto-Sleuth (Bray et al., 2016). Galaxy представляет собой удобную веб-платформу с графическим интерфейсом, однако требует отдельного администрирования сервера и не всегда подходит для индивидуальных исследователей или небольших лабораторий без ИТ-поддержки (Afgan et al., 2018).

Программные платформы Nextflow и Snakemake предлагают высокую гибкость и масштабируемость за счет возможности параллельного запуска задач и поддержки контейнеризации, обеспечивающей полную воспроизводимость анализа, однако использование данных систем требует от исследователей навыков программирования и администрирования среды Linux, что ограничивает их применение биологами без опыта работы с консолью и скриптами (Di Tommaso et al., 2017; Mölder et al., 2021).

Важным аспектом является поддержка анализа дифференциальной экспрессии генов: современные подходы рекомендуют использование пакетов на основе отрицательного биномиального распределения (например, DESeq2), обеспечивающих высокую точность и контроль ложных открытий (Zhao S. et al., 2020; Zhao Y. et al., 2021; Elahimanesh, Najafi, 2024).

Разработанный нами программный пайплайн PipeSeq (см. рис. 1) обеспечивает использование одних из самых быстрых инструментов для выравнивания (HISAT2) и сборки транскриптов (StringTie), которые значительно превосходят предшествующие методы (TopHat и Cufflinks) по скорости и требовательности к вычислительным ресурсам (Kim et al., 2015, 2019), а интеграция DESeq2 позволяет автоматически получать статистически обоснованные результаты дифференциальной экспрессии с контролем FDR (Love et al., 2014). Это ставит его в один ряд с признанными решениями в области анализа данных РНК-сек, такими как связка HISAT2-StringTie-Ballgown, где применяется аналогичный статистический подход, хотя и менее интегрированный (Pertea et al., 2016). Также преимуществом разработанной программы являются возможность обсчета и визуализации данных ПЦР-РВ методом ΔΔCt и проведение сравнительного статистического анализа (с расчетом коэффициента Пирсона, Спирмена, Кендалла) с тремя методами нормализации данных РНК-сек – FPKM, TPM, DESeq2.

Таким образом, разработанная программа PipeSeq обладает следующими особенностями:

1. Взаимодействие пользователя с программой происходит исключительно через графический интерфейс (пять окон).
2. Все этапы обработки данных объединены в рамках одной программы (не требуется подключение дополнительных пакетов), не требующей подключения к сети Интернет.
3. В программе реализован автоматизированный подбор параметров очистки прочтений (удаление адаптеров, по-

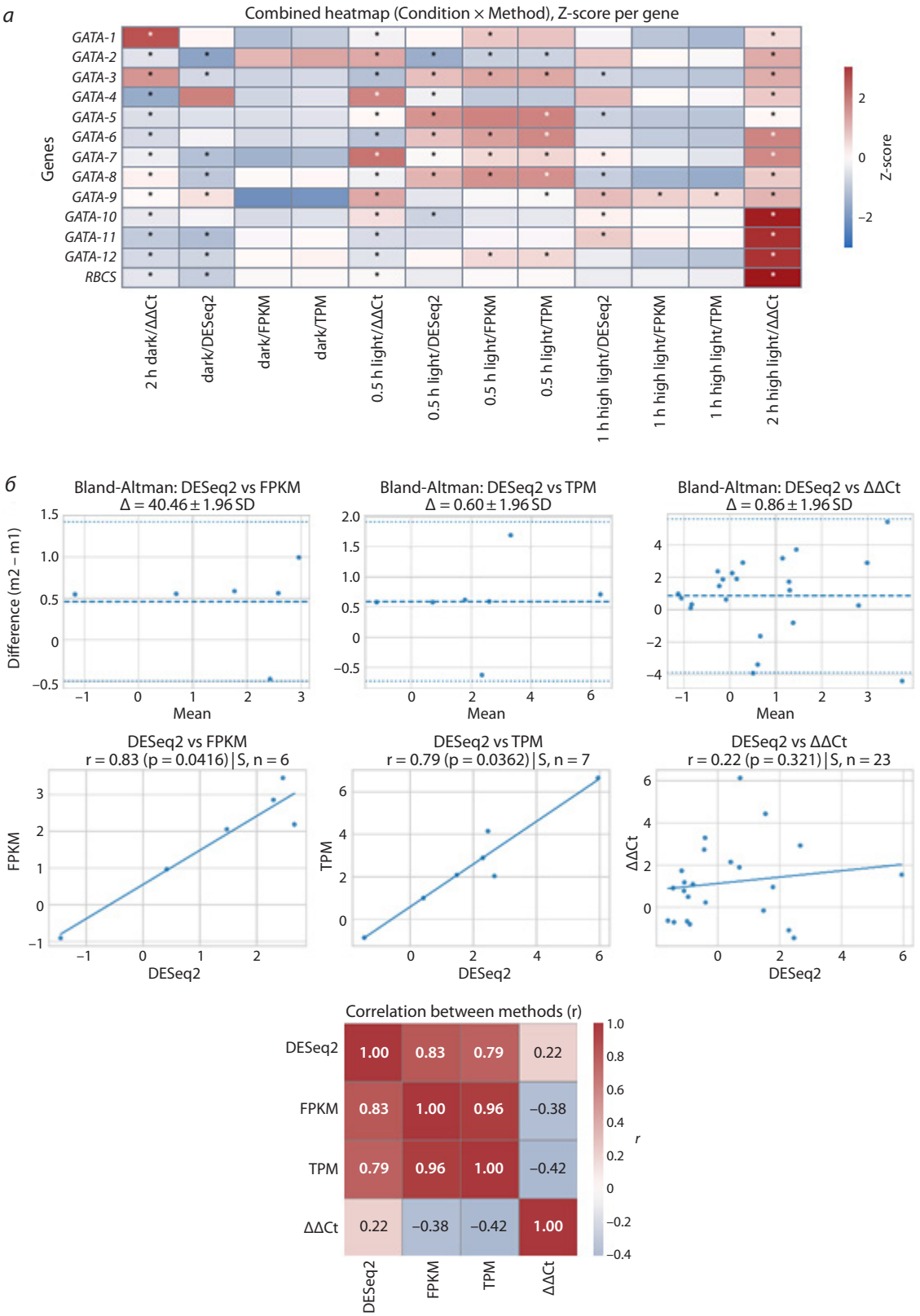


Рис. 4. Сравнительный статистический анализ применения трех методов нормализации данных секвенирования РНК, FPKM, TPM, DESeq2, и метода ΔΔCt данных ОТ-ПЦР-РВ.

а – совмещенная тепловая карта Z-оценок значений логарифма изменения уровня экспрессии (log_2FoldChange) 12 генов, кодирующих факторы транскрипции семейства GATA и гена *RBCS C. reinhardtii* для каждого сочетания условий и метода. Звездочкой отмечены статистически значимые изменения (FDR < 0.05); *б* – диаграммы Бланда–Альтмана, корреляционные диаграммы и матрица статистически достоверных (FDR < 0.05) оценок log_2FoldChange между сравниваемыми методами нормализации данных. Цветовая индикация отражает значение коэффициента Пирсона (r), где красная шкала – положительная корреляция, а синяя – отрицательная. Изображения созданы в программе PipeSeq.

рог качества, минимальная длина прочтения) на основе метрик качества до/после проведения очистки.

4. Встроенный модуль обсчета данных ПЦР-РВ методом ΔΔCt (импорт значений порогового цикла Ct, расчет нормализованной экспрессии, сравнительный статистический анализ по результатам обработки данных РНК-сек и результатов ПЦР-РВ).

Профили экспрессии генов GATA *C. reinhardtii*

Свет – один из основных регуляторов экспрессии генов фотосинтезирующих организмов. По результатам проведенного комплексного исследования мы получили наиболее полные профили экспрессии генов GATA *C. reinhardtii*, дополнив доступные данные РНК-сек собственными экспериментами методом ОТ-ПЦР-РВ. Несмотря на то что в настоящее время результаты транскриптомных исследований в большинстве случаев не требуют дополнительной верификации (Coenye, 2021), мы при поиске интересующих нас данных в открытых базах столкнулись со скудным их количеством (три пакета) и кратким описанием условий выращивания отобранных для работы культур.

Результирующие данные РНК-сек (см. рис. 2) для части генов GATA характеризовались низкой статистической значимостью изменений (FDR > 0.05). Обработка выбранных пакетов данных секвенирования РНК с помощью программы PipeSeq позволила нам обнаружить и подтвердить пару конститутивно экспрессируемых референс-генов *RPL19–RPL32* (Liu et al., 2012), а также профили экспрессии генов интереса в ответ на определенное воздействие с целью дальнейшего применения метода ОТ-ПЦР-РВ для точной количественной оценки этих изменений в расширенном перечне экспериментальных условий.

Полученные нами данные (см. рис. 3) подтверждают, что изменение световых условий является важным фактором, модулирующим экспрессию генов GATA у фотосинтезирующих организмов (Manfield et al., 2007; Naito et al., 2007; Luo et al., 2010; Schröder et al., 2023). Выполненный ранее анализ сети белковых взаимодействий (Virolainen, Chekunova, 2024) выявил три функциональных кластера, в которых предположительно принимают участие 12 факторов GATA *C. reinhardtii*.

Первый функциональный кластер, объединяющий гены *GATA-1*, *GATA-2*, *GATA-10*, связывает фоторецепцию (гены *CHLAMYDOMONAS PHOTOLYASE HOMOLOG 1* (*CPH1*), *SHOC2/SUR8-like LRR* (*CSL*)), циркадную регуляцию (гены семейства *RHYTHM OF CHLOROPLAST* (*ROC*) и метаболизм фосфора (ген *PHOSPHATE STARVATION RESPONSE 1* (*PSR1*)), обеспечивая адаптацию клетки к световому режиму.

Второй кластер (*GATA-9*) связан с функционированием арил-гидрокарбонового рецепторного комплекса и активируется при избытке света, поддерживая детоксикационные процессы в клетке.

Третий кластер (*GATA-3*, *GATA-4*, *GATA-5*, *GATA-6*, *GATA-7*, *GATA-8*, *GATA-10*, *GATA-11*, *GATA-12*) координирует ассимиляцию азота (через связь с геном *NITRATE REDUCTASE* (*NIT1*), ремоделирование хроматина (с генами, кодирующими гистондеацетилазы), репликацию ДНК (с геном, кодирующим хеликазу (*CrRuvBL1*), мембранный транспорт и деление клетки (с генами, кодирующими белок семейства Rab (*RABF1*), кинезин-подобный белок, субъединицу 4 циклосомы), проявляя выраженную светозависимость.

Наблюдаемая динамика экспрессии генов GATA представляет собой скоординированный ответ, интегрирующий ключевые процессы метаболизма клетки в ходе клеточного цикла (Voigt, Münzner, 1987; Müller et al., 2017; Salomé, Merchant, 2019). Ответ всех изученных генов GATA на световой стимул позволяет предположить наличие механизма перекрестной регуляции между тремя функциональными сетями через еще не выявленные или не охарактеризованные гены и белки.

Результаты проведенных исследований позволяют сделать первые значимые предположения относительно функций ФТ GATA у *C. reinhardtii* и зеленых водорослей в целом (Schwechheimer et al., 2022), закладывая необходимую базу для проведения дальнейших исследований.

Сравнительный анализ применения трех методов нормализации данных РНК-сек, FPKM, TPM, DESeq2, и метода ΔΔCt, по данным ОТ-ПЦР-РВ, продемонстрировал наибольшую корреляцию в паре DESeq2–ΔΔCt (см. рис. 4), что подтверждает литературные данные о высокой точности пакетов на основе отрицательного биномиального распределения (Zhao S. et al., 2020; Zhao Y. et al., 2021; Elahimanesh, Najafi, 2024) и демонстрирует надежность метода нормализации ΔΔCt данных ОТ-ПЦР-РВ для точной количественной оценки уровня экспрессии генов интереса (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008; Shi, He, 2014; He et al., 2015; Coenye, 2021; Schröder et al., 2023). В совокупности оба подхода отражают совпадающие тенденции для небольшого количества генов GATA и различия, зависящие от условий, метода и времени, что подчеркивает ценность интеграции массового (РНК-сек) и таргетного (ОТ-ПЦР-РВ) подходов.

Заключение

Разработан и использован программный пайплайн PipeSeq, включающий автоматизированные этапы загрузки, выравнивания и статистического анализа данных РНК-сек, а также позволяющий обсчитывать и визуализировать данные, получаемые методами секвенирования РНК и ПЦР-РВ. Результаты нашего исследования показали, что методы секвенирования РНК и ОТ-ПЦР-РВ дают возможность выявить схожие паттерны изменения экспрессии генов, но демонстрируют различия по оценке степени размера эффекта и чувствительности в детекции изменений.

Полученные данные дают возможность сделать вывод, что ФТ GATA *C. reinhardtii* образуют три функционально специализированные группы (кластеры), согласованная регуляция которых представляет собой ключевой механизм, обеспечивающий корректное прохождение клеточного цикла в изменяющихся условиях внешней среды. Профили экспрессии генов *GATA-2*, *GATA-4*, *GATA-5*, *GATA-6*, *GATA-8*, *GATA-10*, *GATA-11*, *GATA-12* позволяют предположить их участие в регуляции светозависимых процессов метаболизма. Гены *GATA-1*, *GATA-3*, *GATA-7*,

GATA-9 задействованы в переключении метаболизма при переходе свет/темнота и темнота/свет. Дальнейшие исследования ФТ GATA *C. reinhardtii* должны быть направлены на поиск и верификацию генов-мишеней и взаимодействий в регуляторных сетях, а также подтверждение предсказанных функций в ответ на изменение других условий культивирования.

Таким образом, программа PipeSeq продемонстрировала свою эффективность в комплексном исследовании дифференциальной экспрессии генов как инструмент для проведения полного цикла биоинформатического анализа данных РНК-сек с возможностью обработки данных ОТ-ПЦР-РВ и выполнения сравнительного статистического анализа полученных с помощью разных методов результатов. Разработанный пайплайн может быть использован при изучении профилей экспрессии генов любого объекта исследования.

Список литературы / References

Квитко К.В., Борщевская Т.Н., Чунаев А.С., Тугаринов В.В. Петергофская генетическая коллекция штаммов зеленых водорослей (*Chlorella*, *Scenedesmus*, *Chlamydomonas*). В: Культивирование коллекционных штаммов водорослей. Л.: Изд-во Ленингр. ун-та, 1983
[Kvitko K.V., Borschevskaya T.N., Chunaev A.S., Tugarinov V.V. Peterhof genetic collection of green algae strains (*Chlorella*, *Scenedesmus*, *Chlamydomonas*). In: Cultivation of Collection Strains of Algae. Leningrad, 1983 (in Russian)]

Afgan E., Baker D., Batut B., van den Beek M., Bouvier D., Čech M., Chilton J., … Soranzo N., Goecks J., Taylor J., Nekrutenko A., Blankenberg D. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W537-W544. doi 10.1093/nar/gky379

Bray N.L., Pimentel H., Melsted P., Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;4(5):525-527. doi 10.1038/nbt.3519

Coenye T. Do results obtained with RNA-sequencing require independent verification? *Biofilm*. 2021;3:100043. doi 10.1016/j.bioflm.2021.100043

Conesa A., Madrigal P., Tarazona S., Gomez-Cabrero D., Cervera A., McPherson A., Szcześniak M.W., Gaffney D.J., Elo L.L., Zhang X., Mortazavi A. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13. doi 10.1186/s13059-016-0881-8

Derveaux S., Vandesompele J., Hellemans J. How to do successful gene expression analysis using real-time PCR. *Methods*. 2010;50(4):227-230. doi 10.1016/j.ymeth.2009.11.001

Di Tommaso P., Chatzou M., Floden E.W., Barja P.P., Palumbo E., Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. 2017;35(4):316-319. doi 10.1038/nbt.3820

Elahimanesh M., Najafi M. Differentially expressed *genes* of RNA-seq data are suggested on the intersections of normalization techniques. *Biochem Biophys Rep*. 2024;37:101618. doi 10.1016/j.bbrep.2023.101618

Ewels P., Magnusson M., Lundin S., Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi 10.1093/bioinformatics/btw354

Goodstein D.M., Shu S., Howson R., Neupane R., Hayes R.D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar D.S. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(D1):D1178-D1186. doi 10.1093/nar/gkr944

Harris E.H. The *Chlamydomonas* Sourcebook. A Comprehensive Guide to Biology and Laboratory Use. San Diego: Academic Press, 1989

He F., Liu Q., Zheng L., Cui Y., Shen Z., Zheng L. RNA-Seq analysis of rice roots reveals the involvement of post-transcriptional regulation in response to cadmium stress. *Front Plant Sci*. 2015;6:1136. doi 10.3389/fpls.2015.01136

Hunter J.D. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95. doi 10.1109/MCSE.2007.55

Kim D., Langmead B., Salzberg S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360. doi 10.1038/nmeth.3317

Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907-915. doi 10.1038/s41587-019-0201-4

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi 10.1093/bioinformatics/btp352

Li X., Wang C.Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci*. 2021;13:36. doi 10.1038/s41368-021-00146-0

Liao Y., Smyth G.K., Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi 10.1093/bioinformatics/btt656

Liu C., Wu G., Huang X., Liu S., Cong B. Validation of housekeeping genes for gene expression studies in an ice alga *Chlamydomonas* during freezing acclimation. *Extremophiles*. 2012;16:419-425. doi 10.1007/s00792-012-0441-4

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

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

Luo X.M., Lin W.H., Zhu S., Zhu J.Y., Sun Y., Fan X.Y., Cheng M., … Liu L., Zhang M., Xie Q., Chong K., Wang Z.Y. Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in *Arabidopsis*. *Dev Cell*. 2010;19(6):872-883. doi 10.1016/j.devcel.2010.10.023

Manfield I.W., Devlin P.F., Jen C.H., Westhead D.R., Gilmartin P.M. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the *Arabidopsis* GATA gene family. *Plant Physiol*. 2007;143(2):941-958. doi 10.1104/pp.106.090761

Marioni J.C., Mason C.E., Mane S.M., Stephens M., Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. 2008;18(9):1509-1517. doi 10.1101/gr.079558.108

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17(1):10-12. doi 10.14806/ej.17.1.200

McKinney W. pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*. 2011;14(9):1-9.

Merchant S.S., Prochnik S.E., Vallon O., Harris E.H., Karpowicz S.J., Witman G.B., Terry A., … Werner G., Zhou K., Grigoriev I.V., Rokhsar D.S., Grossman A.R. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*. 2007;318(5848):245-250. doi 10.1126/science.1143609

Mölder F., Jablonski K.P., Letcher B., Hall M.B., Tomkins-Tinch C.H., Sochat V., Forster J., … Wilm A., Holtgrewe M., Rahmann S., Nahnsen S., Köster J. Sustainable data analysis with Snakemake. *F1000Res*. 2021;10:33. doi 10.12688/f1000research.29032.2

Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B.J. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621-628. doi 10.1038/nmeth.1226

Müller N., Wenzel S., Zou Y., Künzel S., Sasso S., Weiß D., Prager K., Grossman A., Kottke T., Mittag M. A plant cryptochrome controls key features of the *Chlamydomonas* circadian clock and its life cycle. *Plant Physiol*. 2017;174(1):185-201. doi 10.1104/pp.17.00349

Muzellec B., Teleńczuk M., Cabeli V., Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics*. 2023;39(9):btad547. doi 10.1093/bioinformatics/btad547

Naito T., Kiba T., Koizumi N., Yamashino T., Mizuno T. Characterization of a unique GATA family gene that responds to both light and cytokinin in *Arabidopsis thaliana*. *Biosci Biotechnol Biochem*. 2007;71(6):1557-1560. doi 10.1271/bbb.60692

Pertea M., Pertea G.M., Antonescu C.M., Chang T.C., Mendell J.T., Salzberg S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290-295. doi 10.1038/nbt.3122

Pertea M., Kim D., Pertea G.M., Leek J.T., Salzberg S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650-1667. doi 10.1038/nprot.2016.095

Ren W., Kong L., Jiang S., Ma L., Wang H., Li X., Liu Y., Ma W., Yan X. Genome-wide identification, evolution, and characterization of GATA gene family and GATA gene expression analysis post-MeJA treatment in *Platycodon grandiflorum*. *J Plant Growth Regul*. 2025;44:155-167. doi 10.1007/s00344-024-11468-8

Reyes J.C., Muro-Pastor M.I., Florencio F.J. The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol*. 2004;134(4):1718-1732. doi 10.1104/pp.103.037788

Riechmann J.L., Heard J., Martin G., Reuber L., Jiang C.Z., Keddie J., Adam L., … Broun P., Zhang J.Z., Ghandehari D., Sherman B.K., Yu G.L. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*. 2000;290(5499):2105-2110. doi 10.1126/science.290.5499.2105

Salomé P.A., Merchant S.S. A series of fortunate events: introducing *Chlamydomonas* as a reference organism. *Plant Cell*. 2019;31(8):1682-1707. doi 10.1105/tpc.18.00952

Sanchez-Tarre V., Kiparissides A. The effects of illumination and trophic strategy on gene expression in *Chlamydomonas reinhardtii*. *Algal Res*. 2021;54:102186. doi 10.1016/j.algal.2021.102186

Schmittgen T., Livak K. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc*. 2008;3:1101-1108. doi 10.1038/nprot.2008.73

Schröder P., Hsu B.Y., Gutsche N., Winkler J.B., Hedtke B., Grimm B., Schwechheimer C. B-GATA factors are required to repress high-light stress responses in *Marchantia polymorpha* and *Arabidopsis thaliana*. *Plant Cell Environ*. 2023;46(8):2376-2390. doi 10.1111/pce.14629

Schwechheimer C., Schröder P.M., Blaby-Haas C.E. Plant GATA factors: their biology, phylogeny, and phylogenomics. *Annu Rev Plant Biol*. 2022;73(1):123-148. doi 10.1146/annurev-arplant-072221-092913

