

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

*Основан в 1997 г.**Периодичность 8 выпусков в год**doi 10.18699/vjgb-25-01***Учредители**

Сибирское отделение Российской академии наук

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

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

Главный редактор

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

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)

И. Гроссе – д-р наук, проф. (Германия)

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)

С.В. Костров – чл.-кор. РАН, д-р хим. наук (Россия)

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

И.Н. Лаврик – д-р биол. наук (Германия)

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

Ж. Ле Гуи – д-р наук (Франция)

И.Н. Лебедев – 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)

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

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

С.Р. Мурсалимов – канд. биол. наук (Россия)

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

И.Б. Рогозин – канд. биол. наук (США)

А.О. Рувинский – д-р биол. наук, проф. (Австралия)

Е.Ю. Рыкова – д-р биол. наук (Россия)

Е.А. Салина – 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 – 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)

Ю.Н. Шавруков – д-р биол. наук (Австралия)

Р.И. Шейко – чл.-кор. НАНБ, д-р с.-х. наук (Беларусь)

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

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

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

*Founded in 1997**Published 8 times annually*

doi 10.18699/vjgb-25-01

Founders

Siberian Branch of the Russian Academy of Sciences

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

The Vavilov Society of Geneticists and Breeders

Editor-in-Chief

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

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

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

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

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

A. Börner, Dr. Sci., Germany

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

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

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

I. Grosse, Professor, Dr. Sci., Germany

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

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

E.K. Khlestkina, 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

S.V. Kostrov, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia

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

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

I.N. Lavrik, Dr. Sci. (Biology), Germany

J. Le Gouis, Dr. Sci., France

I.N. Lebedev, 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

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

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

S.R. Mursalimov, Cand. 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

I.B. Rogozin, Cand. Sci. (Biology), United States

A.O. Ruvinsky, Professor, Dr. Sci. (Biology), Australia

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

E.A. Salina, Professor, Dr. Sci. (Biology), Russia

Y.N. Shavruk, Dr. Sci. (Biology), Australia

R.I. Sheiko, Corr. Member of the NAS of Belarus, Dr. Sci. (Agricul.), Belarus

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

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

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

Молекулярная и клеточная биология

- 7 **ОБЗОР**
Белки теплового шока в фолдинге и реактивации белков. Д. Малькеева, Е.В. Киселева, С.А. Фёдорова
- 15 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Создание и характеристика семи линий индуцированных плюрипотентных стволовых клеток от двух пациентов с болезнью Паркинсона, несущих вариант с.1087G>T гена *LGR4*. В.С. Подвысоцкая, Е.В. Григорьева, А.А. Малахова, Ю.М. Минина, Ю.В. Вяткин, Е.А. Хабарова, Дж.А. Рзаев, С.П. Медведев, Л.В. Коваленко, С.М. Закиян
- 26 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Изучение конкатенации баркодированных трансгенов, линейаризованных Cas9. А.В. Смирнов, А.Н. Кораблев, И.А. Серова, А.М. Юнусова, А.А. Муравьёва, Е.С. Валеев, Н.Р. Баттулин (на англ. языке)

Генетика и селекция растений

- 35 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Взаимосвязи основных признаков качества семян люпина узколистного из коллекции ВИР. Т.В. Шеленга, А.В. Саликова, В.С. Попов, Г.П. Егорова, Л.Л. Малышев, М.А. Вишнякова
- 44 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Анализ сходств и различий образцов *Prunus domestica* L. и *P. insititia* L. по размерам эндокарпия и вариациям формы. Т. Милошевич, Н. Милошевич (на англ. языке)
- 55 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Циркадные колебания содержания растворимых сахаров и экспрессии генов инвертаз *TAI*, *LIN6* и транспортера сахаров *STP1* в листьях растения томата (*Solanum lycopersicum* L.). М.А. Филюшин, А.В. Щенникова, Е.З. Кочиева
- 61 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Разнообразие нута, обусловленное полиморфизмом вставок транспозонов. В.А. Станин, М.А. Дук, А.А. Канапин, А.А. Самсонова, С.Ю. Суркова, М.Г. Самсонова

Симбиотические системы

- 72 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Оценка биоразнообразия грибов арбускулярной микоризы при восстановительной сукцессии на песчаных карьерах. А.А. Крюков, А.П. Юрков, А.О. Горбунова, Т.Р. Кудряшова, А.И. Горенкова, Ю.В. Косульников, Ю.В. Лактионов
- 79 **ОБЗОР**
Симбиоз внутриклеточных бактерий *Wolbachia* с насекомыми: некоторые итоги ста лет изучения. О.Д. Шишкина, Н.Е. Грунтенко

Генетика животных

- 92 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Ось «печень-мозг» при хронической инфекции *Opisthorchis felinus* в сочетании с длительной алкоголизацией мышей. Д.Ф. Августиневич, И.В. Чадаева, А.В. Кизименко, А.В. Ковнер, Д.В. Базовкина, Д.В. Пономарёв, В.И. Евсеенко, В.А. Напримеров, М.Н. Львова (на англ. языке)
- 108 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Наименьшее число хромосом в семействе Pteromalidae (Hymenoptera: Chalcidoidea): кариотип и другие генетические особенности *Pachycrepoideus vindemmiae* (Rondani, 1875). В.Е. Гохман, А.С. Рябинин, Р.А. Быков, Ю.Ю. Илинский (на англ. языке)
- 113 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Генетическая характеристика лошадей местных пород по локусам микросателлитов ДНК. Н.В. Блохина, Л.А. Храброва
- 122 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Ассоциация двух миссенс-мутаций в генах *MSS51* и *KAT6B* с массой тела в разном возрасте у коров ярославской породы. А.В. Игошин, Н.С. Юдин, Д.М. Ларкин

Медицинская генетика

- 128 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Особенности полиморфизма генов толл-лайн рецепторов (*TLR-2*, *TLR-3*, *TLR-4* и *TLR-6*) при открытоугольной глаукоме. *А.В. Шевченко, В.Ф. Прокофьев, В.И. Коненков, В.В. Черных, А.Н. Трунов*
- 135 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Роль полиморфизма гена *SELE* при инфаркте миокарда с подъемом сегмента ST. *Н.П. Бабушкина, А.М. Николаева, А.Д. Долбня, В.Е. Шаврак, В.В. Рябов*
- 144 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Сравнительный анализ гаплотипов, несущих патогенные варианты с.1545T>G, с.2027T>A и с.919-2A>G гена *SLC26A4*, у пациентов с потерей слуха из Республики Тыва (Южная Сибирь). *В.Ю. Данильченко, М.В. Зыцарь, Е.А. Панина, К.Е. Орищенко, О.Л. Посух*

Биомедицина

- 153 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
База знаний MiceDEGdb по дифференциально экспрессирующимся генам мыши как модельного объекта биомедицинских исследований. *О.А. Подкодная, И.В. Чадаева, С.В. Филонов, Н.Л. Подкодный, Д.А. Рассказов, Н.Н. Твердохлеб, К.А. Золотарева, А.Г. Богомолов, Е.Ю. Кондратюк, Д.Ю. Ощепков, М.П. Пономаренко*
- 162 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Компьютерная реконструкция и анализ генных сетей, контролирующих уровень тревожности у лабораторных мышей и человека. *Е.Г. Вергунов, В.А. Савостьянов, А.А. Макарова, Е.И. Николаева, А.Н. Савостьянов*

Molecular and cell biology

- 7 **REVIEW**
Heat shock proteins in protein folding and reactivation. *D. Malkeyeva, E.V. Kiseleva, S.A. Fedorova*
- 15 **ORIGINAL ARTICLE**
Generation and characterisation of seven induced pluripotent stem cell lines from two patients with Parkinson's disease carrying the pathological variant c.1087G>T of the *LGR4* gene. *V.S. Podvysotskaya, E.V. Grigor'eva, A.A. Malakhova, J.M. Minina, Y.V. Vyatkin, E.A. Khabarova, J.A. Rzaev, S.P. Medvedev, L.V. Kovalenko, S.M. Zakian*
- 26 **ORIGINAL ARTICLE**
Studying concatenation of the Cas9-cleaved transgenes using barcodes. *A.V. Smirnov, A.N. Korablev, I.A. Serova, A.M. Yunusova, A.A. Muravyova, E.S. Valeev, N.R. Battulin*

Plant genetics and breeding

- 35 **ORIGINAL ARTICLE**
Interrelations between the main seed quality characteristics of narrowleaf lupine from the VIR collection. *T.V. Shelenga, A.V. Salikova, V.S. Popov, G.P. Egorova, L.L. Malyshev, M.A. Vishnyakova*
- 44 **ORIGINAL ARTICLE**
Analysis of similarities and differences of accessions belonging to *Prunus domestica* L. and *P. insititia* L. using endocarp dimensions and shape variations. *T. Milošević, N. Milošević*
- 55 **ORIGINAL ARTICLE**
Diurnal fluctuations in the content of soluble sugars and the expression of the *TAI* and *LIN6* invertase genes and the *STP1* sugar transporter gene in the leaves of the tomato (*Solanum lycopersicum* L.). *M.A. Filyushin, A.V. Shchennikova, E.Z. Kochieva*
- 61 **ORIGINAL ARTICLE**
Chickpea diversity driven by transposon insertion polymorphism. *V.A. Stanin, M.A. Duk, A.A. Kanapin, A.A. Samsonova, S.Yu. Surkova, M.G. Samsonova*

Symbiotic systems

- 72 **ORIGINAL ARTICLE**
Evaluation of the biodiversity of arbuscular mycorrhizal fungi during regenerative succession in quarries. *A.A. Kryukov, A.P. Yurkov, A.O. Gorbunova, T.R. Kudriashova, A.I. Gorenkova, Y.V. Kosulnikov, Y.V. Laktionov*
- 79 **REVIEW**
Symbiosis of intracellular bacteria *Wolbachia* with insects: a hundred years of study summarized. *O.D. Shishkina, N.E. Gruntenko*

Animal genetics

- 92 **ORIGINAL ARTICLE**
The liver-brain axis under the influence of chronic *Opisthorchis felineus* infection combined with prolonged alcoholization in mice. *D.F. Avgustinovich, I.V. Chadaeva, A.V. Kizimenko, A.V. Kovner, D.V. Bazovkina, D.V. Ponomarev, V.I. Evseenko, V.A. Naprimerov, M.N. Lvova*
- 108 **ORIGINAL ARTICLE**
The lowest chromosome number in the family Pteromalidae (Hymenoptera: Chalcidoidea): the karyotype and other genetic features of *Pachycrepoideus vindemmiae* (Rondani, 1875). *V.E. Gokhman, A.S. Ryabinin, R.A. Bykov, Yu.Yu. Ilinsky*
- 113 **ORIGINAL ARTICLE**
Genetic characteristics of local horse breeds by microsatellite DNA loci. *N.V. Blohina, L.A. Khrabrova*
- 122 **ORIGINAL ARTICLE**
Association of two missense mutations in the *MSS51* and *KAT6B* genes with body weight at different ages in cows of the Yaroslavl breed. *A.V. Igoshin, N.S. Yudin, D.M. Larkin*

Medical genetics


- 128 **ORIGINAL ARTICLE**
Features of toll-like receptor genes (*TLR-2*, *TLR-3*, *TLR-4* and *TLR-6*) polymorphism in open-angle glaucoma patients. A.V. Shevchenko, V.F. Prokofiev, V.I. Konenkov, V.V. Chernykh, A.N. Trunov
- 135 **ORIGINAL ARTICLE**
The role of *SELE* gene polymorphism in ST-elevation myocardial infarction. N.P. Babushkina, A.M. Nikolaeva, A.D. Dolbnya, V.E. Shavrak, V.V. Ryabov
- 144 **ORIGINAL ARTICLE**
Comparative analysis of haplotypes carrying pathogenic variants c.1545T>G, c.2027T>A and c.919-2A>G of the *SLC26A4* gene in patients with hearing loss from the Tyva Republic (Southern Siberia). V.Yu. Danilchenko, M.V. Zytsar, E.A. Panina, K.E. Orishchenko, O.L. Posukh

Biomedicine

- 153 **ORIGINAL ARTICLE**
MiceDEGdb: a knowledge base on differentially expressed mouse genes as a model object in biomedical research. O.A. Podkolodnaya, I.V. Chadaeva, S.V. Filonov, N.L. Podkolodnyy, D.A. Rasskazov, N.N. Tverdokhleby, K.A. Zolotareva, A.G. Bogomolov, E.Yu. Kondratyuk, D.Yu. Oshchepkov, M.P. Ponomarenko
- 162 **ORIGINAL ARTICLE**
Computer reconstruction of gene networks controlling anxiety levels in humans and laboratory mice. E.G. Vergunov, V.A. Savostyanov, A.A. Makarova, E.I. Nikolaeva, A.N. Savostyanov

doi 10.18699/vjgb-25-02

Heat shock proteins in protein folding and reactivation

D. Malkeyeva ¹, E.V. Kiseleva ¹, S.A. Fedorova ^{1, 2}¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia² Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia fsveta@bionet.nsc.ru


Abstract. Throughout their lives, cells synthesise new and dispose of the old, denatured proteins and insoluble protein aggregates. An important role in maintaining proteostasis is played by chaperones, which fold various proteins and promote degradation of denatured or misfolded proteins via proteasomes or autophagy. Despite protein folding being an accurate process, as organisms age and experience stress, errors accumulate, which leads to the formation of protein aggregates that can result in pathological changes. In addition, stress factors such as elevated temperature and altered pH can promote protein denaturation that can result in the proteins not only losing their native functions, but also gaining novel cytotoxic properties. With the increase of human average lifespan, more and more cases of proteinopathies – diseases caused by disruptions in proteostasis, e.g. Alzheimer's disease, Huntington's disease etc. – emerge. Therefore, identification of mechanisms preventing the formation of cytotoxic protein aggregates and promoting their clearance is of high importance. Heat shock proteins (HSPs) are the molecular chaperones involved in folding nascent proteins and refolding the denatured ones, leading to their reactivation. Heat shock proteins vary in structure and functions and are found in all prokaryotes and eukaryotes discovered to date. HSPs are constantly synthesised in cells under normal conditions, and a multitude of them are dramatically up-regulated during stress, which includes heat shock (which earned them their name) and metabolic stress caused by the increased numbers of misfolded proteins. In this review, we describe mechanisms of action and functions of members of five heat shock protein families.

Key words: heat shock proteins; molecular chaperones; protein folding; protein quality control; HSP.

For citation: Malkeyeva D., Kiseleva E.V., Fedorova S.A. Heat shock proteins in protein folding and reactivation. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2025;29(1):7-14. doi 10.18699/vjgb-25-02

Funding. E.V. Kiseleva and S.A. Fedorova were supported by the Ministry of Science and Higher Education of the Russian Federation, project No. FWNR-2022-0015.

Белки теплового шока в фолдинге и реактивации белков

Д. Малькеева ¹, Е.В. Киселева ¹, С.А. Фёдорова ^{1, 2}¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия² Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия fsveta@bionet.nsc.ru

Аннотация. В процессе жизнедеятельности в каждой клетке происходят синтез новых белков и удаление старых, денатурированных белков и нерастворимых белковых агрегатов. В поддержании протеостаза значительную роль играют шапероны, которые участвуют в придании правильной конформации (фолдинге) многих белков и способствуют деградации денатурированных или неправильно свернутых белков посредством протеаз или аутофагии. Несмотря на то что фолдинг белков – довольно точный процесс, с возрастом и под воздействием стресса накапливаются ошибки, приводящие к образованию нерастворимых белковых агрегатов, которые могут вызывать различные патологии. Воздействие стрессовых факторов, таких как повышенная температура и изменение кислотности среды, также может способствовать изменению нативной конформации белков, в результате чего они могут не только терять выполняемые в норме функции, но и приобретать новые цитотоксические свойства. В связи с увеличением средней продолжительности жизни человека в мире отмечается рост протеинопатий – заболеваний, связанных с нарушением протеостаза, к которым относятся, например, болезни Альцгеймера, Паркинсона, Хантингтона; поэтому выявление механизмов, препятствующих накоплению и способствующих удалению цитотоксичных агрегатов, стало актуальной задачей. Белки теплового шока (heat shock proteins, HSP) – молекулярные шапероны, принимающие участие как в придании правильной конформации вновь синтезированному белку, так и в рефолдинге денатурированных белков с их последующей реактивацией. HSP разнообразны по структуре и выполняемым функциям и встречаются у всех изученных про- и эукариотических

организмов. HSP синтезируются в клетке постоянно. Выработка множества из них многократно усиливается при стрессах, включая тепловой (за что они и получили свое название) и метаболический стресс, возникающий из-за повышения количества неправильно свернутых белков. В настоящем обзоре описаны механизмы действия и функции представителей пяти семейств HSP в фолдинге и реактивации белков.

Ключевые слова: белки теплового шока; молекулярные шапероны; фолдинг белков; контроль качества белков; HSP.

Introduction

Heat shock proteins (molecular chaperones, HSPs) are a class of conserved proteins, the main function of which is protein quality control (van Leeuwen, Kampinga, 2018). Heat shock proteins are present in all prokaryotes and eukaryotes discovered to date (Lindquist, 1986). The majority of HSPs are synthesised under normal conditions and are up-regulated during stress, e. g. hyper- or hypothermia, hypoxia, oxidative stress, and infection (Sørensen et al., 2003; Kampinga et al., 2009; Sarkar et al., 2011). Under normal conditions, HSPs serve as “molecular chaperones” by folding nascent proteins (Ellis, 1987; Feder, Hofmann, 1999). During stress, the redox balance and hydration of the cell can be disrupted, which leads to protein misfolding (Jolly, Morimoto, 2000). Misfolded proteins, in turn, can gain deleterious functions and tend to form insoluble aggregates (Jolly, Morimoto, 2000). The induction of HSP synthesis allows the cells to maintain homeostasis due to their ability to refold the damaged proteins, prevent aggregation, and to resolubilise the already formed protein aggregates (Jolly, Morimoto, 2000). Mutations in certain HSPs result in disorders, such as myopathies, neuropathies, and lens or retinal diseases (Macario et al., 2005; Kakkar et al., 2014).

Based on their structure and functions, HSPs are classified into five major families, with their names reflecting the members' molecular mass in kilodaltons (kDa): Hsp100 (or Hsp110), Hsp90, Hsp70, Hsp60, and small HSPs (sHSPs, up to 43 kDa) (Sarkar et al., 2011; Bar-Lavan et al., 2016). HSPs cooperate with co-chaperones Hsp40, Hsp10, and NEF, which are sometimes placed into separate HSP families. For human HSP families, a standard nomenclature was suggested in 2009 by Professor H.H. Kampinga et al.: HSPH (Hsp110), HSPC (Hsp90), HSPA (Hsp70), DNAJ (Hsp40), HSPB (sHSPs), and chaperonins HSPD/E (HSP60/HSP10) and CCT (TRiC) (Kampinga et al., 2009).

Hsp100

This family of HSPs (Clp in bacteria, HSPH in humans) includes 100–110 kDa chaperones capable of proteolysis and protein aggregate degradation (Sarkar et al., 2011; Mogk et al., 2015). These chaperones are abundant in prokaryotes and are present in unicellular eukaryotes (e. g. Hsp104 and Hsp78 in budding yeast, *Saccharomyces cerevisiae*); in multicellular organisms, Hsp100 can only be found in mitochondria (Sarkar et al., 2011). Hsp100 proteins belong to the AAA+ (ATPases associated with diverse cellular activities) superfamily and share the AAA domain defined by a region of ~230 amino acid residues containing Walker A, Walker B, sensor-1, and sensor-2 motifs that are necessary for nucleotide binding and hydrolysis and for Hsp100 oligomerisation resulting in ring-like structures (Mogk et al., 2015; Mokry et al., 2015). De-

pending on the number of the AAA domains, Hsp100 proteins can be divided into two classes: Class I with two nucleotide binding domains, and Class II with one nucleotide binding domain (Hodson et al., 2012; Mokry et al., 2015). Class I includes ClpA, ClpB (Hsp104), and ClpC; Class II includes ClpX and HslU (Hodson et al., 2012).

All of the Hsp100 family members have an N-terminal domain contributing to the binding to protein aggregates (Mokry et al., 2015). Certain Hsp100 proteins have an M domain located within the first nucleotide binding domain. The flexible M domains are located on the outer surface of the Hsp100 ring and play an essential role in substrate interaction and protein aggregate reactivation (Mokry et al., 2015). In the presence of ATP, Hsp100 form ring-like homohexamers with a central pore of ~15 Å through which unfolded substrates trapped in protein aggregates are pulled (Hodson et al., 2012; Duran et al., 2017). Inside, the pore is lined with loops containing tyrosine residues that bind the aggregated peptides and translocate them through the channel by performing a rowing motion fuelled by ATP hydrolysis (Saibil, 2013). The substrate is pulled into the Hsp100 central pore as a loop, and not by its terminus, because internal segments of aggregated proteins are preferentially targeted (Avellaneda et al., 2020). Nevertheless, the insertion of the substrate into the channel with its free termini is also possible (Avellaneda et al., 2020).

Some of the Hsp100 family members are associated with proteases, and translocate the unfolded aggregate components to them for degradation (Hodson et al., 2012). For example, the bacterial ClpA, ClpC, and ClpX are associated with ClpP protease, and the HslU (ClpY) chaperone is associated with the HslV (ClpQ) protease (Hodson et al., 2012). The bacterial ClpB, the yeast Hsp78 and Hsp104, and the plant Hsp101 are not connected to any proteases and thus reactivate the aggregated proteins (Hodson et al., 2012).

Hsp100 can perform disaggregation independently; however, their activity is greatly enhanced by the presence of the Hsp70/Hsp40/NEF chaperone system (Mokry et al., 2015). For example, Hsp104 and ClpB are almost incapable of recognising aggregated proteins in the absence of Hsp70 (Mogk et al., 2015). Hsp70 chaperones bind to aggregated peptides and then directly interact with the Hsp100 M domains thereby presenting the substrate to the disaggregase. Next, the substrate is translocated through the Hsp100 central pore and unfolded (Mogk et al., 2015).

Hsp90

Hsp90 proteins (HtpG in bacteria, HSPC in humans) are some of the most abundant chaperones, making up to 1–2 % of total proteins in eukaryotic cells (Sarkar et al., 2011; Li, Buchner, 2013). In addition to folding denatured proteins, Hsp90 facilitates maturation of various *de novo* synthesised peptides

(Bar-Lavan et al., 2016). Bacteria have an Hsp90 protein (HtpG in *Escherichia coli*); no Hsp90 genes were found in Archaea (Li, Buchner, 2013). In yeast, two members of Hsp90 are present, namely, Hsc82 and Hsp82, that are localised to cytosol (Li, Buchner, 2013). Plants possess ch-Hsp90 found in chloroplasts (Li, Buchner, 2013). In the fruit fly *Drosophila melanogaster*, one Hsp90 member, i. e. Hsp83, has been discovered; it is the only HSP-coding gene in *D. melanogaster* that has an intron (Sarkar et al., 2011). In mammals, there are four members of the family, of which two isoforms of Hsp90 (Hsp90 α and Hsp90 β) are localised to the cytosol, Grp94 is located in the endoplasmic reticulum, and Trap-1 is present in the mitochondria (Sarkar et al., 2011; Li, Buchner, 2013).

Hsp90 proteins have three domains: the highly conserved N-terminal domain with the ATP binding site and a charged loop segment, an M domain necessary for substrate binding and regulation of ATP hydrolysis, and a C-terminal domain allowing for the dimerisation of Hsp90 proteins and interaction with several co-chaperones (Bar-Lavan et al., 2016). When it does not bind ATP, the Hsp90 homodimer adopts a V-shaped form, termed “open conformation” (Li, Buchner, 2013). ATP binding leads to a change in the N and M domains orientation and the transition of the homodimer to a “closed conformation”, with the N domains dimerised (Li, Buchner, 2013). After ATP hydrolysis, the N domains dissociate, and Hsp90 returns into the “open conformation” (Li, Buchner, 2013). The transition between the conformations is determined by Hsp90’s interaction with the client proteins and with its various co-chaperones (Li, Buchner, 2013; Bar-Lavan et al., 2016). Interaction with co-chaperones and substrates occurs when Hsp90 is in the “open conformation”; during the transition to the “closed conformation”, the substrate is folded; ADP, phosphate, the substrate, and co-chaperones are then released and Hsp90 returns into the “open conformation” (Li, Buchner, 2013).

Co-chaperone regulation is a conserved trait of the eukaryotic Hsp90 system. More than 20 Hsp90 co-chaperones are known to date (Li, Buchner, 2013). They regulate the Hsp90 function by inhibiting or activating the chaperone’s ATPase activity and recruiting client proteins. Different co-chaperones interact with each other to facilitate the Hsp90 client maturation, and the composition of the co-chaperone complexes depends on the client protein type (Li, Buchner, 2013).

Hsp90 are more specialised than other HSPs. Together with their co-chaperones, Hsp90 play an important role in folding of at least 200 different peptides under normal conditions and in refolding of denatured proteins following stress (Sarkar et al., 2011; Saibil, 2013). Among the Hsp90 clients are signalling proteins participating in cell cycle regulation, kinases, steroid hormone receptors, and the tumour suppressor p53 (Saibil, 2013).

Hsp70

The members of the Hsp70 family (DnaK in prokaryotes, HSPA in humans) have the molecular mass of 70 kDa and are the most conserved of the HSPs. Thus, Hsp70 proteins of all characterized organisms share ~50 % amino acid identity (Sarkar et al., 2011; Bar-Lavan et al., 2016). A distinctive feature of this HSP family is the occurrence of multiple copies

of *Hsp70* genes in the majority of species (Sarkar et al., 2011). For example, the yeast *S. cerevisiae* has 14 *Hsp70* copies (Sarkar et al., 2011); the fruit fly *D. melanogaster* has 6 almost identical stress-inducible *Hsp70* genes, stress-inducible gene *Hsp68*, and several constantly expressed *Hsc70* (Heat shock cognate 70) genes (Tower, 2011; Xiao et al., 2019). In humans, 17 *Hsp70* genes and 30 pseudogenes have been discovered, some of which have up to 90 % sequence similarity (Brocchieri et al., 2008; Radons, 2016).

Hsp70 proteins are found in the cytosol of Archaea and Eubacteria, and in various compartments of eukaryotic cells (i. e. the nucleus, cytoplasm, mitochondria, chloroplasts, and endoplasmic reticulum) (Sarkar et al., 2011; Rosenzweig et al., 2019). The functions of Hsp70 include protein folding, disaggregation of protein aggregates, maintaining aggregation-prone proteins in an unfolded state, and participation in the translocation of proteins across the organelle membranes (Saibil, 2013; Bar-Lavan et al., 2016).

Hsp70 chaperones consist of a conserved N-terminal nucleotide-binding domain of ~44 kDa and a less conserved substrate-binding domain of ~30 kDa, connected by a short conserved hydrophobic linker (Sarkar et al., 2011; Bar-Lavan et al., 2016; Larburu et al., 2020). A disordered C-terminal tail of Hsp70 chaperones has variable length; in eukaryotic nuclear and cytosolic Hsp70 proteins, this region often contains the negatively charged motif Glu-Glu-Val-Asp, which interacts with specific cofactors, including Hsp40 co-chaperones (Rosenzweig et al., 2019). The nucleotide-binding domain consists of four subdomains that are arranged into two lobes separated by a deep crevice, in which an ATP-binding catalytic centre is located (Rosenzweig et al., 2019). The substrate-binding domain contains two subdomains, α and β (Larburu et al., 2020). Subdomain β consists of β -sandwich folds and contains a hydrophobic substrate-binding pocket; subdomain α consists of α -helices and acts as a “lid” that closes over the substrate-binding pocket (Larburu et al., 2020).

The client protein binding and release is mediated by ATP binding and hydrolysis; the speed of these processes is regulated by co-chaperones Hsp40 (DnaJ in prokaryotes) and NEF (nucleotide exchange factor; GrpE in bacteria) (Saibil, 2013; Larburu et al., 2020). The reaction cycle of Hsp70 and its co-chaperones is shown in Figure 1. Binding of ATP to the catalytic centre of Hsp70’s nucleotide-binding domain causes the rotation of the domain’s subunits resulting in the inter-domain linker and the subunits of the substrate-binding domain attaching to the nucleotide-binding domain, opening the substrate-binding pocket (Rosenzweig et al., 2019). The substrate-binding domain of Hsp70 interacts with a short motif of the client protein, consisting of five hydrophobic amino acid residues flanked by charged residues (Larburu et al., 2020). This motif is present in virtually all proteins, which provides a great flexibility in substrate selection (Larburu et al., 2020).

Binding of the client protein induces ATP hydrolysis, which leads to the detachment of the linker and the substrate-binding domain subunits from the nucleotide-binding domain, and to the closure of the substrate-binding pocket with the client protein trapped in it (Rosenzweig et al., 2019; Larburu et al., 2020). Typically, ATP hydrolysis in the Hsp70 catalytic centre

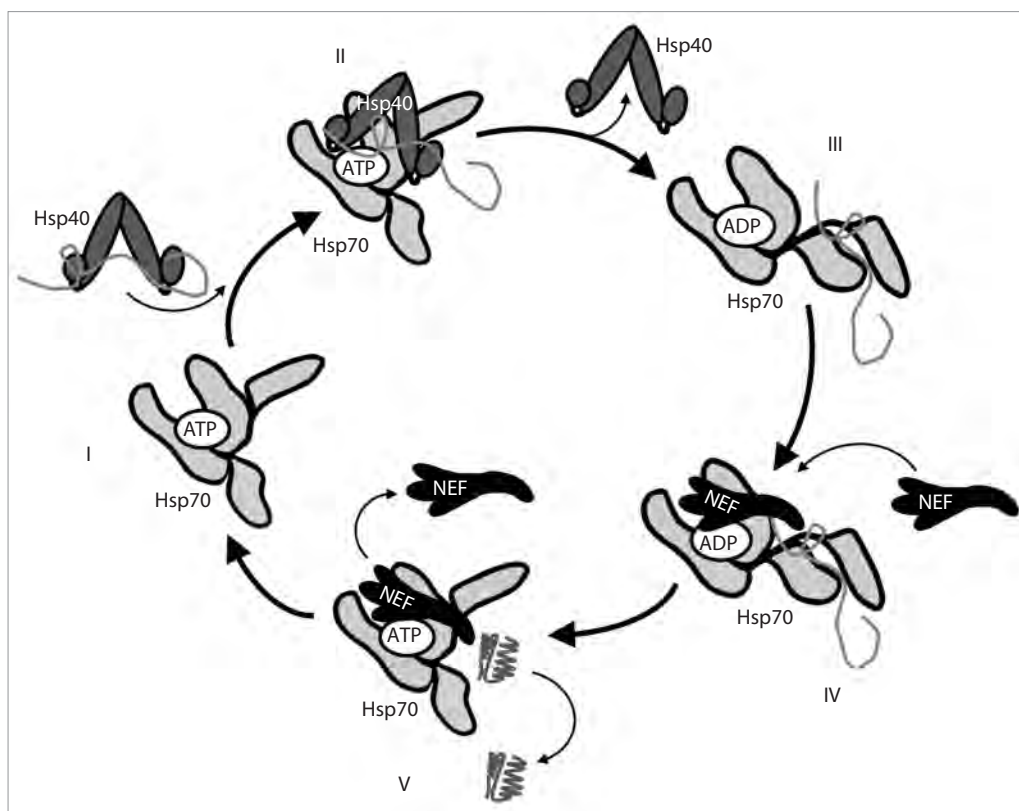


Fig. 1. The chaperoning cycle of Hsp70.

I – in the presence of ATP in the nucleotide-binding domain of Hsp70, the substrate-binding pocket of the chaperone is open and can bind misfolded proteins. II – co-chaperone Hsp40 delivers a client protein and stimulates ATP hydrolysis in the nucleotide-binding domain of Hsp70. III – ATP hydrolysis leads to the conformation change of the chaperone, resulting in the detachment of the substrate-binding domain subunits from the nucleotide-binding domain and their closure around the client protein. IV – co-chaperone NEF exchanges ADP for ATP in Hsp70's nucleotide-binding domain. V – the substrate-binding pocket of Hsp70 opens, releasing the correctly folded client protein.

occurs at the rate of one molecule per 20–30 min; however, the involvement of Hsp40 co-chaperone that delivers the substrate to Hsp70 accelerates the reaction more than 1,000-fold (Bar-Lavan et al., 2016; Larburu et al., 2020). The ADP-to-ATP exchange is stimulated by NEF (Bar-Lavan et al., 2016). This process leads to the opening of the substrate-binding pocket of the chaperone and the release of the client protein (Fig. 1) (Larburu et al., 2020).

There are different hypotheses on the mechanism of protein folding by Hsp70 chaperones. Some models suggest that the binding of Hsp70 to hydrophobic regions of misfolded proteins protects them from aggregation, and upon release, the substrate undergoes folding independently (Bar-Lavan et al., 2016). Other models suggest that clamping of misfolded proteins in the substrate-binding pocket of one or several Hsp70 molecules promotes their unfolding (Bar-Lavan et al., 2016). Given that the Hsp70 chaperones not only prevent aggregation and promote folding, but also reactivate aggregated proteins, models defining unfolding as the main function of Hsp70 may be preferable (Bar-Lavan et al., 2016).

Hsp60

The Hsp60 proteins (GroEL in bacteria, HSPD in humans), also known as chaperonins (Hemmingsen, 1992), are essential

for the majority of organisms not only during stress, but also under normal conditions (Sarkar et al., 2011; Fan et al., 2020). Unlike Hsp70 and Hsp100 chaperones that mainly renature misfolded proteins and resolubilise protein aggregates, Hsp60 chaperonins are crucial for *de novo* protein folding (Saibil, 2013). Approximately 30 % of all *E. coli de novo* synthesised proteins adopt correct conformation through GroEL (Koumoto et al., 2001).

The DNA sequence of Hsp60 chaperonins is highly conserved, which makes it useful for phylogenetic analysis and identification of living organisms (Sarkar et al., 2011). Based on the structure and DNA sequence similarity, the chaperonins are divided into two groups (Saibil, 2013; Bar-Lavan et al., 2016). Group I includes the bacterial GroEL and its co-chaperonin GroES, the mitochondrial Hsp60 and its co-chaperonin Hsp10, and Cpn60 found in chloroplasts, alongside its co-chaperonin Cpn20 (Sarkar et al., 2011; Saibil, 2013; Zhang et al., 2016). Group II includes the archaeal thermosome and CCT (chaperonin-containing TCP1, also known as TriC) found in the cytoplasm of eukaryotes (Saibil, 2013). Hsp60 chaperonins form symmetrical structures comprising two back-to-back rings of 7 (Group I) or 8–9 (Group II) 60 kDa subunits each (Lopez et al., 2015). The Hsp60 subunits of both groups have three domains, namely, the apical,

equatorial, and intermediate. The equatorial domain contains ATP-binding sites and establishes the interactions between the Hsp60 rings; the apical domain binds the substrate and, in Group I chaperonins, also Hsp10 co-chaperonins (Sarkar et al., 2011; Saibil, 2013; Bar-Lavan et al., 2016).

Hsp10 co-chaperonins are homoheptameric structures composed of subunits of ~10 kDa, which bind to the rings of Group I chaperonins, closing their cavities like a lid (Saibil, 2013). In Group II chaperonins, the function of Hsp10 co-chaperonins is carried out by an extra region of the apical domain (Saibil, 2013). The intermediate domain of Hsp60 monomers connects the apical and equatorial domains and undergoes a conformational change upon ATP binding, which facilitates the switching between the “open” state, with the inner surface of the cavity formed by the Hsp60 ring being hydrophobic, and the “closed” state, with the inner surface being hydrophilic (Sarkar et al., 2011; Saibil, 2013; Bar-Lavan et al., 2016).

The reaction cycle of Group I chaperonins GroEL/GroES is described in Figure 2. The folding cycle performed by Group II chaperonins follows the same pattern, except for the cavity closure being performed not by Hsp10 but by the extra regions of the apical domains as a result of ATP hydrolysis, and the rings of Group II chaperonins switching between the “open” and “closed” states synchronously, rather than sequentially (Kumar et al., 2015; Lopez et al., 2015).

The substrates of Hsp60 include peptides with a molecular mass between 35 and 60 kDa, with the maximum substrate size determined by the volume of the Hsp60 ring cavity, which in GroEL is ~175,000 Å³ (Bar-Lavan et al., 2016). The chaperonins perform the folding of such essential proteins as actin and tubulin in eukaryotes, and the RbcL subunit of the RuBisCO enzyme, which participates in the Calvin cycle (Sarkar et al., 2011; Hayer-Hartl, 2017).

Small HSPs

Small HSPs (HSPB in humans) have molecular masses between 12 and 43 kDa (Sarkar et al., 2011). In eukaryotes, they are localised to the cytoplasm, nucleus, mitochondria, chloroplasts, and peroxisomes. Prokaryotes and unicellular eukaryotes typically have one or two cytosolic small HSPs, although some bacteria may have several (for example, bacteria of the genus *Bradyrhizobium* can have up to eight small HSP genes) (Mogk et al., 2019). In multicellular eukaryotes, the number of genes encoding small HSPs ranges from 10 in humans to 50 in higher plants (Mogk et al., 2019).

Unlike other HSP families, small HSPs do not have the ability to refold denatured and aggregated proteins, and they do not hydrolyse ATP when performing their functions, the main of which is to prevent the aggregation of misfolded proteins (Bar-Lavan et al., 2016; Mogk et al., 2019). In addition to the prevention of protein aggregation, small HSPs are involved in several key physiological processes, such as cellular differentiation and apoptosis (Fu, 2015).

A characteristic feature of small HSP family members is the presence of a conserved α -crystallin domain, named after the eye lens protein of vertebrates, α -crystallin (Sarkar et al., 2011; Bar-Lavan et al., 2016; Paul et al., 2016; Mogk et al., 2019). This domain consists of 90–100 amino acid residues

forming a β -sandwich consisting of 7–8 antiparallel β -strands (Mogk et al., 2019). The α -crystallin domain of small HSPs is surrounded by less conserved N- and C-terminal domains (Mogk et al., 2019). The N-terminal domains are especially diverse in amino acid composition and size, consisting of 24–247 amino acid residues (56 on average) (Mogk et al., 2019). The C-terminal domains are up to 20 amino acid residues long (on average, 10) and in 90 % of small HSPs contain the conserved motif Ile-X-Ile/Val (IXI/V) that plays a key role in the oligomerisation of small HSPs (Saji et al., 2008; Mogk et al., 2019).

The oligomers of small HSPs are hollow spheres consisting of 12–32 protomers (Saji et al., 2008; Mogk et al., 2019). Each protomer is a dimer of small HSPs, and the oligomers can be composed of a single type or several types of small HSPs (Mogk et al., 2019). The dimerisation of small HSPs occurs through the interaction of their α -crystallin domains, and the oligomerisation is established by the N- and C-terminal domains (Mogk et al., 2019). Small HSP oligomers are dynamic and constantly exchange dimers (Żwirowski et al., 2017). During stress (e. g. temperature fluctuations, ion balance disruption, pH changes), the equilibrium shifts towards the formation of smaller oligomers and dissociation into dimers, which bind to the hydrophobic regions of misfolded proteins (Żwirowski et al., 2017; Mogk et al., 2019).

Small HSPs have low specificity and interact with a wide range of different substrate proteins (Mogk et al., 2019). When binding to misfolded proteins, small HSPs form complexes consisting of two layers – a stable core containing small HSPs and immobile substrate proteins, and a dynamic shell of small HSPs, which alternately bind to and dissociate from the complex (Fig. 3) (Żwirowski et al., 2017). The size of these complexes is smaller than that of the misfolded protein aggregates, but their molecular mass usually exceeds 1 MDa (Żwirowski et al., 2017).

The size of small HSP-substrate complexes depends on the small HSP-to-substrate ratio; the higher the proportion of small HSPs, the smaller the complex size. At sufficiently high levels of small HSPs, the complexes become soluble (Mogk et al., 2019). At low concentrations of small HSPs, the substrates form dense, insoluble complexes, but unlike insoluble aggregates formed solely by misfolded proteins, the substrates are maintained in a conformation that allows chaperones from other HSP families to extract them from the complex and reactivate them (Mogk et al., 2019).

The reactivation of such denatured proteins is carried out by Hsp70 chaperones (Fig. 3) (Żwirowski et al., 2017). Through competitive substrate binding, Hsp70 chaperones displace small HSPs from the complex shell and perform substrate folding, sometimes in cooperation with Hsp100 (Żwirowski et al., 2017).

The shell of the complexes has several important physiological functions: (1) it maintains the size and solubility of the complex by protecting the hydrophobic regions of the substrate proteins from interacting with other misfolded proteins, thereby preventing their aggregation; (2) it forms a selective barrier that is permeable only to Hsp70 chaperones; (3) it increases the demand for Hsp70 chaperones, in comparison with protein aggregates (Żwirowski et al., 2017).