Shi Y., He M. Differential gene expression identified by RNA-Seq and qPCR in two sizes of pearl oyster (*Pinctada fucata*). *Gene*. 2014;538(2):313-322. doi 10.1016/j.gene.2014.01.031

Virolainen P.A., Chekunova E.M. GATA family transcription factors in alga *Chlamydomonas reinhardtii*. *Curr Genet*. 2024;70(1):1. doi 10.1007/s00294-024-01280-y

Voigt J., Münzner P. The *Chlamydomonas* cell cycle is regulated by a light/dark-responsive cell-cycle switch. *Planta*. 1987;172:463-472. doi 10.1007/BF00393861

Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57-63. doi 10.1038/nrg2484

Wheeler D.L., Barrett T., Benson D.A., Bryant S.H., Canese K., Church D.M., DiCuccio M., … Suzek T.O., Tatusov R., Tatusova T.A., Wagner L., Yaschenko E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2005;33:D39-D45. doi 10.1093/nar/gki062

Zhao S., Ye Z., Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*. 2020;26(8):903-909. doi 10.1261/rna.074922.120

Zhao Y., Li M.C., Konaté M.M., Chen L., Das B., Karlovich C., Williams P.M., Evrard Y.A., Doroshow J.H., McShane L.M. TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. *J Transl Med*. 2021;19(1):269. doi 10.1186/s12967-021-02936-w

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 12.08.2025. После доработки 21.12.2025. Принята к публикации 22.12.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):311-320

doi 10.18699/vjgb-26-35

Анализ вирома растений с помощью высокопроизводительного секвенирования: принципы и подходы

Д.Д. Белкина, С.В. Виноградова

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

Аннотация. В последние годы метагеномный подход, основанный на высокопроизводительном секвенировании, находит все большее применение в диагностике вирусных инфекций растений. Этот метод позволяет изучить видовой состав вирусов, ассоциированных с исследуемым растением, и в том числе обнаружить ранее не описанные виды, охарактеризовать их популяционно-генетическую структуру и разработать генетические тест-системы для рутинной диагностики. Проведение фитосанитарного мониторинга с использованием метагеномного подхода может способствовать определению этиологии неизвестных заболеваний растений, что особенно важно для предотвращения распространения таких патогенов, как вирусы. Кроме того, с учетом невозможности элиминации вирусов растений в полевых условиях комплексная диагностика с помощью высокопроизводительного секвенирования становится эффективным инструментом как в целях соблюдения карантинного законодательства при ввозе импортного материала, так и для получения отечественного посадочного материала высоких категорий. При этом с каждым годом высокопроизводительное секвенирование становится более доступным: расширяется как приборно-техническая, так и аналитическая база. В настоящем обзоре систематизированы ключевые подходы к анализу вирома растений с помощью высокопроизводительного секвенирования. В основной части статьи описаны этапы проведения анализа: от сбора образцов до биоинформатической обработки данных, ее валидации и интерпретации. Подробно рассмотрены особенности современных платформ секвенирования и факторы, влияющие на качество чтения, в том числе контаминация. Охарактеризованы три взаимодополняющих подхода для обработки биоинформатических данных: картирование чтений на референсные последовательности вирусов; сборка и аннотация контигов; таксономическая классификация чтений без дополнительной сборки. Особое внимание уделено необходимости тщательной интерпретации результатов с учетом как биоинформатического анализа, так и валидации идентификации молекулярно-генетическими методами. Обзор будет полезен как для исследователей и специалистов, не имеющих опыта работы с высокопроизводительным секвенированием, так и для тех, кто использовал этот инструмент для выполнения других задач.

Ключевые слова: метагеномика; высокопроизводительное секвенирование; вирусы; вироом растений; биоинформатика

Для цитирования: Белкина Д.Д., Виноградова С.В. Анализ вирома растений с помощью высокопроизводительного секвенирования: принципы и подходы. *Вавиловский журнал генетики и селекции*. 2026;30(2):311-320. doi 10.18699/vjgb-26-35

Финансирование. Работа выполнена при поддержке Российского научного фонда (грант 23-16-00232).

Plant virome analysis by high-throughput sequencing: concepts and approaches

D.D. Belkina, S.V. Vinogradova

Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia
coatprotein@bk.ru

Abstract. The metagenomic approach based on high-throughput sequencing is becoming increasingly prevalent for the detection of viral infections in plants. This method allows us to study the species composition of viruses associated with the plant, including novel species, describe their population genetic structure, and develop genetic test systems for routine diagnostics. A metagenomic approach to phytosanitary monitoring can help to determine the cause of unknown plant diseases, which is particularly important for preventing the spread of pathogens, such as viruses. Furthermore, as it is impossible to eliminate plant viruses in field conditions, comprehensive diagnostics using high-throughput sequencing is becoming an effective tool for complying with quarantine regulations on the import of foreign material, as well as for producing high-quality local planting material. High-throughput sequencing is becoming more affordable every year, with both the instrumentation and analytical capacity improving. This review summarizes key approaches to analyzing plant virome using high-throughput sequencing. The analysis process, from sample collection to bioinformatic data

© Белкина Д.Д., Виноградова С.В., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

processing, validation and interpretation, is described in detail. The features of sequencing platforms and the factors affecting sequencing quality, including contamination, are discussed. Three complementary approaches to processing bioinformatic data are described: mapping reads to reference viral sequences; assembling and annotating contigs; taxonomic classification of reads without assembly. The importance of carefully interpreting the results is emphasized, considering the bioinformatic analysis and the validation by molecular genetic methods. This review will be useful for both researchers and specialists who have no experience with high-throughput sequencing, and those who have used this method for other applications.

Key words: metagenomics; high-throughput sequencing; viruses; plant virome; bioinformatics

For citation: Belkina D.D., Vinogradova S.V. Plant virome analysis by high-throughput sequencing: concepts and approaches. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):311-320. doi 10.18699/vjgb-26-35

Введение

Вирусные заболевания растений наносят значительный ущерб растениеводству по всему миру. В условиях глобализации и изменения климата распространение вирусов может приводить к разрушительным эпифитотиям, угрожающим продовольственной безопасности всего человечества (Jones, 2021). Так, потери урожая томатов и перца из-за инфекции вирусом мозаики томата (tomato mosaic virus, ToMV), по разным оценкам, составляют 25–70 % (Panno et al., 2021), а снижение урожайности винограда от болезни скручивания листьев, вызываемой вирусами из семейства Closteroviridae, может достигать 50 % (Atallah et al., 2012). Ключевое значение в предотвращении распространения вирусных инфекций имеют мониторинг, ранняя диагностика и соблюдение карантинных мер.

В настоящее время основными методами идентификации вирусов растений являются серологические и молекулярные, получившие распространение благодаря высоким скорости проведения анализа, чувствительности и специфичности. Однако их применение позволяет идентифицировать только те виды вирусов, для которых разработаны тест-системы (Maina et al., 2024).

До развития метагеномики наше представление о вирусах растений ограничивалось несколькими сотнями видов (Pappas et al., 2021), поэтому роль вирусов в микробиоме сельскохозяйственных культур оставалась недооцененной. Метагеномный подход позволяет получить комплексное представление о совокупности вирусов растения, т. е. о его вироме. По состоянию на 2025 г. Международным комитетом по таксономии вирусов зафиксировано 2598 вирусов и вироидов растений, причем 206 из них добавлены в перечень за последний год (Rubino et al., 2025). Секвенирование вирома предоставляет уникальные данные о последовательностях геномов вирусов, на основе которых можно провести популяционные исследования и проследить пути распространения вирусов, связать заражение конкретными штаммами с проявлением симптомов, а также разработать специфичные тест-системы для их быстрого обнаружения.

Анализ вирома растений с помощью высокопроизводительного секвенирования – это активно развивающийся метод. Каждый год публикуются новые рекомендации, протоколы и специализированные программы для обработки данных. Поскольку этот анализ еще не вошел в число методов для рутинной диагностики вирусных инфекций растений, у него нет единой методики. Тем не менее основные принципы его проведения уже сформировались.

В настоящей работе систематизированы ключевые подходы к анализу вирома растений с помощью высокопроизводительного секвенирования. Этот обзор будет полезен как для исследователей и специалистов, не имеющих опыта работы с высокопроизводительным секвенированием, так и для тех, кто использовал этот инструмент для выполнения других задач.

Общий план анализа растительного вирома

Структурно анализ вирома можно разделить на два блока. Первый блок включает отбор образцов, пробоподготовку и секвенирование (рис. 1). Эта часть анализа имеет ключевое значение. Если материал для секвенирования будет низкого качества, то на следующих этапах будет невозможно получить достоверные результаты. Во время сбора образцов важно учитывать, что вирусы неравномерно распределены по растительным тканям, поэтому отбирать следует по несколько фрагментов с одного растения. Особенно это актуально для многолетних растений. После сбора образцы должны быть как можно скорее доставлены в лабораторию, чтобы снизить риск деградации нуклеиновых кислот, например, под действием прямых солнечных лучей.

Второй блок анализа растительного вирома включает компьютерную обработку данных секвенирования биоинформатическими программами, валидацию и интерпретацию результатов. В отличие от этапов первой части анализа, биоинформатическую обработку данных в любой момент можно провести повторно и исправить возможные ошибки, применить новые методы или перепроверить результаты других исследований. Поэтому критически важно избегать ошибок при сборе образцов, пробоподготовке и секвенировании библиотек. Для достоверной идентификации вирусов необходимо валидировать результаты биоинформатического анализа молекулярно-генетическими методами.

Исходный материал для анализа

Вирусы растений отличаются от других фитопатогенов многообразием типов геномов. В качестве носителя генетической информации они используют молекулы как ДНК, так и РНК, которые могут быть одноцепочечными или двуцепочечными, линейными или кольцевыми (Lefeuvre et al., 2019). При этом одноцепочечные РНК-вирусы под-

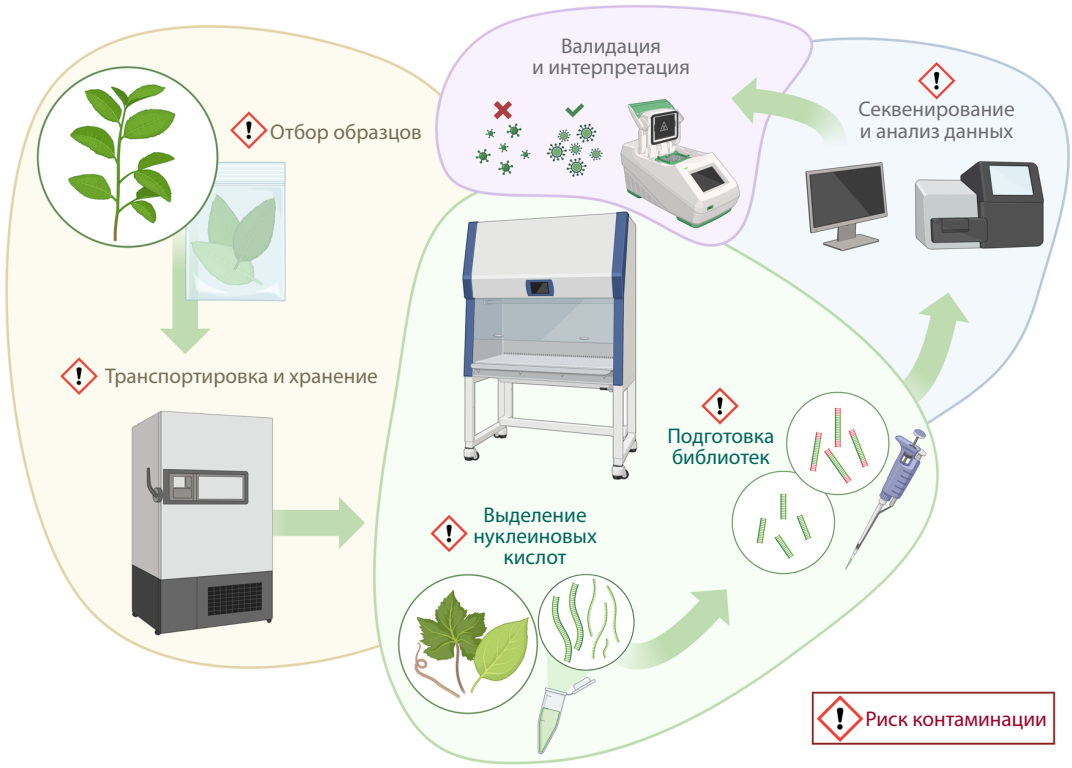


Рис. 1. Основные этапы анализа растительного вирома.

разделяются на вирусы с «позитивным» геномом, когда геномная РНК соответствует мРНК, и с «негативным» геномом, когда геномная РНК комплементарна мРНК. Идентифицировать все вирусы, инфицирующие растение, с их разнообразными геномами, используя только один подход, довольно сложно (Fitzpatrick et al., 2021). Ниже рассмотрены различные подходы к анализу растительных виромов в зависимости от типа используемой матрицы.

Секвенирование тотальной РНК, выделенной из растительных тканей, – это относительно простой, надежный и широко распространенный метод анализа виромов (Lee et al., 2020; Nabeshima, Abe, 2021; Vinogradova et al., 2023). Он позволяет обнаружить не только вирусы с РНК-геномами, но и транскрипты ДНК-вирусов (Pecman et al., 2017; Cobbin et al., 2021), а также вироиды, что делает этот подход оптимальным для диагностики вирусов с различными типами геномных нуклеиновых кислот.

Однако чувствительность этого метода может значительно снизиться из-за наличия в анализируемом образце РНК растения-хозяина, выделяемой совместно с вирусной РНК в ходе пробоподготовки (Maliogka et al., 2018), что создает трудности в обнаружении вирусов с низким титром (Roossinck et al., 2015). Преодолеть эту проблему позволяет обогащение образца нуклеиновыми кислотами, ассоциированными с вирусами (Gaafar, Ziebell, 2020). Как правило, это достигается путем очистки образца от рибосомальной и транспортной РНК или обогащением его полиА-транскриптами. Первый метод обогащения предпочтительнее, поскольку не все группы вирусов имеют полиаденилированный 3′-конец.

Еще один метод анализа растительного вирома – это секвенирование нуклеиновых кислот, ассоциированных с вирионами (Filloux et al., 2015). Преимуществом этого подхода являются секвенирование только вирусных последовательностей и возможность анализа вирусов как с РНК-, так и с ДНК-геномами, в том числе вирусов с низким титром (Moubset et al., 2022). Однако экстракция нуклеиновых кислот, ассоциированных с вирионами, позволяет обнаружить только вирусы, имеющие капсид (который отсутствует у вирусов-сателлитов и вироидов), и требует бóльших усилий, чем выделение тотальной РНК (Maliogka et al., 2018).

Следующий подход предполагает анализ коротких интерферирующих РНК (киРНК), образующихся в клетках растений в результате иммунного ответа на вирусную инфекцию. Эти молекулы представляют собой расщепленные растительными ферментами фрагменты вирусных геномов длиной по 21–24 нуклеотида (Vivek et al., 2020; Zhuravlyov et al., 2022).

Секвенирование киРНК позволяет обнаруживать РНК- и ДНК-вирусы, но небольшой размер этих молекул затрудняет последующую сборку полных геномов и дифференциацию штаммов вирусов (Pecman et al., 2017; Maliogka et al., 2018; Turco et al., 2018).

Многие вирусы во время репликации в качестве промежуточного продукта образуют двуцепочечную РНК (дцРНК), благодаря чему эти молекулы тоже используются в исследованиях растительного вирома. Секвенирование дцРНК позволяет идентифицировать не только вирусы с дцРНК-геномом, но и большинство одноцепочечных

РНК-вирусов, вироиды, а также некоторые ДНК-вирусы (Gallo et al., 2021; Fall et al., 2025). Однако для обнаружения ДНК-вирусов этот метод малоэффективен и показывает положительные результаты только при высоком уровне экспрессии вирусных транскриптов (Gaafar, Ziebell, 2020).

Платформы секвенирования

В основе всех метагеномных исследований лежит высокопроизводительное секвенирование нуклеиновых кислот (high-throughput sequencing, HTS). Эта группа методов объединяет секвенирование второго и третьего поколений, сменившие секвенирование методом Сэнгера, которое остается золотым стандартом только для прочтения последовательностей до 1000 нуклеотидов длиной (Crossley et al., 2020).

Секвенирование второго поколения (next generation sequencing, NGS), также называемое массовым параллельным, позволяет считывать короткие последовательности длиной до нескольких сотен нуклеотидов. С помощью этого подхода можно быстро секвенировать все последовательности в образце, однако небольшая длина чтения усложняет сборку полных геномных последовательностей вирусов на этапе биоинформатического анализа (Maina et al., 2024).

Лидером рынка NGS платформ является компания Illumina (США), выпускающая секвенаторы с низким уровнем ошибок при высокой производительности (Maina et al., 2024). В последние годы конкуренцию Illumina составляют приборы от китайских компаний MGI, GeneMind и Cygnus Biosciences. Кроме того, недавно разработан и выпущен в продажу секвенатор Нанофор СПС российского производства (Kurochkin et al., 2021). Большинство современных NGS секвенаторов основано на флуоресцентном секвенировании путем синтеза в проточной ячейке (Zubov et al., 2021) и имеет схожие протоколы подготовки библиотек и зачастую совместимые реагенты. Процесс подготовки библиотек обычно включает фрагментацию нуклеиновых кислот (например, с помощью ультразвука или ферментов рестрикции), лигирование коротких олигонуклеотидов (адаптеров) с одним или двумя концами фрагментов и амплификацию этих фрагментов (Kutnjak et al., 2021). Адаптеры различаются в зависимости от NGS платформы и необходимы для запуска процесса секвенирования (Lebas et al., 2022). Важно отметить, что секвенаторы второго поколения способны считывать только молекулы ДНК, поэтому, если в качестве матрицы выбрана РНК, после фрагментации проводится обратная транскрипция для синтеза кДНК.

Секвенирование третьего поколения, называемое также одномолекулярным, основано на прочтении последовательностей отдельных молекул и не требует их фрагментации и амплификации (Villamor et al., 2019). Максимальная длина чтения в этом случае исчисляется несколькими сотнями тысяч нуклеотидов, что позволяет целиком считывать небольшие геномы, например геномы вирусов (Liu et al., 2025). Важным преимуществом одномолекулярного секвенирования являются возможность прямого чтения РНК без синтеза кДНК, а также секвенирование в режиме реального времени (Sun et al., 2022).

Ведущие производители секвенаторов третьего поколения – компании Pacific Biosciences (США) и Oxford Nanopore Technologies (Великобритания). Несмотря на то что обе платформы третьего поколения позволяют исключить ошибки, вносимые полимеразой при амплификации, они имеют относительно низкое качество чтения, по сравнению с NGS (Rose et al., 2016; Rang et al., 2018). Впрочем, недавно представленная технология PacBio HiFi обеспечивает точность считывания, превышающую 99.5 % (Han et al., 2024). Нанопорное секвенирование, первоначально разработанное Oxford Nanopore Technologies, тоже с каждым годом становится все более точным и доступным (Javaran et al., 2021). Эту технологию уже освоила китайская компания Qitan Tech (Wang et al., 2022), поэтому можно ожидать, что в будущем одномолекулярное секвенирование придет на смену NGS.

Другие факторы, влияющие на качество секвенирования растительного вирома

Выше рассмотрено, что выбор типа нуклеиновых кислот и платформы секвенирования играет важную роль в анализе вирома растений. На качество секвенирования и количество чтений оказывает влияние еще ряд факторов.

Контаминация – один из главных негативных факторов любого метагеномного анализа (Piombo et al., 2021; Lebas et al., 2022). Загрязнение образцов может произойти на всех этапах: от сбора растительных тканей до непосредственного процесса секвенирования (см. рис. 1). Для отслеживания контаминации на этапах выделения нуклеиновых кислот и подготовки библиотек рекомендуется использовать отрицательные контроли без добавления матрицы (Fitzpatrick et al., 2021). Кроме того, важно придерживаться физического разделения лабораторных этапов, использовать стерильные реагенты и поддерживать в чистоте посуду, оборудование и рабочие поверхности (Kutnjak et al., 2021; Maina et al., 2024). Исчерпывающее руководство о предотвращении контаминации при использовании HTS в диагностике фитопатогенов приведено в публикации S. Massart с коллегами (2022).

Наряду с контаминацией большое влияние на качество секвенирования вирома оказывают используемые методики, реактивы (особенно ферменты) и оборудование. Так, при подготовке библиотек для секвенирования второго поколения используются обратная транскриптаза и ДНК-полимераза, которые могут вносить ошибки в процессе амплификации (Cholet et al., 2020). Поэтому для уменьшения количества артефактов важно применять высокоточные ферменты.

На точность идентификации вирусов влияет количество чтений в наборе данных. При недостаточном количестве чтений вирусы с низким титром в образце, возможно, обнаружить не удастся. Однако из-за избыточного секвенирования чтений существует риск ложноположительной идентификации (Massart et al., 2014). Для определения оптимального значения может быть полезно изучить опубликованные анализы вирома интересующего растения.

Биоинформатический анализ

Биоинформатический анализ данных секвенирования растительного вирома – это многоступенчатый процесс, который можно проводить как на персональном компьютере, так и на высокопроизводительных кластерах. Разница в мощности оборудования влияет на скорость анализа. Полученные в результате секвенирования чтения обрабатываются специализированными программами в порядке, адаптированном для анализа конкретного набора данных.

Большая часть биоинформатических программ разработана для операционных систем Linux и не имеет графического интерфейса, поэтому для их использования необходимо работать в командной строке. Для других операционных систем доступны кроссплатформенные приложения, веб-инструменты и коммерческие программные пакеты. Последние особенно удобны для начинающих пользователей благодаря тому, что позволяют выполнять почти весь анализ вирома в одной среде с интуитивно понятным интерфейсом. Популярные примеры таких приложений – Geneious Prime (Новая Зеландия) и CLC Genomics Workbench (США). Несмотря на удобство коммерческих решений, их функционал ограничен встроенным набором инструментов. В целом его достаточно для рутинного анализа набора виромных данных, но больше возможностей обеспечивает работа с командной строкой Linux.

Как правило, программы, предназначенные для Linux-систем, публикуются в открытом бесплатном доступе. Для пользователей Windows можно воспользоваться подсистемой Windows для Linux (WSL) или виртуальной машиной (например, VirtualBox). Существуют автоматизированные наборы программ, предназначенные для идентификации вирусов и других фитопатогенов: VirusDetect (Zheng et al., 2017), Virtool (Rott et al., 2017), Kodoja (Baizan-Edge et al., 2019), PhytoPipe (Hu et al., 2023) и др. Такие решения позволяют достаточно быстро получить результаты, но могут не учитывать особенности конкретного набора данных. Лучших результатов можно достичь, используя навыки программирования для написания или редактирования собственных программ. В области обработки биологических данных наибольшей популярностью пользуются языки Python и R.

Результаты секвенирования чаще всего предоставляются в виде файлов формата FastQ, в которых содержатся чтения с указанием качества прочтения каждого нуклеотида. Для визуализации этой информации доступна бесплатная программа FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc), формирующая подробные отчеты о качестве секвенирования.

Вне зависимости от типа анализируемых нуклеиновых кислот и платформы секвенирования важным этапом биоинформатического анализа вирома является препроцессинг – набор операций, призванных повысить качество данных. Препроцессинг оказывает определяющее влияние на весь последующий анализ. Некачественная подготовка данных может привести к ошибкам во время сборки геномов вирусов и препятствовать картированию чтений, вследствие чего результаты анализа будут искажены.

В большинстве случаев препроцессинг включает следующие шаги.

– Фильтрация чтений. Этот шаг включает обрезку (тримминг) и удаление низкокачественных и слишком коротких чтений. Необходимый порог качества и минимальная длина чтения задаются пользователем. При обработке данных Illumina часто используются такие значения качества, как Q20 или Q30, что означает точность считывания 99 и 99.9 % соответственно (Kutnjak et al., 2021). Слишком строгие настройки фильтрации могут привести к частичной потере информативных данных, слишком мягкие – к ошибкам во время картирования и сборки геномов. Наряду с удалением низкокачественных чтений, фильтрация включает в себя очистку данных от адаптерных последовательностей, которые были искусственно добавлены в библиотеки для запуска секвенирования. Выполнить эти операции можно с помощью таких программ, как BBDuk (https://sourceforge.net/projects/bbmap), Trimmomatic (Bolger et al., 2014) и другие.
– Объединение парных чтений на основе перекрывающейся области. Этот шаг необходим только для секвенирования с парным чтением концов.
– Удаление дублирующихся чтений для уменьшения объема данных и повышения скорости анализа.
– Еще одним шагом препроцессинга может стать удаление чтений, картирующихся на геном организма-хозяина. Этот шаг позволяет значительно сократить объем и повысить качество данных. Однако следует помнить, что при этом будут удалены последовательности вирусов, интегрированных в геном хозяина (Pappas et al., 2021). Кроме того, сборка генома растения-хозяина может содержать ошибки. Если последовательности некоторых вирусов случайно попали в сборку, удаление растительных чтений исключит такие вирусы из набора данных. Поэтому удаление последовательностей растения-хозяина следует проводить с осторожностью.

Идентификация вирусов в данных секвенирования всегда опирается на сравнение последовательностей с базами данных (Lebas et al., 2022). Качество идентификации прямо пропорционально качеству используемых баз данных. Чем полнее база данных и точнее аннотация последовательностей, тем выше вероятность точной идентификации вирусов. По мере пополнения базы данных новыми последовательностями любой набор данных можно проанализировать повторно.

Существует три основных подхода, которые при совместном использовании могут взаимодополнять друг друга (рис. 2): картирование чтений на референсные последовательности вирусов; сборка и аннотация контигов; таксономическая классификация чтений без дополнительной сборки.

Наиболее быстрый и простой подход к идентификации вирусов в данных HTS состоит в картировании чтений на референсные последовательности вирусов. Картирование – это одна из базовых операций при анализе результатов секвенирования второго поколения. Алгоритм

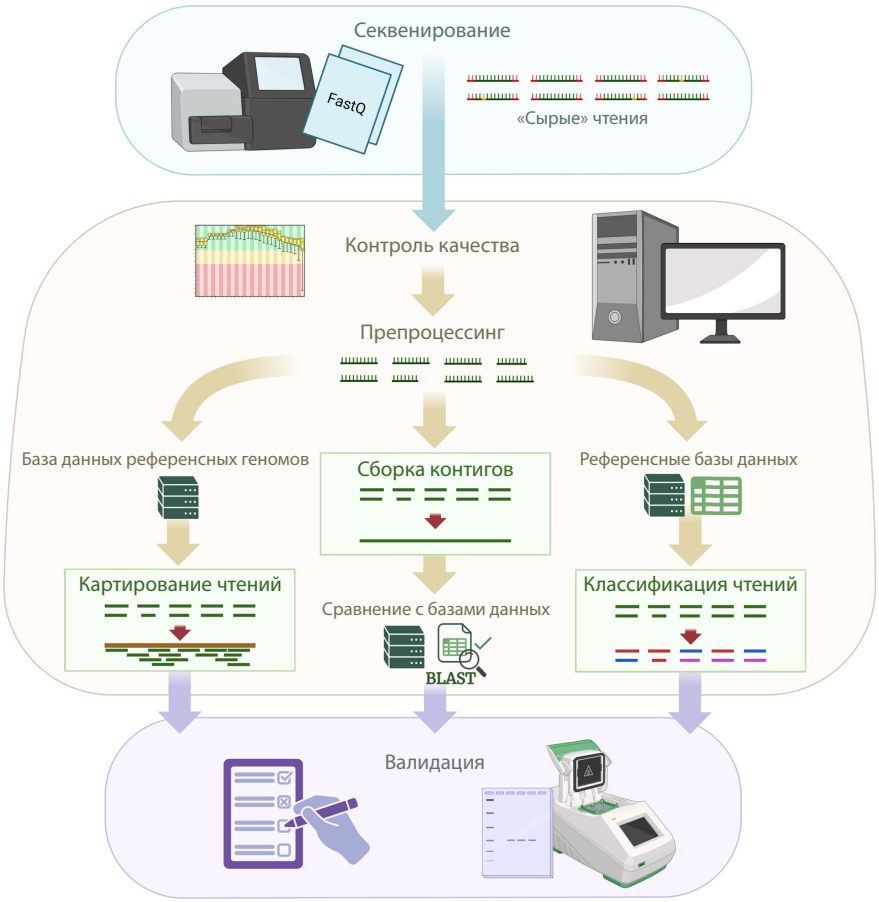


Рис. 2. Схема биоинформатического анализа виромных данных.

выравнивает чтения с референсной последовательностью с учетом заданного порога сходства. Чтения, не удовлетворяющие порогу, отбрасываются, а оставшиеся занимают такие позиции на референсной последовательности, в которых наиболее с ней совпадают.

Чтобы обнаружить вирусы растений с помощью этого подхода, необходимо предварительно подготовить базу данных референсных последовательностей, которая должна содержать по меньшей мере по одному полному геному каждого вируса, способного заражать исследуемую культуру.

Затем препроцессированные чтения картируют на эти референсные последовательности. В результате пользователь получает информацию о доле покрытия референсной последовательности чтениями, глубине покрытия, проценте сходства между референсом и картированными чтениями и другие показатели (Lebas et al., 2022). При достаточном покрытии картированные чтения образуют непрерывную последовательность по всей длине референсного генома.

Для картирования чтений используют такие программы, как BWA (Li, Durbin, 2009), Bowtie2 (Langmead, Salzberg, 2012), Minimap2 (Li, 2018), BBmap (https://sourceforge.net/projects/bbmap), STAR (Dobin, Gingeras, 2015), GraphMap (Sović et al., 2016) и Hisat2 (Kim et al., 2019). Настройки и пороговые значения этих программ имеют большое значение. Следует понимать, что последовательность генома вируса, заразившего исследуемое растение, может значительно отличаться от референсной, поскольку вирусы известны быстрыми темпами накопления мутаций. При слишком строгих настройках картирования будут обнаружены только изоляты, близкие к референсной последовательности (Kutnjak et al., 2021). С другой стороны, если настройки картирования будут недостаточно строгими, на последовательность генома одного вида вируса будут картированы чтения другого, что приведет к ложноположительной идентификации (Roy et al., 2018).

Картирование позволяет эффективно обнаруживать вирусы даже в тех случаях, когда число чтений вируса в анализируемом наборе данных невелико. Однако важно с осторожностью интерпретировать результаты и помнить о риске контаминации во время пробоподготовки и секвенирования. По этой причине рекомендуется устанавливать пороговые значения в количестве чтений или доле покрытия референсного генома, ниже которых образец будет считаться вирус-отрицательным (Belkina et al., 2023).

Второй подход к идентификации вирусов с помощью HTS основан на сборке *de novo* последовательностей из перекрывающихся чтений – контигов, и сравнении их с аннотированными базами данных. Сборка контигов позволяет обнаружить новые виды вирусов, новые растения-хозяева известных вирусов, а также описать различные

варианты их геномов. Кроме того, этот подход исключает риск ложноотрицательных результатов, связанных с возможной неполнотой набора референсных геномов вирусов целевой культуры.

Для сборки контигов могут использоваться различные алгоритмы, каждый из которых имеет свои сильные и слабые стороны (White et al., 2017). При достаточной глубине секвенирования контиги могут соответствовать почти полным геномам вирусов, без коротких 5′- и 3′-концевых участков. Эффективность сборки одним и тем же алгоритмом различается на разных наборах данных, поэтому невозможно выбрать один универсальный (Sutton et al., 2019; Shvets et al., 2022). На качество сборки могут положительно влиять предварительное удаление чтений растения-хозяина и выбор подходящих настроек (Kutnjak et al., 2021).

Наиболее популярными алгоритмами сборки для коротких чтений являются SPAdes (Bankevich et al., 2012), Velvet (Zerbino, Birney, 2008), Geneious (https://www.geneious.com), CLC (https://digitalinsights.qiagen.com), MIRA (Chevreux et al., 1999), ABySS (Simpson et al., 2009), IDBA-UI (Peng et al., 2012), SOAPdenovo2 (Luo et al., 2012); для длинных чтений – Canu (Koren et al., 2017), Falcon (Chin et al., 2016) и Pomoxis (https://github.com/nanoporetech/pomoxis). Некоторые алгоритмы, например SPAdes и Unicycler (Wick et al., 2017), показывают хорошие результаты при сборке геномов из совокупности коротких и длинных чтений (Pappas et al., 2021).

Для получения более надежных результатов рекомендуется использовать несколько алгоритмов сборки. При этом нужно учитывать, что, в зависимости от размера библиотеки и выбранного алгоритма, процесс сборки может занимать продолжительное время и требовать больших объемов оперативной памяти.

После завершения сборки контиги сравнивают с базами данных аннотированных последовательностей. Достоверность результатов во многом зависит от полноты и качества используемых баз данных. Многие из них доступны онлайн, но зачастую бывает удобнее развернуть базу данных на облачном сервере или локально, если имеется достаточное дисковое пространство.

Наиболее удобными, обновляемыми и часто используемыми являются базы данных от NCBI, которые к началу 2025 г. в совокупности содержат 6 млн нуклеотидных и 1 млн аминокислотных последовательностей (Sayers et al., 2025). NCBI предоставляет как курируемую базу данных, содержащую по одному геному каждого организма, – RefSeq (https://www.ncbi.nlm.nih.gov/refseq), так и ежедневно пополняемую исследователями со всего мира базу данных GenBank (https://www.ncbi.nlm.nih.gov/genbank). База данных GenBank содержит преобладающее большинство общедоступных нуклеотидных и аминокислотных последовательностей организмов, но качество последовательностей и правильность аннотации в ней почти не проверяются. Поэтому сравнение контигов с этой базой данных может привести к ошибкам в идентификации. Использование базы данных референсных геномов вирусов (Viral RefSeq) потенциально даст наиболее надежный результат, однако она медленно пополняется новыми записями и может не содержать значительное количество недавно открытых вирусов.

Самым широко используемым инструментом для сравнения контигов с базами данных является BLAST, который объединяет несколько алгоритмов анализа нуклеотидных и аминокислотных последовательностей. Определение того, какому вирусу из базы данных соответствует тот или иной контиг, можно выполнить с помощью самого быстрого алгоритма – megablast. Для поиска новых вирусов лучше использовать blastx или tblastx, которые улавливают отдаленные сходства между последовательностями. Главный недостаток blastx и особенно tblastx – низкая скорость анализа, поэтому при ограниченном времени можно задействовать более быстрый алгоритм DIAMOND (Buchfink et al., 2015). Однако некоторые исследователи считают, что он менее эффективен для идентификации вирусов (Kutnjak et al., 2021).

Для запуска BLAST можно использовать веб-интерфейс на сайте NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Он особенно удобен, когда нужно проверить небольшое число последовательностей. Через сайт осуществляется доступ ко всем базам данных нуклеотидных и белковых последовательностей NCBI, при этом на поиск не расходуются вычислительные ресурсы пользователя. Кроме того, BLAST встроен в программы с графическим интерфейсом, такие как Geneious Prime и CLC Genomics Workbench. При работе с командной строкой Linux, Mac и Windows применяется инструмент BLAST+ (Camacho et al., 2009).

В результате сравнения с базой данных последовательностей вирусов контиги с той или иной достоверностью относят к конкретным видам. Обычно уровень достоверности указывают с помощью E-value – показателя вероятности случайного совпадения между контигом и базой данных (VanderWeele, Ding, 2017). Чем ближе E-value к 0, тем надежнее выполнена идентификация.

Подходы, основанные на таксономической классификации чтений без предварительной сборки или картирования, менее распространены в исследованиях растительных виромов. Однако при наличии необходимых вычислительных ресурсов они позволяют получать результаты намного быстрее рассмотренных выше подходов. Среди таксономических классификаторов чтений популярны такие, как Kraken2 (Wood et al., 2019), Kaiju (Menzel et al., 2016), CLARK-S (Ounit, Lonardi, 2016) и Centrifuge (Kim et al., 2016). Как правило, для работы им требуются обширные базы данных, большие вычислительные мощности и умение пользователя работать в командной строке Linux.

Валидация и интерпретация результатов анализа вирома

После того как любым методом биоинформатического анализа в библиотеке обнаружены чтения того или иного вируса, важно правильно интерпретировать эту информацию. Поскольку качество секвенирования зависит от многих факторов, не существует универсальных критериев того, какое растение следует считать зараженным вирусом: сколько чтений и контигов достоверно свидетельствует

о наличии инфекции (Villamor et al., 2019). Более того, анализ одного набора данных с использованием разных программ или с разными настройками одних и тех же программ может приводить к различающимся результатам.

Поэтому в каждом конкретном случае необходимо всесторонне и критически подходить к анализу полученных данных, а на завершающих этапах анализа вирома большое значение приобретает экспертная роль исследователя, его «биологический взгляд» на биоинформатические данные. В частности, исследователь должен понимать, что не все обнаруженные в образцах вирусы являются патогенами исследуемой культуры. В виромных данных могут присутствовать вирусы насекомых, грибов и бактерий, ассоциированных с растением, а также вирусы человека и животных (Cobbin et al., 2021). Это особенно важно учитывать, когда речь идет об обнаружении ранее не описанных вирусов.

Для достоверной идентификации вирусов растений в метагеномных данных следует подтверждать результаты биоинформатического анализа с помощью молекулярно-генетических методов, например ПЦР или ПЦР с обратной транскрипцией (ОТ-ПЦР), в зависимости от типа генома вируса. Если вирус будет обнаружен в образце в результате как секвенирования вирома, так и молекулярно-биологической диагностики, можно с большой долей вероятности считать идентификацию точной.

Заключение

Анализ вирома с помощью высокопроизводительного секвенирования – это мощный метод, позволяющий комплексно подойти к решению фундаментальных и прикладных задач в области сельского хозяйства. С каждым годом высокопроизводительное секвенирование становится более доступным: расширяется как приборно-техническая, так и аналитическая база. В будущем с появлением правовой основы этот метод станет основанием для принятия решений в сфере государственной продовольственной безопасности.

Список литературы / References

Atallah S.S., Gómez M.I., Fuchs M.F., Martinson T.E. Economic impact of grapevine leafroll disease on *Vitis vinifera* cv. Cabernet Franc in Finger Lakes vineyards of New York. *Am J Enol Vitic*. 2012;63(1):73-79. doi 10.5344/ajev.2011.11055

Baizan-Edge A., Cock P., MacFarlane S., McGavin W., Torrance L., Jones S. *Kodoja*: a workflow for virus detection in plants using *k-mer* analysis of RNA-sequencing data. *J Gen Virol*. 2019;100(3):533-542. doi 10.1099/jgv.0.001210

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., … Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-477. doi 10.1089/cmb.2012.0021

Belkina D., Karpova D., Porotikova E., Lifanov I., Vinogradova S. Grapevine virome of the Don ampelographic collection in Russia has concealed five novel viruses. *Viruses*. 2023;15(12):2429. doi 10.3390/v15122429

Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. doi 10.1093/bioinformatics/btu170

Buchfink B., Xie C., Huson D.H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59-60. doi 10.1038/nmeth.3176

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10(1):421. doi 10.1186/1471-2105-10-421

Chevreux B., Wetter T., Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: Proceedings of the German Conference on Bioinformatics (GCB) 99. Hannover, 1999;45-56

Chin C.-S., Peluso P., Sedlazeck F.J., Nattestad M., Concepcion G.T., Clum A., Dunn C., … Luo C., Ecker J.R., Cantu D., Rank D.R., Schatz M.C. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 2016;13(12):1050-1054. doi 10.1038/nmeth.4035

Cholet F., Ijaz U.Z., Smith C.J. Reverse transcriptase enzyme and priming strategy affect quantification and diversity of environmental transcripts. *Environ Microbiol*. 2020;22(6):2383-2402. doi 10.1111/1462-2920.15017

Cobbin J.C., Charon J., Harvey E., Holmes E.C., Mahar J.E. Current challenges to virus discovery by meta-transcriptomics. *Curr Opin Virol*. 2021;51:48-55. doi 10.1016/j.coviro.2021.09.007

Crossley B.M., Bai J., Glaser A., Maes R., Porter E., Killian M.L., Clement T., Toohey-Kurth K. Guidelines for Sanger sequencing and molecular assay monitoring. *J Vet Diagn Invest*. 2020;32(6):767-775. doi 10.1177/1040638720905833

Dobin A., Gingeras T.R. Mapping RNA-seq reads with STAR. *Curr Protoc Bioinformatics*. 2015;51(1):11.14.1-11.14.19. doi 10.1002/0471250953.bi1114s51

Fall M.L., Xu D., Lemoyne P., Clément G., Moffett P., Ritzenthaler C. An innovative binding-protein-based dsRNA extraction method: comparison of cost-effectiveness of virus detection methods using high-throughput sequencing. *Mol Ecol Resour*. 2025;25(7):e14111. doi 10.1111/1755-0998.14111

Filloux D., Dallot S., Delaunay A., Galzi S., Jacquot E., Roumagnac P. Metagenomics approaches based on virion-associated nucleic acids (VANA): an innovative tool for assessing without a priori viral diversity of plants. In: Lacomme C. (Ed.) Plant Pathology. Methods in Molecular Biology. Vol. 1302. New York: Humana Press, 2015;249-257. doi 10.1007/978-1-4939-2620-6_18

Fitzpatrick A.H., Rupnik A., O'Shea H., Crispie F., Keaveney S., Cotter P. High throughput sequencing for the detection and characterization of RNA viruses. *Front Microbiol*. 2021;12:621719. doi 10.3389/fmicb.2021.621719

Gaafar Y.Z.A., Ziebell H. Comparative study on three viral enrichment approaches based on RNA extraction for plant virus/viroid detection using high-throughput sequencing. *PLoS One*. 2020;15(8):e0237951. doi 10.1371/journal.pone.0237951

Gallo Y., Marín M., Gutiérrez P. Detection of RNA viruses in *Solanum quitoense* by high-throughput sequencing (HTS) using total and double stranded RNA inputs. *Physiol Mol Plant Pathol*. 2021;113:101570. doi 10.1016/j.pmpp.2020.101570

Han Y., He J., Li M., Peng Y., Jiang H., Zhao J., Li Y., Deng F. Unlocking the potential of metagenomics with the PacBio high-fidelity sequencing technology. *Microorganisms*. 2024;12(12):2482. doi 10.3390/microorganisms12122482

Hu X., Hurtado-Gonzales O.P., Adhikari B.N., French-Monar R.D., Malapi M., Foster J.A., McFarland C.D. PhytoPipe: a phytosanitary pipeline for plant pathogen detection and diagnosis using RNA-seq data. *BMC Bioinformatics*. 2023;24(1):470. doi 10.1186/s12859-023-05589-2

Javaran V.J., Moffett P., Lemoyne P., Xu D., Adkar-Purushothama C.R., Fall M.L. Grapevine virology in the third-generation sequencing era: from virus detection to viral epitranscriptomics. *Plants*. 2021;10(11):2355. doi 10.3390/plants10112355

Jones R.A.C. Global plant virus disease pandemics and epidemics. *Plants*. 2021;10(2):233. doi 10.3390/plants10020233

Kim D., Song L., Breitwieser F.P., Salzberg S.L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. 2016;26(12):1721-1729. doi 10.1101/gr.210641.116

Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907-915. doi 10.1038/s41587-019-0201-4

Koren S., Walenz B.P., Berlin K., Miller J.R., Bergman N.H., Phillippy A.M. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res*. 2017; 27(5):722-736. doi 10.1101/gr.215087.116