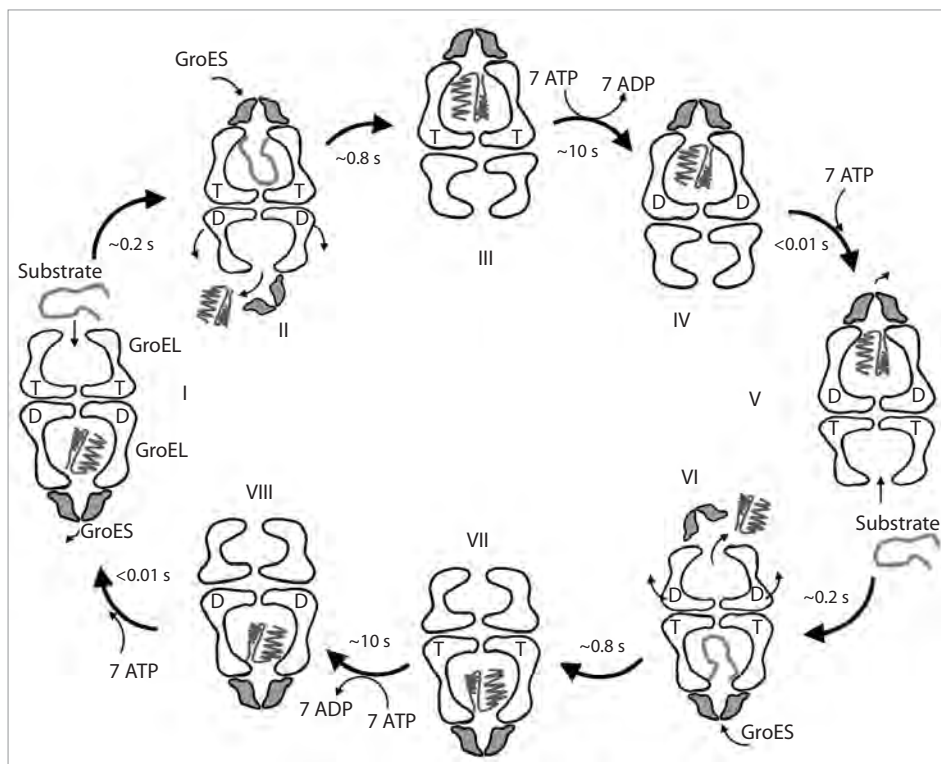


Fig. 2. The reaction cycle of Hsp60 using the example of bacterial GroEL and its co-chaperonin GroES (homohexamers of GroEL and GroES are shown in cross-section).

I – hydrophobic regions of a misfolded client protein interact with the hydrophobic inner surface of the apical domains of the ATP-bound open ring of GroEL (top in the image). II – the GroEL cavity is rapidly (within ~ 0.2 s (Horwich, 2011)) closed by the GroES co-chaperonin, and the substrate becomes “trapped” inside the GroEL cavity. III – the binding of GroES causes the apical domains of GroEL to rotate 100° clockwise (Horwich, 2011; Clare et al., 2012), which results in the inner surface of the GroEL ring becoming hydrophobic, promoting the folding of the substrate. IV – approximately 10 s later, ATP hydrolysis occurs in the equatorial domains of the ring containing the substrate protein, leading to the weakening of the interaction between the two rings allowing the second ring to bind ATP and begin the folding cycle (Horwich, 2011). V – the GroES co-chaperonin dissociates from the first GroEL ring; the folding cycle of a new substrate protein begins in the second ring. VI – the folded (or still misfolded) client protein leaves the now open cavity of the first ring; ADP and phosphate dissociate from the equatorial domains of the first GroEL ring. VII, VIII – the first ring is unable to bind ATP and, therefore, the new substrate protein, until ATP hydrolysis occurs in the second ring. T – ATP; D – ADP.

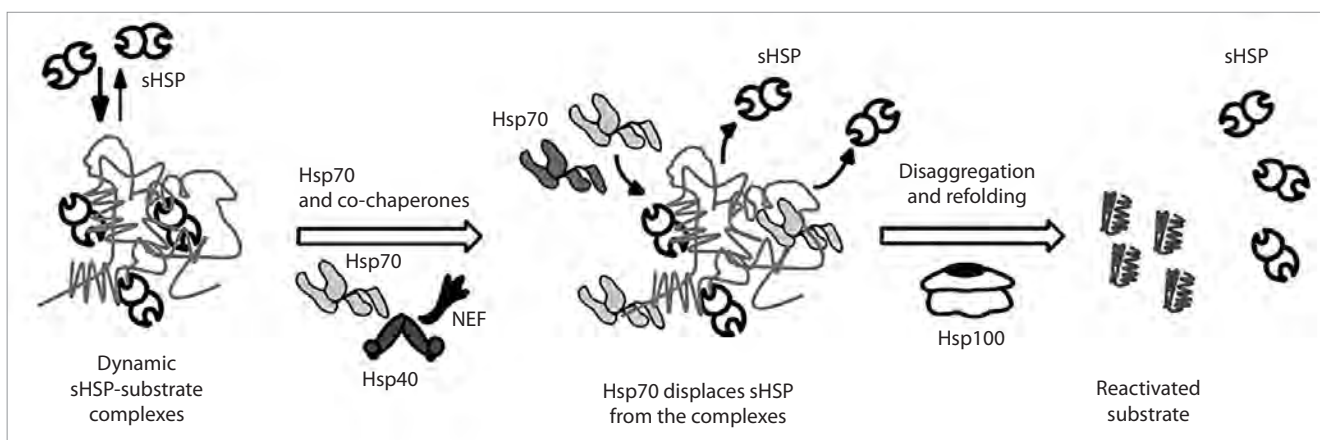


Fig. 3. Schematic model of small heat shock protein release from complexes with substrate proteins by Hsp70 chaperones and the following reactivation of the substrate proteins by Hsp70 and Hsp100 chaperones.

The complexes of small HSPs with substrates consist of a stable core formed by misfolded proteins and the associated small HSPs, and of a dynamic shell, on the surface of which small HSPs bind to and dissociate from the complex (the equilibrium is shifted toward binding). Hsp70 chaperones displace small HSPs from the complexes, releasing them, and then present the substrate proteins to Hsp100 chaperones, which then pull them through their central pore and reactivate the substrates.

Thus, although small HSPs keep aggregation-prone proteins in a state suitable for folding, they impede substrate reactivation at insufficient Hsp70 concentrations. This mechanism may allow for the isolation of misfolded proteins under stress conditions, when the number of available Hsp70 chaperones is limited, for example, during chronic stress (Żwirowski et al., 2017). At the same time, at low Hsp70 concentrations, small HSPs may promote the formation of protein aggregates because they prevent the ubiquitination of substrate proteins in the complex and their degradation (Mogk et al., 2019).

Conclusion

As molecular chaperones, heat shock proteins play a central role in the maintenance of protein homeostasis and are essential for all known prokaryotes and eukaryotes. Members of different HSP families perform various functions, assisting in the correct folding of polypeptides, preventing the aggregation of misfolded proteins, carrying out their refolding, or promoting their degradation. To perform all these functions, molecular chaperones work in tandem with each other and with numerous co-chaperones, maintaining the cell's proteome in working condition. Thus, heat shock proteins are the “first line of defence” for cells against the toxic effects of damaged and misfolded proteins, both under normal conditions and when exposed to various stressors.

References

- Avellaneda M.J., Franke K.B., Sunderlikova V., Bukau B., Mogk A., Tans S.J. Processive extrusion of polypeptide loops by a Hsp100 disaggregase. *Nature*. 2020;578(7794):317-320. doi 10.1038/s41586-020-1964-y
- Bar-Lavan Y., Shemesh N., Ben-Zvi A. Chaperone families and interactions in metazoa. *Essays Biochem*. 2016;60(2):237-253. doi 10.1042/EBC20160004
- Brocchieri L., Conway de Macario E., Macario A.J. Hsp70 genes in the human genome: conservation and differentiation patterns predict a wide array of overlapping and specialized functions. *BMC Evol Biol*. 2008;8(1):19. doi 10.1186/1471-2148-8-19
- Clare D.K., Vasishtan D., Stagg S., Quispe J., Farr G.W., Topf M., Horwich A.L., Saibil H.R. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell*. 2012;149(1):113-123. doi 10.1016/j.cell.2012.02.047
- Duran E.C., Weaver C.L., Lucius A.L. Comparative analysis of the structure and function of AAA+ motors ClpA, ClpB, and Hsp104: common threads and disparate functions. *Front Mol Biosci*. 2017;4:54. doi 10.3389/fmolb.2017.00054
- Ellis J. Proteins as molecular chaperones. *Nature*. 1987;328(6129):378-379. doi 10.1038/328378a0
- Fan F., Duan Y., Yang F., Trexler C., Wang H., Huang L., Li Y., Tang H., Wang G., Fang X., Liu J., Jia N., Chen J., Ouyang K. Deletion of heat shock protein 60 in adult mouse cardiomyocytes perturbs mitochondrial protein homeostasis and causes heart failure. *Cell Death Differ*. 2020;27(2):587-600. doi 10.1038/s41418-019-0374-x
- Feder M.E., Hofmann G.E. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Physiol*. 1999;61(1):243-282. doi 10.1146/annurev.physiol.61.1.243
- Fu X. Insights into how small heat shock proteins bind a great diversity of substrate proteins: a super-transformer model. In: Tanguay R.M., Hightower L.E. (Eds) *The Big Book on Small Heat Shock Proteins*. Cham: Springer Int. Pub., 2015;107-117. doi 10.1007/978-3-319-16077-1_4
- Hayer-Hartl M. From chaperonins to Rubisco assembly and metabolic repair. *Protein Sci*. 2017;26(12):2324-2333. doi 10.1002/pro.3309
- Hemmingsen S.M. What is a chaperonin? *Nature*. 1992;357(6380):650-650. doi 10.1038/357650b0
- Hodson S., Marshall J.J.T., Burston S.G. Mapping the road to recovery: the ClpB/Hsp104 molecular chaperone. *J Struct Biol*. 2012;179(2):161-171. doi 10.1016/j.jsb.2012.05.015
- Horwich A.L. Protein folding in the cell: an inside story. *Nat Med*. 2011;17(10):1211-1216. doi 10.1038/nm.2468
- Jolly C., Morimoto R.I. Role of the heat shock response and molecular chaperones in oncogenesis and cell death. *J Natl Cancer Inst*. 2000;92(19):1564-1572. doi 10.1093/jnci/92.19.1564
- Kakkar V., Meister-Broekema M., Minoia M., Carra S., Kampinga H.H. Barcoding heat shock proteins to human diseases: looking beyond the heat shock response. *Dis Model Mech*. 2014;7(4):421-434. doi 10.1242/dmm.014563
- Kampinga H.H., Hageman J., Vos M.J., Kubota H., Tanguay R.M., Bruford E.A., Cheetham M.E., Chen B., Hightower L.E. Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress Chaperones*. 2009;14(1):105-111. doi 10.1007/s12192-008-0068-7
- Koumoto Y., Shimada T., Kondo M., Hara-Nishimura I., Nishimura M. Chloroplasts have a novel Cpn10 in addition to Cpn20 as co-chaperonins in *Arabidopsis thaliana*. *J Biol Chem*. 2001;276(32):29688-29694. doi 10.1074/jbc.M102330200
- Kumar C.M.S., Mande S.C., Mahajan G. Multiple chaperonins in bacteria-novel functions and non-canonical behaviors. *Cell Stress Chaperones*. 2015;20(4):555-574. doi 10.1007/s12192-015-0598-8
- Larburu N., Adams C.J., Chen C.-S., Nowak P.R., Ali M.M.U. Mechanism of Hsp70 specialized interactions in protein translocation and the unfolded protein response. *Open Biol*. 2020;10(8):200089. doi 10.1098/rsob.200089
- Li J., Buchner J. Structure, function and regulation of the Hsp90 machinery. *Biomed J*. 2013;36(3):106-117. doi 10.4103/2319-4170.113230
- Lindquist S. The heat-shock response. *Annu Rev Biochem*. 1986;55(1):1151-1191. doi 10.1146/annurev.bi.55.070186.005443
- Lopez T., Dalton K., Frydman J. The mechanism and function of group II chaperonins. *J Mol Biol*. 2015;427(18):2919-2930. doi 10.1016/j.jmb.2015.04.013
- Macario A.J.L., Grippo T.M., de Macario E.C. Genetic disorders involving molecular-chaperone genes: a perspective. *Genet Med*. 2005;7(1):3-12. doi 10.1097/01.GIM.0000151351.11876.C3
- Mogk A., Kummer E., Bukau B. Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation. *Front Mol Biosci*. 2015;2:22. doi 10.3389/fmolb.2015.00022
- Mogk A., Ruger-Herreros C., Bukau B. Cellular functions and mechanisms of action of small heat shock proteins. *Annu Rev Microbiol*. 2019;73(1):89-110. doi 10.1146/annurev-micro-020518-115515
- Mokry D.Z., Abrahão J., Ramos C.H.I. Disaggregases, molecular chaperones that resolubilize protein aggregates. *An Acad Bras Cienc*. 2015;87(2 suppl):1273-1292. doi 10.1590/0001-3765201520140671
- Paul A., Rao S., Mathur S. The α -crystallin domain containing genes: identification, phylogeny and expression profiling in abiotic stress, phytohormone response and development in tomato (*Solanum lycopersicum*). *Front Plant Sci*. 2016;7:426. doi 10.3389/fpls.2016.00426
- Radons J. The human HSP70 family of chaperones: where do we stand? *Cell Stress Chaperones*. 2016;21(3):379-404. doi 10.1007/s12192-016-0676-6
- Rosenzweig R., Nillegoda N.B., Mayer M.P., Bukau B. The Hsp70 chaperone network. *Nat Rev Mol Cell Biol*. 2019;20(11):665-680. doi 10.1038/s41580-019-0133-3
- Saibil H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol*. 2013;14(10):630-642. doi 10.1038/nrm3658

- Saji H., Iizuka R., Yoshida T., Abe T., Kidokoro S., Ishii N., Yohda M. Role of the IXI/V motif in oligomer assembly and function of StHsp14.0, a small heat shock protein from the acidothermophilic archaeon, *Sulfolobus tokodaii* strain 7. *Proteins Struct Funct Bioinform.* 2008;71(2):771-782. doi 10.1002/prot.21762
- Sarkar S., Singh M.D., Yadav R., Arunkumar K.P., Pittman G.W. Heat shock proteins: molecules with assorted functions. *Front Biol (Beijing)*. 2011;6(4):312. doi 10.1007/s11515-011-1080-3
- Sørensen J.G., Kristensen T.N., Loeschcke V. The evolutionary and ecological role of heat shock proteins. *Ecol Lett.* 2003;6(11):1025-1037. doi 10.1046/j.1461-0248.2003.00528
- Tower J. Heat shock proteins and *Drosophila* aging. *Exp Gerontol.* 2011;46(5):355-362. doi 10.1016/j.exger.2010.09.002
- van Leeuwen F.W., Kampinga H.H. Heat shock proteins and protein quality control in Alzheimer's disease. In: Wolfe M.S. (Ed.) *The Molecular and Cellular Basis of Neurodegenerative Diseases*. Academic Press, 2018;269-298. doi 10.1016/B978-0-12-811304-2.00010-9
- Xiao C., Hull D., Qiu S., Yeung J., Zheng J., Barwell T., Robertson R.M., Seroude L. Expression of heat shock protein 70 is insufficient to extend *Drosophila melanogaster* longevity. *G3 (Bethesda)*. 2019;9(12):4197-4207. doi 10.1534/g3.119.400782
- Zhang S., Zhou H., Yu F., Bai C., Zhao Q., He J., Liu C. Structural insight into the cooperation of chloroplast chaperonin subunits. *BMC Biol.* 2016;14(1):29. doi 10.1186/s12915-016-0251-8
- Żwirowski S., Kłosowska A., Obuchowski I., Nillegoda N.B., Pi-róg A., Ziętkiewicz S., Bukau B., Mogk A., Liberek K. Hsp70 displaces small heat shock proteins from aggregates to initiate protein refolding. *EMBO J.* 2017;36(6):783-796. doi 10.15252/embj.201593378

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

Received July 11, 2024. Revised September 22, 2024. Accepted October 10, 2024.

doi 10.18699/vjgb-25-03

Generation and characterisation of seven induced pluripotent stem cell lines from two patients with Parkinson's disease carrying the pathological variant c.1087G>T of the *LGR4* gene

V.S. Podvysotskaya^{1, 2} & E.V. Grigor'eva^{1, 3} & A.A. Malakhova^{1, 3}, J.M. Minina¹, Y.V. Vyatkin⁴, E.A. Khabarova^{1, 5}, J.A. Rzaev⁵, S.P. Medvedev^{1, 3}, L.V. Kovalenko^{1, 6}, S.M. Zakian^{1, 3}

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 Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

⁴ NOVEL Ltd., Novosibirsk, Russia

⁵ Federal Neurosurgical Center of the Ministry of Health of the Russian Federation, Novosibirsk, Russia

⁶ Surgut State University, Surgut, Khanty-Mansiysk Autonomous Okrug – Ugra, Russia

✉ evlena@bionet.nsc.ru

Abstract. Parkinson's disease is a neurodegenerative disorder affecting dopaminergic neurons of the substantia nigra pars compacta. The known pathological genetic variants may explain the cause of only 5 % of cases of the disease. In our study, we found two patients with a clinical diagnosis of Parkinson's disease with the genetic variant c.1087G>T (p.Gly363Cys) of the *LGR4* gene. The *LGR4* gene encodes the membrane receptor LGR4 (leucine rich repeat containing G protein-coupled receptor 4) associated with the G protein. We hypothesize that the *LGR4* gene may be either a direct cause or a risk factor for this disease, since it is one of the main participants of the WNT/β-catenin signalling pathway. This signalling pathway is necessary for the proliferation of neurons during their differentiation, which may lead to Parkinson's disease. To study the relationship between this genetic variant and Parkinson's disease, an ideal tool is a cellular model based on induced pluripotent stem cells (iPSCs) and their differentiated derivatives, dopaminergic neurons. We reprogrammed the peripheral blood mononuclear cells of the two patients with the c.1087G>T variant of the *LGR4* gene with non-integrating episomal vectors expressing OCT4, SOX2, KLF4, LIN28, L-MYC and mp53DD proteins. The obtained seven lines of induced pluripotent stem cells were characterised in detail. The iPSCs lines obtained meet all the requirements of pluripotent cells, namely, they stably proliferate, form colonies with a morphology characteristic of human pluripotent cells, have a normal diploid karyotype, express endogenous alkaline phosphatase and pluripotency markers (OCT4, NANOG, SSEA-4 and SOX2) and are capable to differentiate into derivatives of the three germ layers. The iPSC lines obtained in this work can be used as a tool to generate a relevant model to study the effect of the pathological variant c.1087G>T of the *LGR4* gene on dopaminergic neuron differentiation.

Key words: Parkinson's disease; reprogramming; induced pluripotent stem cells; *LGR4* gene.

For citation: Podvysotskaya V.S., Grigor'eva E.V., Malakhova A.A., Minina J.M., Vyatkin Y.V., Khabarova E.A., Rzaev J.A., Medvedev S.P., Kovalenko L.V., Zakian S.M. Generation and characterisation of seven induced pluripotent stem cell lines from two patients with Parkinson's disease carrying the pathological variant c.1087G>T of the *LGR4* gene. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025;29(1):15-25. doi 10.18699/vjgb-25-03

Funding. The study was carried out with the financial support of the Foundation for Scientific and Technological Development of Yugra within the framework of scientific project No. 2023-573-05.

Acknowledgements. Analysis of the results of immunofluorescence staining of preparations and differential staining of chromosomes was performed using the resources of the Center for Collective Use of Microscopic Analysis of Biological Objects of ICG SB RAS (<https://ckp.icgen.ru/ckpmabo/>), supported by the budget project of the Institute of Cytology and Genetics SB RAS (FWNR-2022-0015).

Создание и характеристика семи линий индуцированных плюрипотентных стволовых клеток от двух пациентов с болезнью Паркинсона, несущих вариант с.1087G>T гена *LGR4*

В.С. Подвысоцкая^{1, 2} & Е.В. Григорьева^{1, 3} & А.А. Малахова^{1, 3}, Ю.М. Минина¹, Ю.В. Вяткин⁴, Е.А. Хабарова^{1, 5}, Дж.А. Рзаев⁵, С.П. Медведев^{1, 3}, Л.В. Коваленко^{1, 6}, С.М. Закиан^{1, 3}

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

© Podvysotskaya V.S., Grigor'eva E.V., Malakhova A.A., Minina J.M., Vyatkin Y.V., Khabarova E.A., Rzaev J.A., Medvedev S.P., Kovalenko L.V., Zakian S.M., 2025

& These authors contributed equally to this work

This work is licensed under a Creative Commons Attribution 4.0 License

³ Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия

⁴ ООО «Новые Программные Системы», Новосибирск, Россия

⁵ Федеральный центр нейрохирургии Министерства здравоохранения Российской Федерации, Новосибирск, Россия

⁶ Сургутский государственный университет, Сургут, Ханты-Мансийский автономный округ – Югра, Россия

✉ evlena@bionet.nsc.ru

Аннотация. Болезнь Паркинсона – тяжелое нейродегенеративное заболевание, поражающее дофаминергические нейроны компактной части черной субстанции головного мозга. Известные патологические генетические варианты, ассоциированные с болезнью Паркинсона, объясняют причину всего 5 % случаев заболевания, поэтому исследования в этой области актуальны и активно продолжаются. В данном исследовании мы обнаружили двух пациентов с клиническим диагнозом «болезнь Паркинсона» с генетическим вариантом *LGR4:c.1087G>T* (p.Gly363Cys, rs117543292). Этот ген кодирует мембранный рецептор LGR4 (leucine rich repeat containing G protein-coupled receptor 4), ассоциированный с G-белком, который участвует в регуляции функционирования сигнального пути WNT/ β -катенин. Данный сигнальный путь необходим для пролиферации нейронов во время их дифференцировки, поэтому его дисфункция в результате гетерозиготной мутации *c.1087G>T* в гене *LGR4* нарушает дифференцировку дофаминергических нейронов, что может приводить к болезни Паркинсона. Идеальным инструментом для изучения связи этого генетического варианта с болезнью Паркинсона является клеточная модель на основе индуцированных плюрипотентных стволовых клеток (ИПСК) и их дифференцированных производных – дофаминергических нейронов. В результате репрограммирования эпизомными векторами, экспрессирующими белки OCT4, SOX2, KLF4, LIN28, L-MYC и tp53DD, мононуклеарных клеток периферической крови двух пациентов с вариантом *c.1087G>T* гена *LGR4* нами были получены и детально охарактеризованы семь линий ИПСК. Данные ИПСК отвечают всем требованиям плюрипотентных клеток, а именно: стабильно пролиферируют, образуют колонии с характерной для плюрипотентных клеток человека морфологией, имеют нормальный диплоидный кариотип, экспрессируют щелочную фосфатазу и маркеры плюрипотентности (OCT4, NANOG, SSEA-4 и SOX2) и способны дифференцироваться в производные трех зародышевых листков – энто-, экто- и мезодерму. Полученные в работе линии ИПСК будут в дальнейшем использованы для создания релевантной модели, направленной на исследование эффекта варианта *c.1087G>T* гена *LGR4* на развитие патологического фенотипа дофаминергических нейронов.

Ключевые слова: болезнь Паркинсона; репрограммирование; индуцированные плюрипотентные стволовые клетки; ген *LGR4*.

Introduction

Parkinson's disease is the second most prevalent neurodegenerative disorder after Alzheimer's disease (Wang et al., 2020). It is characterised by the degeneration of dopaminergic neurons in the substantia nigra, which are responsible for regulating movement. The resulting symptoms include tremor, bradykinesia and rigidity. The destruction of these neurons can be attributed to various factors. These include accumulation of alpha-synuclein protein and formation of Lewy bodies, dysfunction of mitochondria and lysosomes, problems of synaptic and vesicle transport. These factors, when combined, lead to an accelerated death of neurons. Some of the signs of Parkinson's disease at the cellular level are oxidative stress resulting from mitochondrial dysfunction (Niu et al., 2021) and endoplasmic reticulum (ER) stress resulting from the accumulation of large amounts of misfolded proteins (Fernandes et al., 2016; Marciniak et al., 2022).

It is currently understood that more than 20 genes are associated with Parkinson's disease, the most prevalent mutations of which are found in the *GBA1*, *LRRK2*, *PRKN*, and *PINK1* genes (Funayama et al., 2023). However, it has been determined that only approximately 5 % of cases of Parkinson's disease are attributable to the inheritance of one of the genes associated with the disease, with the remaining 16–36 % being manifested as a result of non-monogenic inheritance. It can thus be concluded that research into

genetic variants associated with the onset of Parkinson's disease remains an active area of study.

Following a comprehensive analysis of the exomic sequencing results for a cohort of over 70 patients diagnosed with Parkinson's disease, who are under observation at the Federal Neurosurgical Center of the Ministry of Health of the Russian Federation (Novosibirsk), it was found that two patients of different genders, who are not related, carry the *c.1087G>T* variant of the *LGR4* gene (SRA databases, project PRJNA563295). There are currently 48 known single nucleotide substitutions in this gene, three of which are pathogenic missense mutations *c.2531A>G* (p.Asp844Gly), *c.286A>G* (p.Ile96Val) and *c.1087G>T* (p.Gly363Cys) (Mancini et al., 2023).

LGR4 is a G-coupled receptor that plays a role in the function of the WNT/ β -catenin and cAMP/protein kinase A signalling pathways (Shi et al., 2021). Previously, it was shown that the *c.1087G>T* variant of the *LGR4* gene is pathogenic and leads to delayed puberty due to a violation of the WNT/ β -catenin signalling pathway (Mancini et al., 2023), which is also involved in embryogenesis in the development of tissues and organs such as the gonads, kidneys, nervous system, liver and others. It should be noted that this pathway is also required for the development and differentiation of dopaminergic neurons (Marchetti et al., 2020). This correlation, together with the involvement of the LGR4 protein in the WNT/ β -catenin signalling pathway, leads

us to speculate that the genetic variant *LGR4:c.1087G>T* may cause dysfunction or death of dopaminergic neurons, possibly leading to Parkinson's disease. To study the association of this genetic variant with pathogenic processes leading to the death of dopaminergic neurons, the first step is to create a cellular model based on induced pluripotent stem cells (iPSCs) and relevant cell types, dopaminergic neurons, differentiated from them.

The pluripotent state of iPSCs is maintained by adding basic fibroblast growth factor (FGF-basic/bFGF) to the culture medium. To confirm the characteristics of pluripotency and self-renewal of cells, various tests are performed (Grigor'eva et al., 2023), such as qualitative (immunofluorescence staining) and quantitative (real-time PCR) analyses of the expression levels of various pluripotency factors (NANOG, SSEA-4, OCT4, SOX2, TRA1-81, TRA1-60, etc.). It is known that various karyotype abnormalities (polyploidisation, translocations, deletions, aneuploidies, etc.) are possible during prolonged *in vitro* cell culture, so karyotype analysis is one of the key tests for characterising the obtained iPSC lines. The most important test of cell pluripotency is spontaneous differentiation, which is necessary to confirm the ability of iPSCs to give rise to cell derivatives of all three germ layers (ecto-, ento- and mesoderm). In addition, the following shall be performed to characterise the iPSC lines obtained: STR analysis to confirm the origin of the iPSC lines obtained from a peripheral blood mononuclear cell (PBMC) donor; tests to indicate the presence of genetic variants of the PBMC donor in the iPSCs or to confirm the presence of genetic modifications in the case of transgenesis of iPSCs; PCR to eliminate exogenous DNA/episomes that enabled somatic cell reprogramming; and routine tests to ensure the absence of intra-laboratory contamination.

The capacity of iPSCs to yield derivatives of a virtually limitless range of cell types, including neurons, cardiomyocytes, liver cells, endothelial cells and numerous others, renders them a promising tool for generating cell models and investigating the pathogenesis of various inherited diseases in relevant cell types. For instance, by reprogramming PBMCs, culturing the resulting iPSCs and further differentiating them into dopaminergic neurons (Grigor'eva et al., 2023) and astrocytes (Yarkova et al., 2023), it will be possible to obtain *in vitro* models of Parkinson's disease. This platform will facilitate the analysis of the contribution of the genetic variant *LGR4:c.1087G>T* to Parkinson's disease development, as well as the study of its molecular and genetic pathogenesis. It will also support the exploration of new drug targets within diverse signalling pathways, and the prevention of pathogenic processes that result in dopaminergic neuron death. Furthermore, it will enable the evaluation of potential medical drugs.

Materials and methods

Ethics Statement. The study was approved by the Research Ethics Committee of the Federal Neurosurgical Center of the Ministry of Health of the Russian Federation (Novosibirsk), Protocol No. 1, dated 14 March 2017. Peripheral

blood samples from patients were provided by the Federal Neurosurgical Center. Patients signed a voluntary informed consent and an information sheet.

Isolation of PBMCs. PBMCs were isolated in a Ficoll gradient and frozen as previously described (Grigor'eva et al., 2024b).

Reprogramming of patient-specific PBMCs. Reprogramming of PBMCs was performed by transfection with episomal vectors expressing SOX2, OCT4, KLF4, L-MYC, LIN28 and the dominant negative form of mouse p53 protein – mp53DD (Addgene ID No. 41813–14, 41855–57) according to the previously described method (Grigor'eva et al., 2023). Transfection was performed using a Neon Transfection System (Thermo Fisher Scientific), programme: 1,650 V, 10 ms, 3 times. Primary colonies of iPSCs were replated into 1 cm² wells previously coated with mouse embryonic fibroblasts (MEF) inactivated with mitomycin C (Sigma-Aldrich).

Cultivation of patient-specific iPSCs. iPSCs were cultured on MEF as previously described (Yarkova et al., 2024). Colonies of iPSCs were disaggregated using TrypLE Express (Thermo Fisher Scientific) and plated once every 3–4 days at a 1:10 ratio with the addition of a ROCK inhibitor, 2 µM thiazovivin (Sigma-Aldrich).

Histochemical detection of endogenous alkaline phosphatase. Alkaline phosphatase was detected using the SIGMAFAST™ BCIP®/NBT kit (Sigma-Aldrich). The result was analysed using a Nikon Eclipse Ti-E microscope (Nikon) with NIS Elements Advanced Research software version 4.30.

Immunofluorescence staining. Immunofluorescence staining was performed according to the previously described method (Grigor'eva et al., 2024a). The preparations were analysed on a Nikon Eclipse Ti-E microscope using NIS Elements Advanced Research version 4.30 software. The list of antibodies used is shown in Table 1.

Spontaneous differentiation of iPSCs. To determine the potential of the resulting iPSC lines, spontaneous differentiation was performed by the formation of embryoid bodies followed by immunofluorescence staining. iPSCs were grown on MEF in Petri dishes (20 cm²) to a density of 80–90 %, after which the cells were incubated in 0.15 % collagenase type IV (Thermo Fisher Scientific) for 20–40 min. The iPSC colonies were then carefully pipetted to separate them from the feeder cells, centrifuged at 100 g for 5 min, the supernatant was decanted, suspended in iPSC growth medium without bFGF, and the colonies were transferred to a Petri dish (20 cm²) coated with 1 % agarose. After 14 days, embryoid bodies obtained from the colonies were plated on 8-well Chambered Coverglass plates (Thermo Fisher Scientific) coated with Matrigel. After 7–9 days, they were fixed in 4 % paraformaldehyde (Sigma-Aldrich) for 10 min and immunofluorescence analysis was performed for markers of the three germ layers. The list of antibodies is given in Table 1.

Karyotyping of iPSC lines. Karyotype analysis was performed using a previously developed and described method (Yarkova et al., 2023).

Table 1. List of antibodies used in the work

Antibodies	Dilution	Company, Cat No.	RRID
Primary antibodies			
Markers of pluripotency			
Mouse IgG2b anti-OCT-3/4	1:100	Santa cruz Biotechnology, sc-5279	RRID:AB_628051
Mouse IgG3 anti-SSEA-4	1:25	Abcam, ab16287	RRID:AB_778073
Mouse IgG1 anti-NANOG	1:50	Santa cruz Biotechnology, sc-293121	RRID:AB_2665475
Rabbit IgG anti-SOX2	1:400	Cell Signaling, 3579	RRID:AB_2195767
Mesoderm markers			
Mouse IgG2a anti-αSMA	1:200	Dako, M0851	RRID:AB_2223500
Mouse IgG1 anti-CD29	1:100	Thermo Fisher Scientific, 14-0299-82	RRID:AB_1210468
Endoderm markers			
Mouse IgG2a anti-AFP	1:250	Sigma-Aldrich, A8452	RRID:AB_258392
Mouse IgG1 anti-Cytokeratin 18 (KRT18)	1:250	Abcam, ab668	RRID:AB_305647
Ectoderm markers			
Mouse IgG2a anti TUBB3	1:1,000	BioLegend, 801201	RRID:AB_2313773
Chicken IgY anti-MAP2	1:1,000	Abcam, ab5392	RRID:AB_2138153
Secondary antibodies			
Goat anti-Rabbit IgG (H + L) Alexa Fluor 568	1:400	Thermo Fisher Scientific, A11011	RRID:AB_143157
Goat anti-Mouse IgG2b Alexa Fluor 568	1:400	Thermo Fisher Scientific, A21144	RRID:AB_2535780
Goat anti-Mouse IgG3 Alexa Fluor 488	1:400	Thermo Fisher Scientific, A21151	RRID:AB_2535784
Goat anti-Mouse IgG (H + L) Alexa Fluor 488	1:400	Thermo Fisher Scientific, A11008	RRID:AB_143165
Goat anti-Mouse IgG2a Alexa Fluor 568	1:400	Thermo Fisher Scientific, A21134	RRID:AB_2535773
Goat anti-Mouse IgG1 Alexa Fluor 488	1:400	Thermo Fisher Scientific, A-21121	RRID:AB_2535764
Goat anti-Chicken IgY (H + L) Alexa Fluor 488	1:400	Abcam, ab150173	RRID:AB_2827653

Isolation of genomic DNA and RNA. DNA was isolated using QuickExtract™ DNA extraction solution (Lucigen) according to the manufacturer’s protocol. For RNA isolation, cells were grown to an area of 8–12 cm², lysed in 0.5–1 ml TRIzol Reagent (Thermo Fisher Scientific), and RNA was isolated according to the manufacturer’s protocol. RNA concentration was determined using an EzDrop 1000C spectrophotometer (Blue-Ray Biotech). cDNA synthesis was performed using M-MuLV–RH reverse transcriptase (Biolabmix).

Detection of genetic variants in PBMCs and iPSCs. Sequencing of the patients’ exome was performed at Genoanalitika LLC, Moscow. Genomic DNA isolated from PBMCs was fragmented into 300 nucleotide pair fragments using ultrasound on a Covaris S2. Genomic DNA

(800 ng) was used to prepare a library using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs) and Sure Select AllExome V7 (Agilent). Library sequencing was performed on a HiSeq 2500 (Illumina) with paired reads of 150 nucleotides from both ends. Raw exome sequencing results are available in the SRA data-base (project PRJNA563295, samples SAMN42050732, <https://www.ncbi.nlm.nih.gov/biosample/42050732> (PD58), SAMN42050755, <https://www.ncbi.nlm.nih.gov/biosample/42050755> (PD69)).

The presence of the c.1087G>T variant (rs117543292) of the *LGR4* gene was determined by Sanger sequencing of the obtained iPSC lines and patient-derived PBMCs (PD58 and PD69) using primers for the *LGR4* gene (Table 2). DNA from a conditionally healthy donor was used as a control. Reac-

Table 2. List of primers used in the work

Application	Gene/locus	Product size, bp	Forward/reverse primer (5'–3')
Detection of episomal vectors	<i>oriP</i>	544	TTCCACGAGGGTAGTGAACC/ TCGGGGGTGTTAGAGACAAC
Reference gene (RT-qPCR)	<i>B2M</i>	90	TAGCTGTGCTCGCGCTACT/ TCTCTGCTGGATGACGTGAG
Markers of pluripotency (RT-qPCR)	<i>NANOG</i>	116	TTTGTGGCCTGAAGAAAAC/ AGGGCTGTCCTGAATAAGCAG
	<i>OCT4</i>	94	CTTCTGCTTCAGGAGCTTGG/ GAAGGAGAAGCTGGAGCAAA
	<i>SOX2</i>	100	GCTTAGCCTCGTCGATGAAC/ AACCCCAAGATGCACAACCTC
Mycoplasma detection	Ribosomal <i>16S RNA</i> gene	280	GGGAGCAAACAGGATTAGATACCT/ TGCACCATCTGTCACTCTGTTAACCTC
Confirmation of the mutation	<i>LGR4</i>	366	GCGTTTCATGGGATGCCTGATAG/ TCTTAGCTCTGCTTCAACGCTTC

tions were performed on a T100 thermal cycler (Bio-Rad) using a BioMaster HS-Taq PCR-Color (2×) (Biolabmix) with the following program: 95 °C for 3 min; 35 cycles: 95 °C for 30 s, 65 °C for 30 s, 72 °C for 30 s; and 72 °C for 5 min. Sequencing reactions were performed using the Big Dye Terminator V.3.1. Cycle Sequencing Kit (Applied Biosystems) and analysed on an ABI 3130XL Genetic Analyzer at the Genomics Core Facility SB RAS (<http://www.niboch.nsc.ru/doku.php/sequence>).

PCR analysis for detection of reprogramming episomes and mycoplasma contamination. PCR was performed using BioMaster HS-Taq PCR-Color (2×) (Biolabmix) on a T100 thermal cycler (Bio-Rad), programme: 95 °C for 5 min; 35 cycles: 95 °C for 15 s, 60 °C for 15 s, 72 °C for 20 s. DNA fragments of *Mycoplasma* spp. from the BioMaster Myco-visor kit for detection of mycoplasma by RT-PCR (Biolabmix) were used as a positive control. Negative control was H₂O. Primers are listed in Table 2. After electrophoresis, the result was visualised under ultraviolet light using the Gel Doc XR+ System (Bio-Rad).

Quantitative RT-PCR for pluripotency markers. Quantitative PCR (RT-qPCR) was performed on a LightCycler 480 II system (Roche) using BioMaster HS-qPCR SYBR Blue 2× (Biolabmix) with the following programme: 95 °C for 5 min; 40 cycles: 95 °C for 10 s, 60 °C for 1 min. The embryonic stem cell line HUES9 (HVRDe009-A) was used as a positive control for the expression of pluripotency markers (Cowan et al., 2004). The pluripotency marker primers used in this work are listed in Table 2. The results were processed using the $\Delta\Delta CT$ method (Livak, Schmittgen, 2001).

STR analysis. The genetic material of the obtained iPSC lines and PBMCs from patients PD58 and PD69 was sent to Genoanalitika LLC (Moscow) for STR analysis. The STR analysis was performed for 26 loci using the COrDIS EXPERT 26 reagent kit (Russia).

Results and discussion

Obtaining and characterisation of the cell lines

Exomic sequencing analysis of early onset Parkinson's disease patients at the Federal Neurosurgical Center of the Ministry of Health of the Russian Federation (Novosibirsk) revealed two individuals (a 37-year-old woman (PD58) and a 66-year-old man (PD69)) with the variant *LGR4:c.1087G>T* (rs117543292).

PBMCs were isolated from patients' peripheral blood and then reprogrammed to revert to a pluripotency state, using episomal vectors encoding the pluripotency factors OCT4, SOX2, KLF4, LIN28, L-MYC, and the dominant-negative form of the mouse p53 protein, mp53DD (Okita et al., 2013). In this study, 21 cell lines of PD58 and 10 cell lines of PD69 were generated. Primary analysis for the presence/absence of episomal vectors and karyotyping allowed us to select three lines from the first patient (ICGi053-A/PD58-4, ICGi053-B/PD58-7 and ICGi053-C/PD58-14) and four lines from the second one (ICGi054-A/PD69-1/1, ICGi054-B/PD69-2/1, ICGi054-C/PD69-4 and ICGi054-D/PD69-5). All the lines were registered in the Human Pluripotent Stem Cell Registry (hPSCreg, <https://hpscereg.eu>). Information on cell lines can be found in hPSCreg at the following links: <https://hpscereg.eu/cell-line/ICGi053-A>, <https://hpscereg.eu/cell-line/ICGi053-B>, <https://hpscereg.eu/cell-line/ICGi053-C>, <https://hpscereg.eu/cell-line/ICGi054-A>, <https://hpscereg.eu/cell-line/ICGi054-B>, <https://hpscereg.eu/cell-line/ICGi054-C> и <https://hpscereg.eu/cell-line/ICGi054-D>.

All lines from both patients have a morphology characteristic of pluripotent cells, flat monolayer dense colonies with a large nuclear to cytoplasmic ratio, expressing endogenous alkaline phosphatase (Fig. 1a, 2a).

In the karyotyping of the obtained iPSC lines, 56 to 60 metaphases were analysed. From 6 to 12 chromosome

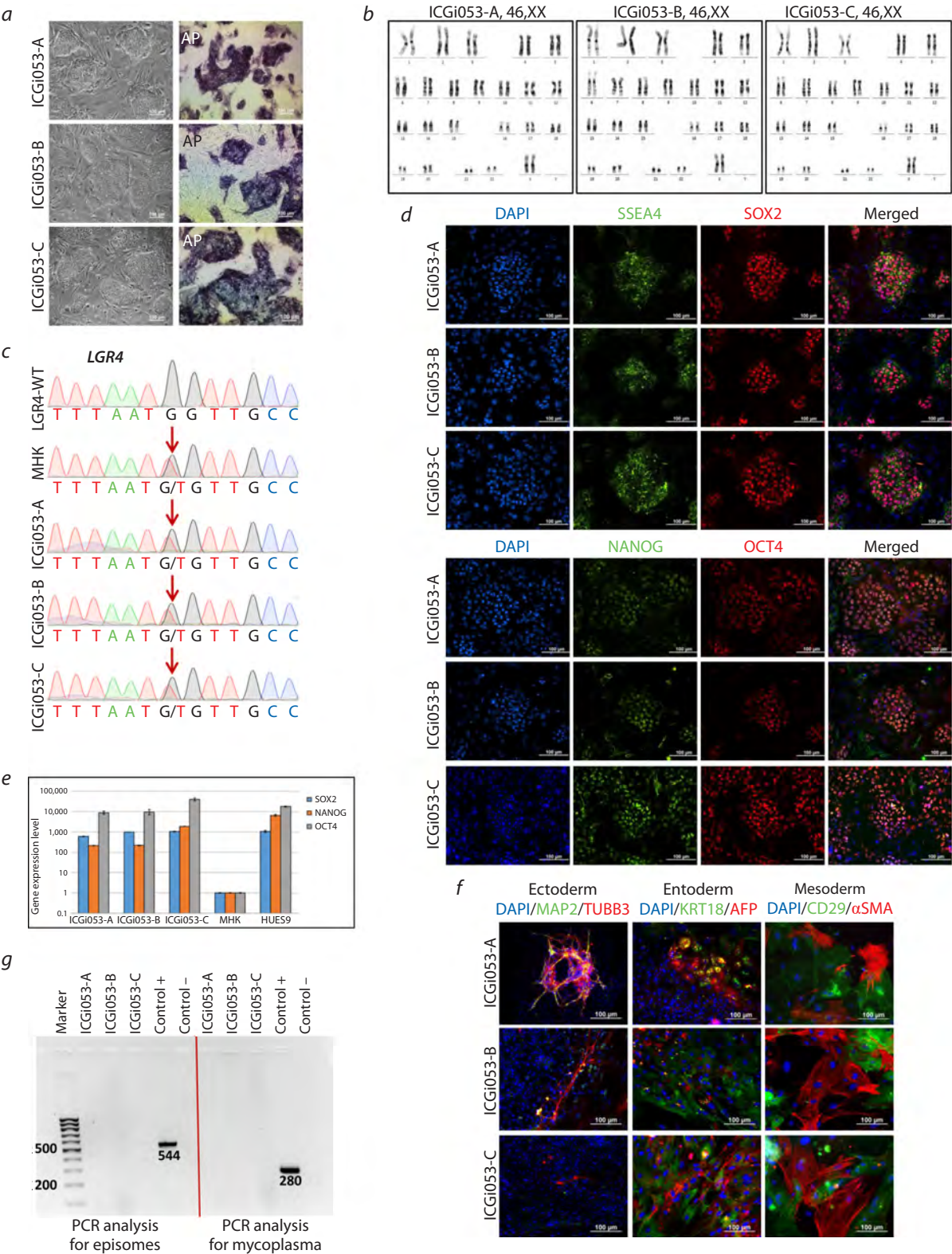


Fig. 1. Characterisation of three iPSC lines ICGi053-A, ICGi053-B and ICGi053-C. *a* – colony morphology and detection of alkaline phosphatase (AP) in iPSCs; *b* – karyotyping of iPSC lines using DAPI banding; *c* – sequenograms of PCR products obtained from genomic DNA of a patient with early onset Parkinson's disease PD58, a healthy donor (LGR4-WT), iPSC lines and PBMCs corresponding to positions 27377174-27377186 on human chromosome 11 (GRCh38 genome assembly). The position of the gene variant studied is *LGR4* NC_000011.10:g.27377180C>A. The nucleotide substitution *LGR4*:c.1087G>T is indicated by the red arrow; *d* – immunofluorescence staining for markers of pluripotency: NANOG (green), OCT4 (red), SSEA-4 (green), SOX2 (red); nuclei stained with DAPI (blue); *e* – quantitative RT-PCR for pluripotency markers (OCT4, SOX2, NANOG) of derived iPSC lines, patient PBMCs and human embryonic stem cell line HUES9; *f* – immunofluorescence staining of spontaneously differentiated cells for markers of ectoderm (TUBB3 (red), MAP2 (green)), endoderm (keratin 18/KRT18 (green), AFP (red)) and mesoderm (α SMA (red), CD29 (green)); nuclei were stained with DAPI (blue); *g* – result of PCR assay for episomal vector elimination and mycoplasma contamination of the cells obtained. Scale bars are 100 μ m.

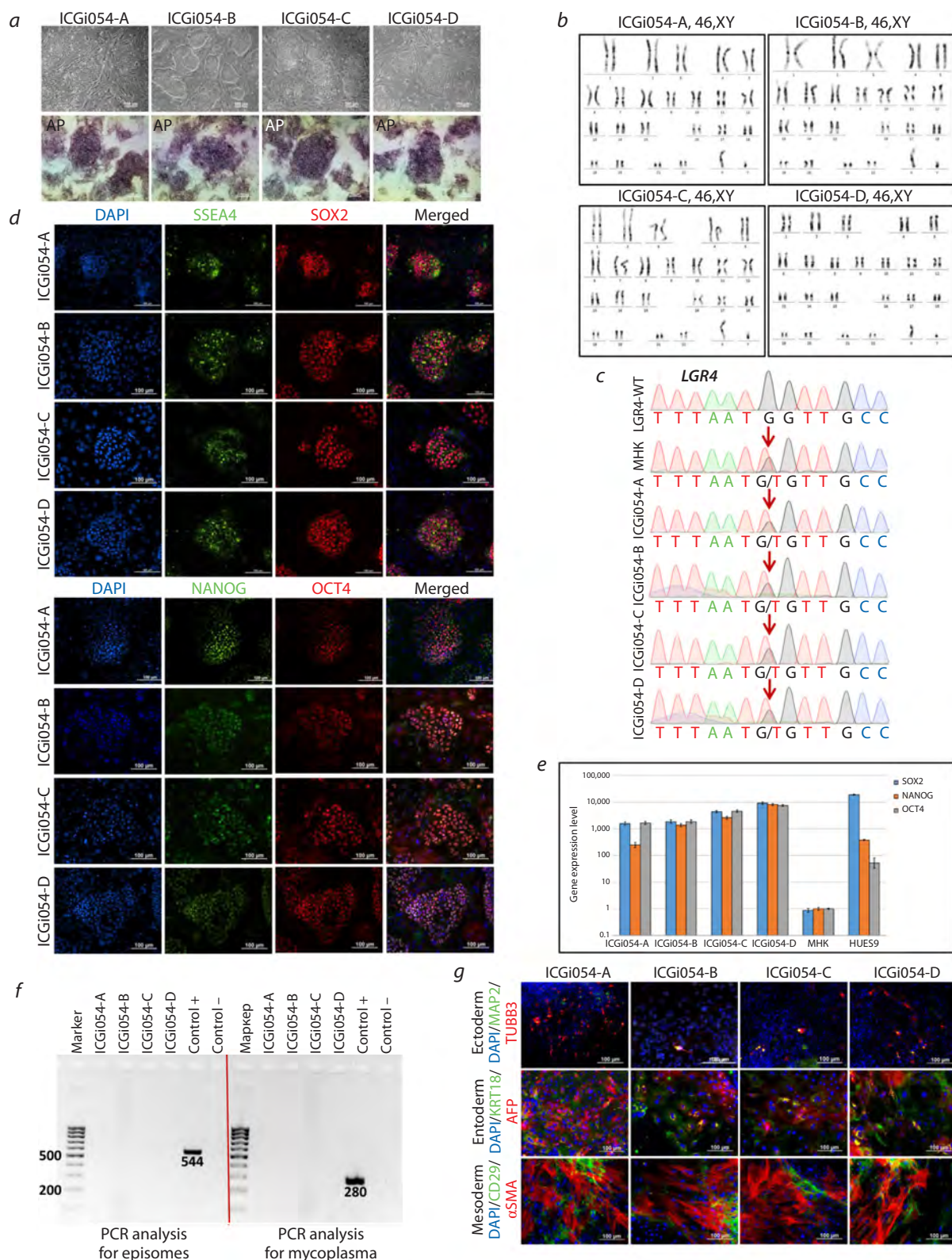


Fig. 2. Characterisation of four patient-specific ICGi054-A, ICGi054-B, ICGi054-C, and ICGi054-D iPSC lines.

a – colony morphology and detection of alkaline phosphatase (AP) in iPSCs; *b* – karyotyping of iPSC lines using DAPI banding; *c* – sequenograms of PCR products obtained from genomic DNA from a PD69 Parkinson's disease patient, a healthy donor (*LGR4*-WT), iPSC lines, and PBMCs corresponding to positions 27377174–27377186 on human chromosome 11 (GRCh38 genome assembly). The position of the gene variant studied is *LGR4* NC_000011.10:g.27377180C>A. The nucleotide substitution *LGR4:c.1087G>T* is indicated by the red arrow; *d* – immunofluorescence staining for markers of pluripotency: NANOG (green), OCT4 (red), SSEA-4 (green), SOX2 (red); nuclei stained with DAPI (blue); *e* – quantitative RT-PCR for pluripotency markers (OCT4, SOX2, NANOG) of derived iPSC lines, patient PBMCs and human embryonic stem cell line HUES9; *f* – immunofluorescence staining of spontaneously differentiated cells for markers of ectoderm (TUBB3 (red), MAP2 (green)), endoderm (keratin 18/KRT18 (green), AFP (red)) and mesoderm (αSMA (red), CD29 (green)); nuclei were stained with DAPI (blue); *g* – result of PCR assay for episomal vector elimination and mycoplasma contamination of the cells obtained. Scale bars are 100 μm.

Table 3. Detailed analysis of the chromosomal composition of the metaphase spreads of the resulting iPSC lines

iPSC line	Number of metaphases with different numbers of chromosomes						Number of metaphases analysed	Karyotype of diploid cells
	42–44	45	46	47	48–50	80–92		
ICGi053-A	4	3	40	8	2	4	58	46,XX
ICGi053-B	9	3	37	4	0	5	58	46,XX
ICGi053-C	5	4	37	3	1	8	58	46,XX
ICGi054-A	4	8	37	5	2	4	60	46,XY
ICGi054-B	7	4	41	3	1	1	57	46,XY
ICGi054-C	15	5	37	6	1	2	60	46,XY
ICGi054-D	8	10	31	4	1	1	56	46,XY

spreads were obtained for each cell line, in accordance with the recommendations based on the European standards for cytogenetic and molecular cytogenetic studies of constitutive and acquired chromosomal abnormalities (Hastings et al., 2012; ISCN, 2020). The appearance of cells with the number of chromosomes ranging from 42 to 45 is the result of the loss of individual chromosomes during preparation. Since the number of such cells was low and they showed an absence of different chromosomes, we did not consider them as the formation of new subclones. It is known that cultured stem cells are characterised by aneuploidy (Menzorov et al., 2016), but monosomal cells are prone to apoptosis or less intensive proliferation, so in cell culture it is common to consider lines with more than 50 % of metaphases with a full set of chromosomes as normal. Detailed analysis showed that all lines have more than 55 % of cells with a normal diploid karyotype – 46,XX for ICGi053 (Fig. 1b) and 46,XY for ICGi054 (Fig. 2b). It should be noted that the analysis of the seven iPSC lines obtained did not reveal any metaphase spreads with structural rearrangements. Detailed information on the composition of the analysed metaphase spreads can be found in Table 3.

Sanger sequencing demonstrated the presence of the c.1087G>T (rs117543292) variant of the *LGR4* gene in a heterozygous state in the obtained cell lines as well as in PBMCs (Fig. 1c, 2c, the genetic variant is indicated by a red arrow). Immunofluorescence analysis of lines ICGi053-A (passage 27), ICGi053-B (passage 26), ICGi053-C (passage 27), ICGi054-A (passage 26), ICGi054-B (passage 22), ICGi054-C (passage 25) and ICGi054-D (passage 24) for pluripotency markers revealed the presence of the SSEA-4 surface antigen and the transcription factors NANOG, SOX2 and OCT4 (Fig. 1d, 2d). Quantitative real-time PCR (RT-PCR) of patient PBMCs, patient-derived iPSC lines, and the control embryonic stem cell line HUES9 showed increased gene expression levels of *OCT4*, *NANOG*, and *SOX2* in the iPSC lines, which was comparable to the expression level

in HUES9 (Fig. 1e, 2e). The test for the ability to spontaneously differentiate into derivatives of three germ layers, which is the main confirmation of the pluripotency status of the obtained cell lines, and further immunofluorescence analysis showed the expression of markers of ectoderm (tubulin $\beta 3$ (TUBB3/TUJ1), microtubule-associated protein 2 (MAP2)), endoderm (alpha-fetoprotein (AFP)), keratin 18 (KRT18)) and mesoderm (α -smooth muscle actin (α SMA), surface marker CD29) (Fig. 1f, 2f). During cultivation, all lines underwent PCR testing for mycoplasma contamination and elimination of episomal vectors (Fig. 1g, 2g). STR analysis at 26 loci showed the identity of the iPSC lines derived from patient PBMCs (data available on request from the authors). The passports of the cell lines derived from PD58 and PD69 patients are shown in Tables 4 and 5, respectively.

Conclusion

As a result of reprogramming PBMCs from two Parkinson’s disease patients carrying the c.1087G>T (rs117543292) variant of the *LGR4* gene, seven iPSC lines – ICGi053-A, ICGi053-B, ICGi053-C, ICGi054-A, ICGi054-B, ICGi054-C and ICGi054-D – were obtained and characterised. The cell lines obtained correspond to human pluripotent cells: they have a characteristic morphology, a normal karyotype, express pluripotency factors, are able to differentiate spontaneously into derivatives of three germ layers and have the same genetic variant *LGR4*:c.1087G>T as donor PBMCs. This work is the first and necessary step to study the effect of the *LGR4*:c.1087G>T genetic variant on the manifestation of Parkinson’s disease. The iPSC lines obtained and their differentiated derivatives represent a unique cell platform on which it will be possible in the future to study at the molecular genetic level the early pathological processes that trigger the cascade of signalling pathways leading to neuronal death, to search for targets that modulate these signalling pathways and to test new potential drugs.

Table 4. Passport of cell lines ICGi053-A, ICGi053-B, ICGi053-C

Unique identifier	ICGi053-A, ICGi053-B, ICGi053-C
Alternative cell line names	PD58-4, PD58-7, PD58-14
Institution	Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
Ethical approval	The study was approved by the Research Ethics Committee of the Federal Neurosurgical Center of the Ministry of Health of the Russian Federation, Novosibirsk, Russia, Protocol No. 1 dated 14 March 2017
Cell type	iPSCs
Origin	Human
Additional information on the origin of the cell line	Age: 37 Gender: F Ethnicity: Caucasian race
Original cell type	Peripheral blood mononuclear cells (PBMCs)
Date of biomaterial collection	2022
Reprogramming method	Non-integrating episomal vectors
Reprogramming factors	OCT4, SOX2, KLF4, LIN28, L-MYC and mp53DD
Clonality	Clonal
Genetic modification	No
Type of genetic modification	N/A
Confirmation of elimination/ silencing of reprogramming transgenes	PCR, not detected
Disease	Parkinson's disease
Age of disease onset	36
Gene/locus	LGR4:c.1087G>T, rs117543292
Morphology	Human pluripotent cell-like monolayer colonies
Pluripotency	Confirmed in tests for the formation of embryoid bodies
Karyotype	46,XX
Check for contamination	Bacteria, fungi and mycoplasma not detected
Scope of application	<i>In vitro</i> model of Parkinson's disease
Cultivation method	On the feeder layer of mitotically inactivated mouse embryonic fibroblasts
Growth medium	85 % KnockOut DMEM, 15 % KnockOut Serum Replacement, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol, 1 % Penicillin-Streptomycin, GlutaMAX-I (all Thermo Fisher Scientific), 10 ng/ml bFGF (SCI Store)
Temperature, °C	37
CO ₂ concentration, %	5
O ₂ concentration, %	20
Passage method	TrypLE Express
Split ratio	1:8–1:10
Cryopreservation	90 % FBS, 10 % DMSO
Storage conditions	Liquid nitrogen
Registry ID	https://hpscreg.eu/cell-line/ICGi053-A https://hpscreg.eu/cell-line/ICGi053-B https://hpscreg.eu/cell-line/ICGi053-C
Date of certification/registration	09.07.2024 (ICGi053-A) 18.07.2024 (ICGi053-B) 22.07.2024 (ICGi053-C)

Table 5. Passport of cell lines ICGi054-A, ICGi054-B, ICGi054-C, ICGi054-D

Unique identifier	ICGi054-A, ICGi054-B, ICGi054-C, ICGi054-D
Alternative cell line names	PD69-1/1, PD69-2/1, PD69-4, PD69-5
Institution	Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
Ethical approval	The study was approved by the Research Ethics Committee of the Federal Neurosurgical Center of the Ministry of Health of the Russian Federation, Novosibirsk, Russia, Protocol No. 1 dated 14 March 2017
Cell type	iPSCs
Origin	Human
Additional information on the origin of the cell line	Age: 66 Gender: M Ethnicity: Caucasian race
Original cell type	Peripheral blood mononuclear cells (PBMCs)
Date of biomaterial collection	2022
Reprogramming method	Non-integrating episomal vectors
Reprogramming factors	OCT4, SOX2, KLF4, LIN28, L-MYC и mp53DD
Clonality	Clonal
Genetic modification	No
Type of genetic modification	N/A
Confirmation of elimination/silencing of reprogramming transgenes	PCR, not detected
Disease	Parkinson's disease
Age of disease onset	41
Gene/locus	LGR4:c.1087G>T, rs117543292
Morphology	Human pluripotent cell-like monolayer colonies
Pluripotency	Confirmed in tests for the formation of embryoid bodies
Karyotype	46,XY
Check for contamination	Bacteria, fungi and mycoplasma not detected
Scope of application	<i>In vitro</i> model of Parkinson's disease
Cultivation method	On the feeder layer of mitotically inactivated mouse embryonic fibroblasts
Growth medium	85 % KnockOut DMEM, 15 % KnockOut Serum Replacement, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol, 1 % Penicillin-Streptomycin, GlutaMAX-I (all Thermo Fisher Scientific), 10 ng/ml bFGF (SCI Store)
Temperature, °C	37
CO ₂ concentration, %	5
O ₂ concentration, %	20
Passage method	TrypLE Express
Split ratio	1:8–1:10
Cryopreservation	90 % FBS, 10 % DMSO
Storage conditions	Liquid nitrogen
Registry ID	https://hpscreg.eu/cell-line/ICGi054-A https://hpscreg.eu/cell-line/ICGi054-B https://hpscreg.eu/cell-line/ICGi054-C https://hpscreg.eu/cell-line/ICGi054-D
Date of certification/registration	26.07.2024 (ICGi054-A) 26.07.2024 (ICGi054-B) 29.07.2024 (ICGi054-C) 09.08.2024 (ICGi054-D)

References

- Cowan C.A., Klimanskaya I., McMahon J., Atienza J., Witmyer J., Zucker J.P., Wang S., Morton C.C., McMahon A.P., Powers D., Melton D.A. Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med*. 2004;350(13):1353-1356. doi 10.1056/NEJMSr040330
- Fernandes H.J.R., Hartfield E.M., Christian H.C., Emmanouilidou E., Zheng Y., Booth H., Bogetofte H., Lang C., Ryan B.J., Sardi S.P., Badger J., Vowles J., Evetts S., Tofaris G.K., Vekrellis K., Talbot K., Hu M.T., James W., Cowley S.A., Wade-Martins R. ER stress and autophagic perturbations lead to elevated extracellular α -synuclein in GBA-N370S Parkinson's iPSC-derived dopamine neurons. *Stem Cell Rep*. 2016;6(3):342-356. doi 10.1016/j.stemcr.2016.01.013
- Funayama M., Nishioka K., Li Y., Hattori N. Molecular genetics of Parkinson's disease: contributions and global trends. *J Hum Genet*. 2023;68(3):125-130. doi 10.1038/s10038-022-01058-5
- 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;24:4437. 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*. 2024a;25:6102. doi 10.3390/ijms25116102
- Grigor'eva E.V., Malakhova A.A., Yarkova E.S., Minina J.M., Vyatkin Y.V., Nadtochy J.A., Khabarova E.A., Rzaev J.A., Medvedev S.P., Zakian S.M. Generation and characterization of two induced pluripotent stem cell lines (ICGi052-A and ICGi052-B) from a patient with frontotemporal dementia with parkinsonism-17 associated with the pathological variant c.2013T>G in the *MAPT* gene. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2024b;28(7):679-687. doi 10.18699/vjgb-24-76
- Hastings R., Howell R., Bricarelli F.D., Kristoffersson U., Cavani S. General guidelines and quality assurance for cytogenetics. *Eur Cytogenet Assoc Newsl*. 2012;29:11-25
- ISCN 2020: An International System for Human Cytogenomic Nomenclature. S. Karger AG, 2020. doi 10.1159/isbn.978-3-318-06867-2
- 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
- Mancini A., Howard S.R., Marelli F., Cabrera C.P., Barnes M.R., Sternberg M.J.E., Leprovots M., Hadjimetriou I., Monti E., David A., Wehkalampi K., Oleari R., Lettieri A., Vezzoli V., Vassart G., Cariboni A., Bonomi M., Garcia M.I., Guasti L., Dunkel L. LGR4 deficiency results in delayed puberty through impaired Wnt/ β -catenin signaling. *JCI Insight*. 2023;5(11):e133434. doi 10.1172/jci.insight.133434
- Marchetti B., Tirolo C., L'Episcopo F., Caniglia S., Testa N., Smith J.A., Pluchino S., Serapide M.F. Parkinson's disease, aging and adult neurogenesis: Wnt/ β -catenin signalling as the key to unlock the mystery of endogenous brain repair. *Aging Cell*. 2020;19(3):e13101. doi 10.1111/ace1.13101
- Marciniak S.J., Chambers J.E., Ron D. Pharmacological targeting of endoplasmic reticulum stress in disease. *Nat Rev Drug Discov*. 2022;21(2):115-140. doi 10.1038/s41573-021-00320-3
- Menzorov A., Pristayzhnyuk I., Kizilova H., Yunusova A., Battulin N., Zhelezova A., Golubitsa A., Serov O.L. Cytogenetic analysis and *Dkl1-Dio3* locus epigenetic status of mouse embryonic stem cells during early passages. *Cytotechnology*. 2016;68(1):61-71. doi 10.1007/s10616-014-9751-y
- Niu Y., Zhang J., Dong M. Nrf2 as a potential target for Parkinson's disease therapy. *J Mol Med (Berl)*. 2021;99(7):917-931. doi 10.1007/s00109-021-02071-5
- Okita K., Yamakawa T., Matsumura Y., Sato Y., Amano N., Watanabe A., Goshima N., Yamanaka S. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells*. 2013;31(3):458-466. doi 10.1002/stem.1293
- Shi S., Li S., Zhang X., Wei Z., Fu W., He J., Hu Y., Li M., Zheng L., Zhang Z. *LGR4* gene polymorphisms are associated with bone and obesity phenotypes in Chinese female nuclear families. *Front Endocrinol*. 2021;12:656077. doi 10.3389/fendo.2021.656077
- Wang M., Ling K.-H., Tan J.J., Lu C.-B. Development and differentiation of midbrain dopaminergic neuron: from bench to bedside. *Cells*. 2020;9(6):1489. doi 10.3390/cells9061489
- Yarkova E.S., Grigor'eva E.V., Medvedev S.P., Pavlova S.V., Zakian S.M., Malakhova A.A. iPSC-derived astrocytes contribute to in vitro modeling of Parkinson's disease caused by the *GBA1* N370S mutation. *Int J Mol Sci*. 2023;25(1):327. doi 10.3390/ijms25010327
- 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;12:744. doi 10.3390/biomedicines12040744

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

Received September 16, 2024. Revised November 20, 2024. Accepted November 21, 2024.

doi 10.18699/vjgb-25-04

Studying concatenation of the Cas9-cleaved transgenes using barcodes

A.V. Smirnov ¹, A.N. Korablev¹, I.A. Serova¹, A.M. Yunusova ¹, A.A. Muravyova², E.S. Valeev¹, N.R. Battulin ^{1, 2} 

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

² Novosibirsk State University, Novosibirsk, Russia

 battulin@gmail.com

Abstract. In pronuclear microinjection, the Cas9 endonuclease is employed to introduce *in vivo* DNA double-strand breaks at the genomic target locus or within the donor vector, thereby enhancing transgene integration. The manner by which Cas9 interacts with DNA repair factors during transgene end processing and integration is a topic of considerable interest and debate. In a previous study, we developed a barcode-based genetic system for the analysis of transgene recombination following pronuclear microinjection in mice. In this approach, the plasmid library is linearized with a restriction enzyme or a Cas9 RNP complex at the site between a pair of barcodes. A pool of barcoded molecules is injected into the pronucleus, resulting in the generation of multicopy concatemers. In the present report, we compared the effects of *in vivo* Cas9 cleavage (RNP+ experiment) and *in vitro* production of Cas9-linearized transgenes (RNP– experiment) on concatenation. In the RNP+ experiment, two transgenic single-copy embryos were identified. In the RNP– experiment, six positive embryos were identified, four of which exhibited low-copy concatemers. Next-generation sequencing (NGS) analysis of the barcodes revealed that 53 % of the barcoded ends had switched their initial library pairs, indicating the involvement of the homologous recombination pathway. Out of the 20 transgene-transgene junctions examined, 11 exhibited no mutations and were presumably generated through re-ligation of Cas9-induced blunt ends. The majority of mutated junctions harbored asymmetrical deletions of 2–4 nucleotides, which were attributed to Cas9 end trimming. These findings suggest that Cas9-bound DNA may present obstacles to concatenation. Conversely, clean DNA ends were observed to be joined in a manner similar to restriction-digested ends, albeit with distinctive asymmetry. Future experiments utilizing *in vivo* CRISPR/Cas9 cleavage will facilitate a deeper understanding of how CRISPR-endonucleases influence DNA repair processes.

Key words: CRISPR/Cas9; pronuclear microinjection; DNA barcoding; transgenic animals; DSB repair; concatemer; homologous recombination (HR); non-homologous end-joining, NHEJ; mouse embryos.

For citation: Smirnov A.V., Korablev A.N., Serova I.A., Yunusova A.M., Muravyova A.A., Valeev E.S., Battulin N.R. Studying concatenation of the Cas9-cleaved transgenes using barcodes. *Vavilovskii Zhurnal Genetiki i Selektcii* = *Vavilov J Genet Breed.* 2025;29(1):26-34. doi 10.18699/vjgb-25-04

Funding. This work was supported by the Russian Science Foundation grant No. 24-74-10013.

Acknowledgements. The experiments involving the visualization of fluorescent embryos were conducted with support from State Project FWNR-2022-0019 at the Institute of Cytology and Genetics SB RAS. The calculations were conducted using the computational resources of the Computational Center of the Novosibirsk State University.

Изучение конкатенации баркодированных трансгенов, линейризованных Cas9

А.В. Смирнов ¹, А.Н. Кораблев¹, И.А.Серова¹, А.М. Юнусова ¹, А.А. Муравьева², Е.С. Валеев¹, Н.Р. Баттулин ^{1, 2} 

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 battulin@gmail.com

Аннотация. При пронукулярной микроинъекции эндонуклеаза Cas9 используется для создания двуцепочечных разрывов ДНК *in vivo* в геномном целевом локусе или внутри донорского вектора для улучшения интеграции трансгена. Взаимодействие Cas9 с факторами репарации ДНК во время обработки концов трансгена и интеграции – актуальная тема. Ранее мы разработали генетическую систему на основе баркодов для анализа рекомбинации трансгенов после пронукулярной микроинъекции у мышей. При этом подходе плазмидная библиотека линейризуется ферментом рестрикции или комплексом Cas9:gRNA на участке между парой баркодов. Пул баркодированных молекул вводится внутрь пронуклеуса, что приводит к образованию многокопийных конкатемеров. В представленной статье мы сравнили эффекты разрезания Cas9 в условиях *in vivo* (эксперимент RNP+) или для трансгенов, линейризованных *in vitro* и очищенных через агарозный гель (эксперимент RNP–). В эксперименте RNP+ было обнаружено два трансгенных однокопийных эмбриона. В экспе-

рименте RNP– было идентифицировано шесть положительных эмбрионов, четыре из них имели низкокопийные конкатемеры. Анализ баркодов методом NGS показал, что 53 % баркодированных концов поменяли свои исходные пары баркодов, что является признаком гомологичной рекомбинации между концами трансгенов. Из 20 слияний «трансген-трансген» 11 не имели мутаций и, по-видимому, были получены путем повторного лигирования тупых концов, полученных с помощью Cas9. Большинство мутировавших соединений содержало асимметричные делеции 2–4 нуклеотидов из-за специфического тримминга концов Cas9. Эти данные указывают на то, что связанная с Cas9 ДНК создает препятствия для конкатенации. В то же время свободные концы ДНК соединяются таким же образом, как и концы, обработанные рестриктазами, но с характерной асимметрией. Будущие эксперименты с пронуклеарными микроинъекциями CRISPR/Cas помогут понять, как CRISPR-нуклеазы влияют на репарацию ДНК.

Ключевые слова: CRISPR/Cas9; пронуклеарная микроинъекция; ДНК-баркодирование; трансгенные животные; репарация двуцепочечных разрывов ДНК; конкатемер; гомологичная рекомбинация; негомологичное соединение концов; эмбрионы мыши.

Introduction

Pronuclear microinjection is a common method for making transgenic animals. At present, CRISPR/Cas nucleases are successfully employed in pronuclear microinjections to produce site-specific knock-ins and mutations. The introduction of site-specific double-strand breaks (DSBs) has been demonstrated to enhance the frequency of editing when single-stranded oligonucleotides or long double-stranded donors are incorporated. Moreover, Cas9 can be utilized to cleave circular donor vectors *in vivo*, thereby providing double-stranded ends for repair (Abe et al., 2020). Approaches such as minicircles (Danner et al., 2021) or HITI/PiCh (Sakuma et al., 2016; Suzuki et al., 2016) are based on non-homologous or microhomology-mediated end-joining (NHEJ, MMEJ) and have been reported to have a high editing frequency due to the fact that NHEJ/MMEJ is active throughout the cell cycle. In contrast to the predominantly utilized clean ends in classical transgenesis, Cas9-cleaved DNA engages with the cellular DNA repair apparatus in a distinctive manner. For example, the binding of the Cas9 complex to the target site results in the local denaturation of the DNA duplex and trimming of the cleaved ends, which then form single-stranded overhangs of varying lengths (Stephenson et al., 2018). Following DNA cleavage, Cas9 remains bound to the cut site for several hours, tethering the two ends together (Richardson et al., 2016). This prevents immediate end processing by DNA repair factors and necessitates the removal of Cas9 by cell factors to seal the break (Clarke et al., 2018; Reginato et al., 2024).

The majority of our knowledge regarding the Cas9 mode of action is derived from cell culture assays utilizing site-specific reporters (Schimmel et al., 2017) and next-generation indel sequencing (NGS) (Schimmel et al., 2017; Taheri-Ghahfarokhi et al., 2018). We elected to investigate the manner in which Cas9-cleaved ends engage with DSB repair systems, employing a novel assay based on molecular barcoding and concatenation in pronuclear microinjection. The concatenation of the injected transgenes into multicopy arrays is a well-documented aspect of transgenesis, occurring with high frequency in pronuclear microinjection. Molecular barcode evidence indicates that concatenation is facilitated by the combined action of homologous recombination (HR) (head-to-tail recombination) and NHEJ (random ligation and initial copy circularization) (Sмирнов et al., 2020). During the process of concatenation, the repair pathways leave specific signatures at the junctions between the transgenes and in the patterns of connection between the barcode and the transgenes. Therefore, the levels

of barcode recombination and terminal truncations may serve as indicators of the accessibility of DNA ends to DNA repair machinery in the presence of the Cas9 nuclease.

To ascertain the impact of Cas9-generated ends on concatenation, a barcoded plasmid library was subjected to Cas9 digestion, either *in vivo* or *in vitro*, and the resulting barcode information was obtained from transgenic embryos through Illumina NGS. Subsequently, an analysis was conducted on the data to evaluate barcode recombination (HR activity), estimate the average transgene copy number (CN), and assess end truncations and mutations at transgene-transgene junction sequences (MMEJ/NHEJ activity).

Materials and methods

Animal ethics statement. The animal housing was implemented with the assistance of the Center for Laboratory Animal Genetic Resources at the Institute of Cytology and Genetics SB RAS, with support from the Ministry of Science and Higher Education of the Russian Federation (unique identifier of the project: RFMEFI62119X0023). The animals were maintained in a standard environment, with a temperature of 24 °C, relative air humidity of 40–50 %, and a 14-hour light/10-hour dark light cycle. The animals had access to food and water at all times. At the conclusion of the experiments, the remaining animals were euthanized via CO₂ asphyxiation. All procedures and technical manipulations involving animals were conducted in accordance with the European Communities Council Directive of November 24, 1986 (86/609/EEC) and approved by the Bioethical Committee at the Institute of Cytology and Genetics SB RAS (Permission No. 45 from November 16, 2018).

DNA cloning. The pCAGGS-mCherry plasmid was a gift from Phil Sharp (Gurtan et al., 2012) (Addgene plasmid #41583). This vector was utilized for the insertion of paired barcodes. The barcoded plasmid library was constructed by assembling a linearized backbone, two barcoded oligonucleotides, and an AAVS1 PCR fragment using the NEBuilder® HiFi DNA Assembly mix (NEB). The sequences of the left and right barcodes were as follows: NNCGANNACTNNATGNNACGNNCTGNNTCANN (left) and NNCTCNGAA NNCGTNNCTANNTCGNNGTANN (right).

To synthesize gRNA against the AAVS1 site, a PCR fragment was amplified from the gRNA_AAVS1-T2 plasmid using a primer with a T7 overhang and treated with the MEGascript™ T7 Transcription Kit (Thermo Fisher, USA). The gRNA_AAVS1-T2 plasmid was a gift from George Church

(Addgene plasmid #41818) (Mali et al., 2013). The resulting gRNA was purified using the RNA Clean & Concentrator™ Kit (Zymo Research, USA).

To assess the *in vitro* activity of Cas9, 1 µg of a selected clone from the plasmid library was incubated with 50–250 nM of the Cas9 protein (Biolabmix, Russia) and an equimolar amount of gRNA in an incubation buffer (20 mM HEPES pH 7.5, 125 mM KCl, 1 mM EDTA, 1 mM DTT, 6 mM MgCl₂, 7 % glycerol). The reaction was incubated at 37 °C for 30 minutes, heat-inactivated at 65 °C, and visualized on a 1 % agarose gel.

Animals. Four-week-old female F1 C57BL/6J×DBA2/J (B6D2F1) mice were utilized for superovulation and oocyte collection due to their high hybrid vigor, while three-month-old male C57BL/6J mice were employed for sperm collection. Two-month-old pseudopregnant female CD-1 mice were utilized for embryo transfer into the oviducts, with isoflurane anesthesia. *In vitro* fertilization and embryo collection were conducted in accordance with the established protocols (Takeo et al., 2008; Takeo, Nakagata, 2010, 2011).

Pronuclear microinjection. Two distinct versions of the DNA injection mixture were prepared and delivered into the pronuclei. The RNP+ mix consisted of 1.65 µM Cas9 (Alt-R® S.p. Cas9 Nuclease V3, IDT, USA), 1.65 µM gRNA against AAVS1, and 8 ng/µl plasmid library diluted in TE buffer (0.01 M Tris–HCl, 0.25 mM EDTA, pH 7.4) (IDT). For the RNP-negative mix (RNP–), plasmid DNA was digested with the same RNP *in vitro* for one hour, gel-purified to remove Cas9, and diluted to 8 ng/µl with TE buffer. A quantity of DNA equivalent to one picoliter, comprising approximately 1,400 copies, was injected into the male pronuclei of *in vitro* fertilized zygotes. Subsequently, the microinjected zygotes were cultured in droplets of mHTF medium covered with mineral oil at 37 °C under 5 % CO₂ in air for approximately 20 hours until the two-cell embryo stage was reached. Two-cell stage embryos were transferred into the oviducts of pseudopregnant CD-1 females at 0.5 days *post-coitum*. Genomic DNA from 13–14-day embryos was subjected to conventional PCR with Q5 polymerase (NEB, USA) and a set of primers spanning the region around transgene-transgene junctions (Tables S1–S2)¹.

Illumina sequencing and data analysis. Two distinct types of PCR products were amplified from the genomes of embryos #31, #34, #35, and #37. To prepare the inverse PCR NGS library, genomic DNA from transgenic embryos was digested overnight with KpnI-HF (NEB), which cuts twice within the transgene-transgene junction. The digested DNA was purified using AMPure XP magnetic beads (Beckman Coulter, USA), and 300 ng of the digested DNA was ligated overnight at 16 °C in a large reaction volume (100 µl) to facilitate self-ligation of transgene copies. The terminal barcodes of the self-ligated DNA fragments were amplified via PCR using Q5 polymerase and primers that spanned the barcode pairs (Table S2). The following PCR program was employed: the temperature was set to 98 °C for 30 seconds, followed by 35 cycles of 98 °C for 15 seconds, 62 °C for 30 seconds, and 72 °C for 30 seconds, with a final extension at 72 °C for 2 minutes. To remove any remaining PCR fragments resulting from undigested DNA,

those of an appropriate size (216 bp) were excised from the agarose gel.

To prepare the PCR library corresponding to transgene-transgene junctions, the internal junctions were amplified with primers, one of which was in close proximity to the junction and the other of which was situated adjacent to the barcode (Tables S1 and S2). The PCR products were generated for both orientations and purified using AMPure XP magnetic beads. The PCR products were prepared with the KAPA HyperPrep Kit (Roche, Switzerland) using two amplification cycles, pooled together, and sequenced on the Illumina HiSeq 2500 platform (Illumina, USA). The quality of the libraries was evaluated using an Agilent 2100 Bioanalyzer (Agilent, USA) and the Qubit dsDNA HS assay Kit (Life Technologies, USA).

The processing of NGS data (Table S3) was conducted in accordance with the methodology described in our previous report (Smirnov et al., 2020) with minor modifications. The analysis of NGS data comprises the following steps. Initially, the data were demultiplexed according to the samples. Metrics based on the Levenshtein distance were employed to identify primer sequences at the 3' and 5' ends of each read pair. Subsequently, barcodes were sought within read pairs exhibiting regular patterns (i.e., head barcode NNCGANN ACTNNATGNNACGNNCTGNNTCANN and tail barcode NNCTCNGAANNCGTNNCTANNTCGNNGTANN). Furthermore, the number and percentage of read pairs sharing identical barcodes and identical barcode pairs were calculated. The resulting set of barcode pairs was subjected to a filtration process to produce the final set of pairs. Finally, the barcode pairs were visualized using the Network module of the vis.js framework (<http://visjs.org/>). All of the aforementioned computations were conducted on a high-throughput computational cluster at Novosibirsk State University. The following Python modules were utilized: the Biopython v1.79 module (Cock et al., 2009, 2010) was utilized. Next-generation sequencing data processing (re v2.2.1) utilizes regular expressions (Levenshtein v0.12.1). The Levenshtein distance was calculated using the Pandas v1.3.3 module for processing table data, while the NetworkX 2.6.2 module was employed to determine the number of pairs.

Droplet digital PCR. The Droplet Digital PCR (ddPCR) method was employed to ascertain the transgene copy number in the embryos. The procedure was conducted using ddPCR Supermix for Probes (no dUTP) and QX100 ddPCR Systems (Bio-Rad, USA), in accordance with the instructions provided by the manufacturer. First, 900 ng of genomic DNA was digested overnight with 20 U of BamHI-HF (NEB) in CutSmart buffer. One microliter of genomic DNA (30 ng) was added to the ddPCR mix (1× ddPCR Supermix, 900 nM primers, 250 nM probes), and the reactions were carried out according to the following program: the reaction was initiated at 95 °C for 10 minutes, followed by 41 cycles of 94 °C for 10 seconds and 61 °C for 1 minute. The final step was conducted at 98 °C for 7 minutes and 20 °C for 30 minutes. All the steps were conducted with a ramp rate of 2 °C/s. ddPCR was performed in two independent technical replicates. The sequences of the primers and probes are provided in Table S4. The data were subsequently analyzed using QuantaSoft (Bio-Rad). The threshold for both genes was set at 5,000.