Kurochkin V.E., Alekseev Y.I., Petrov D.G., Evstrapov A.A. Domestic devices for molecular genetic analysis: developments of the IAP RAS and SINTOL LLC. *Russ Mil Med Acad Rep*. 2021;40(3): 69-74. doi 10.17816/rmmar76918 (in Russian)

Kutnjak D., Tamisier L., Adams I., Boonham N., Candresse T., Chiumenti M., De Jonghe K., … Rollin J., Rott M., Schumpp O., Massart S., Haegeman A. A primer on the analysis of high-throughput sequencing data for detection of plant viruses. *Microorganisms*. 2021;9(4):841. doi 10.3390/microorganisms9040841

Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359. doi 10.1038/nmeth.1923

Lebas B., Adams I., Al Rwahnih M., Baeyen S., Bilodeau G.J., Blouin A.G., Boonham N., … Vicente C.S.L., Vossenberg B.T.L.H., Wetzel T., Ziebell H., Massart S. Facilitating the adoption of high-throughput sequencing technologies as a plant pest diagnostic test in laboratories: a step-by-step description. *EPPO Bull*. 2022;52(2): 394-418. doi 10.1111/epp.12863

Lee H.-K., Kim S.-Y., Yang H.-J., Lee D.-S., Kwon B., Lee D.-Y., Oh J., Lee S.-H. The detection of plant viruses in Korean ginseng (*Panax ginseng*) through RNA sequencing. *Plant Pathol J*. 2020;36(6): 643-650. doi 10.5423/PPJ.NT.07.2020.0137

Lefeuvre P., Martin D.P., Elena S.F., Shepherd D.N., Roumagnac P., Varsani A. Evolution and ecology of plant viruses. *Nat Rev Microbiol*. 2019;17(10):632-644. doi 10.1038/s41579-019-0232-3

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi 10.1093/bioinformatics/bty191

Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi 10.1093/bioinformatics/btp324

Liu S., Rodriguez J.S., Munteanu V., Ronkowski C., Sharma N.K., Alser M., Andreace F., … Ganda E., Davenport E.R., Pop M., Koslicki D., Mangul S. Analysis of metagenomic data. *Nat Rev Methods Primers*. 2025;5(1):5. doi 10.1038/s43586-024-00376-6

Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., … Li Y., Yang H., Wang J., Lam T.-W., Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*. 2012;1(1):18. doi 10.1186/2047-217X-1-18

Maina S., Donovan N.J., Plett K., Bogema D., Rodoni B.C. High-throughput sequencing for plant virology diagnostics and its potential in plant health certification. *Front Hortic*. 2024;3:1388028. doi 10.3389/fhort.2024.1388028

Maliogka V.I., Minafra A., Saldarelli P., Ruiz-García A.B., Glasa M., Katis N., Olmos A. Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. *Viruses*. 2018;10(8):436. doi 10.3390/v10080436

Massart S., Olmos A., Jijakli H., Candresse T. Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Res*. 2014;188:90-96. doi 10.1016/j.virusres.2014.03.029

Massart S., Adams I., Al Rwahnih M., Baeyen S., Bilodeau G.J., Blouin A.G., Boonham N., … van de Vossenberg B.T.L.H., Westenberg M., Wetzel T., Ziebell H., Lebas B.S.M. Guidelines for the reliable use of high throughput sequencing technologies to detect plant pathogens and pests. *Peer Commun J*. 2022;2:e62. doi 10.24072/pcjournal.181

Menzel P., Ng K.L., Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun*. 2016;7(1):11257. doi 10.1038/ncomms11257

Moubset O., François S., Maclot F., Palanga E., Julian C., Claude L., Fernandez E., … Massart S., Ogliastro M., Martin D.P., Filloux D., Roumagnac P. Virion-associated nucleic acid-based metagenomics: a decade of advances in molecular characterization of plant viruses. *Phytopathology*. 2022;112(11):2253-2272. doi 10.1094/PHYTO-03-22-0096-RVW

Nabeshima T., Abe J. High-throughput sequencing indicates novel *Varicosavirus*, *Emaravirus*, and *Deltapartitivirus* infections in *Vitis coignetiae*. *Viruses*. 2021;13(5):827. doi 10.3390/v13050827

Ounit R., Lonardi S. Higher classification sensitivity of short metagenomic reads with CLARK-*S*. *Bioinformatics*. 2016;32(24):3823-3825. doi 10.1093/bioinformatics/btw542

Panno S., Davino S., Caruso A.G., Bertacca S., Crnogorac A., Mandić A., Noris E., Matić S. A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the mediterranean basin. *Agronomy*. 2021;11(11):2188. doi 10.3390/agronomy11112188

Pappas N., Roux S., Hölzer M., Lamkiewicz K., Mock F., Marz M., Dutilh B.E. Virus bioinformatics. In: Encyclopedia of Virology. Elsevier, 2021;124-132. doi 10.1016/B978-0-12-814515-9.00034-5

Pecman A., Kutnjak D., Gutiérrez-Aguirre I., Adams I., Fox A., Boonham N., Ravnikar M. Next generation sequencing for detection and discovery of plant viruses and viroids: comparison of two approaches. *Front Microbiol*. 2017;8:1998. doi 10.3389/fmicb.2017.01998

Peng Y., Leung H.C.M., Yiu S.M., Chin F.Y.L. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28(11):1420-1428. doi 10.1093/bioinformatics/bts174

Piombo E., Abdelfattah A., Droby S., Wisniewski M., Spadaro D., Schena L. Metagenomics approaches for the detection and surveillance of emerging and recurrent plant pathogens. *Microorganisms*. 2021;9(1):188. doi 10.3390/microorganisms9010188

Rang F.J., Kloosterman W.P., de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol*. 2018;19(1):90. doi 10.1186/s13059-018-1462-9

Roossinck M.J., Martin D.P., Roumagnac P. Plant virus metagenomics: advances in virus discovery. *Phytopathology*. 2015;105(6):716-727. doi 10.1094/PHYTO-12-14-0356-RVW

Rose R., Constantinides B., Tapinos A., Robertson D.L., Prosperi M. Challenges in the analysis of viral metagenomes. *Virus Evol*. 2016; 2(2):vew022. doi 10.1093/ve/vew022

Rott M., Xiang Y., Boyes I., Belton M., Saeed H., Kesanakurti P., Hayes S., Lawrence T., Birch C., Bhagwat B., Rast H. Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. *Plant Dis*. 2017;101(8):1489-1499. doi 10.1094/PDIS-03-17-0306-RE

Roy S., Coldren C., Karunamurthy A., Kip N.S., Klee E.W., Lincoln S.E., Leon A., Pullambhatla M., Temple-Smolkin R.L., Voelkerding K.V., Wang C., Carter A.B. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines. *J Mol Diagn*. 2018;20(1):4-27. doi 10.1016/j.jmoldx.2017.11.003

Rubino L., Abrahamian P., An W., Aranda M.A., Ascencio-Ibañez J.T., Bejerman N., Blouin A.G., … Whitfield A.E., Wylie S.J., Yang C., Zerbini F.M., Zhang S. Summary of taxonomy changes ratified by the International Committee on Taxonomy of Viruses from the Plant Viruses Subcommittee, 2025. *J Gen Virol*. 2025;106(7):002114. doi 10.1099/jgv.0.002114

Sayers E.W., Beck J., Bolton E.E., Brister J.R., Chan J., Connor R., Feldgarden M., … Wang J., Ye J., Zellers E., Schneider V.A., Pruitt K.D. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res*. 2025;53(D1):D20-D29. doi 10.1093/nar/gkae979

Shvets D., Sandomirsky K., Porotikova E., Vinogradova S. Metagenomic analysis of ampelographic collections of Dagestan revealed the presence of two novel grapevine viruses. *Viruses*. 2022;14(12): 2623. doi 10.3390/v14122623

Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J.M., Birol İ. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009;19(6):1117-1123. doi 10.1101/gr.089532.108

Sović I., Šikić M., Wilm A., Fenlon S.N., Chen S., Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat Commun*. 2016;7(1):11307. doi 10.1038/ncomms11307

Sun K., Liu Y., Zhou X., Yin C., Zhang P., Yang Q., Mao L., Shentu X., Yu X. Nanopore sequencing technology and its application in plant virus diagnostics. *Front Microbiol*. 2022;13:939666. doi 10.3389/fmicb.2022.939666

Sutton T.D.S., Clooney A.G., Ryan F.J., Ross R.P., Hill C. Choice of assembly software has a critical impact on virome characterisation. *Microbiome*. 2019;7(1):12. doi 10.1186/s40168-019-0626-5

Turco S., Golyaev V., Seguin J., Gilli C., Farinelli L., Boller T., Schumpp O., Pooggin M.M. Small RNA-omics for virome reconstruction and antiviral defense characterization in mixed infections of cultivated *Solanum* plants. *Mol Plant Microbe Interact*. 2018;31(7):707-723. doi 10.1094/MPMI-12-17-0301-R

VanderWeele T.J., Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167(4):268-274. doi 10.7326/M16-2607

Villamor D.E.V., Ho T., Al Rwahnih M., Martin R.R., Tzanetakis I.E. High throughput sequencing for plant virus detection and discovery. *Phytopathology*. 2019;109(5):716-725. doi 10.1094/PHYTO-07-18-0257-RVW

Vinogradova S., Porotikova E., Navrotskaya E., Galbacs Z.N., Massart S., Varallyay E. The first virome of a Russian vineyard. *Plants*. 2023;12(18):3292. doi 10.3390/plants12183292

Vivek A.T., Zahra S., Kumar S. From current knowledge to best practice: a primer on viral diagnostics using deep sequencing of virus-derived small interfering RNAs (vsiRNAs) in infected plants. *Methods*. 2020;183:30-37. doi 10.1016/j.ymeth.2019.10.009

Wang Z., Qin L., Liu J., Jiang L., Zou X., Chen X., Song F., Dai H., Hou Y. Forensic nanopore sequencing of microhaplotype markers using QitanTech's QNome. *Forensic Sci Int Genet*. 2022;57:102657. doi 10.1016/j.fsigen.2021.102657

White D.J., Wang J., Hall R.J. Assessing the impact of assemblers on virus detection in a de novo metagenomic analysis pipeline. *J Comput Biol*. 2017;24(9):874-881. doi 10.1089/cmb.2017.0008

Wick R.R., Judd L.M., Gorrie C.L., Holt K.E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017;13(6):e1005595. doi 10.1371/journal.pcbi.1005595

Wood D.E., Lu J., Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20(1):257. doi 10.1186/s13059-019-1891-0

Zerbino D.R., Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18(5):821-829. doi 10.1101/gr.074492.107

Zheng Y., Gao S., Padmanabhan C., Li R., Galvez M., Gutierrez D., Fuentes S., Ling K.-S., Kreuze J., Fei Z. VirusDetect: an automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology*. 2017;500:130-138. doi 10.1016/j.virol.2016.10.017

Zhuravlyov V.S., Dolgikh V.V., Timofeev S.A., Gannibal F.B. RNA interference method in plant protection against insect pests. *Plant Prot News*. 2022;105(1):28-39. doi 10.31993/2308-6459-2022-105-1-15219

Zubov V.V., Chemeris D.A., Vasilov R.G., Kurochkin V.E., Alekseev Ya.I. Brief history of high-throughput nucleic acid sequencing methods. *Biomics*. 2021;13(1):27-46. doi 10.31301/2221-6197.bmcs.2021-4

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 03.09.2025. После доработки 31.10.2025. Принята к публикации 31.10.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):321-329

doi 10.18699/vjgb-26-36

Применимость инструмента для анализа временных рядов StatFaRmer при цифровом фенотипировании сои (*Glycine max*)

Д.С. Ульянов, А.А. Ульянова, А.А. Кочешкова, А.О. Блинков, А.В. Архипов, Я.С. Меглицкая, Н.Ю. Свистунова, Г.И. Карлов, М.Г. Дивашук

Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва, Россия
uldas1508@gmail.com

Аннотация. Современные исследования в области агробиотехнологий все чаще опираются на методы автоматизированной фиксации и интерпретации морфофизиологических и спектральных характеристик растений – направление, известное как цифровое фенотипирование. Этот подход направлен на обнаружение устойчивых различий между генотипами, культивируемыми в неидентичных условиях среды. Ранее был предложен StatFaRmer – инструмент с открытым кодом, совершенствующийся в рамках настоящей работы для комплексного анализа временных фенотипических наборов данных, преимущественно ориентированный на изучение сельскохозяйственных культур, таких как соя (*Glycine max*). Разработанный инструмент реализует автоматизированные процедуры предобработки данных, включая синхронизацию временны́х меток между образцами и устранение шумовых артефактов и выбросов. Эти функции особенно актуальны для многомесячных экспериментов, связанных с оценкой параметров роста, колебаний площади фотосинтетического аппарата или других биометрических показателей. Поддержка стандартизированных форматов данных (XLSX, CSV) обеспечивает совместимость с распространенными системами фенотипирования, упрощая кроссплатформенную интеграцию. Таким образом, инструмент может поддерживать интеграцию с популярными HTPP-платформами (например, Traitmill, HyperAIxpert, Plant Accelerator), что позволяет использовать данные, полученные из различных источников, в едином аналитическом конвейере. Для экспериментов с соей StatFaRmer предоставляет настраиваемый дисперсионный анализ (ANOVA) с визуализацией диагностических параметров (нормальность распределения, гомогенность дисперсий) и оценкой значимости эффектов между пользовательскими группами. Пример применения включает сравнение параметров роста 20 сортов сои в условиях контролируемого стресса: инструмент автоматически агрегировал данные с неравномерной частотой измерений (от 1 часа до 3 суток), идентифицировал аномалии в динамике удлинения гипокотиля и рассчитал статистическую значимость различий между группами ($p < 0.01$). Инструмент протестирован на масштабных наборах данных (более 2 000 измерений на эксперимент). StatFaRmer реализован в виде веб-приложения на платформе Shiny с пошаговыми инструкциями для установки и запуска в ОС Windows и Linux. Все этапы обработки – от первичных данных до итоговых графиков – документируются, что обеспечивает прозрачность анализа и соответствие требованиям воспроизводимости исследований. Таким образом, StatFaRmer предлагает специализированное решение для статистической верификации гипотез в цифровом фенотипировании сои, сокращая время на подготовку данных и минимизируя риски ошибок при работе с нестационарными временными рядами.

Ключевые слова: высокопроизводительное фенотипирование растений; визуализация фенотипических данных; анализ временных рядов; цифровые платформы фенотипирования; генотип-фенотипический анализ; статистический анализ фенотипических данных; программное обеспечение с открытым исходным кодом; автоматизированный анализ данных

Для цитирования: Ульянов Д.С., Ульянова А.А., Кочешкова А.А., Блинков А.О., Архипов А.В., Меглицкая Я.С., Свистунова Н.Ю., Карлов Г.И., Дивашук М.Г. Применимость инструмента для анализа временных рядов StatFaRmer при цифровом фенотипировании сои (*Glycine max*). *Вавиловский журнал генетики и селекции*. 2026;30(2):321-329. doi 10.18699/vjgb-26-36

Финансирование. Исследование проведено при поддержке Министерства науки и высшего образования Российской Федерации (госзадание FGUM-2025-0010).

Applicability of the StatFaRmer time series analysis tool in soybean (*Glycine max*) digital phenotyping

D.S. Ulyanov, A.A. Ulyanova, A.A. Kocheshkova, A.O. Blinkov, A.V. Arkhipov, Ya.S. Meglitskaya, N.Yu. Svistunova, G.I. Karlov, M.G. Divashuk

All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia
uldas1508@gmail.com

© Ульянов Д.С., Ульянова А.А., Кочешкова А.А., Блинков А.О., Архипов А.В., Меглицкая Я.С., Свистунова Н.Ю., Карлов Г.И., Дивашук М.Г., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

Abstract. Contemporary agrobiotechnology research increasingly relies on automated methods for capturing and interpreting morphophysiological and spectral plant characteristics – a field known as digital phenotyping. This approach aims to identify stable differences between genotypes cultivated under non-identical environmental conditions. We previously introduced StatFaRmer, an open-source tool that we further develop here for comprehensive analysis of temporal phenotypic datasets, with a primary focus on crops such as soybean (*Glycine max*). The tool implements automated data preprocessing procedures, including synchronization of timestamps across samples and removal of noise artifacts and outliers. These features are particularly relevant for multi-month experiments involving assessments of growth parameters, fluctuations in photosynthetic apparatus area, or other biometric indicators. Support for standardized data formats (XLSX, CSV) ensures compatibility with common phenotyping systems, simplifying cross-platform integration. Thus, the tool can integrate with widely used HTPP platforms (e. g., Traitmill, HyperAlxpert, Plant Accelerator), enabling data from diverse sources to be analyzed within a single pipeline. For soybean experiments, StatFaRmer provides customizable analysis of variance (ANOVA) with visualization of diagnostic parameters (normality of distribution, homogeneity of variances) and evaluation of effect significance between user-defined groups. An example application compares growth parameters across 20 soybean cultivars under controlled stress: the tool automatically aggregated data with uneven measurement frequencies (from 1 hour to 3 days), identified anomalies in hypocotyl elongation dynamics, and computed statistical significance between groups ($p < 0.01$).The tool has been tested on large-scale datasets (over 2,000 measurements per experiment). StatFaRmer is implemented as a Shiny-based web application, with step-by-step deployment guides for Windows and Linux. All processing stages – from raw data to final plots – are documented to ensure transparency and compliance with research reproducibility standards. Thus, StatFaRmer offers a specialized solution for statistical hypothesis testing in soybean digital phenotyping, reducing data preparation time and minimizing risks of error when handling non-stationary time series.

Key words: high-throughput plant phenotyping; phenotypic data visualization; time series analysis; digital phenotyping platforms; genotype-phenotype analysis; statistical analysis of phenotypic data; open-source software; automated data analysis

For citation: Ulyanov D.S., Ulyanova A.A., Kocheshkova A.A., Blinkov A.O., Arkhipov A.V., Meglitskaya Ya.S., Svistunova N.Yu., Karlov G.I., Divashuk M.G. Applicability of the StatFaRmer time series analysis tool in soybean (*Glycine max*) digital phenotyping. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):321-329. doi 10.18699/vjgb-26-36

Введение

Несмотря на стремительное развитие геномики, фенотипирование – оценка внешних признаков и характеристик растений – остается ключевым узким местом в ускорении селекционного процесса (Fasoula et al., 2020; Patel et al., 2023). Традиционные методы, основанные на ручных измерениях, характеризуются высокой трудоемкостью, субъективностью, низкой пропускной способностью и необходимостью деструктивного отбора проб, что делает их непригодными для масштабных программ по созданию новых, более продуктивных и устойчивых сортов (Atefi et al., 2021; Li L. et al., 2023).

Выходом из этой ситуации является высокопроизводительное фенотипирование растений (high-throughput plant phenotyping, HTPP) – технология, представляющая собой прорыв в фенотипических исследованиях. Этот подход основан на интеграции современных сенсорных систем (гиперспектральная, тепловизионная или флуоресцентная съемка, 3D-реконструкция) и методов машинного обучения для автоматизированного, неразрушающего и точного анализа растений.

При традиционном ручном фенотипировании существуют такие недостатки, как большие временные и финансовые затраты, низкая масштабируемость из-за ограничения скорости и объема работ, субъективность и вариабельность, а также ограниченность измеряемых признаков (Thrash et al., 2022; Abebe et al., 2023; Anand et al., 2023; Yuan et al., 2023). В то же время цифровое фенотипирование имеет такие преимущества, как существенный прирост эффективности, непрерывность и недеструктивность измерений, улучшенная точность и объективность и возможность измерять десятки признаков одновременно, существенно повышающая полученные объемы данных (Li D. et al., 2021; Anand et al., 2023; Buelvas et al., 2023; Gyan et al., 2023; Li L. et al., 2023; Lu et al., 2023; Wang et al., 2024).

Современные HTP-платформы генерируют огромные объемы гетерогенных данных, включая изображения (RGB, гиперспектральные, тепловые), 3D-модели и показания сенсоров, достигающие терабайтов в день. Ключевыми проблемами являются не только объем, но и многомодальность, пространственно-временной характер данных, а также необходимость хранения информации и управления ей. Отсутствие стандартизированных форматов и сложности обеспечения качества данных еще более осложняют анализ (Coppens et al., 2017; Morota et al., 2019; Danilevicz et al., 2021; Gill et al., 2022; Ninomiya, 2022).

Существующие статистические методы, используемые для данных фенотипирования, включают следующие модели:

- одномерные и многомерные методы анализа для проверки гипотез и изучения взаимосвязей между признаками растений. Включают корреляционный анализ, факторный анализ (FA), главный компонентный анализ (PCA) для уменьшения размерности и выявления важных признаков (Rahaman et al., 2015);
- рандомизированные регрессионные модели и модели многофакторного взаимодействия (например, взаимодействие генотип–среда), применяемые для моделирования динамических фенотипических данных (Morota et al., 2019);
- кластерный анализ (иерархический, k-средних и т. д.) для оценки сходств и различий между растениями и стрессовыми реакциями (Rahaman et al., 2015);

- машинное обучение и глубокое обучение для обработки масштабных и сложных данных, включая классификацию, сегментацию и локализацию признаков на изображениях, а также прогнозирование фенотипических реакций. У таких подходов есть свои преимущества, например возможность обнаружить неочевидную заранее корреляцию; но часто такие методы требуют много обучающих данных, и их интерпретируемость ограничена (Coppens et al., 2017; Ubbens et al., 2025).

Для более комплексного применения вышеописанных методов существуют различные инструменты. Среди коммерческих решений одной из опций является Genstat, предоставляющий комплексные возможности статистического анализа, специально разработанные для приложения в селекции растений. Функции включают пространственный анализ, эксперименты со множественными вариантами условий, геномную селекцию и инструменты анализа стабильности, адаптированные для сельскохозяйственных данных. HortControl – это инструмент, по умолчанию поставляемый для платформы TraitFinder (HortControl – Plant Data Management Software, 2025), с которой работают авторы. Он предлагает интегрированное управление данными для цифрового фенотипирования с совместимостью с BrAPI и снабжен некоторыми инструментами анализа данных. Но для регрессионного анализа пользователю может понадобиться экспортировать файл .csv с полученными данными и проводить дальнейшие анализы сторонними инструментами.

Существуют также инструменты с открытым исходным кодом. К примеру, пакеты R, такие как ASReml-R и pcvr, предоставляют специализированные функции для продольного моделирования роста и пространственного анализа данных фенотипирования (Sumner, 2025). Эти инструменты реализуют передовые модели смешанных эффектов и методы функционального анализа данных, но требуют опыта в программировании. Также существует PIPPA (plant image phenotyping and analysis/фенотипирование и анализ изображений растений), предоставляющий веб-доступ к высокопроизводительным вычислительным ресурсам для конвейеров анализа изображений (Coppens et al., 2017).

Для решения аналогичных задач в области продольных исследований (longitudinal studies – метод научного наблюдения, при котором одна и та же группа объектов изучается многократно в течение длительного времени) нами разработан StatFaRmer – открытый веб-инструмент на базе R с двухэтапной архитектурой: Master Wizard для обработки данных и Main Application для статистического анализа, не требующий навыков программирования. На сегодняшний день функционал StatFaRmer включает:

- Master Wizard: валидацию проектов, DBSCAN кластеризацию, фильтрацию выбросов (Z-критерий/IQR), техническую агрегацию данных;
- Main Application: автоматический выбор моделей (ANOVA/смешанные модели/сплайны), многофакторный анализ с post-hoc тестами, диагностику нормальности и гомогенности дисперсий, анализ величины эффектов и кривых роста;
- интерактивную визуализацию с Plotly, экспорт высокого разрешения (SVG, PNG, PDF);
- специализированные модули: анализ размеров эффектов, сводки роста;
- экспорт результатов в табличных форматах.

Инструмент оптимизирован для продольных исследований в контролируемых условиях, поддерживает двухфакторные взаимодействия в моделях ANOVA и обеспечивает прозрачность статистических выводов. StatFaRmer заполняет нишу между специализированным ПО и потребностями исследователей, предлагая гибкий анализ сложных фенотипических данных.

В сравнении со стандартным HortControl наш инструмент, будучи системой, развернутой на R/Shiny, предлагает, во-первых, большую гибкость в настройке анализа под конкретный эксперимент; во-вторых, – более продвинутую глубину и автоматизацию статистического конвейера, чем HortControl, более фокусирующийся на данных с датчиков и базовых расчетах.

Если сравнивать с уже существующими статистическими пакетами и платформами (Genstat, ASReml-R, pcvr), то они представляют собой мощные, но общего назначения статистические среды. Они могут все, что нужно, но требуют от пользователя глубоких знаний статистики и/или навыков программирования для написания корректных моделей, особенно для сложных пространственных и временных данных. В то же время StatFaRmer автоматически генерирует сложные ANOVA-модели, включая все взаимодействия, что избавляет исследователя от рутинного и подверженного ошибкам кодирования. Также, в отличие от универсальных пакетов, StatFaRmer из коробки специализируется на фенотипических данных (повторные измерения, выбросы). Еще одно положительное отличие – интерактивность и визуализация: возможность интерактивного подмножества данных (по генотипам, обработкам, временным кластерам) и мгновенного получения результатов для проверки гипотез, что является ключевым преимуществом перед традиционными пакетами, где такой анализ требует постоянного переписывания кода.

В нашей работе StatFaRmer (Ulyanov et al., 2025) применен для анализа данных, полученных при цифровом фенотипировании сои (*Glycine max*) – ключевой сельскохозяйственной культуры. Инструмент предусматривает возможность интеграции с платформами, использующими различные сенсорные системы (мультиспектральные сканеры, 3D-камеры), благодаря поддержке стандартизированных форматов данных (XLSX, CSV) и гибкой настройке параметров. Это позволяет адаптировать его для работы с данными, собранными как в контролируемых условиях (например, климатических камерах), так и в полевых экспериментах, где нерегулярность измерений и разнородность источников требуют автоматизированной предобработки. Стандартизация входных данных и модульная архитектура StatFaRmer обеспечивают совместимость с такими системами, как Traitmill или PhytoOracle, что открывает перспективы для кроссплатформенных исследований.

Цель настоящей работы заключалась в представлении возможностей инструмента StatFaRmer, последовательной демонстрации его функционального потенциала для ознакомления читателя, а также в описании усовершенствований, реализованных в нем со времени предыдущей публикации.

Материалы и методы

Экспериментальный дизайн и растительный материал. В исследовании использовано 50 сортов сои (*G. max*) различного происхождения (российская и зарубежная селекция), отличающихся по срокам созревания, типу роста и рекомендованным регионам возделывания. Семена перед посевом подвергали обработке фунгицидом, после чего высевали на глубину 2 см во влажный торф в горшки объемом 500 мл (по четыре семени на горшок); после всходов оставляли три растения на горшок.

Культивация растений осуществлялась в контролируемых условиях климатической камеры на протяжении всего вегетационного цикла. Эксперимент включал два варианта условий, различавшихся режимом освещенности при идентичных остальных параметрах микроклимата (температура, влажность, уровень питательных веществ). Для каждого варианта предусматривалось по две биологические повторности, что в совокупности обеспечило выборку из 200 растений (50 сортов × 2 варианта × 2 повторности). Основной уход включал регулярный полив водой комнатной температуры в первые 7–10 дней, затем – по мере необходимости, а после появления настоящих листьев – ежедневное внесение минерального удобрения. Для целей фенотипирования растения распределялись в случайно сформированные группы по 12 экземпляров, итоговое число групп составило по 9 для каждого варианта опыта. Подробное описание экспериментальных процедур и параметров культивации представлено в работе (Ulyanov et al., 2025).

Цифровое фенотипирование. Сканирование проводили на установке для фенотипирования TraitFinder (Phenospex, Нидерланды) с двумя мультиспектральными сканерами в следующих диапазонах длин волн: красный канал (R): 624–634 нм, зеленый канал (G): 530–540 нм, синий канал (B): 465–485 нм и ближний инфракрасный канал (NIR): 720–750 нм.

Фенотипирование растений осуществляли по схеме, где распределение сортов и вариантов опыта по блокам фиксировалось во встроенном программном обеспечении HortControl и дублировалось в таблицах StatFaRmer для последующего анализа. Растения распределялись по экспериментальной площадке, представленной в виде layout – виртуальной схемы, отражающей физическое расположение блоков и растений. Все растения были разбиты на блоки, каждый из которых соответствует одному столу и имеет уникальный идентификатор; таких блоков получилось 9. Блоки дополнительно делились на единицы (units), соответствующие индивидуальным растениям для сбора данных. В HortControl блоки резервировались для эксперимента и связывались с биологической информацией (генотип, обработка) через присвоение метаданных.

Перед сканированием растения транспортировали из климатических камер на специализированный фенотипический стол, соблюдая идентичное позиционирование в рамках заранее заданных цифровых координат. Процедура выполнялась в затемненном помещении с отключенной системой кондиционирования для исключения артефактов, вызванных воздушными потоками или изменением освещенности. Каждое растение размещали так, чтобы его центр массы совпадал с геометрическим центром сканируемой зоны, минимизируя перекрытие листьев и побегов с соседними образцами. Для стабилизации растений в горшках использовали опорные конструкции из зеленых пластиковых стержней и черной проволоки (спектрально нейтральной в диапазоне датчика PlantEye).

В ходе эксперимента выполнено пять последовательных сканирований, общая продолжительность наблюдения составляла 20 дней. Регистрировали общую высоту растения, а также спектральные индексы (NDVI, PSRI), рассчитанные по данным мультиспектрального сенсора следующим образом:

Средняя высота растения (plant height average, мм): усредненное значение высоты верхних 10 % точек растения, что минимизирует влияние внешних факторов (например, ветра).

Спектральные индексы:

NDVI (normalized difference vegetation index):

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}}.$$

Диапазон значений: от –1 до 1. Высокие значения (>0.66) указывают на здоровые фотосинтезирующие ткани.

PSRI (plant senescence reflectance index):

$$\mathrm{PSRI} = \frac{\mathrm{RED} - \mathrm{BLUE}}{\mathrm{NIR}}.$$

Используется для оценки старения листьев (сенесценции).

Подготовка данных. Пользователи могут подгружать собственные эксперименты, помещая проекты в папку StatFaRmer/data/project_NAME/, предварительно подготовив данные следующим образом:

1. Базовый архив эксперимента. Проект должен содержать архивированный эксперимент в формате ZIP (*_data.zip), включающий CSV-таблицу, импортированную из HortControl с исходными данными.
2. Таблица ручных корректировок. Обязательно наличие файла *_handmade.csv со столбцами:
   – V.T.R – уникальный идентификатор биологического образца (сорт/обработка/повтор);
   – Treatment – параметры обработки (переопределяет соответствующий столбец в исходной таблице);
   – Cultivar – сорт растения.

   Эти CSV-файлы формируются ручным способом для внесения данных о растениях. Напротив уникального идентификатора каждого растения нужно указать его сорт и обработку, также можно внести любые другие данные о каждом растении.
3. Таблица пространственной привязки. Файл *_translation.csv должен включать:

– V.T.R – идентификатор образца;
– T:X:Y – координаты расположения образца, соответствующие столбцу unit в исходных данных.

Эти CSV-файлы формируются ручным способом для того, чтобы программа могла связать импортированные данные из пункта 1 с описанием экспериментальных образцов из пункта 2. Для получения такой таблицы необходимо соотнести идентификатор каждого растения V.T.R с соответствующими ему координатами в базе данных HortControl. Координаты unit кодируются следующим образом:

(номер блока соответствующей группы растений):
(номер ряда, начиная от баркода):
(номер растения в ряду, слева направо).

4. Дополнительная группировка (опционально). Файл groups.xlsx с обязательным столбцом Cultivar.

Прочие столбцы могут использоваться как факторы ANOVA и должны содержать только латинские буквы, цифры и подчеркивания.

Все дополнительные столбцы в указанных таблицах сохраняются и доступны для анализа в качестве независимых факторов.

Настройка анализа осуществляется через графический интерфейс Master Wizard, включая:
– выбор проекта из списка доступных с его валидацией;
– подбор DBSCAN eps (временной интервал между измерениями);
– выбор метода фильтрации/винзоризация/аннотации выбросов (IQR или Z-score) на основе указанной структуры ячеек опыта;
– настройку технической агрегации (медиана или среднее);
– экспорт и импорт конфигураций опыта для прозрачности обработки.

StatFaRmer эффективно обрабатывает высокочастотные временные ряды с неоднородными интервалами измерений, автоматически корректируя артефакты и агрегируя данные для снижения ошибок в продольных исследованиях. Инструмент включает модуль многофакторного дисперсионного анализа (ANOVA) с автоматической проверкой статистических предпосылок. Генерируемые интерактивные визуализации позволяют сосредоточиться на интерпретации результатов, минимизируя технические сложности анализа.

Результаты

Кластеризация, фильтрация выбросов и группировка данных

Кластеризация и фильтрация выбросов – важный этап первоначальной обработки полученных фенотипических данных. Под выбросами (аномалиями) понимаются точки данных, значительно отклоняющиеся от общего паттерна, что может быть вызвано техническими артефактами (например, временным затемнением сенсора) или биологическими факторами (локальные повреждения тканей). Она позволяет избавиться как от шумов, полученных из-за разного времени начала фенотипирования в различных

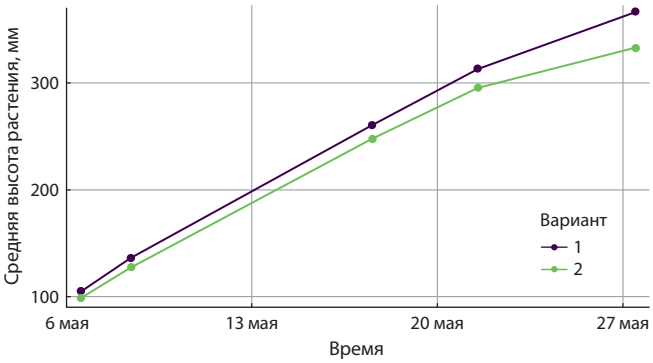


Рис. 1. Динамика изменения параметра «усредненная высота растения» между разными вариантами опыта.

Приведено два графика, показывающих изменение средней высоты растения во времени для вариантов опытов 1 и 2. На рисунке представлены все 50 сортов в двух повторностях для быстрого первичного анализа закономерностей в распределении высоты при вариантах опытов 1 и 2.

экспериментальных блоках, погрешностей сенсоров, а также влияния внешних факторов (например, ветра), так и от крупных случайных выбросов, таких как посторонние объекты в зоне видимости камеры. На данном этапе работы программы также производится кластеризация данных; можно произвести группировку и сравнить данные при разных параметрах – условиях, обработках и других. Фильтрацию выбросов в программе StatFaRmer можно включать и отключать через графический интерфейс, установив или убрав галочку “plot with outliers”, что позволяет пользователю самостоятельно контролировать данный аспект анализа. Этот этап обработки показан на рис. 1.

StatFaRmer предоставляет возможность проводить сравнительный анализ наборов растений, объединяя образцы по обработкам, сортам или любым дополнительным параметрам, указанным пользователем в сопутствующих таблицах данных. Графики демонстрируют, что усредненная высота растений, культивируемых при варианте опыта 1, в среднем превышает таковую при варианте опыта 2. Значительная разница наблюдается уже на начальном этапе, что при высоком количестве образцов (по 100 для каждого варианта опыта) и начале фенотипирования не с первого дня свидетельствует о том, что различия между вариантами опытов проявляются довольно быстро.

ANOVA

StatFaRmer предоставляет автоматический выбор статистических моделей: классический ANOVA для сбалансированных планов (≤ 10 уровней факторов), смешанные модели для несбалансированных планов (11–20 уровней факторов) и сплайновые модели для анализа временных рядов. Программа старается подобрать оптимальную схему анализа с учетом всех заданных факторов и их парных взаимодействий, при этом в продвинутых настройках у пользователя сохраняется возможность фиксировать тип модели.

При анализе рис. 1 возникает важный вопрос: является ли ускоренный рост при варианте опыта 1 положитель-

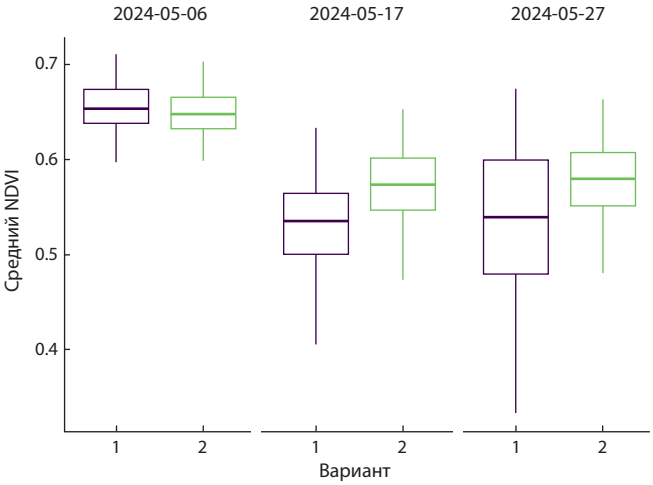


Рис. 2. Сравнение среднего NDVI всех растений в трех временных точках: в начале, середине и конце эксперимента для двух вариантов опыта.

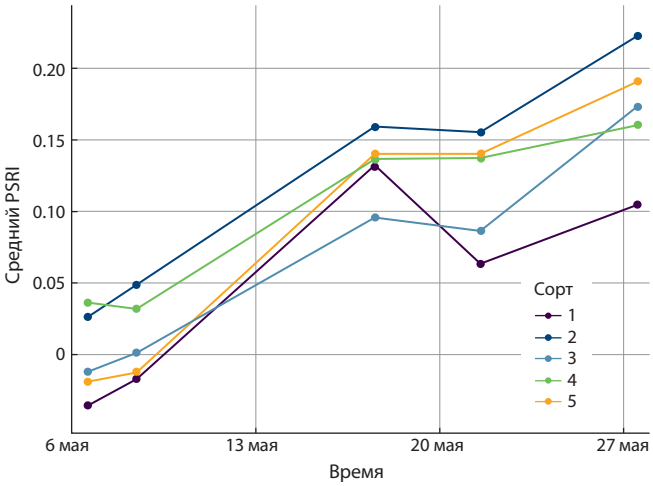


Рис. 3. Изменения показателя среднего PSRI в условиях варианта опыта 1 для пяти сортов.

ным признаком, или свидетельствует о стрессе растений? Для оценки состояния растений можно использовать показатель NDVI, позволяющий приблизительно оценить уровень фотосинтетической активности и общего здоровья растений. Более высокие значения NDVI указывают на лучшее состояние растений и более интенсивный фотосинтез, в то время как низкие значения могут свидетельствовать о стрессе. Изменение среднего NDVI во времени, усредненное по всем сортам, показано на рис. 2.

На рис. 2 NDVI не демонстрирует существенных различий между вариантами опыта в первой временной точке, однако дальше NDVI растений, которые выращивали при варианте опыта 2, значимо выше, как видно из того, что медиана распределения соответствует более высокому NDVI. Далее это также можно будет увидеть из сравнительной таблицы групп в разделе «Экспорт результатов в табличных форматах». Это указывает на более благоприятные условия для роста в условиях варианта опыта 2.

Выборка из данных для анализа временных кластеров

В условиях работы с большими наборами данных пользователям необходима возможность выборочного анализа для самостоятельного выявления закономерностей. В предыдущем примере мы сравнивали значения NDVI в разных вариантах опыта для всех сортов. Теперь предположим, что нам необходимо отобрать из этого набора сорта, более чувствительные и менее чувствительные к варианту опыта. Можно было бы выбрать такие значения полей, чтобы сравнивать один сорт с другим, но проверка таким образом каждого из 50 сортов могла бы оказаться трудоемкой.

Вместо этого можно использовать альтернативный подход. Функция StatFaRmer по созданию выборок данных позволяет, к примеру, сравнивать различные сорта в идентичных условиях, а не один сорт между вариантами опыта. В качестве примера рассмотрим пять сортов и оценим параметр PSRI, отражающий увядание растений. Чем выше этот параметр, тем больше стресса испытывает растение. Результаты этого сравнения для варианта опыта 1 представлены на рис. 3.

Интерактивная визуализация с группировкой по признакам

Интерактивная визуализация играет ключевую роль для пользователей, желающих получить адаптивный инструмент для анализа экспериментов. Для анализа отобранных сортов можно построить графики зависимости PSRI от времени с учетом варианта опыта. При этом для изучения динамики множества групп рационально применять фасетирование – разделение графиков на несколько субграфиков, создаваемых по одному принципу, для подгрупп этой же переменной/переменных. При использовании полей «фактор группировки», «выбранные сорта» и «формула фасетирования» (синтаксический шаблон, задействованный в визуализации данных для описания того, как набор данных разделяется на подгруппы (фасеты) при построении множества связанных графиков; краткое описание принципов построения формулы можно найти на странице github проекта), мы получим отдельный график для каждого сорта (рис. 4).

Такой подход позволяет детально изучить, как изменяется показатель увядания с течением времени для разных сортов. Например, можно заметить, что у некоторых сортов наблюдается резкое изменение динамики показателя – увеличение увядания до середины эксперимента с последующей стабилизацией, тогда как у других этот переломный момент либо слабо выражен, либо проявляется только на завершающих этапах.

Экспорт результатов в табличных форматах

StatFaRmer позволяет представлять результаты анализов в табличном формате и экспортировать графики в высоком разрешении (300 DPI) в форматах SVG, PNG и PDF для публикаций. Ниже представлена таблица с описательной

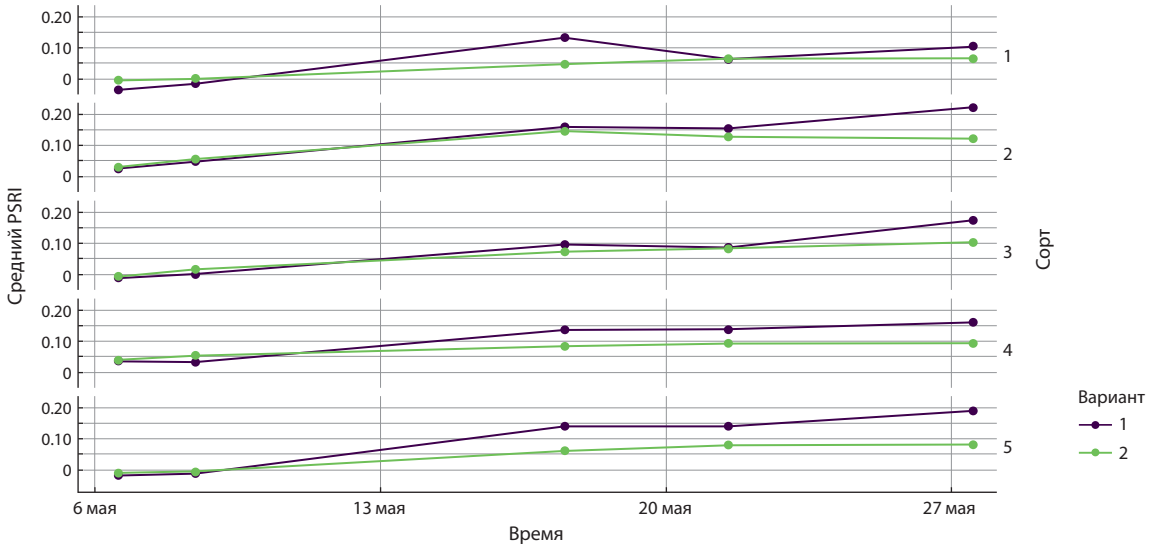


Рис. 4. Изменение среднего PSRI во времени для сортов 1–5.

Описательная статистика для рис. 2.