¹ Supplementary Tables S1–S4 and Figures S1, S2 are available at: <https://vavilov-j-icg.ru/download/pict-2025-29/appx1.pdf>

Results

Generation of transgenic embryos by pronuclear microinjection

For the experiments, we used a barcoded plasmid library, which was based on the pCAGGS-mCherry vector (Fig. 1a, middle section). In brief, the vector contains a pair of 32-base pair barcodes (14 random positions) separated by filler DNA. In the present experiments, a fragment of the human AAVS1

site was utilized as filler DNA due to its well-characterized gRNA sites (Chr19:55115511-55115987, 477 bp) (Mali et al., 2013; Maggio et al., 2014). One of the gRNA sites (5'-GGGGCCACTAGGGACAGGAT-3') was employed as a Cas9 target to linearize plasmid DNA and generate transgenes with barcoded ends (Fig. 1a). *In vitro* tests with the Cas9 protein demonstrated that the synthesized gRNA exhibited robust performance against the plasmid (Fig. 2). To analyze the plasmid library, Illumina paired-end sequencing of the

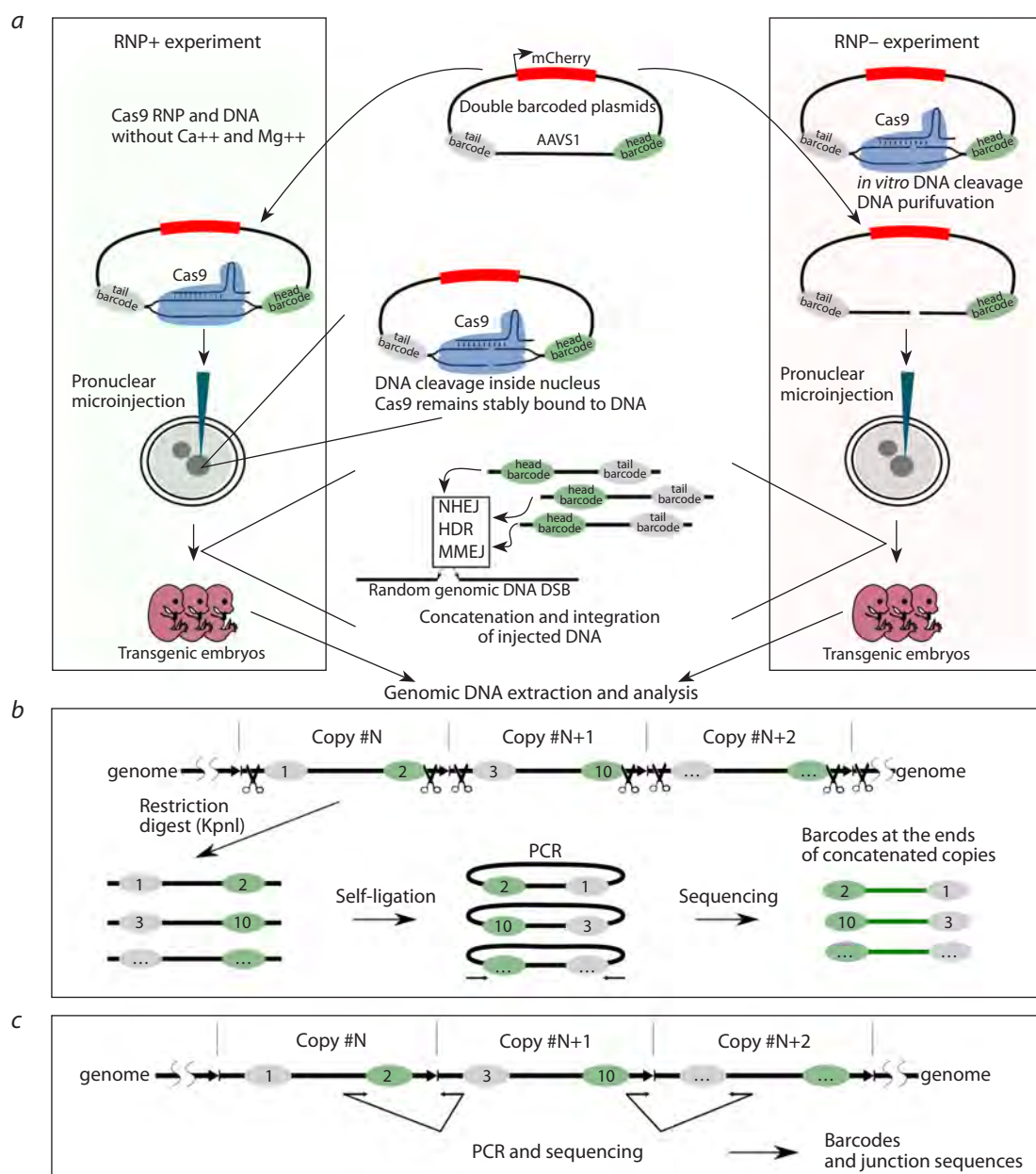


Fig. 1. Outline of the experimental approach.

a – the barcoded plasmid library was prepared for microinjections. A plasmid library comprising a pair of 32 bp barcodes separated by the human AAVS1 fragment was employed to prepare two experimental setups. In the RNP+ experiments, Cas9 and gRNA targeting AAVS1 were co-injected with the plasmid library into pronuclei. *In vivo*, plasmid linearization occurs within the pronucleus. The right-hand illustration depicts the second experimental setup. The plasmid library is digested with RNP *in vitro*, purified via gel electrophoresis, and injected as a purified, linear DNA construct (RNP-). Following the injection, the linear molecules undergo recombination, forming either concatemers or single copies, which are then detected through PCR genotyping; **b** – the barcodes present in the genomic DNA are subjected to analysis through the process of inverse PCR. The concatenated copies are then separated by KpnI (scissors) and re-ligated. PCR products from the newly formed junctions are then subjected to analysis by NGS; **c** – internal junctions between copies are subjected to PCR amplification and subsequent sequencing.

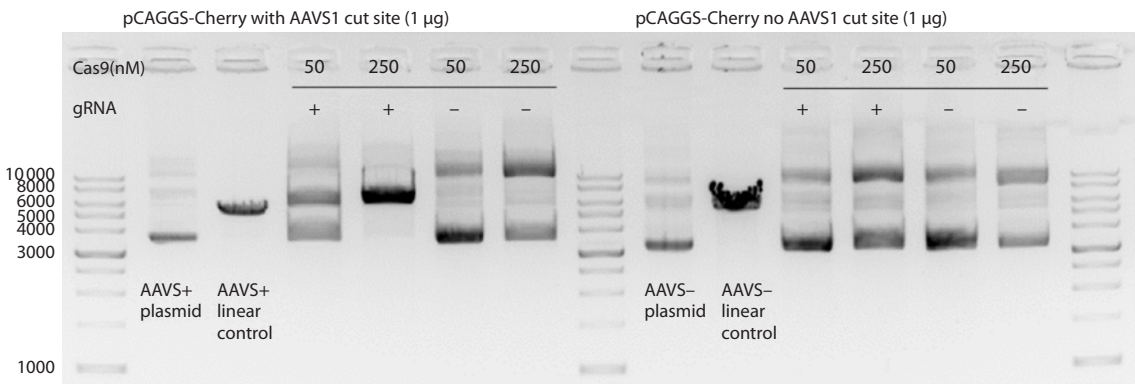


Fig. 2. Preparation of barcoded transgenes for microinjection. *In vitro* cleavage of barcoded plasmid using Cas9 RNP. The image on the left depicts a barcoded clone with an AAVS1 site, while the image on the right illustrates the original backbone without an AAVS1 site. The incubations were maintained at 37 °C for a period of 30 minutes.

barcodes was performed. The plasmid library was found to consist of 102,685 unique barcode pairs. In a typical pronuclear microinjection experiment, approximately one thousand molecules are introduced into the pronucleus. The diversity of the library, comprising approximately 100,000 unique sequences, allows for the injection of a vast number of molecules into a single pronucleus, with minimal likelihood of identical sequences being present.

To understand how the presence of Cas9 at the ends of molecules affects the DSB repair mechanisms involved in concatenation and integration of exogenous DNA, we conducted two experiments on DNA microinjection into pronuclei (Fig. 1a). In one variant, the original circular plasmid library was combined with the Cas9 protein and gRNA, which formed a ribonuclear protein complex (RNP) in standard TE buffer (Fig. 1a, left section, “RNP+”). The absence of Mg²⁺ ions in

the buffer precludes Cas9 endonuclease activity prior to pronuclear microinjection. The concentration of Cas9/gRNA RNP was increased to 1.65 µM, which is above the concentration typically used, in order to ensure robust plasmid linearization. In the second experiment, the DNA was subjected to *in vitro* digestion using Cas9 (Fig. 1a, right section). The linearized DNA was purified via gel electrophoresis and subsequently used for microinjection (henceforth referred to as “RNP–”). In both experiments, approximately 1,400 copies were injected per zygote.

Genomic DNA was extracted from the embryos at the 13–14-day (E13–14) developmental stage. A total of 27 embryos were collected from the RNP+ experiment, and 13 embryos were collected from the RNP– experiment (Fig. 3a). PCR genotyping with primers for the mCherry gene and transgene-transgene junctions revealed the presence of se-

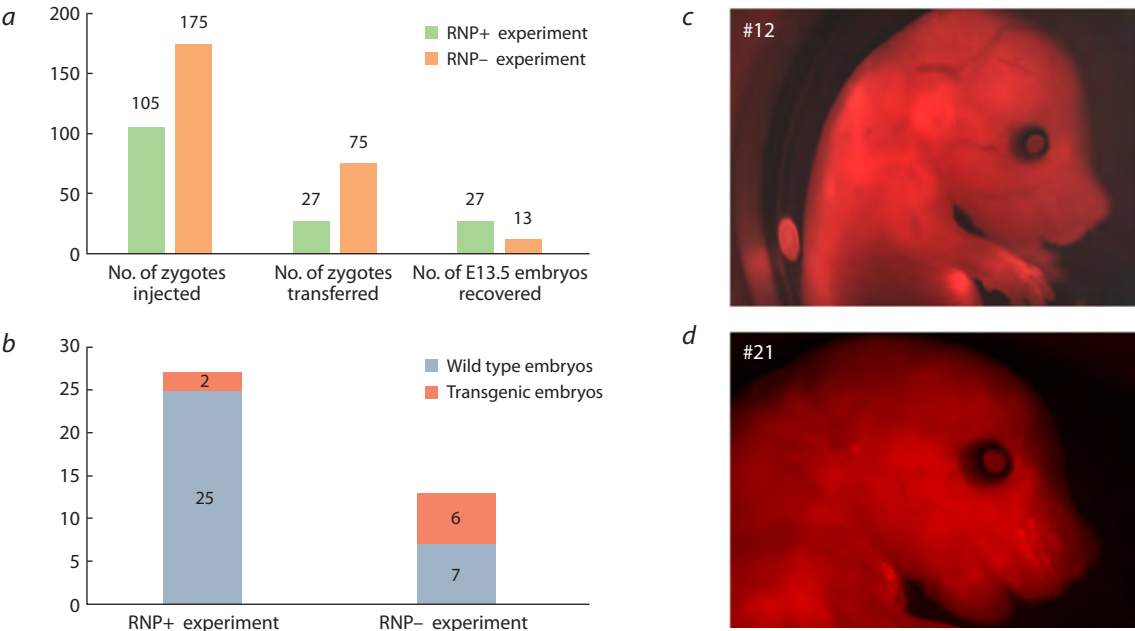


Fig. 3. Screening of the transgenic embryos. *a* – statistical analysis of pronuclear microinjection experiments; *b* – PCR analysis of E13–14 embryos; *c–d* – visualization of mCherry fluorescence through microscopic examination.

veral positive embryos, with 2/27 for RNP+ and 6/13 for RNP– (Fig. 3*b*). Embryos derived from the RNP– experiment exhibited positive results for transgene-transgene junctions, a hallmark of concatemers (Fig. S1). Of the transgenic embryos, only two exhibited mCherry fluorescence (Fig. 3*c–d*).

To ascertain the transgene copy number, droplet digital PCR was conducted using a probe directed against the mCherry region within the transgene (Fig. S2). Embryos from the RNP+ experiment exhibited low CN (0.07, 0.37), whereas embryos from the RNP– experiment demonstrated a higher number of copies on average (CN = 0.06, 0.89, 1.87, 2.8, 5.84, 6.38) (Fig. S2) (see realistic estimates below). It should be noted that in relation to ddPCR, CN < 1 or partial CN values indicate tissue mosaicism, whereby the transgene exists only in a portion of the embryo cells. This could reflect delayed DNA integration.

Barcode analysis

Barcodes derived from concatemers in embryos were sequenced using Illumina paired-end reads. Two types of PCR products were prepared for barcode sequencing, following the methodology described in our previous report (Fig. 1*b–c*) (Sмирнов et al., 2020). In summary, barcode pairs at the ends

of concatenated transgenes were amplified via PCR through self-molecular ligation (Fig. 1*b*). This method facilitates the excision and self-ligation of transgene molecules, which enables the reading of terminal barcodes in the copies. The KpnI sites utilized for the fragmentation of concatemers in inverse PCR are situated between barcodes and yield 216 bp PCR fragments subsequent to ligation (Fig. 4*a*). Additionally, internal junctions, defined as nucleotide sequences situated between the copies, were subjected to sequencing (Fig. 1*c*). In total, barcode data were obtained for four multicopy RNP embryos (via inverse PCR and junction PCR) and two single-copy RNP+ embryos (due to the lack of junctions, only inverse PCR was conducted). To investigate recombination of the barcode sequences, a comparison was made between the barcodes from the transgenic embryos and the list of barcodes from the original plasmid library.

First, we employed NGS barcode data to ascertain the actual CN in the embryos. It is established that the ddPCR results typically exhibit an understated CN due to the mosaic nature of transgenic embryos. By employing barcodes as unique copy identifiers, we were able to calculate the actual CN (Fig. 5). The CN data exhibited a strong correlation with the ddPCR data, with a slight tendency towards an increase



Fig. 4. Generation of PCR fragments for the barcode NGS analysis using inverse PCR.

a – the positions of the PCR primers (indicated by arrows) and the KpnI/Cas9 cut sites at the hypothetical transgene-transgene junction are shown. The empty vector is the original pCAGGS backbone devoid of the AAVS1 element; **b** – generation of PCR products for NGS. G: Untreated genomic DNA; R: Genomic DNA treated with KpnI; L: Genomic DNA after re-ligation for inverse PCR. Inverse PCR entails the digestion of genomic DNA with KpnI and subsequent re-ligation to create self-ligated ends. Consequently, the size of the PCR product is diminished from 699 base pairs to 216 base pairs. The intensity of the bands is proportional to the copy number. Legend: H₂O was used as the negative control, E12–E37 were the transgenic embryos, and WT was the wild-type control. The PCR products selected for NGS analysis are indicated by asterisks. The DNA length marker is 100 base pairs in length.

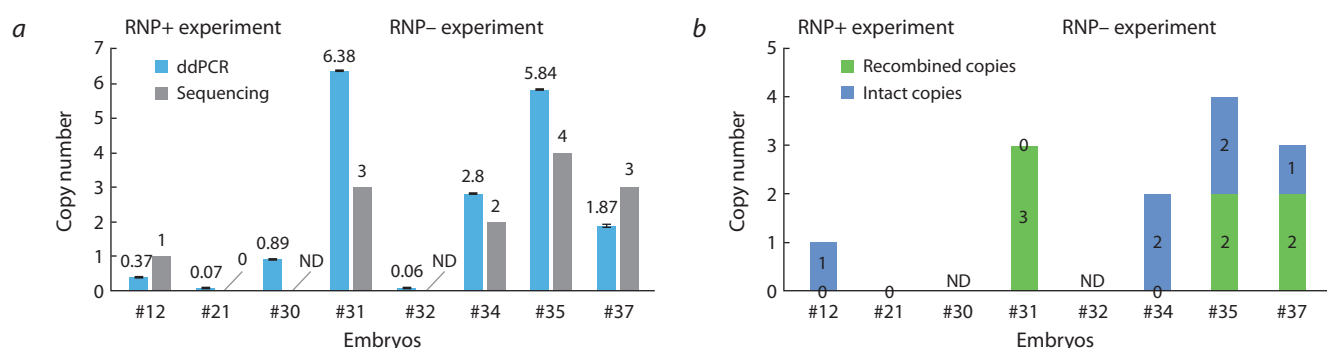


Fig. 5. A comparison of CN measurements derived from ddPCR and NGS barcode analysis.

a – CN estimates derived from ddPCR and NGS in transgenic embryos (#12–37). ND – no data available. A CN value of less than 1 is indicative of tissue mosaicism; **b** – estimation of the proportion of copies with barcode recombination (green).

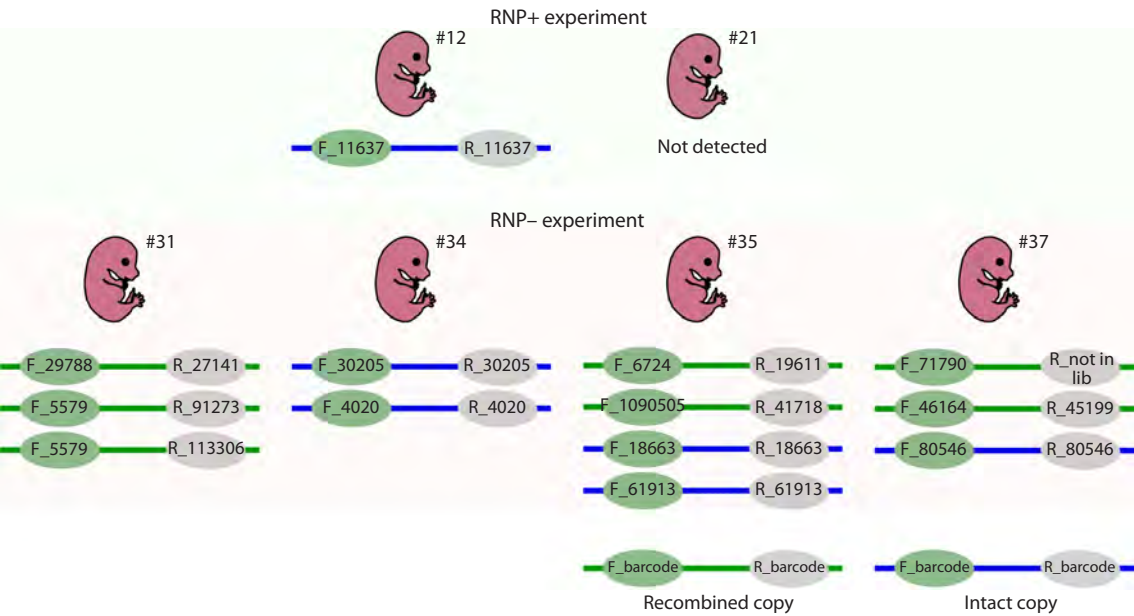


Fig. 6. The barcodes observed in four multicopy RNP– embryos (#31, #34, #35, #37) and one RNP+ embryo (#12). Barcoded transgenes are represented as colored lines with barcode numbers at the ends. Intact transgene copies are indicated in blue (representing the original barcode pairs from the library), while copies with recombined barcodes are shown in green. The top panel depicts embryos from the RNP+ experiment. The bottom panel depicts four RNP– embryos with their corresponding transgene copies. “R_not in lib” indicates that the barcode was not identified within the plasmid library.

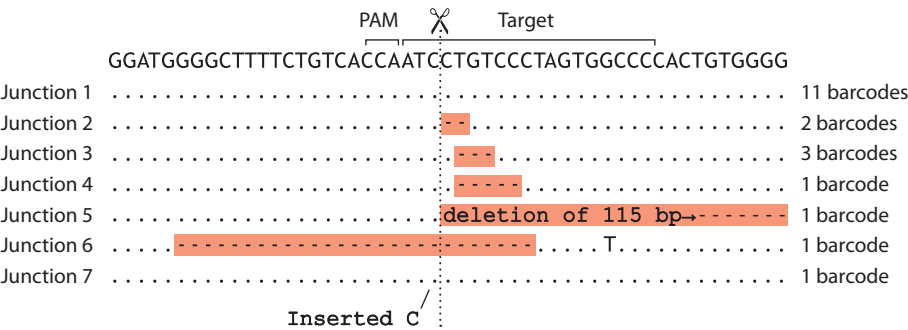


Fig. 7. Nucleotide sequences at the transgene-transgene junctions.

in CN. Embryos with a very low CN (0.06–0.89) as estimated by ddPCR, such as those designated #12/#21 from the RNP+ experiment or #30/#32 from the RNP– experiment, exhibited no transgene-transgene junctions, indicating that they possess a single or truncated copy. It is noteworthy that the truncation of transgene ends during concatenation has the potential to delete the site of a barcode and thereby obscure the calculation of a copy. Nevertheless, in our particular case, an examination of the PCR junctions did not reveal any evidence of terminal truncations, in addition to the presence of a typical transgene-transgene junction (Fig. S1b).

As anticipated, a substantial proportion (53 %) of barcode pairs in concatemers were observed to have undergone shuffling (Fig. 5b and 6). These barcode pairs were not initially present in the plasmid library and appear to have arisen from extrachromosomal end recombination between transgene copies (Smirnov et al., 2020). One barcode was observed to be linked with two distinct partners (Fig. 6, Embryo #31, F_5579), indicating that a single donor molecule was re-

plicated on at least two occasions during the recombination process (Smirnov et al., 2020). This evidence confirms that transgene copies prepared by *in vitro* Cas9 cleavage engage in HR pathways during concatenation, in contrast to the copies from the RNP+ experiment. An internal junction analysis revealed that 11 out of the 20 junctions attributed to specific barcodes did not contain any mutations (Fig. 7). The mutated junction sequences were found to comprise a variety of deletions. Two to five base pair deletions, a large 115 base pair end truncation, and a two-sided 27 base pair deletion (Fig. 7). It should be noted that the DNA ends in question were produced by Cas9 cleavage *in vitro* and were likely trimmed by Cas9 during the incubation period, prior to interaction with the DNA repair machinery.

Discussion

The improvement of efficiency in Cas9-assisted transgenesis will have a profound impact on experiments with a low baseline success rate, such as those involving > 2 kb donor

integration for humanization or genome modifications of farm animals with long reproductive cycles and a lack of scaling. The ability to leverage Cas9 activity hinges on a comprehensive understanding of its interactions with cellular DNA repair pathways. A number of studies have demonstrated that Cas9 forms an unusually stable nuclease-substrate complex and remains bound to the DSB it generates (Jinek et al., 2014; Nishimasu et al., 2014; Stephenson et al., 2018). The precise manner by which cellular repair systems recognize a DSB and the extent to which the Cas9 protein at the DSB influences the choice of repair pathway remain largely unknown. It has been demonstrated that the Cas9 complex can be dislodged from a double-stranded break by RNA polymerase, but only if the single-guide RNA (sgRNA) of Cas9 is annealed to the DNA strand that is used as the template by the RNA polymerase (Clarke et al., 2018).

New evidence indicates that Cas9 should be removed by HLTf to initiate end processing (Reginato et al., 2024), as post-cleavage Cas9 complexes have been shown to impede DNA from MRE11 (Maltseva et al., 2023; Reginato et al., 2024). Concatenation reporter used in the present study has some unique properties. It could be applied in pronuclear microinjection and allows to inspect many end-joining events at once. We utilized the reporter in combination with Cas9 to study three parameters: transgene CN (DNA end accessibility), barcode recombination (DNA end participation in HR), and internal junctions (NHEJ/MMEJ DNA end processing).

The evaluation of the transgene CN by ddPCR and NGS indicated the absence of concatemers in the RNP+ experiments (E12, E21). Two potential explanations for the absence of concatemers in RNP+ can be postulated. Firstly, it is possible that Cas9 may block cleaved ends from processing by DNA repair factors, thereby preventing end resection and concatenation, which represents the primary mechanism of CN increase (Smirnov et al., 2020). Secondly, it may be that Cas9 is less active in cutting plasmid libraries, which would result in a reduction in the number of linear copies that integrate rarely and/or late in the zygote division. However, the RNP concentration used for pronuclear microinjection was higher than the average recommended by D.W. Harms et al. (2014). This concentration should have been sufficient to linearize the library (1400 DNA copies), as most genome-editing experiments using a lesser amount of RNP generally show high rates of genomic modifications.

Given the overall low copy number of the generated concatemers, the barcode recombination assay was not particularly informative. It is noteworthy that some embryos that tested positive by PCR or mCherry fluorescence did not produce barcodes from the NGS analysis. This indicates that the integrated copy/copies are damaged. While the mCherry cassette is located in the middle of the transgene, the ends of the transgenes are more susceptible to truncations that result in the loss of one or both barcodes. Barcode switching was detected (Fig. 6) and a barcode copying event, which also serves as an indicator of HR activity, was observed. The result of barcode switching (53 %) was not significantly different from that of the previous report (80 % recombined copies) (Smirnov et al., 2020). In conclusion, the presented evidence suggests that transgene copies prepared by *in vitro* Cas9

cleavage may engage in HR pathways during concatenation, in contrast to the copies from the RNP+ experiment.

The analysis of internal junctions revealed a typical assortment of Cas9-generated deletions, with occasional instances of nucleotide insertions (Schimmel et al., 2017). It is established that Cas9 produces heterogeneous ends due to endonucleolytic degradation of the DNA by endonuclease domains, with higher activity towards the PAM distal fragment (Stephenson et al., 2018). This may explain the observed asymmetry in deletions at the junctions (Fig. 7, with the PAM distal fragment to the right). It is noteworthy that intact junctions were observed with considerable frequency among the variants. It was not possible to ascertain with certainty whether the junctions originated from undigested vector or *de novo* ligation of blunt DNA ends. However, given that *in vitro* digested DNA was gel-purified (Fig. 2, linear fragment) and should not contain a significant proportion of undigested circular form, it was assumed that the latter was the case. These junction signatures differ from those typically generated by restriction enzymes in standard pronuclear microinjection experiments (Rohan et al., 1990; Dai et al., 2010). For example, in our previous study, we selected the BsmBI enzyme to linearize the barcoded vector and create 4 nt incompatible protruding 5'-ends. Following concatenation, transgene molecules exhibited a loss of approximately 5–10 nucleotides at the junction sites, with no discernible asymmetry (Smirnov et al., 2020).

In comparison to the barcoded library previously described (Smirnov et al., 2020), the pCAGGS-Cherry-based reporter exhibited a tenfold increase in barcode diversity and an extended barcode length (32 bp vs 17 bp). This allows for unambiguous interpretation of complex barcode recombination patterns, including the copying of a single barcode to multiple copies (Fig. 6). Nevertheless, in theory, the relatively low overall CN compared to other experiments (Smirnov et al., 2020) may indicate that the linear copies are impeded in their end recombination. In this instance, the absence of concatemers may be attributed to the augmented barcode length, which impedes homology search and head-to-tail end joining.

Conclusion

In conclusion, our findings demonstrate that the DNA ends generated by Cas9 *in vivo* undergo a distinct processing pathway compared to the “clean” ends. The microinjection of the barcoded library in combination with CRISPR endonucleases represents a fruitful assay that can be augmented further. This could be achieved, for example, by chemical modifications of the DNA ends or by co-injection of the NHEJ/HR factors. It would be beneficial to gain a deeper understanding of how Cas9 affects end recombination during concatenation, with the aim of preventing unwanted concatenation of donor molecules or stimulating end processing at the genomic site in the future.

References

- Abe T., Inoue K., Furuta Y., Kiyonari H. Pronuclear microinjection during S-phase increases the efficiency of CRISPR-Cas9-assisted knockin of large DNA donors in mouse zygotes. *Cell Rep.* 2020; 31(7):107653. doi 10.1016/j.celrep.2020.107653
- Clarke R., Heler R., MacDougall M.S., Yeo N.C., Chavez A., Regan M., Hanakahi L., Church G.M., Marraffini L.A., Merrill B.J. Enhanced bacterial immunity and mammalian genome editing via

- RNA-polymerase-mediated dislodging of Cas9 from double-strand DNA breaks. *Mol Cell*. 2018;71(1):42-55.e8. doi 10.1016/j.molcel.2018.06.005
- Cock P.J.A., Antao T., Chang J.T., Chapman B.A., Cox C.J., Dalke A., Friedberg I., Hamelryck T., Kauff F., Wilczynski B., De Hoon M.J.L. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-1423. doi 10.1093/bioinformatics/btp163
- Cock P.J.A., Fields C.J., Goto N., Heuer M.L., Rice P.M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2010;38(6):1767-1771. doi 10.1093/nar/gkp1137
- Dai J., Cui X., Zhu Z., Hu W. Non-homologous end joining plays a key role in transgene concatemer formation in transgenic zebrafish embryos. *Int J Biol Sci*. 2010;6(7):756-768. doi 10.7150/ijbs.6.756
- Danner E., Lebedin M., De La Rosa K., Kühn R. A homology independent sequence replacement strategy in human cells using a CRISPR nuclease. *Open Biol*. 2021;11(1):200283. doi 10.1098/rsob.200283
- Gurtan A.M., Lu V., Bhutkar A., Sharp P.A. *In vivo* structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. *RNA*. 2012;18(6):1116-1122. doi 10.1261/rna.032680.112
- Harms D.W., Quadros R.M., Seruggia D., Ohtsuka M., Takahashi G., Montoliu L., Gurumurthy C.B. Mouse genome editing using the CRISPR/Cas system. *Curr Protoc Hum Genet*. 2014;83:15.7.1-15.7.27. doi 10.1002/0471142905.hg1507s83
- Jinek M., Jiang F., Taylor D.W., Sternberg S.H., Kaya E., Ma E., Anders C., Hauer M., Zhou K., Lin S., Kaplan M., Iavarone A.T., Charpentier E., Nogales E., Doudna J.A. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. 2014;343(6176):1247997. doi 10.1126/science.1247997
- Maggio I., Holkers M., Liu J., Janssen J.M., Chen X., Gonçalves M.A.F.V. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. *Sci Rep*. 2014;4(1):5105. doi 10.1038/srep05105
- Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M. RNA-guided human genome engineering via Cas9. *Science*. 2013;339(6121):823-826. doi 10.1126/science.1232033
- Maltseva E.A., Vasil'eva I.A., Moor N.A., Kim D.V., Dyrkheeva N.S., Kutuzov M.M., Vokhtantsev I.P., Kulishova L.M., Zharkov D.O., Lavrik O.I. Cas9 is mostly orthogonal to human systems of DNA break sensing and repair. *PLoS One*. 2023;18(11):e0294683. doi 10.1371/journal.pone.0294683
- Nishimasu H., Ran F.A., Hsu P.D., Konermann S., Shehata S.I., Dohmae N., Ishitani R., Zhang F., Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. 2014;156(5):935-949. doi 10.1016/j.cell.2014.02.001
- Reginato G., Dello Stritto M.R., Wang Y., Hao J., Pavani R., Schmitz M., Halder S., Morin V., Cannavo E., Ceppi I., Braunschier S., Acharya A., Ropars V., Charbonnier J.-B., Jinek M., Nussenzweig A., Ha T., Cejka P. HLTf disrupts Cas9-DNA post-cleavage complexes to allow DNA break processing. *Nat Commun*. 2024;15(1):5789. doi 10.1038/s41467-024-50080-y
- Richardson C.D., Ray G.J., DeWitt M.A., Curie G.L., Corn J.E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol*. 2016;34(3):339-344. doi 10.1038/nbt.3481
- Rohan R.M., King D., Frels W.I. Direct sequencing of PCR-amplified junction fragments from tandemly repeated transgenes. *Nucleic Acids Res*. 1990;18(20):6089-6095. doi 10.1093/nar/18.20.6089
- Sakuma T., Nakade S., Sakane Y., Suzuki K.-I.T., Yamamoto T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat Protoc*. 2016;11(1):118-133. doi 10.1038/nprot.2015.140
- Schimmel J., Kool H., Van Schendel R., Tijsterman M. Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *EMBO J*. 2017;36(24):3634-3649. doi 10.15252/embj.201796948
- Smirnov A., Fishman V., Yunusova A., Korablev A., Serova I., Skryabin B.V., Rozhdestvensky T.S., Battulin N. DNA barcoding reveals that injected transgenes are predominantly processed by homologous recombination in mouse zygote. *Nucleic Acids Res*. 2020;48(2):719-735. doi 10.1093/nar/gkz1085
- Stephenson A.A., Raper A.T., Suo Z. Bidirectional degradation of DNA cleavage products catalyzed by CRISPR/Cas9. *J Am Chem Soc*. 2018;140(10):3743-3750. doi 10.1021/jacs.7b13050
- Suzuki K., Tsunekawa Y., Hernandez-Benitez R., Wu J., Zhu J., Kim E.J., Hatanaka F., Yamamoto M., Araoka T., Li Z., Kurita M., Hishida T., Li M., Aizawa E., Guo S., Chen S., Goebel A., Soligalla R.D., Qu J., Jiang T., Fu X., Jafari M., Esteban C.R., Berggren W.T., Lajara J., Nuñez-Delgado E., Guillen P., Campistol J.M., Matsuzaki F., Liu G.-H., Magistretti P., Zhang K., Callaway E.M., Zhang K., Belmonte J.C.I. *In vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*. 2016;540(7631):144-149. doi 10.1038/nature20565
- Taheri-Ghahfarokhi A., Taylor B.J.M., Nitsch R., Lundin A., Cavallo A.-L., Madeyski-Bengtson K., Karlsson F., Clausen M., Hicks R., Mayr L.M., Bohlooly-Y.M., Maresca M. Decoding non-random mutational signatures at Cas9 targeted sites. *Nucleic Acids Res*. 2018;46(16):8417-8434. doi 10.1093/nar/gky653
- Takeo T., Nakagata N. Combination medium of cryoprotective agents containing L-glutamine and methyl- β -cyclodextrin in a preincubation medium yields a high fertilization rate for cryopreserved C57BL/6J mouse sperm. *Lab Anim*. 2010;44(2):132-137. doi 10.1258/la.2009.009074
- Takeo T., Nakagata N. Reduced glutathione enhances fertility of frozen/thawed C57BL/6 mouse sperm after exposure to methyl-beta-cyclodextrin. *Biol Reprod*. 2011;85(5):1066-1072. doi 10.1095/biolreprod.111.092536
- Takeo T., Hoshii T., Kondo Y., Toyodome H., Arima H., Yamamura K., Irie T., Nakagata N. Methyl-beta-cyclodextrin improves fertilizing ability of C57BL/6 mouse sperm after freezing and thawing by facilitating cholesterol efflux from the cells. *Biol Reprod*. 2008;78(3):546-551. doi 10.1095/biolreprod.107.065359

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


Received October 8, 2024. Revised November 21, 2024. Accepted December 10, 2024.

doi 10.18699/vjgb-25-05

Interrelations between the main seed quality characteristics of narrowleaf lupine from the VIR collection

T.V. Shelenga , A.V. Salikova , V.S. Popov  , G.P. Egorova , L.L. Malyshev , M.A. Vishnyakova 

Federal Research Centre the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia

 popovitaly@yandex.ru

Abstract. The widespread use of narrowleaf lupine (NLL, *Lupinus angustifolius* L.) as a feed and food crop requires source material for breeding cultivars with high-quality seeds. The priority criterion for attributing NLL cultivars to the feed or food category is the content of alkaloids. At the same time, equally important seed quality indicators are the protein and oil content, as well as moisture content, which determines the possibility of long-term storage of seeds. For the first time in Russian lupine science, an attempt was made to study the relationships between all the listed characteristics of narrowleaf lupine seeds under the conditions of Northwest Russia (Pushkin town). Sixty-two accessions from the VIR collection were studied in 2019, 2020 and 2022. The range of variability of the studied characteristics was 27.8–37.6 % for protein, 3.9–7.3 % for oil, 1.6–2017.4 mg/100 g of dry matter (D.M.) for alkaloids, and 6.4–7.3 % for moisture. A significant negative correlation between the oil and protein content (–0.33) was observed only in 2019. No significant correlations between the protein and alkaloid content were found in the studied sample. Significant negative relationships were identified between the content of oil and alkaloids only in 2019 and 2020 (–0.38 and –0.27, respectively). In 2022, no correlations were identified. Obviously, the identification of regularities in these correlations requires many years of research taking into account weather conditions. The influence of weather on the concentration of alkaloids in seeds has been proven. The average amount of alkaloids for the sample in 2019 was 504.2 ± 77.7 mg/100 g D.M., 263.7 ± 38.6 mg/100 g D.M. in 2020, and 319.8 ± 51.4 mg/100 g D.M. in 2022. It confirmed the data previously obtained by the authors that the content of alkaloids in seeds increases significantly along with the precipitation deficiency. The temperature regime during this research did not affect this indicator. An increased air temperature contributed to the accumulation of oil, and an increase in precipitation contributed to the accumulation of protein. The most stable indicator independent of environmental conditions was the seed moisture. Accessions with the optimal combination of the main biochemical parameters that determine seed quality have been identified for breeding narrowleaf lupine cultivars in the region in question for feed and food purposes, as well as for green manure.

Key words: narrowleaf lupine; alkaloids; protein; oil; seed moisture; weather conditions.

For citation: Shelenga T.V., Salikova A.V., Popov V.S., Egorova G.P., Malyshev L.L., Vishnyakova M.A. Interrelations between the main seed quality characteristics of narrowleaf lupine from the VIR collection. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025;29(1):35-43. doi 10.18699/vjgb-25-05

Funding. The work was carried out within the framework of the State Assignment to VIR according to the Thematic Plan, Project No. FGEM-2022-0002 "Identifying possibilities in the genetic diversity of leguminous crops to optimize their breeding and diversify uses in various sectors of the national economy".

Взаимосвязи основных признаков качества семян люпина узколистного из коллекции ВИР

Т.В. Шеленга , А.В. Саликова , В.С. Попов  , Г.П. Егорова , Л.Л. Малышев , М.А. Вишнякова 

Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия

 popovitaly@yandex.ru

Аннотация. Широкое использование люпина узколистного (ЛУ) в качестве кормовой и продовольственной культуры возможно при наличии сортов с высоким качеством семян. Приоритетным признаком для отнесения сортов ЛУ к этой категории считается низкое содержание алкалоидов. Однако не менее важные показатели качества семян люпина – содержание белка, масла и влажность. Впервые в отечественном люпиноведении предпринята попытка изучения взаимосвязей между показателями всех перечисленных ингредиентов семян ЛУ в условиях Северо-Запада РФ (г. Пушкин). Шестьдесят два образца из коллекции ВИР были изучены в 2019, 2020 и 2022 гг. Изменчивость содержания белка составила 27.8–37.6 %, масла – 3.9–7.3 %, алкалоидов – 1.6–2017.4 мг/100 г сухо-

го вещества (с.в.), влажности – 6.4–7.3 %. Достоверная отрицательная корреляция между содержанием масла и белка (–0.33) наблюдалась только в 2019 г. Значимых корреляций между содержанием белка и алкалоидов в изученной выборке не обнаружено. Достоверная отрицательная связь между содержанием масла и суммой алкалоидов выявлена только в 2019 и 2020 гг. (–0.38 и –0.27 соответственно). В 2022 г. связь не обнаружена. Очевидно, что для выявления закономерностей, характеризующих эти связи, требуется проведение многолетних экспериментов с учетом погодных условий. Доказано влияние погоды на концентрацию алкалоидов в семенах. Средняя по выборке сумма алкалоидов в 2019 г. составила 504.2 ± 77.7 мг/100 г с.в., в 2020 г. – 263.7 ± 38.6 мг/100 г с.в. и в 2022 г. – 319.8 ± 51.4 мг/100 г с.в. Подтверждены данные, полученные авторами ранее: содержание алкалоидов в семенах значительно возрастает при дефиците осадков. Температурный режим не оказал влияния на этот показатель. Повышенная температура воздуха способствовала накоплению масла, а увеличение осадков – накоплению белка. Влажность семян оказалась наиболее стабильным показателем, не зависящим от условий среды. Выявлены образцы с оптимальным сочетанием основных биохимических показателей, определяющих качество семян, для региональной селекции сортов ЛУ кормового и продовольственного назначения, а также сидеральных сортов.

Ключевые слова: люпин узколистный; алкалоиды; белок; масло; влажность семян; погодные условия.

Introduction

Narrowleaf lupine (NLL) *Lupinus angustifolius* L., a grain legume crop adapted to relatively northern regions, is a vital source of vegetable protein and amino acids, a universal ingredient in feeds for farm animals and a promising component of the human diet. The main limiting factor for the use of NLL for these purposes is the content of quinolizidine alkaloids (QA) in seeds and green matter, which impart bitterness to them. In the 1930s, after the discovery of low-alkaloid mutants (Sengbusch, 1931), the creation of low alkaloid cultivars began.

Both earlier in the USSR and currently in the Russian Federation, these cultivars were and are being created using accessions from the collection of the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR) characterized as sources of valuable traits. Currently, NLL breeding is developing quite actively in this country (Egorova et al., 2017). The State Register of Selection Achievements for 2023 includes 29 cultivars, all of which are domestically bred ones (State Register..., 2023). Breeding is carried out in two directions of cultivar use, i. e. as fodder and as green manure. For example, a low-alkaloid feed cv. ‘Belogorsky 310’ and a high-alkaloid cv. ‘Oligarkh’ for green manure have been created at the Leningrad Research Institute for Agriculture “Belogorka” jointly with the VIR researchers and using the source material from the VIR collection (Egorova et al., 2017).

According to the standards adopted in some European countries and in Australia, the alkaloid content, i. e. alkaloidity of seeds intended for food and feed purposes, should not exceed 0.02 % of their dry weight (D.M.) (Frick et al., 2017). In the Russian Federation, the permissible level of alkaloid content for fodder lupine cultivars is from 0.1 to 0.3 % D.M. (State Standard R 54632-2011), and, according to the existing technical specifications developed at the All-Russian Research Institute of Lupine, it is 0.04 % D.M. for food lupine (Specification No. 9716-004-00668502-2008). In everyday practice, the alkaloid seed content at the level of 0.05 % is considered a borderline value for distinguishing low-alkaloid and high-alkaloid forms (Lee et al., 2007).

Along with the alkaloid content, the quality of NLL seeds is also determined by other metabolites, the main ones being protein and oil. The protein content in the seeds of narrowleaf lupine from the VIR collection was recorded at 34–36 %

(Egorova et al., 2019). The oil content in the NLL collection accessions varies within 6.5–8.4 % (Benken et al., 1993).

High-protein cultivars with low alkaloid content are especially valuable because not only the grain, but also the green matter is eaten by all types of farm animals. Lupine is consumed freshly mown, in the form of compound feeds, silage, haylage, grain haylage, and bran (Kuptsov, Takunov, 2006).

Lupine oil is a functional ingredient and a valuable source of edible fats. The content of oleic, linoleic and linolenic fatty acids, of tocopherols, carotenoids, triglycerides, and triacylglycerols is at a high level. The oil of bitter cultivars contains a large number of phenolic compounds, although it has a lower antioxidant activity compared to that of sweet lupine oil (Siger et al., 2017).

Based on the above, evaluation of a lupine collection only by alkaloid content is not enough for recommending accessions as source material for creating feed and food cultivars. Accessions with a combination of high protein and oil and low alkaloid content are required. An important quality indicator of seeds is also their moisture content, which determines the suitability of seeds for storage and processing (State Standard R 52325-2005; Wang et al., 2001).

To which extent are the optimal values of these features compatible within one genotype, what are the regularity in relationships between them and their variability in the gene pool, and to which extent do they depend on growing conditions? Unfortunately, the answers to these questions in the world scientific literature are fragmentary and ambiguous. Therefore, the objectives of this article were to summarize the results of the biochemical analysis of a set of seed quality features, such as the content of protein, oil, alkaloids, and moisture in a sample of NLL accessions from the VIR collection grown for three years in Pushkin, Pushkinsky District of St. Petersburg (in the Northwest of the Russian Federation), to determine the degree of variability of the studied features, their dependency on weather factors in the region of research, and to identify the source material for breeding for high seed quality.

Materials and methods

Plant materials and cultivation conditions. The object of the study was a sample of 62 NLL accessions from the VIR collection, selected according to the representativeness of

their breeding status and alkaloidity. All the accessions were grown in 2019, 2020 and 2022 in the fields of the “Pushkin and Pavlovsk Laboratories” Scientific and Production Base of VIR (Pushkin, Pushkinsky District of St. Petersburg, 59°42'45.5"N 30°25'05.8"E) according to the methodology adopted for grain legumes (Vishniyakova et al., 2018). The growing area belongs to the Atlantic-continental region of the temperate climate zone.

Weather conditions during the years of research varied greatly (Supplementary Figure S1a, b)¹. The coldest year was 2020, though the temperature in June of this year was slightly higher than the average long-term values. Average monthly air temperatures in June–August 2022 significantly exceeded the long-term average. Uneven precipitation was observed over the years of research. For example, the amount of precipitation in June 2019 was three times less than the average for this month in 2020 and 2022, and almost two times less in August of this year. The total precipitation in July–August 2020 and 2022 exceeded the long-term average. The greatest amount of precipitation over the years of research was recorded in August 2022 (Fig. S1a).

Methods. The alkaloid content was determined on an Agilent 6850 gas chromatograph coupled with an Agilent 5975 mass spectrometer (Agilent Technologies, USA) in extracts obtained by sequentially adding ethyl acetate and an aqueous solution of sodium hydroxide to samples of NLL flour. A caffeine solution was used as an internal standard (Kushnareva et al., 2020; Vishniyakova et al., 2023). The protein, oil and moisture content in the NLL seed flour samples were determined using the developed calibration models by near-infrared spectroscopy (NIRS) on a Matrix-1 IR analyzer (Bruker, Germany) (Popov et al., 2024).

To carry out statistical processing of the results obtained in 2019, 2020 and 2022, the samples of accessions were aligned to 40 units, the accessions were selected randomly. Statistical processing was performed using the Statistica 12.0 software package (StatSoft, Inc. (2019), www.statsoft.com). The statistical processing included calculation of the main descriptive statistics (mean, error of the mean, and coefficient of variation); analysis of variance for assessing the significance of differences between the accessions reproduced in different years with calculation of the least significant difference at $p = 0.05$; calculation of coefficients of rank correlation between the content of alkaloids and other biochemical indicators in different years of plant life to assess the stability of trait manifestation in accessions; factor analysis of the correlation system of biochemical traits.

Results

In addition to the previously obtained results on the alkaloid content in the seeds of the NLL sample studied in 2019–2020, the present study added an assessment of this trait in 2022 (Supplementary Table S1). As we showed earlier (Vishniyakova et al., 2023), this trait is highly weather-dependent, and the results of the third year of research make the picture more objective.

The qualitative composition of QA, characteristic of *L. angustifolius* and determined for the accessions in the studied sample earlier, is stable. The dominant alkaloid lupanine is followed in the descending order by 13-hydroxylupanine, sparteine, angustifoline and isolupanine (Kushnareva et al., 2020). The average long-term values (mg/100 g D.M.) for the entire research period were 252.9 for lupanine, 40.3 for 13-hydroxylupanine, 27.8 for sparteine, 4.6 for angustifoline, 2.9 for isolupanine, and 328.5 for the total alkaloids (Table S1). The variability (CV) of the alkaloid content exceeded 100 % (Table S2).

The average total alkaloids (mg/100 g D.M.) for the sample were maximum in 2019 at 504.2 ± 77.7 , intermediate in 2022 at 319.8 ± 51.4 , and minimum in 2020 at 263.7 ± 38.6 (Table S2). In 2019, the values of individual alkaloids and their total exceeded the average for three years, in 2020 and 2022 they were below the average long-term values. The maximum value of the trait 2017.4 mg/100 g D.M. was also noted in 2019, and the minimum of 1.6 mg/100 g D.M., in 2022. The coefficients of variation for the content of both individual alkaloids and their total were very high; the latter was from 112.4 to 124.6 ($CV > 100\%$, an abnormal range of values) (Table S2). This indicates a very large diversity of the studied accessions of the collection in terms of alkaloid content (Fig. 1a).

Fluctuations in the ranges of alkaloid content variability in NLL seeds over the years of research, taking into account the high CV, make it possible to assume a significant influence of weather conditions on these traits. However, the quantitative ratio of individual alkaloids remained unchanged during the years of research. The highest indicators during all three years of research were demonstrated by lupanine. Average values were typical for 13-hydroxylupanine and sparteine, and minimum ones, for angustifoline and isolupanine.

The content of protein, oil and moisture in seeds found in the studied sample averaged 31.9; 5.2 and 6.9 %, respectively over the entire period of research (Table S3). The protein and moisture indicators of lupine seeds depended little on weather conditions and were characterized by a low degree of variability ($CV < 10\%$). The oil content in NLL seeds had a medium degree of variability ($10 < CV \approx 20\%$). In 2019, the lowest values of protein and oil content were noted along with the highest content of alkaloids. In 2020, the minimum indicators for alkaloids, maximum for protein, and average for oil were recorded. In 2022, the intermediate values of alkaloids corresponded to the intermediate values of protein and maximum values of oil. No visible fluctuations in the moisture content of NLL seeds were detected during the period of research (Fig. 1b, Table S2).

An analysis of the obtained results showed that the seeds of different accessions obtained in 2019, 2020, and 2022 differ significantly in terms of lupanine, sparteine, total alkaloids, protein, and oil content. A difference close to significant was observed for the values of angustifoline and isolupanine. The differences between the NLL accessions in terms of 13-hydroxylupanine and moisture content were not significant. The rank correspondence coefficient (Table S4) showed a reliable match between the alkaloidity of the NLL accessions grown in different years according to their alkaloid status (low- and

¹ Supplementary Figure S1 and Tables S1–S4 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Shelenga_Engl_29_1.pdf

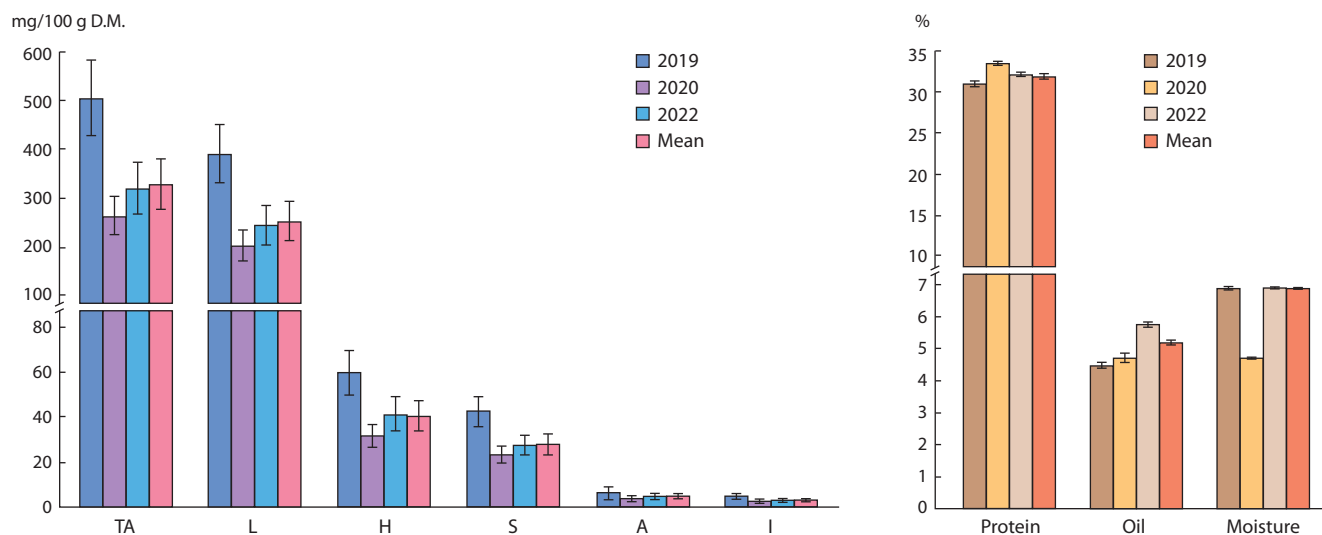


Fig. 1. Total alkaloids and the content of (a) individual alkaloids; (b) protein, oil and moisture content in narrowleaf lupine seeds.

Average, minimum, maximum values of indicators for 2019, 2020 and 2022 and average indicators for three years are presented. Pushkin, Pushkinsky District of St. Petersburg, 2019–2020 and 2022.

high-alkaloid forms). Thus, it can be asserted that the content of individual alkaloids, as well as their total, are features, the order of magnitude of which is determined by the characteristics (genotype) of an accession, and the value of which is determined by the influence of abiotic factors.

The concentration of individual alkaloids and their total content in NLL seeds in 2019 was significantly higher than in 2020 and 2022 (Fig. 1a, Table S4). On the contrary, protein values were significantly higher in 2020, and the oil content in 2022 (Fig. 1b, Table S4). No significant differences in moisture content were found between NLL seeds obtained in different years (Fig. 1b, Table S4). It was found that the values for protein in 2020, and protein and oil in 2022 were higher than the average three-year ones, and those for oil in 2020, and protein and oil in 2019 were lower. The seed moisture content in all the years of research practically corresponded to the average long-term data.

In 2022, a close relationship between the values for individual alkaloids, and between those for individual alkaloids and their total was confirmed in all the years of study ($r = 0.87$ or more). This relationship was previously established for 2019 and 2020 (Vishnyakova et al., 2023). In 2019, a reliable negative relationship of medium strength was observed between the indicators for oil and alkaloids (from -0.35 to -0.42), and for oil and protein (-0.34). In 2020, the direction of the relationship between the content of oil and alkaloids, and that of oil and protein remained, but it became weaker; its reliability was confirmed only for lupanine, sparteine and the total alkaloids (the maximum absolute values of alkaloid concentration). In 2022, an inverse relationship close to reliable remained relevant only for lupanine (-0.23). There was a direct reliable correlation between seed moisture and sparteine in 2019, and between moisture, protein, and oil content in 2020 (0.27 ; 0.30 ; 0.29 , respectively). In 2019, a reliable negative correlation was found between the content of moisture

and lupanine, sparteine, and the total alkaloids (-0.29 ; 0.34 ; -0.29 , respectively), and a negative correlation close to reliable between the content of moisture and 13-hydroxylupanine (-0.25). There were no significant correlations between protein content and alkaloid values (see the Table).

When analyzing the influence of weather conditions on the biochemical characteristics of NLL seeds, the previously established fact (Vishnyakova et al., 2023) that insufficient precipitation contributes to the accumulation of the main alkaloid lupanine and the total alkaloids was confirmed. These two indicators are related by a reliable negative correlation to the total amount of precipitation during the growing season (-0.996 and -0.997 , respectively). A particularly noticeable lack of precipitation within three years was observed in June and August 2019, when 58 and 25 mm fell, respectively. The extremely low amount of precipitation for the region during this period (Fig. S1a) resulted in the maximum accumulation of lupanine and total alkaloids in 2019 (Fig. 1a, Table S2).

Higher air temperatures contributed to an increase in oil content, whereas an increase in precipitation contributed to the accumulation of protein in NLL seeds. Oil content and average air temperatures were linked by a reliable positive correlation (0.96), as were protein content and precipitation (0.96).

The factor analysis of a correlation system allows a reduction of a high-dimensional feature space to a lower-dimensional one, in which the coordinate axes (factors) represent the center of concentration of features that correlate with each other. Factors are hidden variables that influence the observed features. Over three years of research, two factors describing a total of 77.7 % of the variability were identified in the variation structure. The first factor (*Factor 1*, 65.2 % of the variance) is associated with the variation of the indicators characterizing the content of alkaloids, while the second factor (*Factor 2*, 12.5 % of the variance) is negatively correlated with the protein content, and positively with the moisture and oil content

Correlation and factor structure of the variability of biochemical characteristics of seeds of narrowleaf lupine accessions in 2019, 2020, and 2022

Indicator	Abbreviated name*	Year	L	H	S	A	I	TA	Protein	Oil	Moisture
Lupanine	L	2019	1.000	0.968	0.958	0.973	0.945	1.000	-0.064	-0.370	0.199
13-Hydroxylupanine	H		0.001	1.000	0.896	0.956	0.926	0.973	0.027	-0.422	0.193
Sparteine	S		0.001	0.001	1.000	0.939	0.876	0.960	-0.061	-0.354	0.271
Angustifoline	A		0.001	0.001	0.001	1.000	0.908	0.975	-0.057	-0.376	0.196
Isolupanine	I		0.001	0.001	0.001	0.001	1.000	0.944	0.083	-0.421	0.137
Total alkaloids	TA		0.001	0.001	0.001	0.001	0.001	1.000	-0.051	-0.379	0.205
Protein			0.696	0.867	0.711	0.727	0.613	0.755	1.000	-0.335	-0.226
Oil		2020	0.019	0.007	0.025	0.017	0.007	0.016	0.035	1.000	-0.154
Moisture			0.219	0.232	0.091	0.225	0.399	0.204	0.161	0.341	1.000
Lupanine	L		1.000	0.941	0.959	0.956	0.969	0.999	-0.118	-0.289	-0.177
13-Hydroxylupanine	H		0.001	1.000	0.874	0.936	0.938	0.955	-0.152	-0.177	-0.160
Sparteine	S		0.001	0.001	1.000	0.916	0.950	0.961	-0.055	-0.258	-0.183
Angustifoline	A		0.001	0.001	0.001	1.000	0.960	0.962	-0.106	-0.152	-0.067
Isolupanine	I		0.001	0.001	0.001	0.001	1.000	0.975	-0.083	-0.219	-0.130
Total alkaloids	TA		0.001	0.001	0.001	0.001	0.001	1.000	-0.118	-0.271	-0.175
Protein		2022	0.430	0.307	0.713	0.480	0.581	0.428	1.000	-0.066	0.303
Oil			0.049	0.233	0.080	0.308	0.138	0.066	0.662	1.000	0.289
Moisture			0.235	0.282	0.219	0.653	0.384	0.241	0.038	0.049	1.000
Lupanine	L		1.000	0.940	0.935	0.958	0.931	0.998	0.108	-0.229	-0.290
13-Hydroxylupanine	H		0.000	1.000	0.875	0.924	0.889	0.956	0.072	-0.143	-0.250
Sparteine	S		0.000	0.000	1.000	0.898	0.889	0.944	0.144	-0.187	-0.338
Angustifoline	A		0.000	0.000	0.000	1.000	0.936	0.962	0.090	-0.193	-0.208
Isolupanine	I		0.000	0.000	0.000	0.000	1.000	0.935	0.135	-0.179	-0.220
Total alkaloids	TA		0.000	0.000	0.000	0.000	0.000	1.000	0.107	-0.215	-0.290
Protein			0.430	0.599	0.290	0.510	0.323	0.433	1.000	-0.080	-0.189
Oil			0.090	0.292	0.168	0.154	0.187	0.112	0.559	1.000	0.057
Moisture			0.030	0.063	0.011	0.125	0.104	0.030	0.162	0.677	1.000
			Significant positive relationship								
			Close to significant negative relationship								
			Significant negative relationship								
			Diagonal element								

Note. From the top of the diagonal line is the value of the correlation coefficient r , the confidence level p is below.

in NLL seeds (Fig. 2). The alkaloid indicators (lupanine, 13-hydroxylupanine, sparteine, isolupanine, angustiofoline and their total content) are grouped in the right part of the graph. Most likely, this type of grouping is influenced by the fact that these compounds are linked by a single network of metabolic transformations. The indicators of protein, oil and

moisture content in NLL seeds, concentrated in the left part of the figure, show a fairly large spread, which indicates the absence of a close relationship between them.

The eigenvalues of the factors calculated during the analysis for each NLL accession were used to describe the overall pattern of trait variability over the study period (Fig. 3). The dis-

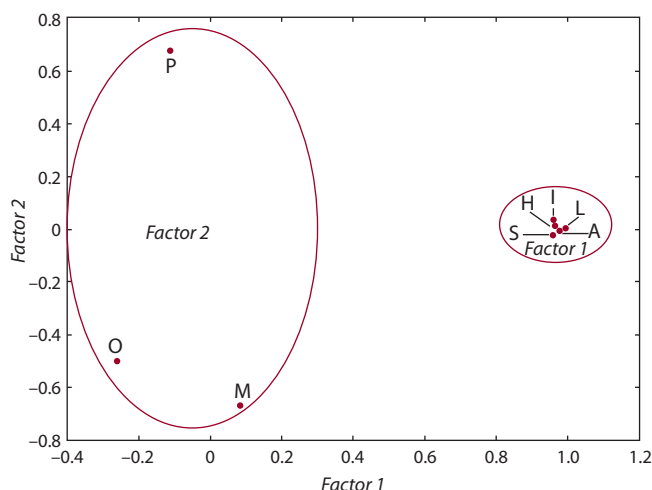


Fig. 2. Factor structure of variability of average values of biochemical indicators determining the seed quality of narrowleaf lupine accessions in 2019, 2020, 2022.

P – protein, O – oil, M – moisture, I – isolupanine, H – 13-hydroxylupanine, S – sparteine, L – lupanine, A – angustifoline.

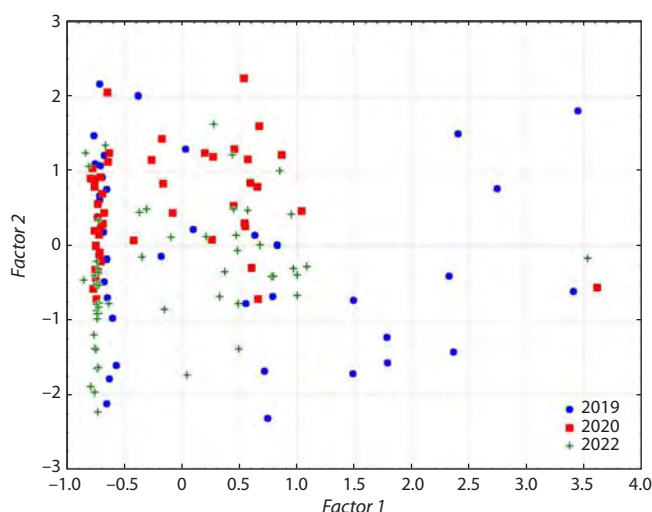


Fig. 3. Position of narrowleaf lupine accessions in the space of selected factors (2019, 2020, 2022).

tribution of accessions was influenced by the above-described factors. In 2019, a higher degree of dispersion was observed for both the first (*Factor 1*) and the second (*Factor 2*) factors. In 2020 and 2022, the spread of samples in the factor space decreased, they grouped mainly in its upper left, middle and partially in the lower left part, which is explained by a decrease in the alkaloid content, a smaller range of their variability and an increase in protein and oil content compared to 2019.

Since obtaining low-alkaloid forms has long been a priority direction of breeding, we paid special attention to the sources of this trait in the studied sample. These included 24 accessions with an alkaloid concentration in seeds of no more than 20–40 mg/100 g D.M. and 7 accessions with that of no

more than 40 mg/100 g D.M. Of these accessions, 16 were characterized by a high protein content (above 30 %), 11 by a high oil content (above 5 %) (Table S1). The combination of a low alkaloid concentration and a high protein and oil content in one genotype characterizes them as sources of high-quality seeds for breeding NLL cultivars in the region for both fodder and food purposes. The group we have identified includes domestic and foreign NLL cultivars and lines: k-3172 ('GL-396', Belarus), k-3329 ('Line 7', Russia), k-3502 ('L-155', Poland), k-3503 ('Mutant 2', Russia), k-3563 ('Rommel', South Africa), k-3627 ('Dikaf-1', Russia), k-3816 ('Ladny 7', Russia), as well as one wild-growing NLL accession k-3457 ('GRC-5008 A', Greece).

The NLL accessions with the total alkaloid content of over 1,000 mg/100 g D.M. in seeds can be recommended as source material for creating green manure cultivars (Table S1). This group is mainly composed of foreign NLL cultivars and lines: k-96 (local, Ukraine), k-1526 (local, Ukraine), k-2183 (IGRIS, Poland), k-3562 ('Slapska', Czechoslovakia), k-3623 (18 86A250-2-4 EX LR2, Australia), and k-3814 ('Oligarkh', Russia).

The high-alkaloid cultivar 'Oligarkh' (k-3814) created at the Leningrad Research Institute for Agriculture "Belogorka" can be used both as raw material for the production of medicines and for creating new pharmacologically significant NLL cultivars. In our study, it demonstrated the following maximum values of individual alkaloid content (mg/100 g D.M.): 1487.3 for lupanine, 338.7 for 13-hydroxylupanine, 142.8 for sparteine, 38.6 for angustifoline, and 21.0 for isolupanine (Table S1). As is known, NLL can be a producer of alkaloids for the use in pharmacology and medicine (Vishnyakova et al., 2020).

Discussion

The three-year data on the main biochemical characteristics that determine the quality of seeds in NLL accessions from the VIR collection grown in conditions of the Northwest of the Russian Federation (Leningrad Province) revealed significant variability in the alkaloid content. The data analysis shows that this is due to both genotypic determination and the influence of environmental conditions. The dependency of alkaloid content on weather factors during cultivation in these conditions was shown earlier (Vishnyakova et al., 2023). In our research, the protein content is characterized by low and oil content by medium variability. Seed moisture was the most stable characteristic. Seed moisture reaches about 15 % during harvesting, while before storing, its value should not exceed 10 % according to (State Standard R 52325-2005).

The statistical processing data confirmed that the alkaloid status (low and high alkaloid forms) was maintained by the NLL accessions reproduced in different years, i.e. the values of the individual alkaloids content, their total and ratio are determined by the genotype, while their variability is determined by the weather conditions of the year of reproduction. The previously established fact of an increase in the amount of alkaloids in seeds under dry conditions (Vishnyakova et al., 2023) was confirmed. A reliable negative correlation between the alkaloid content and the amount of precipitation was established.

The influence of weather conditions on the alkaloid content in lupine seeds has been shown in different regions. In the Russian Federation, this was noted in the conditions of the Yaroslavl Province (Taran, Tsvik, 2017) and the Southwest zone of the Central Region (Ageeva, Pochutina, 2018). In Denmark, when three NLL cultivars were exposed to drought, the amount of alkaloids in the green matter at the stage of plant vegetation (before flowering) clearly increased, and then the genotypes responded differently to drought (Christiansen et al., 1997) by both increasing and decreasing the amount of alkaloids. Our opinion is that a comparison of these results can only be correct if the same plant organs (vegetative matter or seeds), the content of alkaloids in which is different, are assessed, and if the ontogenetic stages at which the study is conducted are the same.

It is generally accepted that QA is synthesized in NLL in the chloroplasts of young leaves (Wink, 1991; 1993; Wink et al., 1995). The most intensive accumulation begins at the budding stage (Maknickienė, Asakavičiūtė, 2008). It has been shown that the expression of the identified to date seven candidate genes involved in the QA synthesis, and the alkaloids themselves, are detected in all plant tissues at the budding stage (Czepiel et al., 2021). However, the presence of alkaloids in plant tissues is detected as early as in young seedlings, into which the alkaloids pass from the germinating seeds (De Cortes Sánchez et al., 2021). That is, there is no unambiguous answer yet about the place and time of the onset of alkaloid synthesis in NLL. It is only obvious that the overwhelming fraction of QA is formed in green, above-ground tissues (Frick et al., 2017) with a small contribution from roots (Lee et al., 2007). By the time of beans formation, alkaloids enter the reproductive organs via the phloem (see a review by Vishnyakova, Krylova, 2022). It follows from this that the onset of the multi-stage synthesis of QAs in lupine and the time of their accumulation in seeds are separated in space and time. Numerous enzymes, transporters, and regulators are involved in this process. However, the expression of genes involved in the synthesis of alkaloids is no longer observed in mature seeds containing alkaloids (Czepiel et al., 2021). That is, only the accumulation process occurs in them. Therefore, it is obvious that the most vulnerable periods for the impact of stressors on the content of alkaloids are the synthesis and transport of the latter. The effect of early drought on a sharp increase in alkaloids in NLL has already been well proven (Frick et al., 2018). In the present research, the driest conditions in 2019 occurred in June, the time of budding (synthesis), and August (alkaloids delivery to the forming seeds). In general, the entire period of research has demonstrated a reliable negative relationship between the amount of precipitation during the entire vegetative period and the content of the main alkaloid lupanine, as well as the total alkaloids in NLL seeds (–0.996, –0.997, respectively).

An increase in alkaloid concentration under the influence of drought is used for their industrial production from producer plants, i.e. representatives of the genus *Nicotiana*, *Papaver somniferum* and *Catharanthus roseus* (Waller, Nowacki, 1978; Szabó et al., 2003; Jaleel et al., 2007; Amirjani, 2013). They

are purposefully exposed to drought stress to increase the yield of alkaloids (Kleinwächter, Selmar, 2015).

It should be noted that the almost 2-fold average increase in alkaloid concentration recorded for the sample in 2019 compared to 2020 is characteristic only of high- and moderately alkaloid accessions. The accessions with the alkaloid content of less than 0.05 % showed a relatively little change in this indicator in all three years of study. The low-alkaloid accessions were stable in the manifestation of the trait and did not transform into high-alkaloid ones under the influence of weather conditions.

The temperature factor did not affect the alkaloid content in NLL seeds in our research; at the same time, it was mentioned above that the elevated air temperature contributed to the accumulation of oil, and precipitation contributed to the accumulation of protein.

The studied sample demonstrated a reliable negative correlation between the protein and oil content (–0.34) only in 2019. A strong negative relationship ($r = -0.96$, $p < 0.01$) between the content of these metabolites in seeds was shown in the work of Australian scientists who studied six NLL cultivars in 55 locations in Western Australia (Cowling, Tarr, 2004). However, a study of other lupine species from the VIR collection reproduced in the Northwest of the Russian Federation demonstrated a positive relationship between the protein and oil content (Egorova et al., 2019). It can be assumed that these relationships manifest themselves differently in different weather conditions.

No significant correlations were found between the protein and alkaloid content in the sample studied in the present research. Meanwhile, among 126 samples of white lupine (*L. albus* L.) accessions from the collection at Pullman (Washington, USA), ranked into six classes by the degree of alkaloidity, a higher protein content was found in seeds from the group of high-alkaloid accessions (Staples et al., 2017). Reliable negative relationships between the oil and alkaloid content were found only in 2019 and 2020 (–0.38 and –0.27, respectively). In 2022, no relationships were found. We believe that the search for regularities in these relationships requires long-term research with weather conditions duly accounted for. The most stable indicator in our research, independent of environmental conditions, was seed moisture.

Conclusion

The relationships between the main NLL seed quality indicators (content of alkaloids, protein, oil, and moisture) and the influence of weather conditions on them were studied for the first time in the conditions of the Northwest of the Russian Federation in the Leningrad Province. The limits of variability of these traits are shown. The absence of significant correlations between the content of protein and oil, and protein and alkaloids was noted. Reliable negative relationships were found between the content of oil and alkaloids only in 2019 and 2020. It is obvious that identifying the regularities in these relationships requires long-term research with weather conditions duly accounted for. The influence of weather on the concentration of alkaloids in seeds has been proven, namely, its

significant increase in dry conditions. The temperature regime during the present research did not affect this indicator. An increased air temperature contributed to the accumulation of oil, whereas precipitation contributed to the accumulation of protein. The studied sample was found to contain accessions that combine the necessary indicators of the main (protein and oil) and secondary metabolites (alkaloids) that determine the NLL seed quality for the use as source material when creating new regionally adapted cultivars for food, fodder, green manure and pharmaceutical purposes.

References

- Ageeva P.A., Pochutina N.A. Results of the narrowleafed lupin variety testing. *Zernobovoye i Krupanye Kul'tury = Legumes and Groat Crops*. 2018;27(3):77-81. doi 10.24411/2309-348X-2018-11037 (in Russian)
- Amirjani M. Effects of drought stress on the alkaloid contents and growth parameters of *Catharanthus roseus*. *J Agric Biol Sci*. 2013; 8(11):745-750
- Benken I.I., Kurlovich B.S., Kartuzova L.T., Nikishkina M.A., Vlasov V.A., Kutuzova E.A., Nazarova N.S., Pilipenko S.I., Rybnikova V.A. Narrow-leaved lupine – *Lupinus angustifolius* L.: Biochemical characterization of specimens. In: VIR World Collection Catalog. Iss. 637. St. Petersburg, 1993 (in Russian)
- Christiansen J.L., Jørgsgård B., Buskov S., Olsen C. E. Effect of drought stress on content and composition of seed alkaloids in narrow-leaved lupin, *Lupinus angustifolius* L. *Eur J Agron*. 1997;7:307-314. doi 10.1016/S1161-0301(97)00017-8
- Cowling W., Tarr A. Effect of genotype and environment on seed quality in sweet narrow-leaved lupin (*Lupinus angustifolius* L.). *Aust J Agric Res*. 2004;55(7):745-751. doi 10.1071/AR03223
- Czepiel K., Krajewski P., Wilczura P., Bielecka P., Świącicki W., Kroc M. Expression profiles of alkaloid-related genes across the organs of narrow-leaved lupin (*Lupinus angustifolius* L.) and in response to anthracnose infection. *Int J Mol Sci*. 2021;22:2676. doi 10.3390/ijms22052676
- De Cortes Sánchez M., Altares P., Pedrosa M.M., Burbano C., Cuadrado C., Goyoaga C., Muzquiz M., Jiménez-Martínez C., Dávila-Ortiz G. Alkaloid variation during germination in different lupin species. *Food Chem*. 2005;90(3):347-355. doi 10.1016/j.foodchem.2004.04.008
- Egorova G.P., Solovyova A.E., Proskuryakova G.I. Genetic resources of the VIR lupine collection for breeding. In: Abstracts of the Int. Sci.-Pract. Conf. "New Varieties of Lupine, Technology for their Cultivation and Processing, Adaptation to Farming Systems and Livestock Breeding", 3–4 July, 2017. Bryansk, Russia, 2017;13-23 (in Russian)
- Egorova G.P., Shelenga T.V., Proskuryakova G.I. Biochemical characterization of of lupin (*Lupinus* L.) seeds from the VIR collection. *Zernobovoye i Krupanye Kul'tury = Legumes and Groat Crops*. 2019;31(3):79-87. doi 10.24411/2309-348X-2019-11118 (in Russian)
- Frick K., Kamphuis L., Siddique K., Singh K., Foley R. Quinolizidine alkaloid biosynthesis in lupins and prospects for grain quality improvement. *Front Plant Sci*. 2017;8:1-12. doi 10.3389/fpls.2017.00087
- Frick K., Foley R., Kamphuis L.G., Siddiqu K., Gar G., Singh K.B. Characterization of the genetic factors affecting quinolizidine alkaloid biosynthesis and its response to abiotic stress in narrow-leaved lupin (*Lupinus angustifolius* L.). *Plant Cell Environ*. 2018;41:2155-2168. doi 10.1111/pce.13172
- Jaleel C.A., Manivannan P., Kishorekumar A., Sankar B., Gopi R., Somasundaram R. Alterations in osmoregulation, antioxidantenzymes and indole alkaloid levels in *Catharanthus roseus* exposed to water deficit. *Colloids Surf B Biointerfaces*. 2007;59:150-157. doi 10.1016/j.colsurfb.2007.05.001
- Kleinwächter M., Selmar D. New insights explain that drought stress enhances the quality of spice and medicinal plants: potential applications. *Agron Sustain Dev*. 2015;35(1):121-131. doi 10.1007/s13593-014-0260-3
- Kuptsov N.S., Takunov I.P. Lupine: Genetics, Breeding, Heterogeneous Cultivation. Bryansk, 2006 (in Russian)
- Kushnareva A.V., Shelenga T.V., Perchuk I.N., Egorova G.P., Malyshev L.L., Kerv Yu.A., Shavarda A.L., Vishnyakova M.A. Selection of an optimal method for screening the collection of narrow-leaved lupine held by the Vavilov Institute for the qualitative and quantitative composition of seed alkaloids. *Vavilov J Genet Breed*. 2020;24(8):829-835. doi 10.18699/VJ20.680
- Lee M.J., Pate J.S., Harris D.J., Atkins C.A. Synthesis, transport and accumulation of quinolizidine alkaloids in *Lupinus albus* L. and *L. angustifolius* L. *J Exp Bot*. 2007;58:935-946. doi 10.1093/jxb/erl254
- Maknickienė Z., Asakavičiūtė R. Alkaloid content variations in lupin (*Lupinus* L.) genotypes and vegetation periods. *Biologija*. 2008;54: 112-115. doi 10.2478/v10054-008-0023-7
- Popov V.S., Salikova A.V., Perchuk I.N., Konkova N.G., Egorova G.P., Vishnyakova M.A., Shelenga T.V. Rapid assessment of the main economic value indicators in lupine flour samples using infrared spectroscopy. *Trudy po Prikladnoi Botanike, Genetike i Selekcii = Proceedings on Applied Botany, Genetics and Breeding*. 2024;185(1): 99-108. doi 10.30901/2227-8834-2024-1-99-108 (in Russian)
- Sengbusch R. Bitterstoffarme lupinen. II. *Züchter*. 1931;4:93-109
- Siger A., Grygier A., Czubinski J. Comprehensive characteristic of lipid fraction as a distinguishing factor of three lupin seed species. *J Food Compos Anal*. 2017;115:104945. doi 10.1016/j.jfca.2022.104945
- Specification No. 9716-004-00668502-2008. Food Lupin. Available at: <https://e-ecolog.ru/crc/57.01.01.000.%D0%A2.000230.05.08?ysclid=14sa0dtvbn325210024> (Accessed June 24, 2022) (in Russian)
- Staples K.D., Hamama I.A.A., Knight-Mason R., Bhardwaj H.L. Alkaloids in white lupin and their effects on symbiotic N fixation. *J Agric Sci*. 2017;9(6):13. doi 10.5539/jas.v9n6p13
- State Register of Selection Achievements Authorized for Use for Production Purposes. Vol. 1. Plant Varieties (official publication). Moscow, 2023 (in Russian)
- State Standard R 54632-2011. Fodder Lupine. Specification. 2013. Available at: <https://docs.cntd.ru/document/1200093158?ysclid=l4s80m228h216628534> (Accessed May 7, 2024) (in Russian)
- State Standard R 52325-2005. Seeds of Agricultural Plants. Varietal and Sowing Characteristics. General Specifications. Available at: <https://internet-law.ru/gosts/gost/4709/?ysclid=lwgrcx2jpt453276006> (Accessed May 21, 2024) (in Russian)
- Szabó B., Tyihák E., Szabó G., Botz L. Mycotoxin and drought stress-induced change of alkaloid content of *Papaver somniferum* plantlets. *Acta Bot Hung*. 2003;45:409-417. doi 10.1556/ABot.45.2003.3-4.15
- Taran T.V., Tsvik G.S. The influence of growing conditions on the chemical composition of narrow-leaved lupine. In: Abstracts of the Int. Sci.-Pract. Conf. "New Varieties of Lupine, Technology for their Cultivation and Processing, Adaptation to Farming Systems and Livestock Breeding", 3–4 July, 2017. Bryansk, Russia, 2017;35-40 (in Russian)
- Vishnyakova M.A., Seferova I.V., Buravtseva T.V., Burlyaeva M.O., Semenova E.V., Filipenko G.I., Aleksandrova T.G., Egorova G.P., Yankov I.I., Bulyntsev S.V., Gerasimova T.V., Drugova E.V. VIR Global Collection of Grain Legume Crop Genetic Resources: Replenishment, Conservation and Studying. Methodological Guidelines. St. Petersburg: VIR, 2018. doi 10.30901/978-5-905954-79-5 (in Russian)
- Vishnyakova M.A., Krylova E.A. Prospects for obtaining low-alkaloid and adaptive forms of narrow-leaved lupine based on the genome

- and transcriptome resources of the species. *Biotehnologiya i Seleksiya Rastenij* = *Plant Biotechnology and Breeding*. 2022;5(2): 5-14. doi 10.30901/2658-6266-2022-2-o1 (in Russian)
- Vishnyakova M.A., Kushnareva A.V., Shelenga T.V., Egorova G.P. Alkaloids of narrow-leaved lupine as a factor determining alternative ways of the crop's utilization and breeding. *Vavilov J Genet Breed*. 2020;24(6):625-635. doi 10.18699/VJ20.656
- Vishnyakova M.A., Salikova A.V., Shelenga T.V., Egorova G.P., Novikova L.Y. Alkaloid content variability in the seeds of narrow-leaved lupine accessions from the VIR collection under the conditions of the Russian Northwest. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2023;27(2):119. doi 10.18699/VJGB-23-17
- Waller G.R., Nowacki E.K. *Alkaloid Biology and Metabolism in Plants*. New York: Plenum Press, 1978. doi 10.1007/978-1-4684-0772-3
- Wang X.-F., Jing X.-M., Zheng G.-H. Effect of seed moisture content on seed storage longevity. *Acta Bot Sinica*. 2001;43:551-557
- Wink M. Plant breeding: Low or high alkaloid content. In: Proc. of the 6th Int. Lupin Conf. 25–30 Nov. 1990. Temuco-Pucón, Chile. Int. Lupin Association, 1991;326-334
- Wink M. Allelochemical properties or the raison d'être of alkaloids. *The Alkaloids: Chemistry and Pharmacology*. 1993;43:1-118. doi 10.1016/S0099-9598(08)60134-0
- Wink M., Meißner C., Witte L. Patterns of quinolizidine alkaloids in 56 species of the genus *Lupinus*. *Phytochemistry*. 1995;38:139-153. doi 10.1016/0031-9422(95)91890-d

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

Received May 27, 2024. Revised September 3, 2024. Accepted September 3, 2024.

doi 10.18699/vjgb-25-06

Analysis of similarities and differences of accessions belonging to *Prunus domestica* L. and *P. insititia* L. using endocarp dimensions and shape variations

T. Milošević ¹ , N. Milošević ²

¹ Department of Fruit Growing and Viticulture, Faculty of Agronomy in Čačak, University of Kragujevac, Čačak, Republic of Serbia

² Department of Pomology and Fruit Breeding, Fruit Research Institute, Čačak, Republic of Serbia

 tomomilosevic@kg.ac.rs

Abstract. The endocarp or stone is the most stable morphological feature of the genus *Prunus*. However, the identification of plum types, groups and/or genotypes based on endocarp is complicated because of a wide range of variation and morphological transitional states. From this point of view, knowledge on the degree of variability within and between plum species or cultivars is a *sine qua non* for taxonomists and also for pomologists. In this study, different endocarp morphological traits, such as SW, linear dimensions (L, W and T), D_a , D_g , S, V and shape indexes (ϕ , SI, E, RS, RO, DE and PI) were determined using analysis of variance and multivariate analysis (correlations and PCA). Results showed significant differences among accessions for all properties evaluated but with high overlaps in values. In most cases, the examined parameters were positively or negatively correlated with each other, indicating developmental relationships between them. Indeed, positive correlations were recorded for most variables, especially related to SW and endocarp linear dimensions. These results showed that the above properties could be a powerful indicator for selecting adequate endocarp size and shape in accessions, which may be used in taxonomic analysis. With an account of these correlations, PCA was employed to correctly estimate the endocarp size and shape and distribution, segregation and dispersion of accessions. All linear measurements and index values showed a normal or low variability at the individual level in most cases, with the exception of SW, V and PI in both European and Damson plums and S in Damson plums. Of the 15 examined parameters, European plum had significantly higher SW, L, T, D_a , D_g , S, E, RO and PI values than Damson plum. In contrast, Damson plum had higher SI, RS and DE values, while W, V and ϕ were similar.

Key words: endocarp; European and Damson plums; morphological properties; size; shape indexes.

For citation: Milošević T., Milošević N. Analysis of similarities and differences of accessions belonging to *Prunus domestica* L. and *P. insititia* L. using endocarp dimensions and shape variations. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov J Genet Breed*. 2025;29(1):44-54. doi 10.18699/vjgb-25-06

Funding. Financial support for this research was provided by the Ministry of Science, Technological Development and Innovations of the Republic of Serbia (grants 451-03-66/2024-03/200088 and 451-03-66/2024-03/200215).

CRedit authorship contribution statement. Tomo Milošević: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. Nebojša Milošević: Data curation, Formal analysis, Investigation, Software, Validation, Writing – original draft, Writing – review & editing.