№	Временная точка	Вариант опыта	*n*	Среднее	Медиана	Стандартное отклонение	Стандартная ошибка среднего	min	max	25-й квартиль	75-й квартиль	Интер-квартильный размах	Коэффициент вариации	Коэффициент асимметрии	Куртозис
1	1	1	100	0.655	0.653	0.0259	0.00259	0.596	0.731	0.637	0.674	0.0369	3.96	–0.00113	0.0668
2	1	2	100	0.648	0.647	0.0268	0.00268	0.56	0.724	0.633	0.665	0.0315	4.14	–0.0214	1.3
3	4	1	100	0.532	0.535	0.0467	0.00467	0.395	0.633	0.5	0.564	0.0636	8.77	–0.411	0.263
4	4	2	100	0.57	0.573	0.0457	0.00457	0.421	0.653	0.547	0.601	0.0542	8.02	–0.823	1.16
5	6	1	100	0.528	0.539	0.0838	0.00838	0.333	0.674	0.479	0.6	0.12	15.9	–0.477	–0.537
6	6	2	100	0.579	0.58	0.0396	0.00396	0.48	0.664	0.551	0.607	0.0567	6.84	–0.0886	–0.274

Примечание. В колонке «Временная точка» отображены анализируемые группы, где 1, 4 и 6 – это порядковый номер временной точки, т. е. начало, середина и конец эксперимента.

статистикой для рис. 2, в которой указан набор параметров, характеризующий каждую из наблюдаемых групп в каждый момент времени.

Производительность

Производительность StatFaRmer протестирована на ноутбуке со следующей конфигурацией: процессор AMD Ryzen 5 5500U с Radeon Graphics (2.10 ГГц) и 8 ГБ оперативной памяти.

- Master Wizard: обработка проекта project_NO3 (58380 строк, 49 столбцов) занимает 8–16 секунд с медианной агрегацией, в зависимости файла конфигурации;
- Main Application: запуск и анализ происходят практически мгновенно благодаря кэшированию обработанных данных;
- визуализация оптимизирована для больших датасетов с использованием карт плотности для > 2000 измерений.

Реализованы логирование (logger.R) и бенчмаркинг (benchmark.R) затрачиваемых памяти и времени при работе модулей.

Таким образом, применение автоматизированных методов анализа временных рядов позволяет выявить устойчивые динамические паттерны в данных сои. В отличие от средств, не поддерживающих анализ временных рядов, StatFaRmer справляется с этой задачей, будучи при этом гибким инструментом, не требующим навыков в программировании. При анализе длительных экспериментов StatFaRmer демонстрирует эффективность в обработке высокочастотных временных рядов с неоднородными интервалами измерений. Кроме того, инструмент автоматически выполняет коррекцию артефактов и агрегацию данных. Эти процедуры критически важны для минимизации ошибок в продольных исследованиях. StatFaRmer также включает модуль многофакторного дисперсионного анализа (ANOVA) с автоматической диагностикой статистических предпосылок (например, проверкой нормальности распределения и однородности дисперсий). Интерактивные визуализации высокого разрешения, генерируемые инструментом, выгодно отличают его от большинства других платформ, где графики обычно статичны. В результате

исследователи могут сосредоточиться на интерпретации полученных результатов, а не на технических деталях анализа.

Обсуждение

В отличие от ряда существующих решений, таких как HTPheno (Hartmann et al., 2011) и IAP (Yang et al., 2020), ориентированных преимущественно на обработку изображений, StatFaRmer обеспечивает расширенную интеграцию с внешними системами фенотипирования. Это позволяет исследователям комбинировать данные из разнородных источников, минимизируя ручной труд. Системы HTPheno и IAP имеют ограниченную поддержку автоматизированного анализа нестационарных временных рядов большого объема, поэтому для полноценной статистической обработки таких данных зачастую требуются дополнительные инструменты. StatFaRmer, напротив, успешно решает задачу автоматизации анализа сложных фенотипических временных рядов, генерируемых высокопроизводительными платформами фенотипирования.

Параллельно с развитием StatFaRmer появились и другие специализированные инструменты, решающие смежные задачи цифрового фенотипирования. Например, пакет AllInOne Pre-processing (Najafabadi et al., 2023) сосредоточен на предобработке полевых фенотипических данных и предоставляет функции нормализации изображений, пространственного анализа и базовой визуализации. Хотя AllInOne эффективно повышает качество исходных данных и способен обрабатывать большие наборы снимков, он не обеспечивает полноценного анализа временных рядов или статистической проверки гипотез. StatFaRmer дополняет такие решения, предлагая углубленный статистический анализ очищенных данных, включая сравнение групп по нескольким факторам и выявление динамических закономерностей во времени.

Современным подходом, ориентированным на масштабируемость, является также платформа PhytoOracle (Gonzalez et al., 2023) – модульный конвейер обработки фенотипических данных, оптимизированный для высокопроизводительных вычислительных кластеров. PhytoOracle способен параллельно обрабатывать мультимодальные данные (например, RGB-изображения, тепловые карты, 3D-точечные облака) и эффективно масштабируется на большие объемы информации. Однако высокая сложность и требование специализированной инфраструктуры делают PhytoOracle менее доступным для широкого круга биологов. StatFaRmer, напротив, ориентирован на доступность: он реализован в виде веб-приложения с интуитивно понятным интерфейсом, что упрощает его применение на стандартных настольных системах. При этом StatFaRmer поддерживает обработку крупных массивов численных данных и интеграцию разнородных источников информации, обеспечивая баланс между мощностью анализа и удобством использования.

Для быстрой оценки возможностей программы был создан демонстрационный датасет на основе эксперимента с азотным стрессом у злаковых культур. Он выложен в открытом доступе на GitHub (https://github.com/Stathmin/StatFaRmer) и развернут для интерактивных тестов – на shinyapps.io (https://stathmin.shinyapps.io/StatFaRmer/), где можно ознакомиться с данными и протестировать рабочую копию StatFaRmer. Кроме того, мы провели масштабное тестирование на разнообразных датасетах, включая пшеницу, тритикале, сахарную свеклу, кукурузу, подсолнечник и сою. Программа показала себя как надежный и эффективный инструмент для анализа больших объемов фенотипических данных в различных экспериментальных условиях.

Заключение

Современные методы цифрового фенотипирования растений генерируют колоссальные объемы данных, включая высокочастотные временные ряды, структурированные метаданные (сорт, обработки, повторности, вариант опыта) и внешние параметры, интегрируемые из сторонних источников. Это, а также многообразие измеряемых платформой признаков, требует специализированных решений для их интерпретации. Разработанный нами инструмент StatFaRmer решает эту задачу благодаря:

- интерактивной визуализации данных с поддержкой временных срезов и фильтрации артефактов;
- многофакторному сравнению групп (по сортам, обработкам, пользовательским тегам) с применением различных статистических анализов;
- гибкой настройке срезов данных для самостоятельного поиска закономерностей, доступного среднему пользователю.

Таким образом, StatFaRmer упрощает обработку сложных фенотипических данных, сокращая время на выявление закономерностей адаптации сои и других культур при различных вариантах опыта, стрессах и других воздействиях.

Список литературы / References

Abebe A.M., Kim Y., Kim J., Kim S.L., Baek J. Image-based high-throughput phenotyping in horticultural crops. *Plants*. 2023;12(10):2061. doi 10.3390/plants12102061

Anand K.J., Nagre S.P., Shrivastava M.K., Amrate P.K., Patel T., Katara V.K. Enhancing crop improvement through synergistic integration of advanced plant breeding and proximal remote sensing techniques: a review. *Int J Plant Soil Sci*. 2023;35(19):121-138. doi 10.9734/ijpss/2023/v35i193533

Atefi A., Ge Y., Pitla S., Schnable J. Robotic technologies for high-throughput plant phenotyping: contemporary reviews and future perspectives. *Front Plant Sci*. 2021;12:611940. doi 10.3389/fpls.2021.611940

Buelvas R.M., Adamchuk V.I., Lan J., Hoyos-Villegas V., Whitmore A., Stromvik M.V. Development of a quick-install rapid phenotyping system. *Sensors* (*Basel*). 2023;23(9):4253. doi 10.3390/s23094253

Coppens F., Wuyts N., Inzé D., Dhondt S. Unlocking the potential of plant phenotyping data through integration and data-driven approaches. *Curr Opin Syst Biol*. 2017;4:58-63. doi 10.1016/j.coisb.2017.07.002

Danilevicz M.F., Bayer P.E., Nestor B.J., Bennamoun M., Edwards D. Resources for image-based high-throughput phenotyping in crops and data sharing challenges. *Plant Physiol*. 2021;187(2):699-715. doi 10.1093/plphys/kiab301

Fasoula D., Ioannides I., Omirou M. Phenotyping and plant breeding: overcoming the barriers. *Front Plant Sci*. 2020;10:1713. doi 10.3389/fpls.2019.01713

Gill T., Gill S.K., Saini D., Chopra Y., de Koff J.P., Sandhu K. A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics*. 2022;2(3):156-183. doi 10.1007/s43657-022-00048-z

Gonzalez E.M., Zarei A., Hendler N., Simmons T., Zarei A., Demieville J., Strand R., … Swetnam T.L., Merchant N., Michelmore R.W., Lyons E., Pauli D. PhytoOracle: scalable, modular phenomics data processing pipelines. *Front Plant Sci.* 2023;14:1112973. doi 10.3389/fpls.2023.1112973

Gyan F., Cudjoe D., Sadeghi-Tehran P., Virlet N., Riche A., Castle M., Greche L., Mohareb F., Simms D., Mhada M., Hawkesford M. Machine learning methods for automatic segmentation of images of field- and glasshouse-based plants for high-throughput phenotyping. *Plants*. 2023;12(10):2035. doi 10.3390/plants12102035

Hartmann A., Czauderna T., Hoffmann R., Stein N., Schreiber F. HTPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics*. 2011;12:148. doi 10.1186/1471-2105-12-148

HortControl – Plant Data Management Software. *PHENOSPEX*. URL: https://www.phenospex.com/products/plant-phenotyping/science-hortcontrol-data-management-software/ (accessed 8.28.25)

Li D., Quan C., Song Z., Li X., Yu G., Li C., Muhammad A. High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Front Bioeng Biotechnol*. 2021;8:623705. doi 10.3389/fbioe.2020.623705

Li L., Hassan M.A., Song J., Xie Y., Rasheed A., Yang S., Li H., Liu P., Xia X., He Z., Xiao Y. UAV-based RGB imagery and ground measurements for high-throughput phenotyping of senescence and QTL mapping in bread wheat. *Crop Sci*. 2023;63(6):3292-3309. doi 10.1002/csc2.21086

Lu Y., Wang J., Fu L., Yu L., Liu Q. High-throughput and separating-free phenotyping method for on-panicle rice grains based on deep learning. *Front Plant Sci.* 2023;14:1219584. doi 10.3389/fpls.2023.1219584

Morota G., Jarquín D., Campbell M.T., Iwata H. Statistical methods for the quantitative genetic analysis of high-throughput phenotyping data. In: Lorence A., Medina Jimenez K. (Eds) High-Throughput Plant Phenotyping. Methods in Molecular Biology. Vol. 2539. Humana, New York, 2019;269-296. doi 10.1007/978-1-0716-2537-8_21

Najafabadi M.Y., Heidari A., Rajcan I. AllInOne pre-processing: a comprehensive preprocessing framework in plant field phenotyping. *SoftwareX*. 2023;23:101464. doi 10.1016/j.softx.2023.101464

Ninomiya S. High-throughput field crop phenotyping: current status and challenges. *Breed Sci*. 2022;72(1):3-18. doi 10.1270/jsbbs.21069

Patel T., Babbar A., Behera K., Katara V.K., Anand K.J., Vyshnavi R.G., Pachori S., Bichewar N. Exploring the potential of proximal remote sensing in plant stress phenotyping: a comprehensive review. *Int J Environ Clim Change*. 2023;13(9):2602-2621. doi 10.9734/ijecc/2023/v13i92511

Rahaman M., Chen D., Gillani Z., Klukas C., Chen M. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front Plant Sci*. 2015;6:619. doi 10.3389/fpls.2015.00619

Sumner J. Longitudinal Growth Modeling Options [WWW Document]. 2025. URL: https://cran.r-project.org/web/packages/pcvr/vignettes/longitudinal.html (accessed 28.08.25)

Thrash T., Lee H., Baker R.L. A low-cost high-throughput phenotyping system for automatically quantifying foliar area and greenness. *Appl Plant Sci*. 2022;10(6):e11502. doi 10.1002/aps3.11502

Ubbens J., Stavness I., Pound M.P., Guo W. Deep learning in plant phenotyping: the first ten years. *Plant Phenomics*. 2025;7(4):100062. doi 10.1016/j.plaphe.2025.100062

Ulyanov D.S., Ulyanova A.A., Litvinov D., Kocheshkova A., Kroupina A.Yu., Syedina N.M., Voronezhskaya V.S., Vasilyev A.V., Karlov G.I., Divashuk M. StatFaRmer: cultivating insights with an advanced R shiny dashboard for digital phenotyping data analysis. *Front Plant Sci*. 2025;16:1475057. doi 10.3389/fpls.2025.1475057

Wang Z., Hao J., Shi X., Wang Q., Zhang W., Li F., Mur L.A.J., Han Y., Hou S., Han J., Sun Z. Integrating dynamic high-throughput phenotyping and genetic analysis to monitor growth variation in foxtail millet. *Plant Methods*. 2024;20(1):168. doi 10.1186/s13007-024-01295-z

Yang W., Feng H., Zhang X., Zhang J., Doonan J., Batchelor W., Xiong L., Yan J. Crop phenomics and high-throughput phenotyping: past decades, current challenges and future perspectives. *Mol Plant*. 2020;13(2):187-214. doi 10.1016/j.molp.2020.01.008

Yuan H., Song M., Liu Y., Xie Q., Cao W., Zhu Y., Ni J. Field phenotyping monitoring systems for high-throughput: a survey of enabling technologies, equipment, and research challenges. *Agronomy*. 2023;13(11):2832. doi 10.3390/agronomy13112832

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 22.04.2025. После доработки 20.10.2025. Принята к публикации 01.11.2025.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):330-338

doi 10.18699/vjgb-26-37

Описание морфологических характеристик колосьев пшеницы в базе данных SpikeDroidDB в виде цифрового паспорта

Е.Г. Комышев[1, 2], Ю.В. Кручинина[1, 2], В.С. Коваль[1, 2], А.А. Потешкина[1, 3], Н.В. Петраш[1, 3], В.В. Пискарев[1, 3], Н.П. Гончаров[1], Д.А. Афонников[1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[3] Сибирский научно-исследовательский институт растениеводства и селекции – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, р. п. Краснообск, Новосибирская область, Россия
ada@bionet.nsc.ru

Аннотация. При анализе структуры урожая пшеницы неоднократно показано, что его выраженность зависит от продуктивности колоса. Основные характеристики колоса, которые связаны с продуктивностью, это прежде всего масса 1000 зерен, число зерен и колосков в колосе, его размеры, наличие/отсутствие остей и другие. В современных генетических исследованиях для идентификации локусов, контролирующих признаки продуктивности колоса, требуется морфометрия сотен и тысяч колосьев. С другой стороны, современные коллекции генетических ресурсов содержат тысячи образцов, которые также требуют своего детального описания. Все это обусловливает необходимость развития цифровых технологий описания признака колоса пшеницы, которые могут быть достигнуты на основе методов анализа изображений. Эти методы позволяют автоматически получать значения набора признаков, которые могут служить основой формирования цифровых коллекций растений. В настоящей работе предложено цифровое описание колоса пшеницы на основе характеристик, полученных как вручную, так и на основе анализа цифровых изображений, а также характеристик растений, у которых был взят колос. Эти данные положены в основу обновленной версии базы данных SpikeDroidDB (http://spikedroid.biores.cytogen.ru/). Цифровое описание колоса состоит из двух блоков. Блок загружаемых данных содержит описание растения и включает 5 таблиц: коллекция, сортообразец (год выращивания (вегетация), посевной номер, таксономическую информацию и др.), место выращивания, характеристики колоса растений, оцененные экспертом вручную (длину, ширину фронтальной и боковой проекций, тип и цвет колоса и др.). Блок извлекаемых характеристик включает признаки колоса, полученные в результате цифрового фенотипирования, и состоит из шести таблиц: характеристики контура колоса на изображении; характеристики модели четырехугольников, компоненты цвета колоса, доминантные цвета колоса, текстурные характеристики колоса на изображении. Было проведено выделение наиболее наглядных и информативных характеристик колоса, которые позволили сформировать цифровой паспорт колоса, включающий признаки размера, формы и цвета, определенные на основе анализа цифровых изображений. Проведено сравнение признаков, формирующих цифровой паспорт, у двух видов пшеницы, *T. aethiopicum* и *T. carthlicum*. Показано, что признаки цифрового паспорта позволяют наглядно представить модель колоса, а также выявить достоверно различающиеся параметры (цвет колоса и остей, округлость фронтальной проекции колоса). В интерфейс базы данных также добавлена возможность пакетной загрузки данных о характеристиках растения и колоса, а также их изображений.

Ключевые слова: пшеница; колос; морфометрия; цифровое фенотипирование; база данных; коллекция

Для цитирования: Комышев Е.Г., Кручинина Ю.В., Коваль В.С., Потешкина А.А., Петраш Н.В., Пискарев В.В., Гончаров Н.П., Афонников Д.А. Описание морфологических характеристик колосьев пшеницы в базе данных SpikeDroidDB в виде цифрового паспорта. *Вавиловский журнал генетики и селекции*. 2026;30(2):330-338. doi 10.18699/vjgb-26-37

Финансирование. Работа поддержана грантом РНФ № 23-14-00150.

Благодарности. Размещение сайта базы данных и обработка изображений выполнены с использованием ресурсов Центра коллективного пользования «Биоинформатика» при поддержке бюджетного проекта FWNR-2026-0023.

Description of morphological characteristics of wheat spike as a digital certificate in the SpikeDroidDB database

E.G. Komyshev[1, 2], Yu.V. Kruchinina[1, 2], V.S. Koval[1, 2], A.A. Poteshkina[1, 3], N.V. Petrash[1, 3], V.V. Piskarev[1, 3], N.P. Goncharov[1], D.A. Afonnikov[1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[3] Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk region, Russia
ada@bionet.nsc.ru

© Комышев Е.Г., Кручинина Ю.В., Коваль В.С., Потешкина А.А., Петраш Н.В., Пискарев В.В., Гончаров Н.П., Афонников Д.А., 2026
Контент доступен под лицензией Creative Commons Attribution 4.0

Abstract. It has been repeatedly shown that spike productivity is the main component of wheat yield. The main spike parameters related to productivity are size, the number of grains and spikelets per spike, and the presence or absence of awns. In modern genetic research, morphometric analysis of hundreds and thousands of spikes is required to determine the loci that control spike productivity traits. On the other hand, thousands of accessions in modern collections of wheat genetic resources need detailed description. These considerations motivate the development of digital technologies for describing spike traits in wheat, which can be achieved through image analysis methods. These methods allow for automated acquisition of trait values that can serve as the basis for digital plant collections. Here we propose an extended set of spike characteristics obtained both manually and through digital image analysis and present plant characterization. These data form the basis of the updated version of the SpikeDroidDB database (http://spikedroid.biores.cytogen.ru/). The digital description of the spike consists of two blocks. The block of uploaded data includes a description of the plant and contains five tables: collection; variety sample (year of cultivation (vegetation), sowing identifier, taxonomic information, etc.), planting site, and characteristics of the spike determined manually (length, width of frontal and lateral views, type and color of the spike, etc.) The block of extracted features includes spike characteristics obtained by digital phenotyping and contains six tables: characteristics of the spike outline in the image; characteristics of the quadrangle model, values of the color components of the spike, dominant colors of the spike, and texture characteristics of the spike in the image. The most illustrative and significant features of the spike have been identified, allowing for the formation of the spike digital certificate, which includes size, shape, and color features derived from the digital images. The features forming the digital certificate have been compared between two wheat species, *T. aethiopicum* and *T. carthlicum*. It is shown that the features of the digital certificate allow for a clear representation of the spike model and the identification of distinct parameters: colors of the spike and awns and roundness of the frontal view of the spike. The database interface has been supplemented with the ability to upload data on plant and spike characteristics, as well as their images, in the batch mode.

Key words: wheat; spike; morphometry; digital phenotyping; database; collection

For citation: Komyshev E.G., Kruchinina Yu.V., Koval V.S., Poteshkina A.A., Petrash N.V., Piskarev V.V., Goncharov N.P., Afonnikov D.A. Description of morphological characteristics of wheat spike as a digital certificate in the SpikeDroidDB database. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):330-338. doi 10.18699/vjgb-26-37

Введение

При анализе структуры урожая пшеницы неоднократно показано, что его выраженность зависит от продуктивности колоса (Моисеева, 2017; Романов, Пимонов, 2018; Демина, 2022; Шуклина и др., 2022). Наиболее значимы для урожайности такие его характеристики, как число колосков и зерен, размеры и степень компактности. При обработке пшеницы также важны наличие или отсутствие остей, пленчатость, окраска колоса и т. д. Изучение генов, которые контролируют эти признаки, находится в фокусе интереса генетиков и селекционеров (Guo et al., 2018; Garland-Campbell, 2022). В таких исследованиях значима оценка разнообразия признаков колосьев для большого количества образцов, что позволяет проводить поиск ассоциаций между вариациями характеристик колоса и нуклеотидными заменами в геноме пшеницы (Wu et al., 2012; Xu et al., 2024).