Анализ сходств и различий образцов *Prunus domestica* L. и *P. insititia* L. по размерам эндокарпия и вариациям формы

Т. Милошевич , Н. Милошевич ²

¹ Кафедра садоводства и виноградарства, Факультет агрономии в г. Чачак, Крагуевацкий университет, Чачак, Республика Сербия

² Кафедра помологии и селекции фруктов, Научно-исследовательский институт фруктов, Чачак, Республика Сербия

 tomomilosevic@kg.ac.rs

Аннотация. Эндокарпий, или косточка, является наиболее стабильной морфологической особенностью рода *Prunus*. Однако идентификация типов, групп и/или генотипов сливы на основе эндокарпия осложнена из-за широкого спектра вариаций и морфологических переходных состояний. С этой точки зрения знание степени изменчивости внутри и между видами или сортами сливы является неперенным условием для таксономистов и помологов. В нашем исследовании различные морфологические признаки эндокарпия, такие как SW, линейные размеры (L, W и T), D_a , D_g , S, V и индексы формы (ϕ , SI, E, RS, RO, DE и PI), были определены с помощью дисперсионного анализа и многомерного анализа (корреляции и PCA). В результате обнаружены значительные различия между образцами по всем оцененным свойствам, но с высоким перекрытием значений. В большинстве случаев исследованные параметры положительно или отрицательно коррелировали друг с другом, что указывало на связи развития между ними. Действительно, положительные корреляции были зарегистрированы для большинства переменных, особенно связанных с SW и линейными размерами эндокарпия. Эти результаты показали, что вышеперечисленные свойства могут служить мощным индикатором для выбора адекватного размера

и формы эндокарпия образцов, которые могут быть использованы в таксономическом анализе. Выявленные корреляции были использованы для проведения анализа методом главных компонент (PCA), что позволило правильно оценить не только размеры и форму эндокарпия, но и диапазоны варьирования признаков, включая дисперсию, а также распределение образцов по группам. Все линейные измерения и значения индекса показали нормальную или низкую изменчивость на индивидуальном уровне в большинстве случаев, за исключением SW, V и PI у европейской сливы и сливы Damson и S у сливы Damson. Из 15 исследованных параметров для европейской сливы были получены значительно более высокие значения SW, L, T, D_a, D_g, S, E, RO и PI по сравнению со сливой Damson. Напротив, слива Damson отличалась более высокими значениями SI, RS и DE, тогда как значения W, V и ф были схожими.

Ключевые слова: эндокарпий; слива европейская и чернослива; морфологические свойства; размер; индексы формы.

Introduction

Prunus spp. has been grown throughout the world for centuries. Among commercial species, the plum is the most commonly cultivated for its fruits (Milošević T., Milošević N., 2018). The genus originates from five centers in general: Europe for *Prunus domestica* L. (European plum), Western Asia for *P. insititia* L. (Damson plum), Western and Central Asia for *P. cerasifera* Ehrh. (cherry plum), North America for *P. americana* Marsh. (American plum) and China for *P. salicina* Lindl. (Japanese plum) (Watkins, 1976). Among them, genotypes belonging to *P. domestica* L. and *P. salicina* Lindl. are the most important. The evaluation of plum diversity may be essential; for example, for on-farm conservation schemes, utilization of genetic resources for sustainable agriculture and future breeding programs (Ropelewska, 2022). Also, cultivar differentiation is important for farming due to adaptation to climatic conditions and disease resistance, especially under global warming conditions (Milošević T., Milošević N., 2018).

Serbia is a small country on the Balkan Peninsula, but a major world producer of plums, as it ranks third or fourth in the world behind China, USA or Romania (FAOSTAT, 2024, <https://www.fao.org/faostat/en/#data/QCL>). Apart from a large number of commercial cultivars, Serbia is known for growing autochthonous (primitive, folk) genotypes mostly originating from *P. domestica* L. (European plum) and *P. insititia* L. (Damson plum) (Milošević T., Milošević N., 2012). Their fruits are mainly used for the production of a traditional Serbian alcoholic drink known as “Šljivovica” or “Prepečenica”, which is included in the UNESCO representative list of the intangible cultural heritage of humanity as an element of the intangible cultural heritage of Serbia (Source: UNESCO Bulletin, 2022) (Milošević T. et al., 2023). We have earlier described the properties of their trees and fruits (Milošević T., Milošević N., 2012; Milošević N. et al., 2017). Also, a great diversity of types belong to *P. cerasifera* Ehrh. (cherry plum), and *P. spinosa* L. (sloe) can be found in this country. Along this line, it can be said that Serbia is a very rich source of the biological diversity of plums.

The fruit of representatives of the *Prunus* genus consists of an epicarp (outer layer), a mesocarp (flesh), and an endocarp (stone). The stone of a plum consists of the seed covered with a hull. When used fresh or during processing, the flesh and skin of plums are the main raw materials, and the stones are generated as by-products. The plum seed or kernel may be a source of useful substances for food, cosmetics (e. g. personal care products), pharmaceutical industries (González-García et al., 2014; Plainfossé et al., 2019). Also, the seeds can be used to obtain generative rootstocks in horticultural practice

Abbreviations

SW, stone (endocarp) weight
L, stone length
T, stone thickness
W, stone width
D_a, arithmetic mean diameter
D_g, geometric mean diameter
S, surface area
V, stone volume
φ, sphericity
SI, shape index
E, elongation
RS, relative slenderness
RO, roundness
DE, Behre's index
PI, Pollmann's index
PCA, principal component analysis

and in breeding programs (Milošević T., Milošević N., 2018). The endocarp is the innermost layer of the pericarp, which directly surrounds the seeds. It may be very hard and non-edible as in drupes (also called stone fruits) such as members of the *Prunus* genus, i. e. plums, peaches, apricots and cherries (Carrillo-López, Yahia, 2019).

In pomological research, the stone of the *Prunus* genus represents a very stable feature of plum genotypes and serves for the determination and classification of cultivars (Behre, 1978; Woldring, 2000; Milošević T., Milošević N., 2018). However, during the last 50 years, with a few exceptions, in very complex botanical studies, stone dimensions, size and shape in the plum and other members are a crucial component for classification of systematic categories due to taxonomic complexity of the *Prunus* taxa within the Rosaceae family (Depypere et al., 2007, 2009; Burger et al., 2012; Bawari et al., 2022; Ropelewska, 2022). Behre (1978) reported that endocarp dimensions are very useful for the identification of *P. domestica* L., *P. insititia* L. and *P. spinosa* L. In general, reliable discriminating characters for species and subspecies of *Prunus* taxa identification are missing (Nielsen, Olrik, 2001).

The aim of this study was to characterize and evaluate the diversity of endocarps (stones) within and among traditional European and Damson autochthonous plums that are cultivated in Serbia and other western Balkan countries using essential morphological data (weight, dimensions, size, shape) in

order to provide experimental evidence for the implementation of measures to safeguard this agricultural biodiversity. The secondary goal of this work was to reliably determine the degree of the possibility of identifying plum genotypes using the stone (endocarp) through multivariate statistical analysis.

Material and methods

Study area, plant material and measurements. Analysis was performed using a combination of weight, dimension and shape parameters of 5,500 endocarps (stones) belonging to two closely related Eurasian plum taxa (*P. domestica* L. and *P. insititia* L.). The analysis involved 55 genotypes i.e. accessions [43 accessions (78.18 %) belonging to *P. domestica* and 12 accessions (21.82 %) belonging to *P. insititia*]. Their name, series number and code were presented in Table 1.

Ripe fruit and stone samples (25 fruits or stones in 4 replicates, $n = 100$ per one accession) were taken from a private orchard in the Prislonica village (43°33' N, 16°21' E) near Čačak city (western Serbia) established in 1998. Whole ripe fruits of each accession were harvested individually, manually and randomly in 2007. After this, the fruits were cut in half to extract the stones (endocarps). The removed stones were washed and cleaned of their flesh. After air-drying for 40 days

at room temperature (20 °C in the shade), the undamaged and dry stones were placed in glass jars with hermetic closures and stored in a refrigerator at +4 °C. Stones (endocarps) were subjected to measuring in 2024.

The SW (g) was measured using MAULsteel 5000 G digital balance (Jakob Maul GmbH, Bad König, Germany). In order to determine the endocarp size, three major perpendicular dimensions i.e. L (mm), W (mm) and T (mm) were determined using a digital caliper Starrett 727 (Athol, NE, USA) with the accuracy of ± 0.01 mm. The position of the measurements for L, T and W proposed by Van Zeist, Woldring (2000) was illustrated in Figure 1.

Arithmetic mean diameter (D_a , mm) and geometric mean diameter (D_g , mm) were computed from geometrical dimensions by Eq. (1) and Eq. (2) (Mohsenin, 1986):

$$D_a = (L + T + W)/3, \tag{1}$$

$$D_g = \sqrt[3]{LWT}. \tag{2}$$

The S (cm²) is a measure of the total area that the surface of the object occupies and was determined by approximating its shape to a sphere with the same geometric mean diameter by using Eq. (3) (Mohsenin, 1986):

$$S = \pi D_g^2. \tag{3}$$

Table 1. Name and code of autochthonous plum accessions

Accession (local name)	Accession code	Accession (local name)	Accession code
1. Arapka	ARP	29. Piskavica	PIS
2. Bela Požegača	BPZ	30. Šarica	SAR
3. Belošljiva	BEL	31. Trnošljiva	TRS
4. Cerovački Piskavac	CPI	32. Turgonja	TURa
5. Crnošljiva	CRN	33. Dronga	DRO
6. Crvena Ranka*	CRB	34. Magareška Crna Šljiva	MCS
7. Crvena Ranka**	CRD	35. Beluvra	BEV
8. Čokešinka	COK	36. Trnošljiva – M	TRA
9. Kapavac	KAP	37. Magareška	MAG
10. Maričevka	MAR	38. Crna Petrovka	CPT
11. Metlaš	MET	39. Panadjurka	PAN
12. Mudara	MUD	40. Zimna	ZIM
13. Obični Piskavac	OPI	41. Modra Šljiva	MSI
14. Petrovača	PET	42. Gurgutka	GUR
15. Požegača	POZ	43. Banska Šljiva	BAS
16. Trnovača	TRN	44. Korajka	KOR
17. Turgulja	TUR	45. Bosanka	BOS
18. Moravka	MOR	46. Bilka Rana	BIR
19. Crnica	CRI	47. Julka	JUL
20. Plavoača	PLA	48. Dobojska Rana	DRA
21. Volujevača	VOL	49. Banjalučka Bjelica	BAB
22. Gorka Bula	GBU	50. Sitnica	SIT
23. Bjelica	BJL	51. Slatkulja	SLA
24. Bjelošljiva	BJS	52. Miškovačka Rana	MIR
25. Car Dušan	CDU	53. Kaurka	KAU
26. Durgulja	DUR	54. Ružica	RUZ
27. Mednica	MED	55. Podsedlinka	POD
28. Mudovalj	MUV		

Note. Accessions belonging to *P. insititia* L. are marked in bold-italic. * 'Crvena Ranka' var. 'Bardaklija' (*P. domestica* L.). ** 'Crvena Ranka' var. 'Derosavka' (*P. domestica* L.).

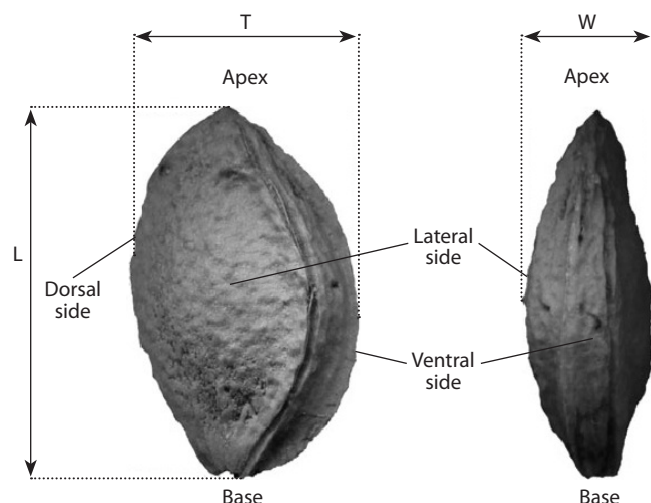


Fig. 1. Overview of basic linear endocarp measurements.
Left: lateral view; right: ventral view (Van Zeist, Woldring, 2000).

The volume (V , mm^3) of the endocarp was calculated by Eq. (4) (Mansouri et al., 2017), which is based on the assumption that plum endocarps are similar to a scalene ellipsoid where $L > T > W$ (Munder et al., 2017):

$$V = LWT\phi. \quad (4)$$

Sphericity (ϕ) is defined as the ratio of the surface area of the sphere having an equivalent volume to that of the endocarp and the surface area of the endocarp. It is a measure of how spherical an object is. It was estimated based on the isoperimetric property of a sphere by Eq. (5) (de Figueiredo et al., 2011):

$$\phi = D_g/L. \quad (5)$$

Shape index (SI) and elongation ratio (E) were calculated using Eqs. (6) (Mohsenin, 1986) and (7) (Firatligil-Durmuş et al., 2010):

$$SI = (W + T)/2L, \quad (6)$$

$$E = L/W. \quad (7)$$

Other indexes of endocarp shape were also calculated according to Behre (1978), Van Zeist, Woldring (2000) and Schmidt-Tauscher et al. (1996, cited in Pollmann et al., 2005). In order to have a more pronounced relationship between individual dimensions, the numbered values were multiplied by 100.

They can be represented as relative slenderness (RS) (Eq. (8)), roundness index (RO) (Eq. (9)) and Behre's index (DE) (Eq. (10)) proposed by Behre (1978) and modified by Van Zeist, Woldring (2000):

$$RS = 100 \times T/L, \quad (8)$$

$$RO = 100 \times T/W, \quad (9)$$

$$DE = 100 \times W/L. \quad (10)$$

Schmidt-Tauscher et al. (1996, cited in Pollmann et al., 2005) introduced a fourth index value which is calculated using Eq. (11). It was named Pollmann's index (PI) because Pollmann et al. (2005) demonstrated its usefulness in differentiating the stones of modern plum cultivars:

$$PI = L^2/(T + W). \quad (11)$$

Variation at different taxonomic levels was studied by analyzing the coefficients of variation (CV, %), which were interpreted following Rasch (1988, cited in Hübner and Wissemann, 2004), i.e. $CV < 10\%$, low variability; $10\% < CV < 20\%$, normal variability; $CV > 25\%$, high variability of the character studied.

Statistical analysis. Data were subjected to analysis of variance (ANOVA) using the Microsoft Office Excel software (Microsoft Corporation, Redmond, WA, USA) procedure followed by least significant difference (LSD) Fisher's test at $p \leq 0.05$ significance level. Pearson's rank correlation matrix ($\alpha = 0.05$) was performed using the R corrplot package version 4.0.2 (R Core Team, 2021). Principal components analysis (PCA) was performed and a biplot was designed using the XLSTAT v. 7.5 software package (Addinsoft, Paris, France).

Results and discussion

Evaluation of endocarp dimensions and shapes

Data in Table 2 showed that SW significantly varied among accessions. High intra- and inter-variability between plum types was observed. The highest and statistically similar values were observed in 'MUD' and 'CPT' (both belonging to *P. domestica*), and the lowest, in 'TRN' and 'CRI' (both belonging to *P. insititia*). Twelve accessions (21.82 % of the total number) had $SW > 1$ g, whereas only four accessions (7.27 % of the total number) had $SW < 0.5$ g. The most numerous (70.91 %) were the accessions, the SW of which ranged between 0.5 and 1 g.

It is known that the former Yugoslavia, i.e. the Western Balkan region, is a very rich source of plum germplasm, especially that of European and Damson plums, cherry plums and sloes (Milošević T., Milošević N., 2012). Considering this, a large number of researchers highlight data related to their biological, agronomic and pomological characteristics (Milošević T. et al., 2023). Thus, Milošević N. et al. (2017) and Glišić et al. (2023) reported SW between 0.57 and 1.49 g or 0.57 and 2.39 g, respectively, for local genotypes belonging to European plum in the Čačak area (western Serbia). Drkenda, Kurtović (2012) reported values between 0.84 and 1.21 g for six local cultivars from Bosnia and Herzegovina. For accessions belonging to European and Damson plums from Montenegro, Jaćimović et al. (2011) and Šebek (2013) reported that SW varied from 0.46 to 2.20 g and 0.16 to 2.20 g, respectively. For nine domestic and local plums grown in Turkey, Gunes (2003) noted SW from 0.26 to 0.99 g. Our data for SW are within the limits of the results of the mentioned authors. However, in taxonomic description and morphometric analysis, SW is not a reliable indicator for the determination, systematization and segregation of members of the *Prunus* genus (Depypere et al., 2007; Mijnsbrugge et al., 2013; Sarigu et al., 2017; Heidari et al., 2022; Kosina, 2023) due to the negative "effect of controlled moisture" (Sheikh et al., 2021) and state of the seed (embryo or kernel) inside the endocarp (das Graças Souza et al., 2016; Sheikh et al., 2021). All of the above authors favored dimensional measurements of the endocarps.

The ANOVA showed significant differences among the accessions for endocarp L, T and W (Table 2). These data are in agreement with the results of Woldring (2000), Van Zeist, Woldring (2000), Depypere et al. (2007) and Heidari

Table 2. Stone weight, stone dimensions and size and shape indexes of plum accessions

Accession code	SW, g	L, mm	T, mm	W, mm	D _a , mm	D _{gr} , mm	S, cm ²	V, cm ³	φ	SI	E	RS	RO	DE	PI
ARP	0.60 q-v	21.20 e-n	10.80 tu	6.00 n-s	12.67 nop	11.12 vw	3.88 r-v	0.72 u-z	0.524 kl	0.396 z	3.53 f	50.94 stu	180.00 ijk	28.30 E	6.94 c-f
BPZ	0.79 i-p	22.95 b-f	12.25 ijk	6.60 i-q	13.93 h	12.29 k	4.74 ij	0.97 j-m	0.535 x-A	0.411 y	3.48 g	53.38 qrs	185.61 e-h	28.76 D	6.51 d-j
BEL	0.90 g-m	18.00 r-w	12.00 klm	6.90 h-n	12.30 qrs	11.42 pqr	4.10 o-s	0.78 q-v	0.635 fg	0.525 ef	2.61 x	66.67 de	173.91 lmn	38.33 k	3.91 u-x
CPI	0.37 xy	21.75 d-k	11.25 qrs	6.00 n-s	13.00 l	11.37 r	4.06 p-s	0.77 r-v	0.522 BCD	0.397 z	3.63 d	51.72 st	187.50 d-g	27.59 FG	7.01 cde
CRN	0.57 s-w	21.30 e-m	11.00 r-u	6.00 n-s	12.77 mn	11.20 tu	3.94 q-u	0.74 s-y	0.526 ABC	0.399 z	3.55 f	51.64 st	183.33 ghi	28.17 EF	6.87 def
CRB	0.75 l-s	24.30 abc	11.80 lmn	6.10 m-s	14.07 gh	12.05 l	4.56 jk	0.92 k-n	0.495 F	0.368 C	3.98 a	48.56 uvw	193.44 b	25.10 H	8.20 b
CRD	0.55 t-x	21.85 c-k	10.80 tu	6.60 i-q	13.08 kl	11.59 o	4.22 m-q	0.82 o-s	0.530 ZAB	0.398 z	3.31 m	49.43 tuv	163.64 rst	30.21 wx	6.70 d-h
COK	0.76 k-r	21.40 e-l	11.95 klm	5.90 o-s	13.08 kl	11.47 p	4.13 n-r	0.79 p-u	0.536 xyz	0.417 wx	3.63 d	55.84 n-q	202.54 a	27.57 G	6.49 d-j
KAP	0.42 v-y	17.80 s-w	10.70 uv	6.70 h-p	11.73 vwz	10.85 y	3.69 uvw	0.67 w-C	0.609 j	0.489 hi	2.66 w	60.11 g-j	159.70 t-w	37.64 l	4.42 r-v
MAR	0.81 i-o	20.85 f-p	11.70 mno	6.80 h-o	13.12 kl	11.84 m	4.40 k-n	0.87 n-q	0.568 pqr	0.444 p	3.07 q	56.11 m-q	172.06 mno	32.61 r	5.46 k-q
MET	0.64 o-u	19.80 h-u	11.40 opq	6.00 n-s	12.40 q	11.06 w	3.84 s-v	0.71 u-z	0.559 rst	0.339 D	3.80 c	57.57 j-n	190.00 b-e	30.30 wx	5.73 t-o
MUD	1.90 a	19.00 l-w	14.00 c	9.50 a	14.17 fg	13.62 d	5.83 c	1.32 cd	0.717 a	0.618 a	2.00 E	73.68 ab	147.37 BC	50.00 b	2.71 z
OPI	0.61 q-u	19.60 h-w	11.80 lmn	7.05 g-m	12.82 m	11.77 n	4.35 k-o	0.85 n-r	0.500 F	0.481 jk	2.78 u	60.20 ghi	167.38 pqr	35.97 n	4.62 q-v
PET	1.01 e-h	17.00 w	12.00 klm	7.00 g-m	12.00 tu	11.26 st	3.98 p-t	0.75 s-x	0.662 d	0.559 d	2.43 z	70.59 c	170.93 nop	41.18 g	3.44 w-z
POZ	0.74 l-s	23.00 b-f	12.95 fg	6.80 h-o	14.25 ef	12.65 hi	5.75 cd	1.30 cde	0.550 tuv	0.429 st	3.38 jk	56.30 l-q	190.44 bcd	29.57 zA	6.01 f-m
TRN	0.13 z	14.20 x	10.40 v	6.20 l-s	10.27 A	9.71 F	2.96 zA	0.48 FG	0.684 c	0.584 c	2.29 B	73.24 abc	167.74 o-r	43.66 e	3.13 xyz
TUR	1.50 b	24.00 a-d	12.00 klm	7.00 g-m	14.33 de	12.63 i	5.01 ghi	1.06 hij	0.526 ABC	0.396 z	3.43 hi	50.00 tu	171.43 nop	29.17 BC	6.86 def
MOR	0.53 u-x	20.50 g-r	11.60 nop	6.30 k-s	12.80 mn	11.44 pq	4.11 o-s	0.78 q-v	0.558 st	0.437 q	3.25 no	56.58 l-o	184.13 f-i	30.73 uv	5.75 h-o
CRI	0.31 yz	17.50 t-w	8.40 yz	5.50 s	10.47 z	9.32 G	2.73 A	0.42 G	0.532 yzA	0.397 z	3.18 p	48.00 uvw	152.73 zA	31.43 st	6.63 d-i
PLA	1.25 cd	21.08 e-o	13.80 c	7.50 d-i	14.46 d	13.17 e	5.07 fgh	1.24 def	0.597 k	0.482 jk	2.94 r	62.50 fg	184.00 f-i	33.97 q	4.71 p-u
VOL	0.66 o-u	17.41 u-w	10.50 v	5.70 qrs	11.20 y	10.14 D	3.62 vw	0.64 z-D	0.582 no	0.465 lm	3.05 q	60.31 ghi	184.21 f-i	32.74 r	5.06 m-s
GBU	0.58 r-v	19.45 j-w	9.38 w	5.95 n-s	11.59 x	10.28 C	3.32 xy	0.57 DEF	0.529 z-C	0.394 z	3.27 n	48.23 uvw	157.65 vwz	30.54 vw	6.78 d-g
BJL	0.95 g-j	19.53 l-w	12.18 jk	6.34 k-s	12.68 m-p	11.48 p	4.13 n-r	0.79 p-u	0.512 DE	0.374 B	3.08 q	62.37 fgh	192.11 bc	32.46 r	4.94 n-u
BJS	0.91 g-l	19.02 l-w	11.45 opq	7.25 f-k	12.57 p	11.64 o	4.26 l-p	0.83 n-s	0.591 k-n	0.492 h	2.62 x	60.20 ghi	157.93 u-x	38.12 k	4.36 r-w
CDU	1.04 efg	22.11 b-h	12.43 ij	8.13 c-f	14.22 efg	13.07 f	5.37 e	1.17 fg	0.543 vwz	0.465 lm	2.72 v	56.22 l-q	152.89 yzA	36.77 m	4.84 o-u
DUR	0.89 g-m	21.92 c-j	11.76 lmn	6.9 h-n	13.41 j	11.90 lm	4.45 klm	0.88 m-p	0.583 mno	0.417 vw	3.35 kl	53.65 o-s	179.82 ijk	29.84 yz	6.25 e-k
MED	0.70 n-u	19.67 h-v	12.05 kl	6.46 j-r	12.70 mno	11.48 p	4.14 n-r	0.79 p-u	0.715 a	0.468 l	3.08 q	61.26 f-i	188.87 cde	32.44 r	7.84 bc
MUV	1.33 bc	19.99 h-t	15.16 a	9.63 a	14.93 c	14.29 b	6.41 b	1.53 b	0.503 EF	0.620 a	3.08 q	75.84 a	157.42 wxy	48.17 c	2.74 z

Table 2 (end)

Accession code	SW, g	L, mm	T, mm	W, mm	D _a , mm	D _g , mm	S, cm ²	V, cm ³	φ	SI	E	RS	RO	DE	PI
PIS	0.55 t-x	20.69 f-q	9.48 w	5.76 p-s	11.98 tu	10.41 B	3.41 x	0.59 B-E	0.547 uvw	0.368 C	3.59 e	45.82 wx	164.58 qrs	27.84 FG	4.43 f-v
SAR	0.95 g-j	21.18 e-n	11.22 qrs	6.56 i-q	12.99 l	11.59 o	4.51 jkl	0.90 l-o	0.609 j	0.420 uv	3.23 o	52.97 rs	171.04 nop	30.97 u	6.09 e-l
TRS	0.78 j-q	18.56 o-w	10.95 stu	7.1 g-l	12.20 s	11.30 s	4.01 p-t	0.76 r-w	0.586 lmn	0.486 ij	2.61 x	59.00 i-m	154.22 x-A	38.25 k	4.43 f-v
TURa	1.33 bc	23.52 e	13.45 d	8.28 cde	15.08 b	13.78 c	5.96 c	1.37 c	0.585 lmn	0.462 mn	2.84 t	57.19 k-n	162.44 stu	35.20 o	4.97 n-t
DRO	1.01 e-h	21.08 e-o	12.02 kl	7.4 e-j	13.50 j	12.33 k	4.77 ij	0.98 i-l	0.521 CD	0.461 mn	2.85 t	57.02 k-n	162.43 stu	35.10 o	5.00 n-s
MCS	0.62 p-u	19.90 h-u	9.59 w	5.81 p-s	11.77 vw	10.35 B	3.36 x	0.58 CDE	0.511 DE	0.387 A	3.42 hi	48.19 uvw	165.06 qrs	29.20 zBC	7.11 cde
BEV	0.97 f-i	21.21 e-n	11.06 rst	7.20 f-k	13.16 k	11.91 l	4.45 klm	0.88 m-p	0.705 b	0.430 rst	2.95 r	52.15 st	153.61 x-A	33.95 q	5.65 j-p
TRA	0.39 wxy	11.96 x	8.67 y	5.79 p-s	8.81 B	8.44 H	2.23 B	0.31 H	0.645 e	0.605 b	2.07 D	72.49 bc	149.74 AB	48.41 q	2.85 yz
MAG	1.47 b	19.98 h-t	12.77 gh	8.38 bcd	13.71 i	12.68 hi	5.21 efg	1.12 gh	0.583 mne	0.529 e	2.38 A	63.91 ef	152.39 zA	41.94 f	3.73 v-y
CPT	1.72 a	25.63 a	14.46 b	9.00 abc	16.36 a	14.94 a	7.01 a	1.75 a	0.513 D	0.458 ns	2.85 t	56.42 l-p	160.67 s-w	35.12 o	5.05 n-s
PAN	0.57 s-w	19.32 k-w	9.00 x	5.59 rs	11.30 y	9.91 E	3.08 yz	0.51 EFG	0.478 G	0.378 B	3.46 gh	46.58 vmx	161.00 s-v	28.93 CD	7.42 bcd
ZIM	0.59 r-v	22.04 c-i	9.35 w	5.69 qrs	12.36 qr	10.54 A	3.49 wx	0.61 A-D	0.498 F	0.341 D	3.87 b	42.42 y	164.32 qrs	25.72 H	9.17 a
MSI	0.52 u-x	18.77 m-w	7.90 A	5.50 s	12.72 mno	9.34 G	2.74 A	0.43 G	0.593 kl	0.357 C	3.41 ij	42.09 y	143.64 C	29.30 zB	8.11 b
GUR	0.85 h-n	18.84 l-w	11.41 opq	6.48 j-r	12.24 rs	11.17 uv	3.92 r-u	0.73 t-z	0.569 pq	0.425 tu	2.91 st	60.56 ghi	176.08 klm	34.39 p	4.80 o-u
BAS	0.69 n-u	20.07 h-s	11.87 lmn	6.24 l-s	12.73 mn	11.41 qr	4.09 o-s	0.78 q-v	0.569 pg	0.451 o	3.22 o	59.14 i-l	190.22 bcd	31.09 tu	5.44 k-q
KOR	0.85 h-n	24.50 ab	13.10 ef	7.20 f-k	14.93 c	13.22 e	5.49 de	1.21 efg	0.540 wxy	0.414 xy	3.40 ijk	53.47 p-s	181.94 hij	29.39 zB	6.36 e-k
BOS	0.78 j-q	23.10 b-f	13.00 efg	7.30 f-j	14.47 d	12.99 g	5.30 ef	1.15 fgh	0.562 qrs	0.439 pq	3.16 p	56.28 l-q	178.08 jkl	31.60 s	5.62 j-p
BIR	0.69 n-u	18.50 p-w	11.00 rst	8.30 b-e	12.60 op	11.91 l	4.45 klm	0.88 mp	0.644 ef	0.522 f	2.23 C	59.46 h-k	132.53 D	44.86 d	3.75 v-y
JUL	1.14 def	21.00 e-p	11.60 no	7.90 d-g	13.50 j	12.44 j	3.67 uvw	0.66 x-D	0.592 klm	0.464 lm	2.66 w	55.24 n-r	146.84 BC	37.62 l	4.81 o-u
DRA	0.82 i-o	18.50 p-w	12.50 hi	6.80 h-o	12.60 op	11.63 o	4.24 l-p	0.82 o-s	0.629 gh	0.522 f	2.72 v	67.57 d	183.82 ghi	36.76 m	4.03 t-x
BAB	0.73 l-t	18.20 q-w	8.20 zA	9.20 ab	11.87 uv	11.11 vw	3.88 r-v	0.72 u-z	0.611 j	0.478 k	1.98 E	45.05 xy	89.13 F	50.55 a	4.39 f-w
SIT	0.78 j-q	18.00 r-w	11.00 r-u	7.10 g-l	12.03 t	11.20 tu	3.94 q-u	0.74 s-y	0.622 hi	0.503 g	2.54 y	61.11 f-i	154.93 xyz	39.44 i	4.15 s-w
SLA	1.16 cde	22.90 b-g	13.00 efg	6.90 h-n	14.27 ef	12.71 h	5.07 fgh	1.07 hi	0.555 stu	0.434 qrs	3.32 lm	56.77 k-n	188.41 c-f	30.13 xy	5.85 o-u
MIR	0.85 h-n	17.20 vw	11.30 pqr	6.70 h-p	11.73 vmx	10.92 x	3.74 t-w	0.68 v-B	0.635 fg	0.523 f	2.57 y	65.70 de	168.66 opq	38.96 j	3.91 u-x
KAU	0.72 m-t	18.70 n-w	8.70 xy	7.60 d-h	11.67 wx	10.73 z	3.62 vw	0.65 y-D	0.574 op	0.436 qr	2.46 z	46.52 vmx	114.47 E	40.64 h	5.29 l-r
RUZ	0.94 g-k	20.20 h-s	13.30 de	7.10 g-l	13.53 j	12.40 j	4.83 hij	1.00 ijk	0.614 ij	0.505 q	2.84 t	65.84 de	187.32 d-g	35.15 o	4.32 s-w
POD	0.85 h-n	21.10 e-n	11.00 r-u	7.20 f-k	13.10 kl	11.87 lm	3.76 t-w	0.69 v-A	0.562 qrs	0.410 rs	2.76 rs	45.90 st	126.62 zA	36.25 pq	6.01 jp

Note. Codes of members of *P. insititia* L. are marked in bold-italic. Mean values with different letters in a column differ significantly at $p \leq 0.05$ by LSD test. Means are initially designated by small letters and afterwards by capital letters due to great variation in means.

et al. (2022) who reported that all three linear dimensions of endocarps significantly varied among plum genotypes. The highest endocarp L was observed in 'CPT', 'KOR', 'CRB' (all belonging to European plum), 'TUR' and 'TURA' (both Damson plums) with no significant differences among them, whereas the lowest L was found in 'TRA' (Damson plum). Out of the total number of accessions, 50.91 % had $L > 20$ mm.

Regarding endocarp T, 'MUV' had the highest value, whereas the lowest was found in 'MSI'. In the case of endocarp W, a very high number of the largest and/or smallest values overlapped. The highest values were observed in the set of accessions such as 'MUV', 'MUD', 'BAB' and 'CPT' with no significant differences among them. In contrast, the lowest and similar values were found in 17 accessions (30.91 %) i. e. 'ARP', 'CPI', 'CRN', 'COK', 'MET', 'TRN', 'CRI', 'VOL', 'GBU', 'BJL', 'PIS', 'MCS', 'TRA', 'PAN', 'ZIM', 'MSI' and 'BAS'. Some of them, such as 'ARP', 'CPI', 'CRN' and 'MET', had identical mean W values.

Taking into account the absolute values of the three linear dimensions, the descending order was $L > T > W$, which is in accordance with the recommendations of morphometric analysis of plum endocarp proposed by Behre (1978) and Woldring (2000). Also, the values of endocarp dimensions obtained in our study were within the limits described by Van Zeist, Woldring (2000).

Following the procedure proposed by Caillavet and Souty (1950), values of all three dimensions (L, T and W) were transformed into the parameter denominated "size" or D_a and/or D_g . In the present study (Table 2), both D_a and D_g values significantly varied among accessions.

A similar finding applies to endocarp S and V, respectively. The lowest value for all four parameters was found at 'TRA' belonging to *P. insititia*. The highest endocarp "size", calculated as D_a and D_g , S and V, was observed in 'CPT' belonging to European plum. ANOVA results showed that the differences among accessions for these properties were clear and significant. Only two accessions (3.64 %) had endocarp S between 6 and 7 cm², whereas nine accessions (16.36 %) had S between 5 and 6 cm². Other accessions (80 %) had $S < 5$ cm². Otherwise, the knowledge of a specific surface area (S) could be a relevant tool to determine the shape. Other authors also report large variations in the endocarp size, S and V in plums (Sheikh et al., 2021), cherries (Pérez-Sánchez et al., 2010; Ganopoulos et al., 2015; Khadivi et al., 2022), peaches (das Graças Souza et al., 2016) and apricots (Gezer et al., 2002).

The fruit or stone (endocarp) shape is determined in terms of its ϕ . Moreover, ϕ is an expression of the shape of a solid related to that of a sphere of the same volume. In the current study, 'MUD' and 'MED' had similar and the highest ϕ values, whereas the lowest was observed in 'PAN' (Table 2).

In general, ϕ is used for determining the similarity of a fruit or a stone to a sphere. Hence, higher values of ϕ indicate the tendency of endocarps towards sphericity (Sheikh et al., 2021). The ϕ value more than 0.7 i. e. 70 % is assumed to be spherical (Garnayak et al., 2008). However, average ϕ values for accessions were much lower than 0.7 or 70 % with the exception of 'MUD', 'MED' and 'BEV'.

The shape parameters such as SI and E indicate the shape tendency of the endocarps. Both indexes in the present study showed high variability among accessions (Table 2). 'MUV'

and 'MUD' had the highest and similar SI values whereas the lowest were observed in 'MET'. The highest E value was in 'CRB', and the lowest, in 'BAB'. Lower values of these shape parameters indicate the tendency of endocarps to being flat and oblong in shape as previously reported by Sheikh et al. (2021) for plum kernels. Based on the values of the above endocarp shape indexes in this study and their comparison with the IBPGR recommendations on stone shapes (lateral view) (Cobianchi, Watkins, 1984), it can be said that elongated and ovate shapes dominate, and sporadically, rounded ones appear.

Endocarp shape indexes proposed by Behre (1978), Van Zeist, Woldring (2000) and Pollmann et al. (2005) of the evaluated plum accessions were presented in Table 2.

The ANOVA showed significant differences among accessions for all four indexes. The highest RS index (endocarps in lateral view) denominated as relative slenderness was found in 'MUD', 'MUV' and 'TRN' with no significant differences between them. The lowest and statistically similar values of this index were discovered in 'ZIM' and 'MSI', both belonging to *P. domestica*. Our values of this index for 45 accessions (81.82 %) were within the limits described previously (Van Zeist, Woldring, 2000; Pollmann et al., 2005), while 10 accessions (18.18 %) had slightly lower values than the minimum described by the above authors. Otherwise, the more slender the stone, the lower the index value (Depypere et al., 2007). The high variability of this index was previously described by Van Zeist, Woldring (2000).

With regard to RO index, which expresses the roundness of the endocarp in apical view, the highest value was observed in 'COK', and the lowest, in 'BAB', with 2.27-fold difference between them. The minimum and maximum values according to the descriptors proposed by Van Zeist, Woldring (2000) and Pollmann et al. (2005) for this index are 112.15 and 225.45, respectively, which was the case in our study with the exception of 'BAB' that had a much lower value than the minimum limit. In general, endocarps with strongly domed sides show a low RO value, while in rather flat stones, this value is relatively high and always more than 100 (Van Zeist, Woldring, 2000).

Similarly to previous indexes, DE index, which represents endocarps in ventral view, varied among and within accessions. It was the highest in 'BAB', and the lowest and statistically similar one was observed in 'ZIM' and 'CRB', all belonging to *P. domestica*. Our values were generally closer to the minimum values proposed by Van Zeist, Woldring (2000), which varied from 26.30 to 106.32.

PI significantly varied among and within accessions, 'ZIM' being the accession with the highest value, whereas the smallest value was found in 'MUD' with 3.38-fold difference between them. According to Pollmann et al. (2005), minimal and maximal values of this index ranged between 1.27 and 7.68, which was confirmed by our data. Depypere et al. (2007) reported that the PI index value was highly variable for *P. domestica* and *P. insititia*.

Evaluation of variability and mean values of properties between plum types

With regard to the variability of mean values of properties evaluated by means of the coefficients of variability (CV, %), the results showed that L, D_a , D_g and somewhat ϕ had a low

Table 3. Intraspecific variability for *P. domestica* and *P. insititia* expressed by means of the coefficients of variability (CV, %) and mean values for each property evaluated

Parameter	<i>P. domestica</i> L.		<i>P. insititia</i> L.	
	CV, %	Mean \pm SE	CV, %	Mean \pm SE
SW	36.61	0.84 \pm 0.15 a	55.90	0.78 \pm 0.22 b
L	9.88	20.48 \pm 1.00 a	18.57	19.28 \pm 1.79 b
T	12.88	11.49 \pm 0.74 a	17.22	11.20 \pm 0.96 b
W	13.80	6.92 \pm 0.44 a	17.92	6.80 \pm 0.50 a
D _a	8.04	13.01 \pm 0.52 a	15.83	12.42 \pm 0.98 b
D _g	8.88	11.71 \pm 0.51 a	15.79	11.31 \pm 0.89 b
S	18.77	4.32 \pm 0.41 a	30.53	4.14 \pm 0.63 b
V	29.37	0.86 \pm 0.13 a	45.43	0.82 \pm 0.19 a
ϕ	10.15	0.577 \pm 0.03 a	9.22	0.571 \pm 0.03 a
SI	13.64	0.45 \pm 0.03 b	17.84	0.47 \pm 0.04 a
E	16.18	3.03 \pm 0.24 a	15.47	2.97 \pm 0.23 b
RS	13.04	56.35 \pm 3.67 b	16.82	59.09 \pm 4.97 a
RO	13.14	167.87 \pm 11.03 a	7.19	165.91 \pm 5.96 b
DE	17.33	34.08 \pm 2.95 b	20.20	35.88 \pm 3.62 a
PI	25.99	5.56 \pm 0.72 a	29.09	4.84 \pm 0.70 b

Note. CV > 25 % indicates high variability; 10 % < CV < 20 % indicates normal variability; CV < 10 % indicates low variability (following Rasch, 1988; cited in Hübner, Wissemann, 2004). Mean values with different letters in a row differ significantly at $p \leq 0.05$ by LSD test.

variability (CV < 10 %) in *P. domestica*, and only ϕ and RO, in *P. insititia* (Table 3).

Parameters T, W, S, SI, E, RS, RO and DE in *P. domestica* and L, T, W, D_a, D_g, SI, E, RS and DE in *P. insititia* had a low-to-normal variability (10 % < CV < 20 %). Parameters SW, V and PI in both European and Damson plums, and S and DE in *P. insititia* were moderate to highly variable (CV > 25 %). It appears that the mean value of the coefficients of variation of all morphological levels for accessions belonging to *P. domestica* was significantly smaller (CV = 16.51 %) compared to the accessions belonging to *P. insititia* (CV = 22.20 %) (data not shown). Our results were in good agreement with the results found by Depypere et al. (2007) for several indexes such as W, T, DE, RS and PI.

Regarding mean values of properties evaluated for both European and Damson plums, the results from Table 2 showed that there were significant differences between them with the exception of W, V and ϕ . These three values were statistically similar for both plum types. Accessions belonging to *P. domestica* had higher mean values for SW, L, T, D_a, D_g, S, E, RO and PI than accessions belonging *P. insititia*. On the contrary, accessions belonging to *P. insititia* had higher SI, RS and DE mean values than representatives of European plum.

Correlations among variables and principal component analysis (PCA)

Relationships among 15 endocarp parameters were studied and Pearson's correlations were calculated and were presented graphically (Fig. 2). Significant correlations were found among most of the studied traits, but high values were noted only in some cases.

SW was significantly correlated with all parameters and indexes with the exception of RS and RO, indicating that accessions with a big stone tend to have greater endocarp dimensions and higher indexes values in general and *vice versa*. Hence, all parameters can be used to predict each other. Similar tendencies were observed in cherry plum (Heidari et al., 2022). However, the intensity of correlations between some parameters differed. Namely, strong positive correlations were observed between SW and T, W, D_a, D_g, S and V, whereas other correlations were weak, which shows that endocarps have some very similar properties and that their values are not greatly influenced by genotype. In addition, SW was negatively correlated with E and PI. These findings are in good agreement with the results obtained on hazelnut (Milošević T., Milošević N., 2017).

L, W and T were significantly correlated with endocarp indexes with the exception of L vs W, T vs E and/or DE. The absence of significant correlation between L and W is in agreement with the results of Kosina (2023) for *P. spinosa*. So, these traits were considered to be independent. Strong correlations were observed between L and D_a and between T and D_a, D_g, S and V, indicating that endocarps with higher L and T tend to have a greater endocarp size. Both D_a and D_g showed very strong mutual correlation, and also with endocarp S and V. However, relationships of D_a and D_g with other indexes were small and not significant. ϕ was strongly positively correlated with SI, RS and DE, but strongly negatively correlated with E and PI. Both S and V were significantly correlated with SW, endocarp dimensions and size parameters, whereas correlations with all indexes were minor and not significant.

There was an extremely strong mutual correlation between S and V indicating that an endocarp with higher S values tends

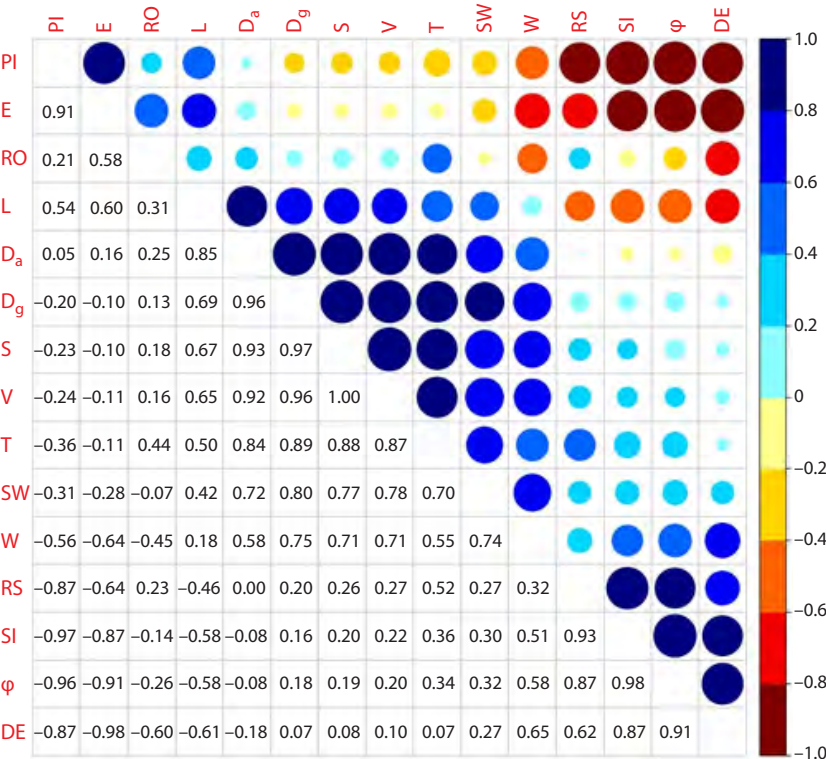


Fig. 2. Correlation matrix of Pearson's correlation coefficients (*r*) between the mean values of the endocarp parameters evaluated.

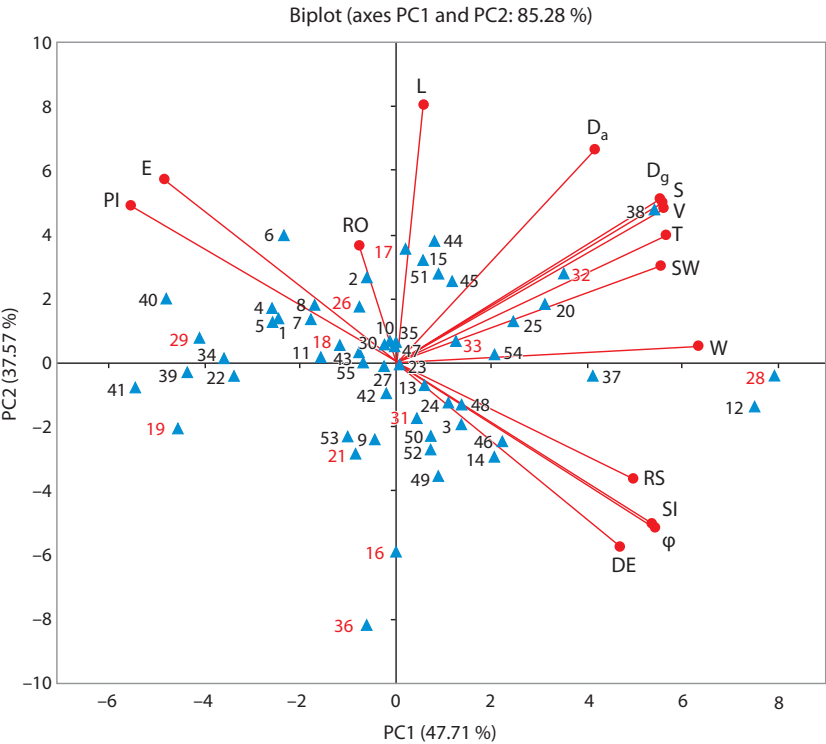


Fig. 3. Segregation of European and Damson plum accessions according to endocarp properties (linear dimensions, size and shapes) and their projections on the first (PC1) and second factors (PC2) of the component analysis.
See Table 1 for accession series numbers and accession codes; the series number in red represents accessions belonging to Damson plum.

to have higher *V* values and *vice versa*. On the other hand, there were no significant correlations between *V* and all shape indexes. *SI* was strongly negatively correlated with *E* and *PI*, and positively with *RS* and *DE*. Similar trends have been reported for European plum kernels (Sheikh et al., 2021). *E* vs *RS* and *DE* were negatively correlated, and positively correlated with *RO* and *PI*. *RS* showed negative correlations with *PI*, and positive with *DE*, whereas both *RO* and *PI* were negatively correlated with *DE*.

Principal component analysis (PCA), as a statistical tool, is performed to reduce the number of effective traits and to identify groups. In the current study, using the 15 analyzed parameters, the first three principal components accounted for 96.67 % of the total variance. PC1 explained 47.71 % of the total variation, while PC2 explained 37.57 % and PC3 explained 11.39 % (Fig. 3). According to the correspondence between the PCA and the original properties and eigenvectors, *SW*, *T*, *D_g*, *phi*, *S*, *V*, and *SI* made the largest contributions to PC1 with positive values, while *PI* had a negative contribution. As a result, genotypes such as ‘MUD’, ‘PLA’, ‘CDU’, ‘MUV’, ‘TURa’, ‘DRO’, ‘MAG’, ‘CPT’ and ‘RUZ’ tended to exhibit higher *SW*, *T*, *D_g*, *phi*, *S*, *V*, and *SI* values but lower *PI* values. In contrast, accessions like ‘ARP’, ‘CPI’, ‘CRN’, ‘CRD’, ‘MET’, ‘MOR’, ‘CRI’, ‘GBU’, ‘PIS’, ‘MCS’, ‘PAN’, ‘ZIM’ and ‘MSI’ displayed the opposite trend.

For PC2, positive values were associated with *L*, *D_a*, and *E*, whereas *DE* contributed negatively. This suggests that genotypes like ‘BPZ’, ‘CRB’, ‘COK’, ‘MAR’, ‘POZ’, ‘TUR’, ‘DUR’, ‘KOR’, ‘BOS’, and ‘SLA’ exhibited higher *L*, *D_a*, and *E* values, while accessions such as ‘BEL’, ‘KAP’, ‘OPI’, ‘PET’, ‘TRN’, ‘VOL’, ‘BJS’, ‘TRS’, ‘TRA’, ‘GUR’, ‘BIR’, ‘SIT’ and ‘MIR’ showed lower values of these parameters.

Finally, *RS* and *RO* contributed to the positive values of PC3, whereas *W* contributed negatively, indicating that genotypes such as ‘BJL’, ‘MED’, ‘BAS’, and ‘DRA’ were predisposed to higher *RS* and *RO* values, while accessions like ‘SAR’, ‘BEV’, ‘JUL’, ‘BAB’, ‘KAU’, and ‘POD’ tended to have lower *W* values.

Conclusions

The stones of accessions belonging to *P. domestica* L. and *P. insititia* L. showed characteristic differences in size and shape features, which greatly facilitate the identification of genotypes or accessions. Each of

them could be identified by means of dimensions and morphological features of the endocarps. In the present study, most endocarp parameters were found to be very useful for further taxonomic research, based on their low variability in both *P. domestica* and *P. insititia*. However, some parameters such as SW, V and PI exhibited a high variability and we suggest omitting their use for taxonomic purposes in some cases or for them to be used in a limited way. In general, the examined parameters varied less in accessions belonging to European plum compared to Damson plum genotypes. In addition, the mean values of SW, L, T, D_a, D_g, S, E, RO and PI were higher in *P. domestica* type compared to *P. insititia*, while the mean values of W, V and ϕ were similar. Others, such as SI, RS and DE were higher in Damson plum. However, due to overlapping ranges in most cases within and between plum types and accessions, the use of one or two endocarp parameters is not satisfactory for discrimination between Eurasian plum taxa. The multivariate analysis as a statistical tool can be useful for higher quality dispersion, segregation and determination of plum accessions, but in these analyses, the overlapping of values of endocarp morphological parameters also occurs. Finally, based on our results obtained on dry endocarps and the results of other researchers who experimented with fresh stones, we recommend full hydration of dried endocarps, as this restores the original dimensions and shape of the sampled endocarps.

References

- Bawari S., Sah A.N., Tewari D. Discovering the antiurolithiatic potential of wild Himalayan cherry through *in vitro* and preclinical investigations. *S Afr J Bot.* 2022;145:218-227. doi 10.1016/j.sajb.2021.01.020
- Behre E.-K. Formenkreise von *Prunus domestica* L. von der Wikingerzeit bis in die frühe Neuzeit nach Fruchtsteinen aus Haithabu und Alt-Schleswig. *Ber Deutsch Bot Ges.* 1978;91(1):161-179. doi 10.1111/j.1438-8677.1978.tb03640.x (in German)
- Burger P., Terral J.-F., Raus M.-P., Ivorra S., Picq S. Assessing past agrobiodiversity of *Prunus avium* L. (Rosaceae): a morphometric approach focused on the stones from the archaeological site Hôtel-Dieu (16th century, Tours, France). *Veget Hist Archaeobot.* 2012; 20(5):447-458. doi 10.1007/s00334-011-0310-6
- Caillaud H., Souty J. Monographie des principales variétés de pêcheurs. Bordeaux: Sté bordelaise d'imprimerie, 1950
- Carrillo-López A., Yahia E.M. Morphology and anatomy. In: Yahia E.M. (Ed.) Postharvest Physiology and Biochemistry of Fruits and Vegetables. Woodhead Publishing; Elsevier, 2019;113-130. doi 10.1016/B978-0-12-813278-4.00006-3
- Cobianchi D., Watkins R. (Eds) Descriptor List for Plum and Allied Species. IBPGR (International Board for Plant Genetic Resources), 1984
- das Graças Souza A., Smiderle O.J., Spinelli V.M., de Souza R.O., Bianchi V.J. Correlation of biometrical characteristics of fruit and seed with twinning and vigor of *Prunus persica* rootstocks. *J Seed Sci.* 2016;38(4):322-328. doi 10.1590/2317-1545v38n4164650
- de Figueiredo A.K., Baümler E., Riccobene C., Nolasco S.M. Moisture-dependent engineering properties of sunflower seeds with different structural characteristics. *J Food Eng.* 2011;102(1):58-65. doi 10.1016/j.jfoodeng.2010.08.003
- Depypere L., Chaerle P., Mijnsbrugge K.V., Goetghebeur P. Stony endocarp dimension and shape variation in *Prunus* section *Prunus*. *Ann Bot.* 2007;100(7):1585-1597. doi 10.1093/aob/mcm260
- Depypere L., Chaerle P., Mijnsbrugge K.V., Goetghebeur P. Classification trees and plant identification: a case study of European *Prunus* section *Prunus* taxa. *Belg J Bot.* 2009;142(2):163-176. doi 10.2307/41427184
- Drkenda P., Kurtović M. Pomological characteristics of autochthonous plum genotypes (*Prunus domestica*) from Bosnia and Herzegovina. *Acta Hort.* 2012;968:81-83. doi 10.17660/ActaHortic.2012.968.10
- Fıratlıgil-Durmuş E., Šárka E., Bubník Z., Schejbal M., Kadlec P. Size properties of legume seeds of different varieties using image analysis. *J Food Eng.* 2010;99(4):445-451. doi 10.1016/j.jfoodeng.2009.08.005
- Ganopoulos I., Moysiadias T., Xanthopoulou A., Ganopoulou M., Avramidou E., Aravanopoulos F.A., Tani E., Madesis P., Tsafaris A., Kazantzis K. Diversity of morpho-physiological traits in worldwide sweet cherry cultivars of GeneBank collection using multivariate analysis. *Sci Hort.* 2015;197:381-391. doi 10.1016/j.scienta.2015.09.061
- Garnayak D.K., Pradhan R.C., Nalk S.N., Bhatnagar N. Moisture-dependent physical properties of jatropha seed (*Jatropha curcas* L.). *Ind Crops Prod.* 2008;27(1):127-129. doi 10.1016/j.indcrop.2007.09.001
- Gezer İ., Haciseferoğlu H., Demir F. Some physical properties of Hacıhaliloğlu apricot pit and its kernel. *J Food Eng.* 2002;56(1): 49-57. doi 10.1016/S0260-8774(02)00147-4
- Glišić I., Milošević N., Tomić J., Milinković M., Đorđević M., Marić S., Radičević S., Popović B. Biological and pomological characteristics of autochthonous plum cultivars collected in western Serbia. *J Pomol.* 2023;57(215-216):7-16. doi 10.18485/pomology.2023.57.215_216.1
- González-García E., Marina M.L., García M.C. Plum (*Prunus domestica* L.) by-product as a new and cheap source of bioactive peptides: extraction method and peptides characterization. *J Funct Foods.* 2014;11:428-437. doi 10.1016/j.jff.2014.10.020
- Gunes M. Some local plum varieties grown in Tokat province. *Pak J Appl Sci.* 2003;3(5):291-295. doi 10.3923/jas.2003.291.295
- Heidari P., Sahebi M., Azadvari E., Lawson S., Rezaei M., Khadivi A. Morphological variability of indigenous cherry plum (*Prunus divaricata* Ledeb.) accessions. *Eur J Hort Sci.* 2022;87(1):1-8. doi 10.17660/eJHS.2022/002
- Hübner S., Wissemann V. Morphometrische Analysen zur Variabilität von *Prunus spinosa* L. – Populationen (*Prunoideae*, Rosaceae) im Mittleren Saaletal, Thüringen. *Forum Geobot.* 2004;1:19-51. doi 10.3264/FG.2004.1215 (in German)
- Jačimović V., Božović Đ., Lazović B., Adakalić M., Ljutica S. Autochthonous plum cultivars grown in the region of Gornje Polimlje. *J Pomol.* 2011;45(175-176):109-115 (in Serbian with English abstract)
- Khadivi A., Mirheidari F., Moradi Y., Paryan S. Morphological and pomological diversity of wild *Prunus microcarpa* Boiss. germplasm. *BMC Plant Biol.* 2022;22(1):185. doi 10.1186/s12870-022-03572-2
- Kosina R. A contribution to the variability of *Prunus spinosa* L. in the vicinity of the mediaeval Castle Kolno, S-W Poland. *Biodiv Res Conserv.* 2023;70:13-18. doi 10.14746/biocr.2023.70.2
- Mansouri A., Mirzabe A.H., Ráufi A. Physical properties and mathematical modeling of melon (*Cucumis melo* L.) seeds and kernels. *J Saudi Soc Agric Sci.* 2017;16(3):218-226. doi 10.1016/j.jssas.2015.07.001
- Mijnsbrugge K.V., Depypere L., Chaerle P., Goetghebeur P., Breyne P. Genetic and morphological variability among autochthonous *Prunus spinosa* populations in Flanders (northern part of Belgium): implications for seed sourcing. *Plant Ecol Evol.* 2013;146(2):193-202. doi 10.5091/plecevo.2013.762
- Milošević N., Glišić I., Lukić M., Đorđević M. Biological and pomological characteristics of autochthonous plum cultivars in agroecological conditions of Čačak. In: Proceedings of 22nd Symposium on Biotechnology with International Participation, Faculty of Agronomy. Čačak, 2017;271-280 (in Serbian with English abstract)

- Milošević T., Milošević N. Phenotypic diversity of autochthonous European (*Prunus domestica* L.) and Damson (*Prunus insititia* L.) plum accessions based on multivariate analysis. *Hortic Sci.* 2012; 39(1):8-20. doi 10.17221/99/2011-HORTSCI
- Milošević T., Milošević N. Determination of size and shape features of hazelnuts using multivariate analysis. *Acta Sci Pol Hortorum Cultus.* 2017;16(5):49-61. doi 10.24326/asphc.2017.5.6
- Milošević T., Milošević N. Plum (*Prunus* spp.) breeding. In: Al-Khayri J.M., Jain M.S., Johnson D.V. (Eds) *Advances in Plant Breeding Strategies: Fruits*. Springer International Publishing AG, 2018;165-215. doi 10.1007/978-3-319-91944-7_5
- Milošević T., Milošević N., Mladenović J. Diversity of plums belonging to *P. domestica* L., *P. insititia* L. and *Prunus* × *rossica* Erem. Tree vigour, yielding and fruit quality attributes. *Sci Hortic.* 2023;320: 112220. doi 10.1016/j.scienta.2023.112220
- Mohsenin N.N. *Physical properties of plants and animal materials*. New York: Gordon & Breach Science Publishers, 1986;5-8;616-647
- Munder S., Argyropoulos D., Muller J. Class-based physical properties of air-classified sunflower seeds and kernels. *Biosyst Eng.* 2017; 164:124-134. doi 10.1016/j.biosystemseng.2017.10.005
- Nielsen J., Olrik D.C. A morphometric analysis of *Prunus spinosa*, *P. domestica* ssp. *insititia*, and their putative hybrids in Denmark. *Nord J Bot.* 2001;21(4):349-363. doi 10.1111/j.1756-1051.2001.tb00778.x
- Pérez-Sánchez R., Gómez-Sánchez M.A., Morales-Corts M.R. Description and quality evaluation of sweet cherries cultured in Spain. *J Food Qual.* 2010;33(4):490-506. doi 10.1111/j.1745-4557.2010.00339.x
- Plainfossé H., Burger P., Verger-Dubois G., Azoulay S., Fernandez X. Design methodology for the development of a new cosmetic active based on *Prunus domestica* L. leaves extract. *Cosmetics.* 2019;6:8. doi 10.3390/cosmetics6010008
- Pollmann B., Jacomet S., Schlumbaum A. Morphological and genetic studies of waterlogged *Prunus* species from the Roman *vicus Tasgetium* (Eschenz, Switzerland). *J Archaeol Sci.* 2005;32(10):1471-1480. doi 10.1016/j.jas.2005.04.002
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2021. Available at: <https://www.R-project.org/>. Accessed on June 13, 2024
- Ropelewska E. Diversity of plum stones based on image texture parameters and machine learning algorithms. *Agronomy.* 2022;12:762. doi 10.3390/agronomy12040762
- Sarigu M., Oscar G., Lo Bianco M., Uccesu M., d'Hallewin G., Loi M.C., Venora G., Bacchetta G. Phenotypic identification of plum varieties (*Prunus domestica* L.) by endocarps morpho-colorimetric and textural descriptors. *Comput Electron Agr.* 2017;136:25-30. doi 10.1016/j.compag.2017.02.009
- Šebek G. Selection of some autochthonous plum cultivars seedling rootstocks in the region of North Montenegro. *J Hyg Eng Des.* 2013;4:117-121
- Sheikh M.A., Saini C.S., Sharma H.K. Computation of design-related engineering properties and fracture resistance of plum (*Prunus domestica*) kernels to compressive loading. *J Agric Food Res.* 2021;3: 100101. doi 10.1016/j.jafr.2021.100101
- Van Zeist W., Woldring H., Plum (*Prunus domestica* L.) varieties in late- and post-medieval Groningen: the archaeobotanical evidence. *Palaeohistoria.* 2000;39/40:563-576
- Watkins R. Cherry, plum, peach, apricot and almond. In: Simmonds N.W. (Ed.) *Evolution of Crop Plants*. London: Longman, 1976;242-247
- Woldring H. On the origin of plums: a study of sloe, damson, cherry plum, domestic plums and their intermediate forms. *Paleohistoria.* 2000;39/40:535-562

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


Received July 16, 2024. Revised October 10, 2024. Accepted October 14, 2024.

doi 10.18699/vjgb-25-07

Diurnal fluctuations in the content of soluble sugars and the expression of the *TAI* and *LIN6* invertase genes and the *STP1* sugar transporter gene in the leaves of the tomato (*Solanum lycopersicum* L.)

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

Federal Research Center "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia

 michel7753@mail.ru

Abstract. The content of hexoses (fructose, glucose) essential for the fruit of the tomato (*Solanum lycopersicum* L.) is regulated by the joint activity of sucrose hydrolysis enzymes (including invertases), invertase inhibitors, and sugar transporters. In addition to fruit taste, soluble sugars are closely related to the stress resistance of the tomato plant. In this work, we determined the diurnal dynamics of the content of soluble sugars (sucrose, fructose and glucose) and the expression of genes for sucrose hydrolysis enzymes (vacuolar invertase *TAI*, cell wall invertase *LIN6*) and the hexose transporter (*STP1*) in the leaves of the tomato variety Korneevsky. It was shown that both the amount of sugars and the level of transcripts of the *TAI*, *LIN6* and *STP1* genes depend on the circadian rhythm and correspond to the biological processes occurring in the plant at different periods of the day. The content of sucrose and hexoses changes in a similar way during the day. At the beginning of the light phase, the concentration of sugars is minimal, at the end it has the highest daily values; at the beginning of the dark phase, it shows a residual increase and then decreases towards the end of the phase. *In silico* analysis of organ-specific expression of *TAI*, *LIN6* and *STP1* in *S. lycopersicum* cv. Micro-Tom showed the presence of mRNA of all three genes in all tissues. The *TAI* gene was expressed most strongly in ripe fruits, while the level of *LIN6* and *STP1* transcripts was extremely low. The level of *TAI* mRNA in the leaves was ~2 times higher than that of *LIN6* and ~27 times higher than that of *STP1*. Analysis using qRT-PCR of the diurnal dynamics of *TAI*, *LIN6* and *STP1* expression in the cv. Korneevsky leaves showed that all three genes were expressed at all points analyzed. Fluctuations in their expression levels occur in a similar manner: mRNA levels reach peak values in the middle of the light and dark phases. The results obtained are important for understanding the functions of invertases and sugar transporters in the tomato plant, and can be used in predicting the stress resistance of plants in tomato breeding.

Key words: tomato; *Solanum lycopersicum* L.; soluble sugars; invertases; hexose transporter; gene expression; circadian rhythm.


For citation: Filyushin M.A., Shchennikova A.V., Kochieva E.Z. Diurnal fluctuations in the content of soluble sugars and the expression of the *TAI* and *LIN6* invertase genes and the *STP1* sugar transporter gene in the leaves of the tomato (*Solanum lycopersicum* L.). *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025;29(1):55-60. doi 10.18699/vjgb-25-07

Funding. The work was financially supported by the Russian Science Foundation (grant No. 24-16-00043, biochemical and expression analysis) and the Ministry of Science and Higher Education of the Russian Federation (obtaining seed material and preparing plants for analysis).

Циркадные колебания содержания растворимых сахаров и экспрессии генов инвертаз *TAI*, *LIN6* и транспортера сахаров *STP1* в листьях растения томата (*Solanum lycopersicum* L.)

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

Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия

 michel7753@mail.ru

Аннотация. Содержание основных для плода томата (*Solanum lycopersicum* L.) гексоз (фруктозы, глюкозы) регулируется совместной активностью ферментов гидролиза сахарозы (включая инвертазы), ингибиторов инвертаз и транспортеров сахаров. Кроме вкусовых качеств, растворимые сахара тесно связаны со стрессоустойчивостью растений томата. В настоящей работе была определена суточная динамика содержания растворимых сахаров (сахарозы, фруктозы и глюкозы), а также экспрессия генов ферментов гидролиза сахарозы (вакуолярной ин-

вертазы *TAI*, инвертазы клеточной стенки *LIN6*) и транспортера гексоз (*STP1*) в листьях растений томата сорта Корнеевский. Было показано, что количество сахаров и уровень транскриптов генов *TAI*, *LIN6* и *STP1* зависят от циркадного ритма и соответствуют биологическим процессам, протекающим в растении в разное время суток. Содержание сахарозы и гексоз в течение суток изменяется сходным образом. В начале световой фазы концентрация сахаров минимальна, в конце – имеет наибольшие дневные значения; в начале темновой фазы демонстрирует остаточный рост и затем снижается к концу фазы. Анализ *in silico* органоспецифичной экспрессии *TAI*, *LIN6* и *STP1* у сорта Микро-Том *S. lycopersicum* показал наличие их транскриптов во всех тканях. Ген *TAI* экспрессировался наиболее активно в спелых плодах, тогда как уровень транскриптов *LIN6* и *STP1* в этих органах носил следовой характер. В листьях уровень мРНК *TAI* был выше, чем таковой *LIN6* и *STP1*, в ~2 и ~27 раз соответственно. Анализ с помощью ПЦР-РВ суточной динамики экспрессии генов *TAI*, *LIN6* и *STP1* в листьях растений сорта Корнеевский показал, что гены экспрессируются во всех проанализированных временных точках. Колебания уровня экспрессии генов происходят сходным образом: уровень мРНК достигает пиковых значений в середине световой и темновой фаз. Полученные результаты важны для понимания функций инвертаз и транспортеров сахаров в растении томата и могут быть использованы в селекции при прогнозировании стрессоустойчивости растений.

Ключевые слова: томат; *Solanum lycopersicum* L.; растворимые сахара; инвертазы; транспортер гексоз; экспрессия генов; циркадный ритм.

Introduction

During photosynthesis, the plant accumulates assimilates – vital organic compounds utilized for respiration, maintenance of cell metabolism, growth and development. The sucrose is the main transport form of photoassimilates during distribution throughout the plant (Lemoine et al., 2013). The signals for distribution are provided by sucrose and glucose molecules, the number of which influences the regulation of genes active at a particular stage of plant development (Koch, 2004; González et al., 2005; Roland et al., 2006).

After delivery to storage organs (flowers, fruits, tubers, etc.), sucrose is utilized by being broken down into glucose and fructose by sucrose synthases (reversible hydrolysis) or invertases (irreversible hydrolysis); the functions of the latter are highly variable and closely related to localization in different cellular compartments (Roitsch, González, 2004). The expression level of invertase genes depends on the type of tissue/organ, stage of plant development, and external stimuli, including exposure to stress factors, phytohormones, elicitors, etc. (Roitsch, González, 2004; Koch, 2004; Proels, Roitsch, 2009). Cell wall invertases are involved in the distribution of sucrose in plant tissues and organs and signal transduction, while vacuolar invertases are involved in sugar accumulation and osmoregulation (Roitsch, González, 2004; González et al., 2005). Hexoses formed during sucrose hydrolysis enter the cells of storage tissues via hexose transporters (Proels, Roitsch, 2009).

Tomato (*Solanum lycopersicum* L.) is one of the most popular vegetable crops in the world. Tomato fruits accumulate glucose and fructose during ripening (Beckles et al., 2012). These hexoses affect the degree of fruit sweetness, and their amount is regulated by the combined activity of sucrose synthases (reversible hydrolysis of sucrose), invertases (irreversible hydrolysis of sucrose), invertase inhibitors, and sugar transporters (Kawaguchi et al., 2021; Wang B. et al., 2021). In addition to determining an important fruit quality trait, soluble sugars significantly contribute to the regulation of stress resistance of tomato plants during growth and development (Proels, Roitsch, 2009). Increased carbohydrate influx to the stressed area provides energy for protective reactions, stimulation of carbohydrate accumulation and modulation of the expression of the corresponding genes,

including genes for invertases and sugar transporters (Proels, Roitsch, 2009). The coordinated induction of the monosaccharide transporter and cell wall invertase genes observed under biotic stress (Fotopoulos et al., 2003; Voegelé et al., 2006) supports the important role of apoplastic sucrose degradation in mediating defense responses. Regardless of the process, both the metabolism and distribution of sugars, and thus, the expression of the genes involved, are controlled by circadian rhythms, in particular diurnal variations in the intensity of biological processes (González et al., 2005; Roland et al., 2006).

Among tomato invertases, the most significant roles belong to the cell wall invertase LIN6 (Wiv-1) (Proels, Roitsch, 2009) and the vacuolar invertase TAI (other names AI, PAIN1) (Elliott et al., 1993). The LIN6 enzyme is important for plant growth and response to various stress factors, and is also under the control of key circadian oscillator factors (Proels, Roitsch, 2009; Zhang et al., 2013). TAI activity is associated mainly with sucrose hydrolysis in the tomato fruit (Slugina et al., 2017). No data on the possible dependence of *TAI* gene expression on circadian rhythms in tomato plants have been found, but the dependence is assumed, since it has been shown using the example of vacuolar invertase of sugar beet *Beta vulgaris* (González et al., 2005).

Hexose transporters in tomato include the most well-known proteins STP1 and STP2. Knockdown of the genes encoding them reduces the amount of glucose and fructose in the roots, which reduces the plant's sensitivity to nematode infestation (Warnock et al., 2016). Of particular note is the *STP1* gene, which is considered a target of domestication in the tomato genome; lack of *STP1* expression negatively affects the efficiency of fruiting and the amount of sugars in the fruit (Wang Y. et al., 2023). The available literature does not mention the presence of a dependence of *STP* expression on the circadian oscillator; however, a connection is assumed, as for invertases.

In this study, we analyzed the dependence of the expression levels of vacuolar invertase *TAI*, cell wall invertase *LIN6* and hexose transporter *STP1* genes, as well as the content of soluble sugars (sucrose, glucose, fructose), on the diurnal rhythm during tomato plant growth. The results obtained are important for understanding the functions of invertases and sugar transporters in tomato plants.

Material and methods

The study was carried out on cv. Korneevsky *S. lycopersicum*, bred at the Federal Scientific Vegetable Center (FSVC, Moscow Region, Russia). The cultivar is mid-season, suitable for greenhouse conditions, produces fruits with high sugar content, and is resistant to various stress factors, including fluctuations in temperature and photoperiod (accession number 8262334, <https://gossortrf.ru/registry/>).

Tomato plants of cv. Korneevsky were grown to the fruiting stage in 2023 in greenhouse conditions of the FSVC. The collected seeds were used (2024) to obtain seedlings at the 5–7 leaf phase (experimental climate control facility, Federal Research Center of Biotechnology, Russian Academy of Sciences) under conditions of a long photoperiod and optimal temperature (day/night – 16 h/8 h, 23 °C/21 °C; light phase from 7:00 to 23:00; illumination 190 µM/(m²·s)). Leaf samples (two plants for each analysis point) were collected during the day at six time points: 1 h before (6:00) and after (8:00) the onset of the light phase; in the middle of the light phase (15:00); 1 h before (22:00) and after (24:00) the onset of the dark phase; in the middle of the dark phase (3:00). The tissue was ground in liquid nitrogen and used for analysis of the content of soluble sugars and the expression level of invertase (*TAI*, Solyc03g083910; *LIN6*, Solyc10g083290) and hexose transporter (*STP1*, Solyc02g079220) genes.

The concentration (mg/100 g of fresh weight (FW)) of soluble sugars (sucrose, glucose, fructose) was determined using the Enzytec™ Liquid Sucrose/D-Glucose and Enzytec™ Liquid D-Glucose/D-Fructose tests (R-Biopharm AG, Germany).

A preliminary profiling of *TAI*, *LIN6*, and *STP1* expression in different tomato plant organs was performed *in silico* using transcriptome data for the cv. Micro-Tom *S. lycopersicum* (TomExpress database; <http://tomexpress.toulouse.inra.fr/login/>) (Zouine et al., 2017). Data were visualized using online HeatMapper (<http://www2.heatmapper.ca/expression/>) based on FPKM (Fragments per kilobase of transcript per million mapped fragments; TomExpress) values.

To analyze gene expression using quantitative real-time PCR (qRT-PCR), total RNA was isolated from 0.2–0.5 g of collected leaf material and purified from DNA impurities (RNeasy Plant Mini Kit, RNase-free DNase set, QIAGEN, Germany). Based on total RNA preps, cDNA was synthesized (GoScript Reverse Transcription System, Promega, USA). The concentration of RNA and cDNA preparations was determined fluorimetrically (Qubit® Fluorometer, Thermo Fisher Scientific, USA; Qubit RNA HS Assay Kit, Invitrogen, USA). Primers for qRT-PCR were designed by structural analysis of the *S. lycopersicum* genes and their transcripts (available in the databases: <https://www.ncbi.nlm.nih.gov/>; <https://solgenomics.net/>) using NCBI-BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and MEGA 7.0 (<https://www.megasoftware.net/>). Primers for qRT-PCR were designed for the *LIN6* (5'-ttcc gatgctcaaggtcaag-3', 5'-cacgttttctctccagcacca-3') and *STP1* (5'-tgctcagaatgtgtctatgctc-3', 5'-gtgctctctgtattgtatgg-3') genes. For the *TAI* gene, we used the primers developed earlier (Slugina et al., 2017). The qRT-PCR reaction mixture included cDNA (3 ng), specific primers and “2.5 × Reaction mixture for qRT-PCR in the presence of SYBR Green I and

ROX” (Synthol LLC, Russia). qRT-PCR was performed in a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, USA); program: 5 min, 95 °C, 40 cycles (15 s, 95 °C; 40 s, 60 °C). Data were normalized to the expression of two reference genes, *Expressed* (SGN-U346908) and *actin2/7* (NM_001330119.1) (Efremov et al., 2020). The analysis was carried out in two biological and three technical replicates.

The results of the analysis of sugar content (mg/100 g FW) and gene expression were statistically processed using GraphPad Prism v. 8 (GraphPad Software Inc., USA; <https://www.graphpad.com/scientific-software/prism/>). The significance ($p < 0.05$) of differences between the values obtained for the time points was determined using Two-way ANOVA (“multiple comparisons, corrected with the Bonferroni test”).

Results

In this study, using the cv. Korneevsky tomato (*S. lycopersicum*) as an example, we characterized daily changes in the content of soluble sugars, as well as the expression pattern of genes of two key invertases (vacuolar, *TAI*; cell wall, *LIN6*) and sugar transporter (*STP1*) in the leaves of a plant in the active stage of vegetative growth and development (5–7 leaves).

Since *S. lycopersicum* is day-neutral species (Lifschitz, Eshed, 2006), a long photoperiod (16 h/8 h – day/night), typical for summer, was used in the work. Time points for measuring the target indicators were selected considering the boundaries between daily phases (one hour before and after the onset of light (points 6:00, 8:00) and darkness (22:00, 24:00)), as well as the middle of the dark (3:00) and light (15:00) periods.

At these six points, the content of soluble sugars (glucose, fructose and sucrose) was measured (Fig. 1). It was shown that the amount of all analyzed sugars is minimal at the beginning of the light phase. By the middle of the day (15:00) it increases by ~1.2–2.0 times, and by the end of the photoperiod (22:00), it sharply rises by ~15 times (*vs.* 8:00), reaching the daily maximum. At the beginning of the dark phase (24:00), the content of hexoses continues to grow (by ~1.3–1.6 times *vs.* 22:00); however, in the second half of the dark period (3:00, 6:00), it decreases (by ~1.5–2.0 times *vs.* 24:00). At the beginning of the light phase (8:00), the number of hexoses decreases even more sharply (by ~50–60 times *vs.* 6:00) (Fig. 1).

The sucrose content changes during the day in a manner similar to that of hexoses, with the exception of the 24:00 point (a decrease of ~2 times *vs.* 22:00) and a smoother decrease compared to hexoses at the 8:00 point (~18 times *vs.* 6:00) (Fig. 1).

Thus, it was shown that the content of the analyzed soluble sugars is minimal at the beginning and maximal at the end of the photoperiod, while in the dark phase their amount is more constant.

Next, the expression pattern of the *TAI*, *LIN6*, and *STP1* genes was characterized. A preliminary analysis of the organ-specific expression pattern of these genes was performed *in silico* (Fig. 2). It was shown that transcripts of all three genes are present in vegetative tissues and in the growing fruit (including the breaker (BR) stage of reaching the final size and the beginning of the color change). In the ripening fruit

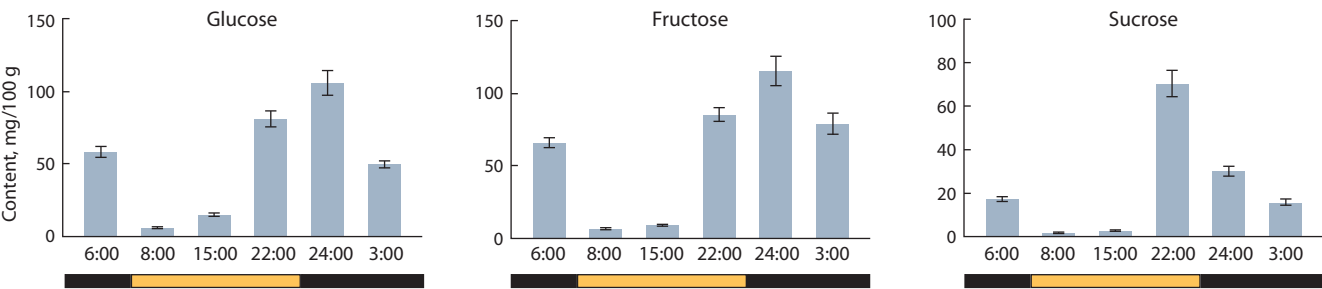


Fig. 1. Diurnal dynamics of the content (mg/100 g FW) of glucose, fructose and sucrose in the leaves of a cv. Korneevsky tomato plant (*S. lycopersicum*). The values of sugar concentration at the analyzed time points differ ($p < 0.05$), with the exception of glucose (6:00 vs. 3:00; 8:00 vs. 15:00), fructose (6:00 vs. 22:00; 8:00 vs. 15:00; 6:00 vs. 3:00; 22:00 vs. 3:00), and sucrose (6:00 vs. 3:00; 8:00 vs. 15:00).

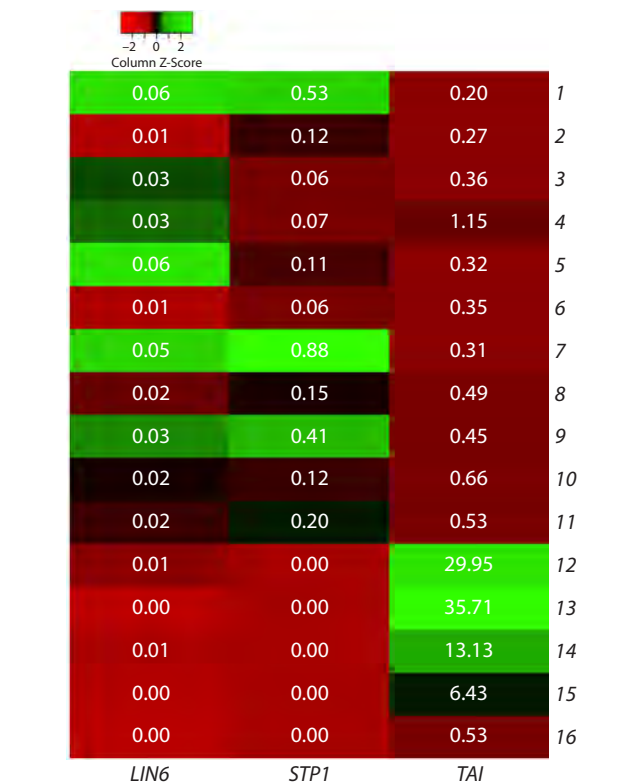


Fig. 2. Graphical visualization (heat map) of *TAI*, *LIN6*, and *STP1* gene expression data in cv. Micro-Tom tomato (*S. lycopersicum*) constructed using TomExpress transcriptome data (Zouine et al., 2017).