Изучение разнообразия пшеницы и ее сородичей также опирается на идентификацию различных признаков колоса, прежде всего таких, которые позволяют установить таксономическую принадлежность растения, провести классификацию биологического материала (Spagnoletti Zeuli, Qualset, 1987; Al Khanjari et al., 2008; Гончаров, 2009) и апробацию семеноводческих посевов (Методика…, 2019). К признакам такого типа относятся форма колоса (Konopatskaia et al., 2016), опушенность колоса (Goncharov et al., 2007) и наличие/отсутствие остей (Смоленская и др., 2022), цвет колоса и остей (Ляпунова, 2017; Смоленская и др., 2022), компактоидность (Vavilova et al., 2017). Эти морфологические признаки используются генетиками и селекционерами при анализе коллекций пшениц и в селекционных программах, а их оценка требует больших трудозатрат, поскольку объемы анализируемого материала могут составлять сотни и тысячи образцов (Пакуль, Шерина, 2009; Пискарев и др., 2018; Зуев и др., 2019).

Для повышения эффективности определения признаков растений в современной генетике и селекции используют технологии фенотипирования на основе анализа цифровых изображений, которые позволяют автоматизировать процесс фенотипирования, снижая трудозатраты и позволяя получать оценки признаков растений для тысяч образцов (Афонников и др., 2016; Meraj et al., 2024), в том числе и с использованием методов машинного обучения (Murphy et al., 2024). Для колосьев пшеницы методы фенотипирования, основанные на анализе цифровых изображений, в последнее время также активно разрабатываются. Они позволяют оценивать прежде всего морфологические характеристики, такие как длина и ширина, плотность колоса (Genaev et al., 2019). Изображения на основе рентгеновской томографии позволяют детально описать трехмерную структуру колоса вплоть до каждого зерна (Ling et al., 2023). Использование методов машинного обучения позволяет автоматически определить число (Генаев и др., 2026) и форму (Niu et al., 2024) колосков в колосе, оценить наличие/отсутствие опушения колоса (колосковой чешуи) (Artemenko et al., 2024).

Цифровое фенотипирование позволяет изучать количество характеристик растений гораздо большее, чем раньше мог исследовать селекционер (Афонников и др., 2016). Некоторые из этих характеристик имеют прямую связь с урожайностью, например, длина колоса или число колосков в колосе (Моисеева, 2017; Романов, Пимонов, 2018; Демина, 2022; Шуклина и др., 2022). Часть характеристик связана с ней косвенно, например, массу зерна можно

оценить по площади его проекции на изображении. Часть полученных цифровых характеристик трудно интерпретировать с точки зрения агробиологии, однако они могут быть эффективно использованы для решения задач классификации колосьев методом машинного обучения (Bi et al., 2010). Выбор цифровых характеристик растений важен при создании цифровых коллекций – баз данных, которые описывают их фенотип (Conejo-Rodríguez et al., 2024). С одной стороны, они должны быть привычны селекционеру, работающему с этой базой данных (Методика…, 2019), с другой, достаточно представительны, чтобы на их основе можно было бы решать и задачи классификации образцов.

Ранее нами была разработана база данных SpikeDroidDB, в которой хранится информация об образцах колоса пшеницы, включая описание признаков, оцененных экспертом вручную (Генаев и др., 2018). В настоящей работе показаны расширение и систематизация набора цифровых признаков для описания колосьев в этой базе данных, таких как параметры модели четырехугольников (Genaev et al., 2019), а также характеристики цвета колоса и остей. Интерфейс базы дополнен возможностью пакетной загрузки информации, включающей характеристики образцов пшеницы, оцененные экспертом, и изображения колосьев. Это позволило расширить количество описаний образцов колосьев в БД до более чем 1700 штук.

Материалы и методы

Растительный материал. База данных была дополнена описаниями образцов колосьев растений из коллекции сортов мягкой пшеницы «Сибирский генофонд» СибНИИРС/ИЦиГ СО РАН (Пискарев и др., 2018), выращенных в 2019–2020 гг. на экспериментальных полях СибНИИРС (п. Краснообск, Новосибирская обл.).

Цифровое фенотипирование. Для получения изображений колосьев использовали протокол «на прищепке», описанный ранее (Генаев и др., 2018; Genaev et al., 2019). Колос размещался на прищепке вертикально на фоне листа синей бумаги. Съемка проводилась цифровой камерой Canon 350D, объектив EF-S 18-55mm f/3.5-5.6. Для оценки масштаба и калибровки цвета был использован цветовой маркер X-Rite Mini ColorChecker Classic, который располагался в области кадра. Колосья снимали в четырех проекциях: фронтальной (наиболее широкой), обратной к ней и двух боковых. Для получения цифровых характеристик колоса изображения колосьев сегментировали методом глубокого машинного обучения (Artemenko et al., 2024) на фон, цветовую палитру, тело колоса и ости. Размеченное изображение подавалось на вход программе WERecognizer (Genaev et al., 2019) для расчета морфометрических и цветовых характеристик колосьев. Морфологические характеристики колоса включали: набор признаков для модели четырехугольников, симметризованной относительно оси колоса (Komyshev et al., 2024) (рис. 1); набор признаков контура тела колоса на изображении: длину, периметр, площадь, округлость, закругленность, целостность, шероховатость, площадь остей. Всего проанализированы 19 морфологических признаков колоса и остей.

Для колоса и его остей рассчитывали ряд цифровых характеристик цвета. За основу взят метод, использованный ранее для анализа зерен (Afonnikov et al., 2022). Для всех пикселей тела колоса оценивали средние значения компонент цвета в четырех цифровых пространствах RGB, Lab, HSV и YCrCb (Комышев и др., 2020), всего 12 характеристик. Дополнительно выявляли три доминантных цвета и соответствующие значения их компонент. Доминантные цвета определяются как центроиды кластеров в пространстве компонент RGB, полученных в результате метода k-средних, $k = 3$ (Afonnikov et al., 2022). Они характеризуют три наибольшие группы пикселей колоса, однородные по цвету и достаточно полно описывают неоднородность окраски биологических объектов на изображении. Как правило, первый кластер (доминантный цвет) имеет значения компонент, близкие к средним на изображении колоса. Аналогично рассчитывали цветовые характеристики для пикселей остей.

Дополнительно рассчитывали текстурные признаки области колоса на изображении. Эти признаки характеризуют неоднородность структуры объекта на изображении, которая может возникать из-за неравномерной окраски или структуры поверхности колосковой чешуи, например опушения. Отличительной чертой простой текстуры является регулярность, повторяемые или частично воспроизводимые элементы на некоторой поверхности или объекте. Рассчитывали два типа структурных признаков, основанных на матрице совпадения уровней серого (GLCM, Grey Level Co-occurrence Matrix, 10 характеристик) и матрице длин серий уровней серого (GLRM, Gray Level Run-length Matrix, 6 характеристик) (Комышев и др., 2020).

Структура БД SpikeDroid. Описанные выше признаки колоса, которые получены в результате анализа цифровых изображений, в совокупности с их характеристиками, которые определяет эксперт, и параметрами растения, хранятся в обновленной версии БД SpikeDroid, логическая структура которой представлена на рис. 2.

Структура обновленной версии БД включает два блока данных. Блок загружаемых данных включает 5 таблиц и описывает коллекцию и растение, колос которого был проанализирован. Таблица описания коллекции (Collection) содержит текстовое поле с названием коллекции, а также поле описания коллекции. Таблица растения (Plant) включает 20 полей. К ним относятся год выращивания (вегетация), посевной номер, номер каталога в коллекции, номер репродукции, аллель *Q* гена (Vavilova et al., 2020), информация о родительских растениях, таксономическая информация и информация о месте происхождения сорта/линии. Таблица с описанием места выращивания растения (Environment) содержит информацию о времени и месте выращивания, включая адрес организации-оригинатора. Таблица с описанием характеристик колоса (Spike), определенных экспертом вручную, включает столбцы: номер колоса у растения, длину, ширину фронтальной и боковой проекций, тип колоса, цвет колоса, количество колосков, цвет и тип остей, характеристики опушения колоса (колосковой чешуи), признаки ломкоколосости и осыпания зерна, тип присоединения остей. Помимо вышеперечис-

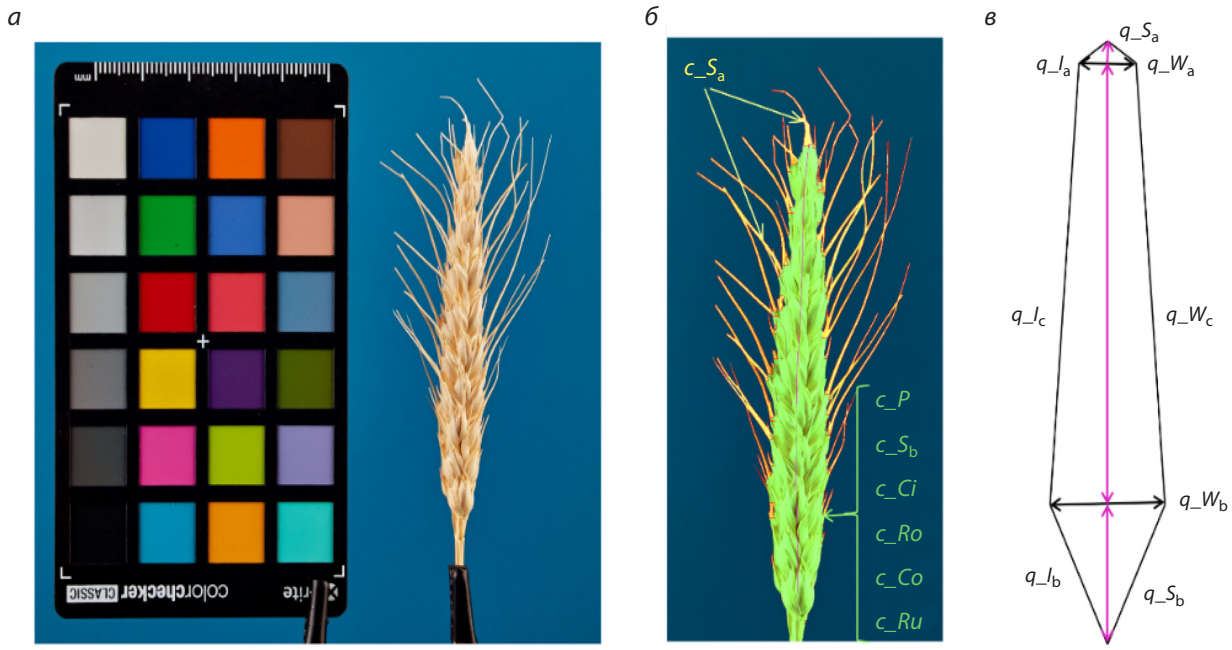


Рис. 1. Морфометрические характеристики колоса: *а* – изображение колоса, расположенного вертикально на прищепке, фронтальная проекция; *б* – результаты сегментации изображения на тело колоса (светло-зеленый цвет), ости (желтый цвет) и фон (синий цвет). Ось колоса показана розовой линией; *в* – модель четырехугольников для колоса; *б*, *в* – обозначения признаков представлены согласно табл. S1.

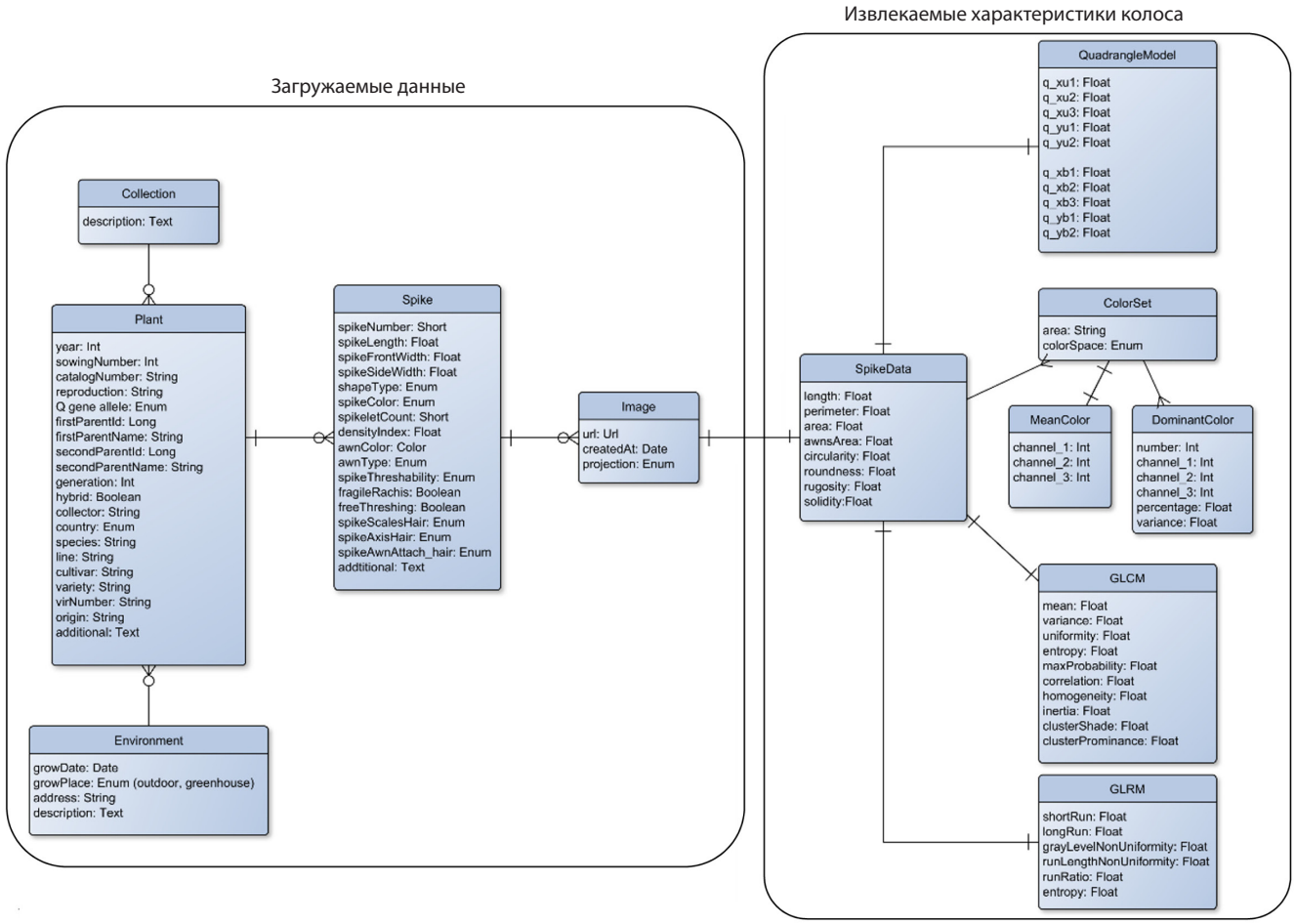


Рис. 2. Логическая структура обновленной версии БД SpikeDriodDB. Справа – блок данных, который вводит пользователь. Слева – блок данных, полученных в результате цифрового фенотипирования изображения.

ленного, этот блок содержит информацию об изображении колоса (Image), для которого приводят номер проекции, дату получения и ссылку на файл изображения.

Блок извлекаемых характеристик включает признаки колоса, полученные в результате цифрового фенотипирования (см. выше). Он включает 7 таблиц: характеристики контура колоса на изображении (SpikeData); характеристики модели четырехугольников (QuadrangleModel), цвет колоса (MeanColor), доминантные цвета колоса (DominantColor), параметры текстуры для матриц GLCM и GLRM для различных цветовых пространств, описанных в таблице ColorSet.

Цифровой паспорт колоса. Среди большого количества характеристик формы и цвета колоса, представленных в БД SpikeDroid (см. выше), можно выделить ряд ключевых, которые могут быть использованы для достоверной оценки сходства/различий формы и размера колоса, а также его цвета. С другой стороны, такие характеристики могут быть использованы для наглядной визуализации сходства или различий колосьев разных образцов из коллекции растений. Ранее мы показали, что цифровые характеристики упрощенной геометрической модели колоса на основе четырехугольников, характеристики контура колоса на изображении могут быть использованы для поиска как межвидовых, так и внутривидовых различий колосьев (Komyshev et al., 2024). Они позволяют классифицировать растения на виды с высокой точностью. Эти признаки были взяты нами за основу при формировании цифрового паспорта колоса: наиболее наглядных признаков формы/размера/цвета колоса, полученных на основе анализа цифровых изображений, достаточных для решения задач классификации. Они были дополнены характеристиками цвета колоса и остей. Список характеристик, формирующих цифровой паспорт, приведен в Приложении[1]. Для одного изображения колоса он включает: 11 параметров модели четырехугольников (Komyshev et al., 2024), 7 признаков контура колоса (Genaev et al., 2019), 3 параметра цвета колоса (средние компонент R, G, B пикселей тела колоса) и 3 параметра цвета остей (средние компонент R, G, B пикселей остей).

Для большинства колосьев в базе данных хранится как минимум 4 проекции колосьев, поэтому для каждого колоса цифровой паспорт включает усредненные данные характеристик для фронтальной/обратной стороны колоса и боковых характеристик. В случае если для колоса на изображении представлена только одна проекция, как правило, это фронтальная проекция по протоколу «на столе» (Генаев и др., 2018), паспорт содержит только характеристики, полученные на основе одного изображения.

Реализация БД. База данных реализована на основе системы управления контентом Drupal (https://www.drupal.org). В качестве хранилища данных используют реляционную базу данных под управлением MySql СУБД, которая развернута на сервере баз данных ЦКП «Биоинформатика» под управлением CentOS Linux.

[1] Табл. S1 и S2 Приложения см. по адресу:
https://vavilovj-icg.ru/download/pict-2026-30/appx19.pdf

Результаты

Пример визуализации колосьев на основе характеристик, представленных в цифровом паспорте, показан на рис. 3 для двух видов пшеницы из коллекции Н.П. Гончарова: *Triticum aethiopicum* Jacubz. (номер каталога ВИР к-19301/2, 6 растений) и *T. cartlicum* Nevski (номер каталога ВИР к-32496, 12 растений).

Рисунок демонстрирует близость двух видов пшениц по длине колоса. Однако форма колоса несколько различается. У *T. aethiopicum* форма во фронтальной проекции выглядит более округлой, а у *T. carthlicum* более вытянутой. Хорошо заметны различия по размеру и количеству остей у двух видов на изображении: у *T. aethiopicum* остей меньше и они более короткие, чем у *T. carthlicum*. Цвет колосьев и остей колосьев также различается: у *T. aethiopicum* колосья и ости светлее, а у *T. carthlicum* темнее.

Мы провели тест на достоверность различий между средними значениями параметров цифрового паспорта колосьев двух видов. Достоверные различия для фронтальной и боковой проекций обнаружены для пяти признаков формы (все они связаны с характеристиками контура, но не модели) и всеми характеристиками цвета – как колосьев, так и остей. Результаты для них приведены в табл. S2. Они хорошо согласуются с визуальной оценкой (см. рис. 3). Для фронтальной проекции значимые различия обнаружены для признаков площади остей (c_S_a), округлости (c_Ci), закругленности (c_Ro), целостности (c_So). Для боковой проекции наблюдаются достоверные различия только по площади остей. Достоверные различия для всех компонент цвета наблюдаются для обеих проекций колоса и остей. Таким образом, признаки, формирующие цифровой паспорт колоса, позволяют наглядно отразить сходства и различия характеристик колосьев для двух видов пшениц, а также определить, какие из них достоверно различаются.

Возможность загрузки данных сразу для множества колосьев облегчает пользователю работу с базой данных. Интерфейс пользователя для импорта данных в этом случае представлен на рис. 4. Он показывает пошаговую реализацию загрузки множества изображений и описания колосьев к ним.

В качестве примера на рис. 5 приведена статистика загруженных данных образцов «Сибирской коллекции пшениц», выращенных в 2019 г. Всего было выращено и проанализировано 696 образцов. Их распределение по странам происхождения, т. е. странам, откуда получен конкретный сортообразец, 10 наиболее представленных в проанализированной коллекции, и разновидностям, наиболее представленным в анализируемом материале, показано на рис. 5.

Обсуждение

Развитие современных методов фенотипирования колоса позволяет оцифровать его характеристики для последующего компьютерного анализа. Ранее мы показали, что цифровые характеристики формы и размера колоса коррелируют с эквивалентными им по биологическому

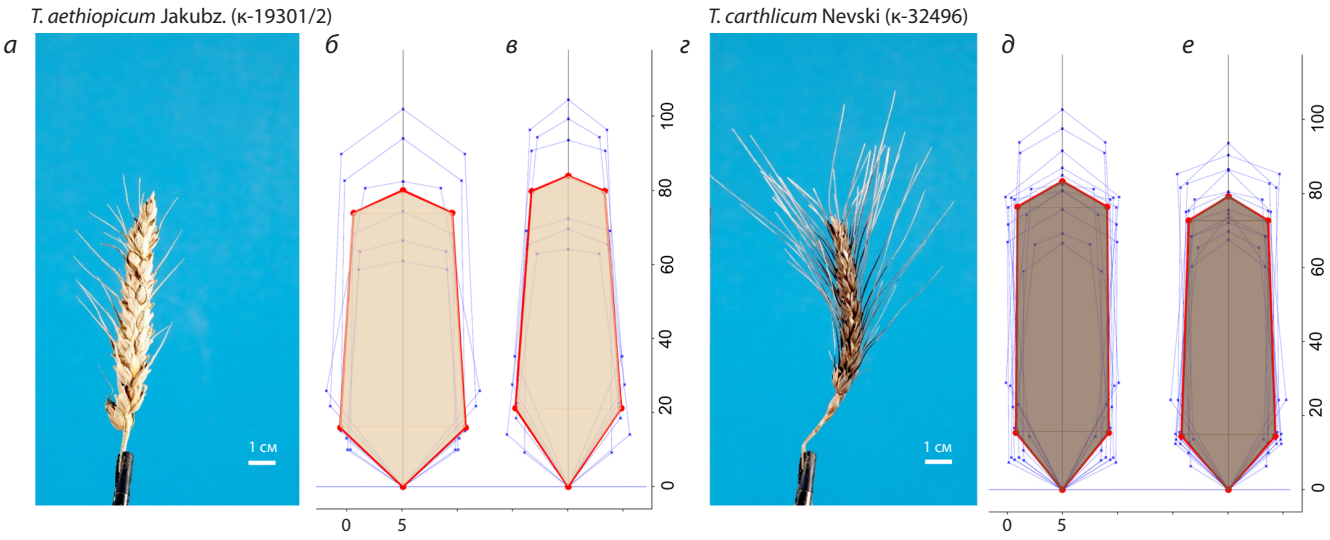


Рис. 3. Представление в графическом виде параметров цифрового паспорта для двух видов пшениц *T. aethiopicum* Jacubz. (*а–в*) и *T. carthlicum* Nevski (*г–е*). Цифровые изображения колосьев показаны на панелях *а*, *г*; модели колосьев для фронтальной проекции показаны на панелях *б*, *д*; модели колосьев для боковых проекций показаны на панелях *в*, *е*. Синими тонкими линиями показаны геометрические модели колосьев для индивидуальных растений, красными жирными линиями – многоугольники моделей для средних значений параметров колоса. Цвет заливки красных многоугольников соответствует усредненному цвету пикселей контура колоса. Масштабы длины (в мм) по осям *X* и *Y* для моделей колосьев отличаются в 5 раз для более наглядного отображения особенности формы колосьев.

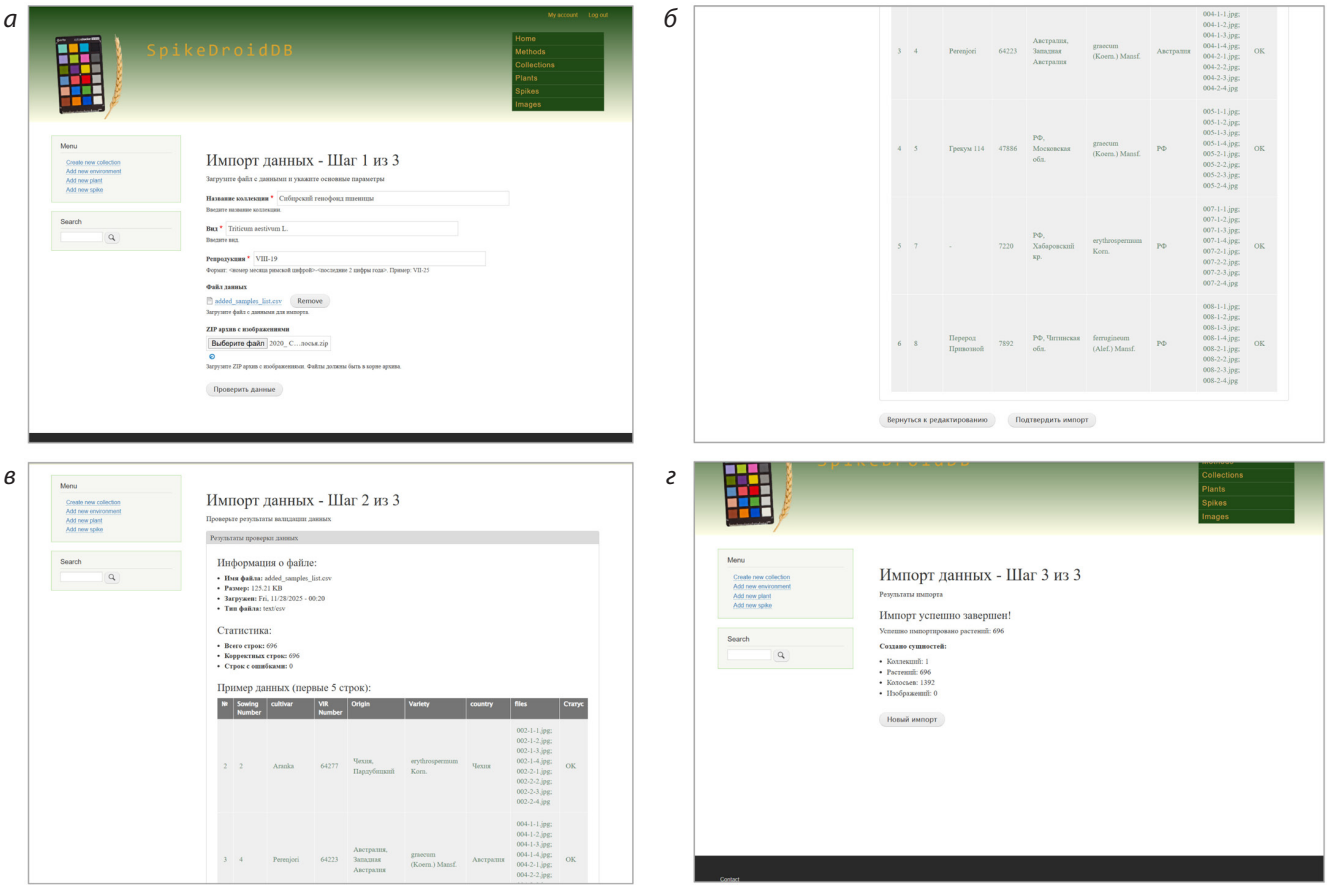


Рис. 4. Визуализация интерфейса пользователя при выполнении основных шагов пакетной загрузки данных о колосьях и их изображениях в БД SpikeDroidDB: *а* – ввод основных характеристик пакета данных: вид пшеницы, номер репродукции, поле для ввода файла с описанием коллекции; *б* – визуализация списка загруженных данных; *в* – визуализация информации о загружаемых данных; *г* – завершение импорта данных.



Рис. 5. Статистика образцов мягкой пшеницы (репродукция 2019 г.) в БД SpikeDroidDB, распределение образцов по странам (*а*) и разновидностям (*б*).

смыслу характеристиками, полученными вручную (Komyshev et al., 2024). В этой же работе было показано, что совместное использование цифровых и классических характеристик позволяет классифицировать растения на виды с высокой точностью. Более того, такое представление позволяет выявлять признаки, которые достоверно различаются у разных образцов одного и того же вида (Komyshev et al., 2024). Все это демонстрирует, что использование как классических, так и цифровых признаков колоса может быть полезным генетикам и селекционерам для описания характеристик колоса. Мы учли это при создании цифрового паспорта колоса, который включает описание характеристик растений, места их произрастания, признаков колосьев, полученных на основе оценок эксперта, а также цифровых характеристик, полученных на основе анализа изображений. Такое представление позволяет более полно описать свойства колоса. К цифровым характеристикам колоса относятся параметры модели четырехугольников, свойства контура колоса на изображении, набор характеристик цвета и текстуры. Таким образом, цифровые параметры фенотипа совместно с наборами классических признаков позволяют более точно описывать растения, классифицировать их (Conejo-Rodríguez et al., 2024).

Цифровое представление фенотипических характеристик растений является важным шагом для развития современных коллекций генетических ресурсов растений (Thormann et al., 2012; Heberling, 2022). С одной стороны, это позволяет быстро оценить огромное разнообразие фенотипических признаков растений в коллекции. С другой, цифровые данные коллекций могут быть более эффективно использованы в селекционных программах (Dipta et al., 2023). Это демонстрирует преимущество цифрового описания свойств коллекций генбанков и рабочих коллекций селекционеров для селекционно-генетических исследований. Дальнейшее развитие методов фенотипирования, выбор наиболее важных цифровых характеристик даст возможность более полного понимания проявления фенотипических признаков, следовательно, может повысить эффективность идентификации генов, которые их контролируют.

Заключение

В результате проведенного исследования разработана комплексная система цифрового описания колоса пшеницы, объединяющая как традиционные методы ручной оценки, так и современные технологии цифрового фенотипирования. Результатом стало обновление структуры БД SpikeDroidDB для комплексного представления свойств колоса. Обновленная структура включает два основных блока данных. Блок загружаемых данных содержит фундаментальную информацию о коллекции и характеристиках конкретного сортообразца, представленную в пяти взаимосвязанных таблицах. Особое внимание уделено агробиологической и таксономической информации, условиям выращивания и базовым характеристикам колоса, определяемым экспертом вручную. Блок извлекаемых характеристик представляет собой расширенный набор параметров, полученных посредством цифрового анализа изображений. Он включает шесть категорий данных: характеристики контура колоса, параметры модели четырехугольников, цветовые характеристики (включая доминантные цвета), а также текстурные особенности. Важным функциональным дополнением стала реализация возможности пакетной загрузки данных, существенно упрощающая процесс внесения информации о растениях, загрузку их изображений, а также последующую статистическую обработку.

Выделены наиболее наглядные и информативные характеристики колоса, которые позволили сформировать цифровой паспорт колоса, включающий признаки размера, формы и цвета, оцененные на основе анализа цифровых изображений. Проведено сравнение признаков, формирующих цифровой паспорт у двух видов пшеницы, *T. aethiopicum* и *T. carthlicum*. Показано, что признаки цифрового паспорта позволяют наглядно представить модель колоса, а также определить достоверно различающиеся параметры: цвет колоса и остей, округлость фронтальной проекции колоса. Таким образом, повышается эффективность описания признаков колоса для современных исследований в области селекции и генетики, которая позволяет осуществлять комплексный анализ морфологических характеристик растений.

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/S0016675816070024

[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]

Генаев М.А., Комышев Е.Г., Фу Хао, 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):132140. doi 10.18699/VJ18.340

[Genaev M.A., Komyshev E.G., Fu Hao, Koval V.S., Goncharov N.P., Afonnikov D.A. SpikeDroidDB: an information system for annotation of morphometric characteristics of wheat spike. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2018; 22(1):132-140. doi 10.18699/VJ18.340 (in Russian)]

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

[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 Selektsii = Vavilov J Genet Breed*. 2026;30(1):27-35. doi 10.18699/vjgb-26-09]

Гончаров Н.П. Определитель разновидностей мягкой и твердой пшениц. Новосибирск: СО РАН, 2009

[Goncharov N.P. Manual book of common and durum wheat varieties. Novosibirsk: SB RAS Publ. House, 2009 (in Russian)]

Демина И.Ф. Адаптивность сортов и линий яровой пшеницы по урожайности и элементам ее структуры в условиях лесостепи Среднего Поволжья. *Международный сельскохозяйственный журнал*. 2022;65(5):530-534. doi 10.55186/25876740_2022_65_5_530

[Demina I.F. Adaptability of spring wheat varieties and lines in terms of productivity and elements of its structure under forest-steppe conditions of the Middle Volga region. *Mezhdunarodnyy Sel'skokhozyajstvennyy Zhurnal = International Agricultural Journal*. 2022; 65(5):530-534. doi 10.55186/ 25876740_2022_ 65_5_530 (in Russian)]

Зуев Е.В., Брыкова А.Н., Кудрявцева Е.Ю. Результаты анализа паспортной базы данных «Местные сорта яровой мягкой пшеницы в коллекции ВИР». *Труды по прикладной ботанике, генетике и селекции.* 2019;180(1):7-11. doi 10.30901/2227-8834-2019-1-7-11

[Zuev E.V., Brykova A.N., Kudryavtseva E.Yu. Results of analyzing the passport database 'Spring bread wheat landraces in the VIR collection'. *Proc Appl Bot Genet Breed*. 2019;180(1):7-11. doi 10.30901/2227-8834-2019-1-7-11 (in Russian)]

Комышев Е.Г., Генаев М.А., Афонников Д.А. Анализ цветовых и текстурных характеристик зерен злаков на цифровых изображениях. *Вавиловский журнал генетики и селекции*. 2020;24(4): 340-347. doi 10.18699/VJ20.626

[Komyshev E.G., Genaev M.A., Afonnikov D.A. Analysis of color and texture characteristics of cereals on digital images. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2020;24(4):340-347. doi 10.18699/VJ20.626]

Ляпунова О.А. Внутривидовая классификация пшеницы твердой: новые ботанические разновидности и формы. *Вавиловский журнал генетики и селекции*. 2017;21(2):152-157. doi 10.18699/VJ17.233

[Lyapunova O.A. Intraspecific classification of durum wheat: new botanical varieties and forms. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2017;21(2):152-157. doi 10.18699/VJ17.233 (in Russian)]

Методика государственного сортоиспытания сельскохозяйственных культур. Вып. 1. Общая часть. М.: 2019. (Минсельхоз России. ФГБУ «Госсорткомиссия»). URL https://gossortrf.ru/upload/2019/08/metodica_1.pdf (Дата обр. 04.12.2025)

[Methodology of State Variety Testing of Agricultural Crops. Issue 1. General Part. Moscow: 2019. (Ministry of Agriculture of the Russian Federation. Federal State Budgetary Institution «Gossortkomissiya»). URL https://gossortrf.ru/upload/2019/08/metodica_1.pdf (Accepted 04.12.2025) (in Russian)]

Моисеева К.В. Продуктивность сортов озимой пшеницы. *Аграрный вестник Урала*. 2017;(9):30-34

[Moiseeva K.V. The productivity of winter wheat vareities. *Agrar Bull Urals*. 2017;(9):30-34 (in Russian)]

Пакуль В.П., Шерина А.Н. Сравнительное изучение образцов яровой пшеницы коллекции ВИР по комплексу агробиологических свойств и признаков. *Достижения науки и техники АПК*. 2009; 9:7-8

[Pakul' V.P., Sherina A.N. Comparative study of spring wheat samples from the VIR collection according to the complex of agrobiological properties and characteristics. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology of AIC*. 2009;9:7-8 (in Russian)]

Пискарев В.В., Зуев Е.В., Брыкова А.Н. Исходный материал для селекции яровой мягкой пшеницы в условиях Новосибирской области. *Вавиловский журнал генетики и селекции.* 2018;22(7): 784-794. doi 10.18699/VJ18.422

[Piskarev V.V., Zuev E.V., Brykova A.N. Sources for the breeding of soft spring wheat in the conditions of Novosibirsk region. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2018; 22(7):784-794. doi 10.18699/VJ18.422 (in Russian)]

Романов Б.В., Пимонов К.И. Феномогеномика продукционных признаков видов пшеницы. Персиановский, 2018

[Romanov B.V., Pimanov K.I. Phenogenomics of production traits of wheat varieties. Persianovsky, 2018 (in Russian)]

Смоленская С.Э., Ефимов В.М., Кручинина Ю.В., Немцев Б.Ф., Чепурнов Г.Ю., Овчинникова Е.С., Белан И.А., Зуев Е.В., Чжоу Чэньси, Пискарев В.В., Гончаров Н.П. Скороспелость и морфотип сортов мягкой пшеницы Западной и Восточной Сибири. *Вавиловский журнал генетики и селекции*. 2022;26(7):662-674. doi 10.18699/VJGB-22-81

[Smolenskaya S.E., Efimov V.M., Kruchinina Yu.V., Nemtsev B.F., Chepurnov G.Yu., Ovchinnikova E.S., Belan I.A., Zuev E.V., Zhou Chenxi, Piskarev V.V., Goncharov N.P. Earliness and morphotypes of common wheat cultivars of Western and Eastern Siberia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2022;26(7):662-674. doi 10.18699/VJGB-22-81]

Щуклина О.А., Завгородний С.В., Аленичева А.Д., Иванова Л.П., Квитко В.Е., Пыльнев В.В., Упелниек В.П. Связь элементов структуры колоса с продуктивностью растений образцов × *Trititrigia cziczinii* Tzvel. *Известия ТСХА*. 2022;(5):57-69. doi 10.26897/0021-342X-2022-5-57-69

[Shchuklina O.A., Zavgorodniy S.V., Alenicheva A.D., Ivanova L.P., Kvitko V.E., Pylnev V.V., Upelniek V.P. Relationship between spike structure elements and plant productivity of × *Trititrigia cziczinii* Tzvel samples. *Izvestiya of Timiryazev Agricultural Academy*. 2022; (5):57-69. doi 10.26897/0021-342X-2022-5-57-69 (in Russian)]

Afonnikov D.A., Komyshev E.G., Efimov V.M., Genaev M.A., Koval V.S., Gierke P.U., Börner A. Relationship between the characteristics of bread wheat grains, storage time and germination. *Plants.* 2022;11(1):35. doi 10.3390/plants11010035

Al Khanjari S., Filatenko A.A., Hammer K., Buerkert A. Morphological spike diversity of Omani wheat. *Genet Resour Crop Evol.* 2008; 55(8):1185-1195. doi 10.1007/s10722-008-9319-9

Artemenko N.V., Genaev M.A., Epifanov R.U., Komyshev E.G., Kruchinina Y.V., Koval V.S., Goncharov N.P., Afonnikov D.A. Image-based classification of wheat spikes by glume pubescence using convolutional neural networks. *Front Plant Sci.* 2024;14:1336192. doi 10.3389/fpls.2023.1336192

Bi K., Jiang P., Lei L., Shi B., Wang C. Non-destructive measurement of wheat spike characteristics based on morphological image processing. *Trans Chin Soc Agric Eng*. 2010;26(12):212-216. doi 10.3969/j.issn.1002-6819.2010.12.036

Conejo-Rodriguez D.F., Gonzalez-Guzman J.J., Ramirez-Gil J.G., Wenzl P., Urban M.O. Digital descriptors sharpen classical descriptors, for improving genebank accession management: A case study on *Arachis* spp. and *Phaseolus* spp. *PLoS One*. 2024;19(5):e0302158. doi 10.1371/journal.pone.0302158

Dipta B., Sood S., Devi R., Bhardwaj V., Mangal V., Thakur A.K., Kumar V., Pandey N.K., Rathore A., Singh A.K. Digitalization of potato breeding program: Improving data collection and management. *Heliyon*. 2023;20;9(1):e12974. doi 10.1016/j.heliyon.2023.e12974

Garland-Campbell K.A. Club Wheat – A review of club wheat history, improvement, and spike characteristics in wheat. *Plant Breed Rev*. 2022;46:421-465. doi 10.1002/9781119874157.ch7

Genaev M.A., Komyshev E.G., Kruchinina Y.V., Goncharov N.P., Afonnikov D.A., Smirnov N.V. Morphometry of the wheat spike by analyzing 2D images. *Agronomy*. 2019;9(7):390. doi 10.3390/agronomy9070390

Goncharov N.P., Bannikova S.V., Kawahara T. Wheat artificial amphiploids involving *Triticum timopheevii* genome: their studies, preservation and reproduction. *Genet Resour Crop Evol*. 2007;54(7):1507-1516. doi 10.1007/s10722-006-9141-1

Guo Z., Zhao Y., Röder M.S., Reif J.C., Ganal M.W., Chen D., Schnurbusch T. Manipulation and prediction of spike morphology traits for the improvement of grain yield in wheat. *Sci Rep*. 2018;8(1):14435. doi 10.1038/s41598-018-31977-3

Heberling J.M. Herbaria as big data sources of plant traits. *Int J Plant Sci.* 2022;183(2):87-118. doi 10.1086/717623

Komyshev E.G., Genaev M.A., Kruchinina Yu.V., Koval V.S., Goncharov N.P., Afonnikov D.A. Evaluation of the spike diversity of seven hexaploid wheat species and an artificial amphidiploid using a quadrangle model obtained from 2D images. *Plants.* 2024;13(19):2736. doi 10.3390/plants13192736

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

Ling Y., Zhao Q., Liu W., Wei K., Bao R., Song W., Nie X. Detection and characterization of spike architecture based on deep learning and X- ray computed tomography in barley. *Plant Methods*. 2023;19(1):115. doi 10.1186/s13007-023-01096-w

Meraj T., Sharif M.I., Raza M., Alabrah A., Kadry S., Gandomi A.H. Computer vision-based plants phenotyping: A comprehensive survey. *iScience*. 2024;27(1):108709. doi 10.1016/j.isci.2023.108709

Murphy K.M., Ludwig E., Gutierrez J., Gehan M.A. Deep learning in image-based plant phenotyping. *Annu Rev Plant Biol*. 2024;75(1):771-795. doi 10.1146/annurev-arplant-070523-042828

Niu Z., Liang N., He Y., Xu C., Sun S., Zhou Z., Qiu Z. A novel method for wheat spike phenotyping based on instance segmentation and classification. *Appl Sci.* 2024;14:6031. doi 10.3390/app14146031

Spagnoletti Zeuli P.S., Qualset C.O. Geographical diversity for quantitative spike characters in a world collection of durum wheat. *Crop Sci*. 1987;27(2):235-241. doi 10.2135/cropsci1987.0011183X002700020022x

Thormann I., Gaisberger H., Mattei F., Snook L., Arnaud E. Digitization and online availability of original collecting mission data to improve data quality and enhance the conservation and use of plant genetic resources. *Genet Resour Crop Evol*. 2012;59(5):635-644. doi 10.1007/s10722-012-9804-z

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

Vavilova V., Konopatskaia I., Blinov A., Kondratenko E.Ya., Kruchinina Y.V., Goncharov N.P. Genetic variability of spelt factor gene in *Triticum* and *Aegilops* species. *BMC Plant Biol*. 2020;20(Suppl. 1):310. doi 10.1186/s12870-020-02536-8

Wu X., Chang X., Jing R. Genetic insight into yield-associated traits of wheat grown in multiple rain-fed environments. *PLoS One*. 2012;7(2):e31249. doi 10.1371/journal.pone.0031249

Xu H., Wang Z., Wang F., Hu X., Ma C., Jiang H., Xie C., Gao Y., Ding G., Zhao C., Qin R., Cui D., Sun H., Cui F., Wu Y. Genome-wide association study and genomic selection of spike-related traits in bread wheat. *Theor Appl Genet.* 2024;137(6):131. doi 10.1007/s00122-024-04640-x

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 30.11.2025. После доработки 15.01.2026. Принята к публикации 15.01.2026.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» до 2011 г. выходил под названием «Информационный вестник ВОГиС»/ "The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Сетевое издание «Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» – реестровая запись СМИ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Белый список (уровень 1), Russian Science Citation Index, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:
русскоязычная версия – на сайте https://vavilovj-icg.ru/
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440

англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

При перепечатке материалов ссылка обязательна.

✉ email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.
Дата выхода в свет 30.03.2026. Формат 60 × 84 1/8. Уч.-изд. л. 24.2.