Organs analyzed: root (1); leaf (2); bud (3); flower at the anthesis stage (4); fruit, 4 days post anthesis (dpa) (5); pulp (6) and skin (7) of the fruit (10 dpa); pulp (8) and skin (9) of the fruit (35 dpa); pulp (10) and skin (11) of the fruit (38 dpa, BR); pulp (12) and skin (13) of the fruit (41 dpa, OR); pulp (14) and skin (15) of the fruit (44 dpa, RR); mature seeds (16). The rectangles show FPKM values rounded to the second decimal place.

(the orange (OR) and red ripe (RR) fruit stages), only trace numbers of *LIN6* and *STP1* transcripts (0.002–0.0129 FPKM) were detected.

At the same time, *TAI* was expressed most intensely in these tissues. The peak of *TAI* expression (26.95–35.71 FPKM) corresponded to the OR stage of fruit ripening, where the level of gene transcripts was ~2 and ~27–36 times higher than in the fruit at the RR stage (6.43–13.13 FPKM) and in vegetative tissues/growing fruit (including the BR stage; 0.20–1.15 FPKM),

respectively. In other reproductive tissues – buds, flowers and RR fruit seeds, the number of *TAI* transcripts was ~6–12, 20–40 and 50 times higher, respectively, compared to *LIN6* and *STP1* (Fig. 2).

Despite the obvious specificity of *TAI* activity to the ripe fruit, its expression level in vegetative organs (0.20–1.15 FPKM) was, on average, higher than that of *LIN6* (0.01–0.06 FPKM) and *STP1* (0.06–0.88 FPKM). In whole leaf tissue, the number of *TAI* transcripts was ~2 and ~27 times higher than that of *LIN6* and *STP1*, respectively. The *LIN6* expression level in the leaves was the lowest (~12 times lower than *STP1*) (Fig. 2).

Thus, *in silico* profiling of gene expression showed that in all plant organs, the activity of the vacuolar invertase gene *TAI* significantly exceeds that of the cell wall invertase gene *LIN6* and the sugar transporter gene *STP1*. Moreover, *TAI* expression is highest in the reproductive organs, especially in the storage tissues of the ripe fruit at the OR and RR stages.

Next, in the same leaf samples of cv. Korneevsky used for the analysis of sugar content, the expression levels of the *TAI*, *LIN6* and *STP1* genes were determined (using qRT-PCR) at six time points during the day. As a result, it was shown that all three genes are expressed at all six time points. On average, the highest relative transcript levels were observed for *LIN6*, and the lowest, for *STP1*. The expression level of *TAI* (contrary to expectations based on *in silico* data, Fig. 2) was an order of magnitude lower than that of *LIN6*, and only ~3–4 times higher than the level of *STP1* transcripts (Fig. 3).

Overall, the change in the diurnal expression dynamics was similar for all three genes: *TAI*, *LIN6* and *STP1*. The transcript level increased significantly (by ~7 (*TAI*), ~7,000 (*LIN6*) and ~128 (*STP1*) times) from the 8:00 to the 15:00 point of the light phase. Then it decreased less sharply towards its end (~1.2–2.5 times, 22:00 vs. 15:00) and the beginning of the dark phase (~2–3.5 times, 24:00 vs. 22:00). By the middle of the night, the gene expression was upregulated (by ~2–5 times, 3:00 vs. 24:00), and by the end of the dark phase, it decreased by ~2.5 (*TAI*), ~7,000 (*LIN6*), and ~183 (*STP1*) times (6:00 vs. 3:00) (Fig. 3).

Thus, the diurnal dynamics of the *TAI*, *LIN6* and *STP1* expression level was similar, but fluctuations in the case of *LIN6* and *STP1* were significantly more pronounced than those of the *TAI* gene. Nevertheless, the pre-dawn and early afternoon expression level of *LIN6* and *STP1* was extremely low, while that of *TAI* was relatively constant and notable.

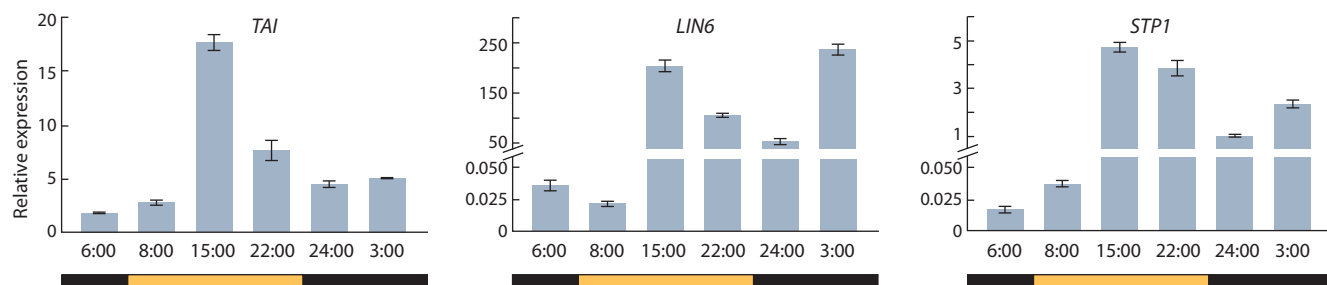


Fig. 3. Expression patterns of the *TAI*, *LIN6* and *STP1* genes based on qRT-PCR data.

The relative transcript levels for each gene at the analyzed time points differ significantly ($p < 0.05$), with the exception of *TAI* (6:00 vs. 8:00; 24:00 vs. 3:00), *LIN6* (6:00 vs. 8:00), and *STP1* (6:00 vs. 8:00).

Discussion

Soluble mono- and disaccharides have a significant impact on plant growth and development (Proels, Roitsch, 2009; Lemoine et al., 2013). Their content is regulated (besides glycan synthesis/degradation) through sucrose hydrolysis by sucrose synthases, invertases and invertase inhibitors, as well as through transfer between tissues using transporters (Kawaguchi et al., 2021; Wang B. et al., 2021).

Soluble sugars play an important role in all developmental processes in plant species, including tomato plants (Proels, Roitsch, 2009). Moreover, under stress, the influx of carbohydrates to the affected areas increases, which provides energy for a protective response, including the coordinated stimulation of carbohydrate accumulation and the expression of invertase and sugar transporter genes (Fotopoulos et al., 2003; Voegelé et al., 2006; Proels, Roitsch, 2009; Bolouri Moghaddam, Van den Ende, 2013).

In any process occurring in a plant involving soluble sugars, both the sugar content and the intensity of expression of the corresponding genes are characterized by synchronous cyclical oscillations during the day under the control of a circadian oscillator (González et al., 2005; Rolland et al., 2006).

In this study, we analyzed the diurnal dynamics of the concentration of soluble sugars (sucrose, glucose, and fructose) in the leaves of cv. Korneevsky tomato seedlings. The measurement points covered the border periods between the dark and light phases (6:00, 8:00, 22:00, 24:00) and the middle of the phases (15:00, 3:00). At the same points, we determined the expression of the genes of vacuolar invertase (*TAI*), cell wall invertase (*LIN6*), and hexose transporter (*STP1*), the role of which in tomato sugar metabolism is most important (Elliott et al., 1993; Proels, Roitsch, 2009; Warnock et al., 2016; Slugina et al., 2017).

The resulting diurnal profile of sugar content (Fig. 1) is consistent with known active sucrose synthesis in the daytime phase of photosynthesis, as well as with the diurnal cycle of sugar accumulation/intake due to the day/night synthesis/degradation of transient starch (Haydon et al., 2011). During the light phase, sucrose, glucose and fructose gradually accumulate (Fig. 1). Some of the glucose is presumably utilized for the synthesis of transient starch, and at the same time, sugars are released from the leaves (as sources of sugars) into storage organs (in our case, into the roots of the seedlings). By the end of the day, the amount of sugars reaches its highest

values, and in the dark phase it tends to decrease and then is maintained at a more or less constant level (Fig. 1), due to the pause in sucrose synthesis and the activation of the transient starch degradation (Koch, 2004; Haydon et al., 2011).

The main result of the *in silico* characterization of *TAI*, *LIN6*, and *STP1* expression (Fig. 2) is the confirmation of the important role of the vacuolar invertase gene *TAI* in sucrose hydrolysis in the ripe fruit as a storage organ, shown earlier (Elliott et al., 1993; Slugina et al., 2017). Furthermore, the higher transcript level of *TAI* compared to *LIN6* (Fig. 2) assumes a greater importance of *TAI* (than that of *LIN6*) for sucrose hydrolysis in vegetative tissue as well. Nevertheless, the significant transcript number of the cell wall invertase gene *LIN6* in vegetative tissues indicates the known importance of *LIN6* for plant vegetative growth (Proels, Roitsch, 2009; Zhang et al., 2013). Also, the presence of transcripts of the *LIN6* gene and the sugar transporter gene *STP1* in vegetative tissues (Fig. 2) is consistent with the previously shown involvement of these genes in the plant stress response (Proels, Roitsch, 2009; Warnock et al., 2016). At the same time, the extremely low number of *STP1* transcripts in ripe fruits of cv. Micro-Tom (Fig. 2), given the high sugar content in the fruits of this cultivar and the shown direct relationship between the expression level of this gene and the amount of sugars in the fruits (Wang Y. et al., 2023), suggests that even low activity of the *STP1* gene is sufficient to implement this relationship.

Subsequent analysis of the diurnal dynamics of *TAI*, *LIN6*, and *STP1* expression in the leaves of cv. Korneevsky plants showed that the levels of all three genes change in a similar manner and in association with the circadian rhythm (Fig. 3). The diurnal dynamics of *TAI* expression is consistent with the previously demonstrated diurnal dynamics for the *B. vulgaris* vacuolar invertase gene (González et al., 2005): both genes reach peak expression in the middle of the light phase.

Unlike *TAI*, the cell wall invertase gene *LIN6* has another expression maximum – in the middle of the night (Fig. 3). Moreover, contrary to *in silico* data (Fig. 2), the level of *LIN6* transcripts was an order of magnitude higher than that of *TAI* (Fig. 3). A possible reason for the discrepancy may be older plants (compared to our study) and, therefore, older leaves with large vacuoles in the cells taken into the transcriptome analysis, the results of which are presented in the *in silico* database we used (Zouine et al., 2017). We tested seedlings at the 5–7 leaf stage, the young leaves of which contained small

vacuoles in the cells, which suggests more active apoplastic processes of sucrose hydrolysis and sugar transport. This is also supported by significantly more pronounced (by an order of magnitude) diurnal fluctuations in the expression of *LIN6* and *STP1* compared to *TAI* (Fig. 3).

Conclusion

In this study, the diurnal dynamics of the content of soluble sugars and the expression of genes encoding sucrose hydrolysis enzymes (invertase genes *TAI* and *LIN6*) and sugar transfer proteins (*STP1* transporter gene) in tomato seedlings of cv. Korneevsky was determined. It was shown that both sugars and the transcript level of *TAI*, *LIN6* and *STP1* depend on the circadian rhythm and correspond to biological processes occurring in the plant at different periods of the day. The results obtained are important for understanding the functions of invertases and sugar transporters in the tomato plant, and can be used to predict plant stress resistance in tomato breeding.

References






- Beckles D.M., Hong N., Stamova L., Luengwilai K. Biochemical factors contributing to tomato fruit sugar content: a review. *Fruits*. 2012;67:49-64. doi 10.1051/fruits/2011066
- Bolouri Moghaddam M.R., Van den Ende W. Sweet immunity in the plant circadian regulatory network. *J. Exp. Bot.* 2013;64(6):1439-1449. doi 10.1093/jxb/ert046
- Efremov G.I., Slugina M.A., Shchennikova A.V., Kochieva E.Z. Differential regulation of phytoene synthase PSY1 during fruit carotenogenesis in cultivated and wild tomato species (*Solanum* section *Lycopersicon*). *Plants*. 2020;9:1169. doi 10.3390/plants9091169
- Elliott K.J., Butler W.O., Dickinson C.D., Konno Y., Vedvick T.S., Fitzmaurice L., Mirkov T.E. Isolation and characterization of fruit vacuolar invertase genes from two tomato species and temporal differences in mRNA levels during fruit ripening. *Plant Mol. Biol.* 1993;21(3):515-524. doi 10.1007/BF00028808
- Fotopoulos V., Gilbert M.J., Pittman J.K., Marvier A.C., Buchanan A.J., Sauer N., Hall J.L., Williams L.E. The monosaccharide transporter gene, *AtSTP4*, and the cell-wall invertase, *Atβfruct1*, are induced in *Arabidopsis* during infection with the fungal biotroph *Erysiphe cichoracearum*. *Plant Physiol.* 2003;132(2):821-829. doi 10.1104/pp.103.021428
- González M.C., Roitsch T., Cejudo F.J. Circadian and developmental regulation of vacuolar invertase expression in petioles of sugar beet plants. *Planta*. 2005;222(2):386-395. doi 10.1007/s00425-005-1542-4
- Haydon M.J., Bell L.J., Webb A.A. Interactions between plant circadian clocks and solute transport. *J. Exp. Bot.* 2011;62(7):2333-2348. doi 10.1093/jxb/err040
- Kawaguchi K., Takei-Hoshi R., Yoshikawa I., Nishida K., Kobayashi M., Kusano M., Lu Y., Ariizumi T., Ezura H., Otagaki S., Matsumoto S., Shiratake K. Functional disruption of cell wall invertase inhibitor by genome editing increases sugar content of tomato fruit without decrease fruit weight. *Sci. Rep.* 2021;11(1):21534. doi 10.1038/s41598-021-00966-4
- Koch K. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr. Opin. Plant Biol.* 2004;7:235-246. doi 10.1016/j.pbi.2004.03.014
- Lemoine R., La Camera S., Atanassova R., Dédaldéchamp F., Allario T., Pourtau N., Bonnemain J.L., Laloi M., Coutos-Thévenot P., Maurousset L., Faucher M., Girousse C., Lemonnier P., Parrilla J., Durand M. Source-to-sink transport of sugar and regulation by environmental factors. *Front. Plant Sci.* 2013;4:272. doi 10.3389/fpls.2013.00272
- Lifschitz E., Eshed Y. Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J. Exp. Bot.* 2006;57(13):3405-3414. doi 10.1093/jxb/erl106
- Proels R.K., Roitsch T. Extracellular invertase LIN6 of tomato: a pivotal enzyme for integration of metabolic, hormonal, and stress signals is regulated by a diurnal rhythm. *J. Exp. Bot.* 2009;60(6):1555-1567. doi 10.1093/jxb/erp027
- Roitsch T., González M.-C. Function and regulation of plant invertases: sweet sensations. *Trends Plant Sci.* 2004;9:606-613. doi 10.1016/j.tplants.2004.10.009
- Rolland F., Baena-Gonzalez E., Sheen J. Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu. Rev. Plant Biol.* 2006;57:675-709. doi 10.1146/annurev.arplant.57.032905.105441
- Slugina M.A., Shchennikova A.V., Kochieva E.Z. TAI vacuolar invertase orthologs: the interspecific variability in tomato plants (*Solanum* section *Lycopersicon*). *Mol. Genet. Genomics*. 2017;292(5):1123-1138. doi 10.1007/s00438-017-1336-y
- Voegelé R.T., Wirsén S., Möll U., Lechner M., Mendgen K. Cloning and characterization of a novel invertase from the obligate biotroph *Uromyces fabae* and analysis of expression patterns of host and pathogen invertases in the course of infection. *Mol. Plant Microbe Interact.* 2006;19(6):625-634. doi 10.1094/MPMI-19-0625
- Wang B., Li N., Huang S., Hu J., Wang Q., Tang Y., Yang T., Asmutola P., Wang J., Yu Q. Enhanced soluble sugar content in tomato fruit using CRISPR/Cas9-mediated *SlINVINH1* and *SlVPE5* gene editing. *PeerJ*. 2021;9:e12478. doi 10.7717/peerj.12478
- Wang Y., Shi C., Ge P., Li F., Zhu L., Wang Y., Tao J., Zhang X., Dong H., Gai W., Wang F., Ye Z., Grierson D., Xu W., Zhang Y. A 21-bp InDel in the promoter of *STP1* selected during tomato improvement accounts for soluble solid content in fruits. *Hortic. Res.* 2023;10(3):uhad009. doi 10.1093/hr/uhad009
- Warnock N.D., Wilson L., Canet-Perez J.V., Fleming T., Fleming C.C., Maule A.G., Dalzell J.J. Exogenous RNA interference exposes contrasting roles for sugar exudation in host-finding by plant pathogens. *Int. J. Parasitol.* 2016;46(8):473-477. doi 10.1016/j.ijpara.2016.02.005
- Zhang Y.L., Zhang A.H., Jiang J. Gene expression patterns of invertase gene families and modulation of the inhibitor gene in tomato sucrose metabolism. *Genet. Mol. Res.* 2013;12(3):3412-3420. doi 10.4238/2013.January.24.1
- Zouine M., Maza E., Djari A., Lauvernier M., Frasse P., Smouni A., Pirrello J., Bouzayen M. TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. *Plant J.* 2017;92(4):727-735. doi 10.1111/tpj.13711

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

Received July 15, 2024. Revised September 17, 2024. Accepted September 17, 2024.


doi 10.18699/vjgb-25-08

Chickpea diversity driven by transposon insertion polymorphism

V.A. Stanin¹, M.A. Duk ^{1, 2}, A.A. Kanapin ¹, A.A. Samsonova ¹, S.Yu. Surkova¹, M.G. Samsonova ¹ 

¹ Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

² Ioffe Institute of the Russian Academy of Sciences, St. Petersburg, Russia

 m.g.samsonova@gmail.com

Abstract. Chickpea is the second most important legume crop, which is used as a food by people in different parts of the world due to its high nutritive value. Omics technologies have revolutionized the characterization of chickpea genetic diversity by considering single-nucleotide polymorphisms, while structural variants and transposons have been overlooked. The specific contribution of transposons to the phenotypic diversification of crop species is still poorly documented, therefore its characterization is important. We focused on landraces collected before the “green revolution”, as they are a valuable source of species diversity and can be used to broaden the genetic base of modern cultivars. Analyzing 190 chickpea genomes, we found 42,324 new transposon insertion sites from 83 families and showed that such sites are highly polymorphic. Most insertions were caused by mobilization of retrotransposons (67 % of insertions); among DNA transposons, the highest number of insertions was found for the superfamilies *MuDR*, *PIF*, *hAT*, *CMC*, and *TcMar*. We also demonstrated an uneven distribution of insertion sites along chromosomes. Analysis of the localization of transposon insertion sites relative to genes and their structural elements has shown that the largest number of insertions in all transposon superfamilies falls on introns and the smallest, on exons. We also showed that transposon insertion sites, which until recently have been overlooked by population genomics, are an important factor that diversifies phenotypes and can be used in GWAS as markers replacing SNPs. Comparative analysis of landraces collected in different geographic regions showed that the Ethiopian accessions have many unique transposon insertion sites. Our results highlight the unique role of transposon mobilization in chickpea diversification and have important implications for breeding improved chickpea varieties adapted to global climate change.

Key words: chickpea; transposons; polymorphism; landraces; GWAS; adaptation.

For citation: Stanin V.A., Duk M.A., Kanapin A.A., Samsonova A.A., Surkova S.Yu., Samsonova M.G. Chickpea diversity driven by transposon insertion polymorphism. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2025;29(1): 61-71. doi 10.18699/vjgb-25-08

Funding. This research was funded by Russian Science Foundation, grant number 22-46-02004.


Acknowledgements. The authors would like to thank the Peter the Great St. Petersburg Polytechnic University Centre for Supercomputing (scc.spbstu.ru) for providing excellent computational resources and support for this project.

Разнообразие нута, обусловленное полиморфизмом вставок транспозонов

В.А. Станин¹, М.А. Дук ^{1, 2}, А.А. Канапин ¹, А.А. Самсонова ¹, С.Ю. Суркова¹, М.Г. Самсонова ¹ 

¹ Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

² Физико-технический институт им. А.Ф. Иоффе Российской академии наук, Санкт-Петербург, Россия

 m.g.samsonova@gmail.com

Аннотация. Нут – важная зернобобовая культура, которая используется народонаселением разных частей света в пищу в силу высокой ценности. Применение омических технологий позволило охарактеризовать генетическое разнообразие нута, обусловленное однонуклеотидными полиморфизмами, тогда как структурные варианты и инсерции транспозонов выпали из поля зрения исследователей. Поэтому характеристика состава мобилома индивидуальных сортов нута и оценка его влияния на фенотипическую изменчивость и адаптацию актуальны. В фокусе нашего внимания были староместные сорта, собранные до «зеленой революции», поскольку они являются ценным источником видового разнообразия и могут быть использованы для расширения генетической базы современных сортов. Проанализировав 190 геномов нута, мы обнаружили 42 324 сайта инсерции транспозонов 83 семейств. Большинство инсерций (67 %) вызваны мобилизацией ретротранспозонов. Из ДНК-транспозонов наибольшее число инсерций найдено для суперсемейств *MuDR*, *PIF*, *hAT*, *CMC* и *TcMar*. Продемонстрирована неравномерность распределения сайтов инсерции вдоль хромосом. Анализ локализации сайтов инсерции транспозонов относительно генов показал, что наибольшее количество вставок у всех суперсемейств транспозонов приходится на интроны, наименьшее – на экзоны. Мы также показали, что сайты встройки транспозонов, которые до недавнего времени находились вне поля зрения популяционной геномики, являются важным фактором, диверсифицирующим фенотипы, что позволяет использовать их в полногеномном поиске ассоциаций в качестве маркеров наряду с однонуклеотидными полиморфизмами. Сравнительный анализ мобиломов сортов из

разных географических регионов выявил существенное отличие эфиопских образцов от образцов других групп, собранных в Индии, Узбекистане, Турции, Средиземноморье, на юге России и в Ливане. Совокупность полученных нами данных и результатов – ценный ресурс, который может быть использован в качестве отправной точки для селекции улучшенных сортов нута, адаптированных к различным климатическим условиям.

Ключевые слова: нут; транспозоны; полиморфизм; староместные сорта; GWAS; адаптация.

Introduction

Chickpea is one of the most important food legumes grown in many parts of the world, including Asia, Africa, North and South America and Europe. It accounts for 15 % of the world's legume yield (Abbo et al., 2003; Jain et al., 2013). Chickpea is an important component of the diet for millions of people all over the world, providing protein, dietary fiber, unsaturated fatty acids, vitamins, macro and micronutrients.

Chickpea is grown mainly in arid and semi-arid regions on poor soils (de la Peña, Pueyo, 2012). In these regions, various abiotic stresses such as water scarcity, extreme temperatures, short growing season affect chickpea productivity. For example, drought reduces global chickpea yields by 50 % and losses due to temperature extremes go up to 20 % (Kaloki et al., 2019). In such a scenario, identification and/or development of high-yielding genotypes is critical. These new chickpea varieties need to be resilient to climate change and adapted to changing consumer demands, agricultural practices and a wider climatic range. However, current elite chickpea varieties have low genetic diversity and do not contain useful alleles associated with tolerance to biotic and abiotic stresses. Hence, a broader genetic base is required for continuous production of new varieties.

More primitive landraces collected before the “green revolution” are a valuable source of crop species diversity. Their use in plant breeding can lead to the development of resistant varieties with stable characteristics under unfavorable conditions. In the early 20th century, N.I. Vavilov systematically collected chickpea landraces, which are now stored at the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR) in St. Petersburg, Russia. This collection has been explored earlier to identify associations between SNPs and phenotypic traits using a single-locus genome-wide association study (Sokolkova et al., 2020).

Although the application of omics technologies has enabled large-scale characterization of germplasm, our understanding of the mechanisms underlying chickpea diversity is still limited. This situation is partly explained by the fact that until recently, for technical reasons, such studies have focused on the functional role of single nucleotide polymorphisms and short insertions/deletions (Varshney et al., 2019), while larger structural variants can account for a significant proportion of interspecific differences in DNA sequences. Most structural variants arise from the mobilization of transposons. Transposons constitute a significant part of the plant genome (Quesneville, 2020; Mhiri et al., 2022), and their movement leads to genome rearrangement, epigenetic silencing, and rewiring of gene networks (Bourque et al., 2018). Moreover, transposons are not randomly distributed in the genome and can serve as material for the emergence of new protein-coding genes and non-coding RNAs (Pulido, Casacuberta, 2023).

Transposons are a highly heterogeneous group that can be divided into two main classes depending on the mode of

transposition (Bourque et al., 2018; Quesneville, 2020). Class I transposons (retrotransposons) propagate via RNA intermediates, and their “copy-and-paste” transposition mechanism results in the doubling of element copies with each transposition cycle (Mhiri et al., 2022). As a result, retrotransposons with long terminal repeats (LTRs) can account for up to 80–90 % of the total transposon content and are the most abundant in plant genomes. Class II transposons (DNA transposons) are predominantly mobilized by a “cut-and-paste” mechanism, which usually does not result in an increase in transposon copy number. However, transposons such as *Helitrons* and *MITEs* can achieve high copy numbers in some genomes.

The distribution and accumulation of transposons is shaped by genetic drift and selection (Mhiri et al., 2022). New insertions usually have deleterious effects and are removed from the population. However, transposons can also undergo positive selection and promote adaptation (Niu et al., 2019). Transposons peak during periods of stress, allowing genomes to rearrange and rapidly diversify (Schrader and Schmitz, 2019).

Although associations of transposons with numerous agronomic traits are well documented (Catlin, Josephs, 2022), their contribution to crop phenotypic variability remains poorly understood (Akakpo et al., 2020; Alioto et al., 2020). Here, we investigated the chickpea mobilome composition by analyzing transposon insertions in 190 genomes of chickpea accessions from the VIR collection.

Material and methods

Plant material. 190 chickpea accessions from the collection of N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR, St. Petersburg, Russia) were used in this work. Of these, 22 accessions were elite varieties, and the remaining accessions were landraces collected by N.I. Vavilov during his expeditions in the 1920–1930s. Based on the geographical proximity of the collection sites, landraces were divided into seven groups: accessions collected in the Mediterranean (MED), Lebanon (LEB), southern Russia (RUS), Turkey (TUR), Uzbekistan (UZB), India (IND) and Ethiopia (ETH) (Fig. 1a).

Bioclimatic variables. We used latitude and longitude coordinates for chickpea sample collection regions to obtain values of nineteen bioclimatic variables (Supplementary Table S1)¹. Bioclimatic variables represent annual, seasonal and monthly averages and extremes of temperature and precipitation and are widely used in biogeographic analysis, climate change studies and ecological modelling. Data were downloaded from the WorldClim database version 1.4 (Hijmans et al., 2005), which contains information on climatic conditions recorded between 1960 and 1990. Values of the variables were extracted using the ‘raster’ package in R (<https://raster.org/raster/>) at

¹ Supplementary Tables S1–S13 are available at:
<https://vavilovj-icg.ru/download/pict-2025-29/appx3.xlsx>

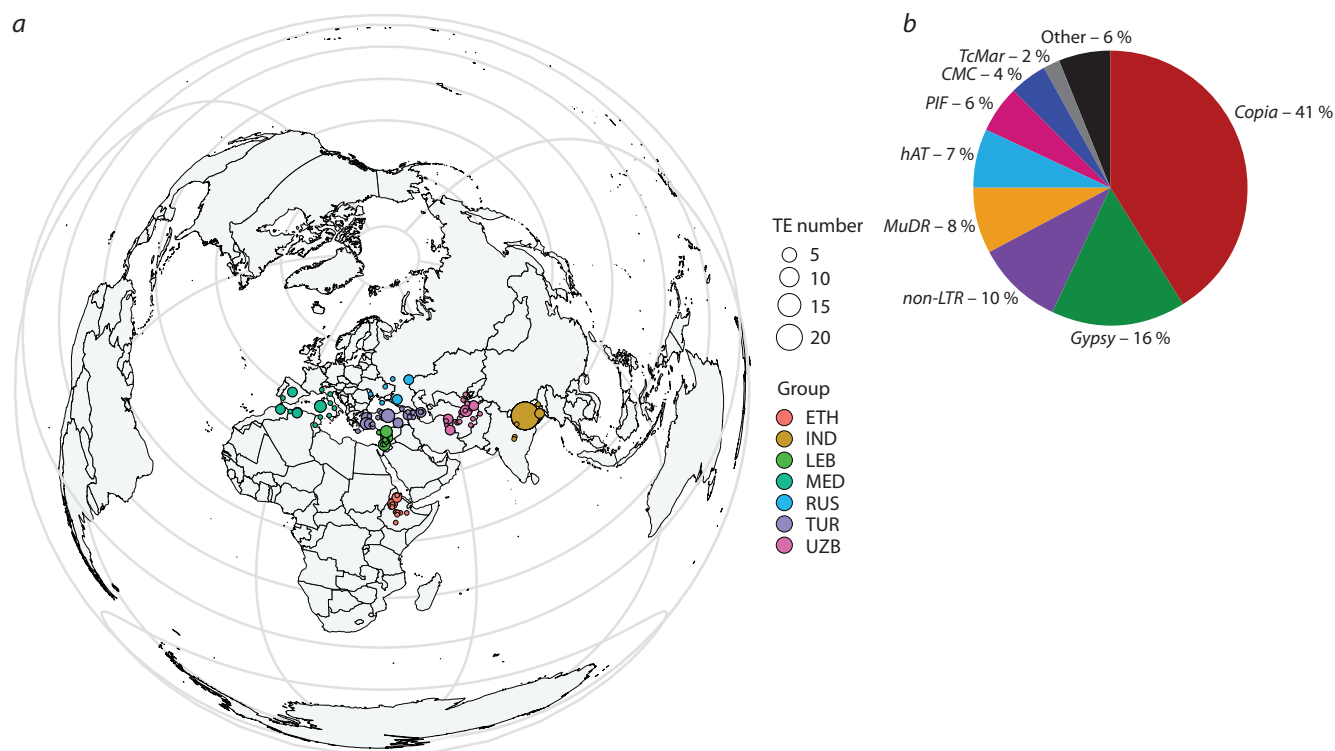


Fig. 1. Collection sites of landraces (a); proportion of insertion sites for the most represented transposon superfamilies (b).

a spatial resolution of 30 angular seconds, corresponding to approximately 1 square kilometer at the equator.

DNA sequencing and search for transposon insertions. The DNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA) was used to isolate DNA from leaves. DNA samples were sequenced at the Beijing Genomics Institute (BGI, China) using the Illumina protocol, generating 150-bp paired-end reads. A total of 7,700 Gbytes of raw data were obtained, comprising about 26 billion reads with an average of 25-fold coverage or about 37 Gbytes per sample. Reads were processed and aligned to the chickpea genome reference assembly ASM33114v1 (Varshney et al., 2013) using the bwa-mem software with the default parameters (Li H., Durbin, 2009). The search for transposon insertion sites and assessment of their representation were performed using the PoPoolationTE2 program (Kofler et al., 2011, 2016). PoPoolationTE2 requires reads mapped to a reference genome with masked transposon sequences and a set of such sequences. The transposon sequences can be either consensus sequences of families (e.g. from RepBase), or sequences that have been masked in the reference genome, or both. When reads are aligned to such a modified genome, transposon insertions will result in groups of discordant paired ends, where one read maps to the reference genome, and the other, to the transposon sequence, while correctly aligned paired ends indicate the absence of an insertion. Based on the position of the matched ends of the paired fragments, a physical stack track (pile-up) is generated. The physical coverages of overlapping paired ends are summed, resulting in a physical coverage track, the height of which reflects the number of paired ends that overlap the given position. Transposon insertion signatures are determined using a sliding window method, by scanning for

peaks in the physical coverage that confirm the presence of an insertion.

PoPoolationTE2 implements two fundamentally different analysis modes. In the separate mode, each sample/population is processed separately. This is similar to running the PoPoolationTE2 pipeline several times, for each bam file separately. In joint analysis (joint mode), the physical pileup tracks of different samples are combined and a joint pileup track is created. Transposon insertion signatures are identified on this joint pileup track. When identifying insertion signatures using the identifySignatures utility, a minimum average physical coverage parameter of three was used. Further, in the separate analysis, the signatures were filtered by the maximum frequency of other transposons in a given site and the maximum frequency of structural variants (rearrangements) in a given site. Both parameters were set equal to zero. In the joint analysis, filtering was performed only by the maximum frequency of other transposons in a given site, equal to 0.05. Validation of key insertions was performed using the Integrative Genomics Viewer program, which allows visualization of read alignment at the insertion site.

The search for hotspots of transposon insertions was performed by the PrimatR program (<https://github.com/daewoooo/primatR>). The hotspotter function was used, which compares (within a 50 kb window) the density distributions of randomly scattered points, the number of which is equal to the number of transposons in the genome and the transposon location densities obtained from the experiment. The higher the value of transposon density in a given region, the more extreme it is for the distribution of “random” densities, and, therefore, the lower the *p*-level of significance. Hot spots were defined by *p*-values less than $1e-8$.

Genetic data analysis. The population structure of the data was assessed using the ADMIXTURE v.1.3.0 program (Alexander et al., 2009). The Mann–Whitney–Wilcoxon test (Mann, Whitney, 1947) was used to compare groups.

Genome-wide association studies. Phenotyping of chickpea accessions was carried out at two VIR experimental stations, in Kuban and Astrakhan, as described earlier (Duk et al., 2024). 12 phenological and morphological traits were measured: plant height (PH), height of first pod (HFP), number of first order branches (NPB), number of second order branches (NSB), plant dry weight with pods (PWwP), pod weight per plant (PoW), pod number per plant (PoNP), 100 seeds weight (100SW), leaf size (LS), number of days from germination to flowering (DFst), flowering duration (DF), number of days from germination to full maturity (Dmat) (Table S2).

Phenotypic data from two experimental stations were quantile normalized. GWAS was performed using the FarmCPU, Blink, SUPER and MLM programs of the GAPIT3 package for R with parameters MAF = 0.05 and FDR = 0.9. In addition, the IIIVmrMLM program in Single_env mode with parameters svpal = 0.01 (Li M. et al., 2022a, b) was also used for association studies. The IIIVmrMLM model was designed to address methodological shortcomings in detecting all types of interactions between alleles, genes, and environments, and to unbiasedly estimate their genetic effects. As a multilocus MLM model, IIIVmrMLM estimates the effects of all genes and the effects of all interactions simultaneously. However, IIIVmrMLM is less computationally complex, since the calculation of a large number of variance components has been replaced by the calculation of only three estimates. In addition, all effects in IIIVmrMLM are estimated within a single multilocus model using the Bayesian expectation-maximization algorithm, and all non-zero effects are further assessed using a likelihood ratio test for significant associations. All this actually guarantees accurate detection of insertion regions, unbiased estimation of their effects and makes IIIVmrMLM a good choice for detecting associations between markers, traits and the environment.

Information on the coordinates of candidate genes, containing markers in genes or in 1-kb flanking regions, was obtained from the GFF file version 1 Cicer_arietinum_GA_v1.0.gene.gff, and functional description of genes was obtained from the Pulse Crop Database (<https://www.pulsedb.org/Analysis/1869759>).

Results

Composition of the chickpea mobilome

A total of 105 transposon families have been annotated in the chickpea reference genome (Varshney et al., 2013). To characterize novel transposon insertions in individual chickpea accessions, we analyzed whole-genome sequencing data from 190 samples, represented by 22 cultivated varieties and 168 landraces, which were divided into seven groups based on the sampling location (Fig. 1). A total of 42,324 new transposon insertion sites not represented in the reference genome were identified, with most sites being polymorphic and present in multiple accessions.

Transposons of polymorphic insertion sites belong to 83 families and thus likely constitute the majority of the

chickpea mobilome. Most insertions are due to mobilization of *Copia* (41 %) and *Gypsy* (16 %) retrotransposons (Fig. 1b, Supplementary Figure S1a)² and 10 % account for insertions due to mobilization of *non-LTR* retrotransposons. Five superfamily groups – *MuDR* (8 %), *PIF* (6 %), *hAT* (7 %), *CMC* (4 %) and *TcMar* (2 %) – make the main contribution to the number of insertions caused by DNA transposons (Fig. S1b–f, Table S3).

Polymorphic insertion sites are distributed unevenly along chromosomes (Fig. 2a, Table S4) and form 47 hotspots, with the lowest number of hotspots found in chromosomes 5 and 8. Sixteen hotspots contain exclusively retrotransposon insertions. *Copia* retrotransposon insertions were observed in all hotspots, and *hAT*, *MuDR*, *PIF*, and *CMC* DNA transposon insertions were observed in 60, 53, 23, and 47 % of the hotspots.

Chickpea mobilome landscape

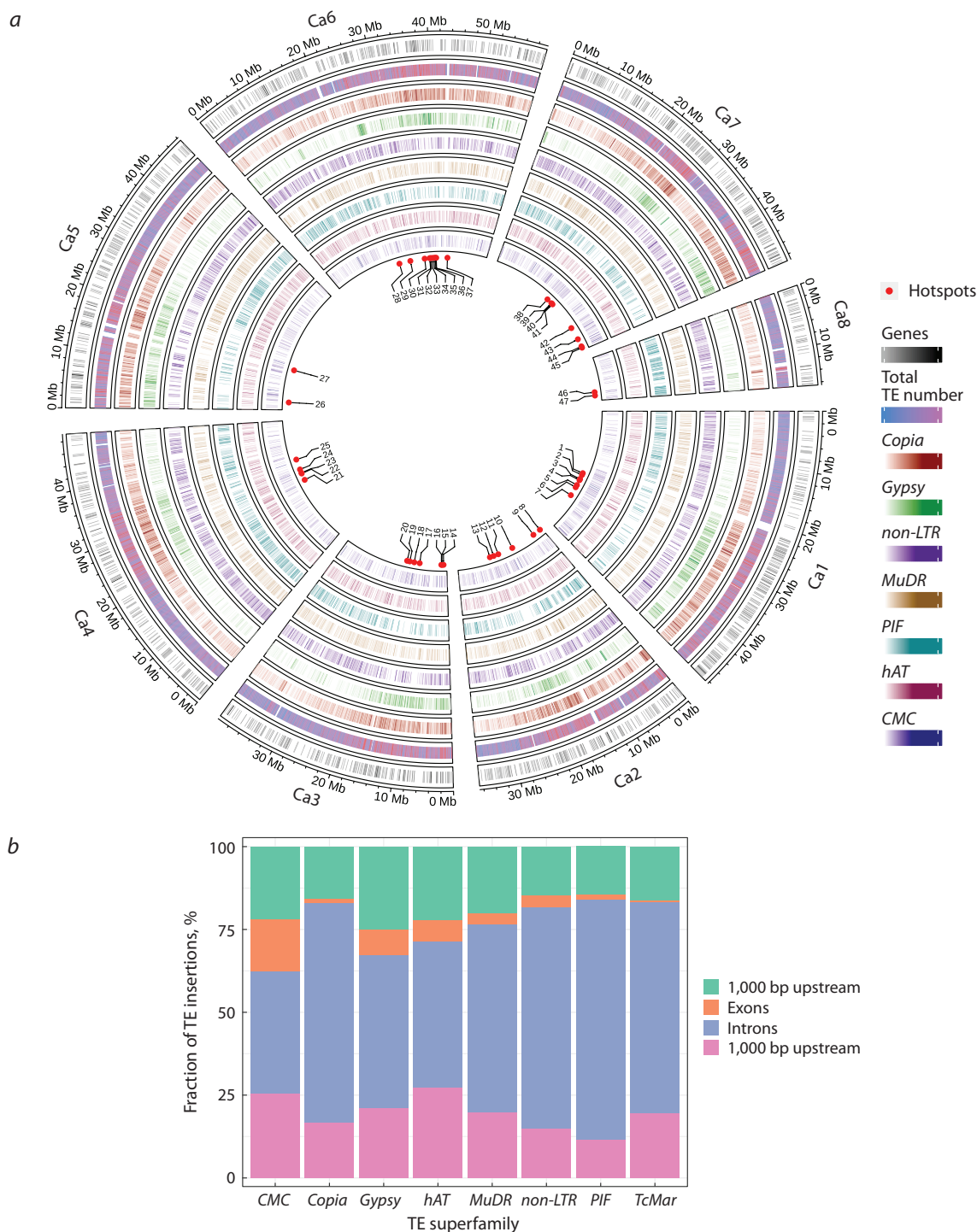
From 15 to 22 % of the insertion sites of *Copia* elements, as well as elements of the *MuDR*, *CMC*, and *hAT* superfamilies, are located in genes or within 1 kb-flanking regions of genes (Fig. 2b, Table S5). In *non-LTR* retrotransposons, such insertions are about a third of the total number (35.33 %), and in DNA transposons of the *TcMar* and *PIF* superfamilies, they constitute half of the total number of such insertions (44.11 % and 57.93 %, respectively) (Table S5). The highest number of insertions in all transposon superfamilies occurs in introns (Fig. 2b), and the lowest, in exons. The largest excess of insertions in introns compared to exonic insertions was observed for the *Copia* and *PIF* superfamilies (42 times), the smallest, for *Gypsy* (6 times) and *CMC* (2.36 times). In *TcMar*, almost all gene-specific insertion sites fall into introns, and the flanking regions contain 30 times fewer insertions compared to introns. The greatest excess of transposon insertions in the flanking regions of genes compared to exons was observed for elements of the *Copia*, *MuDR*, and *PIF* superfamilies (10, 6, and 4 times, respectively) (Table S5).

Transposon insertion site polymorphism as a new source of phenotypic variability

To more systematically assess whether polymorphic insertion sites are a potentially important source of phenotypic variability, we used them as markers in the search for associations with agronomically important traits assessed at Astrakhan and Kuban VIR experiment stations in 2022 (Duk et al., 2024). GWAS was performed separately for each trait measured at each station using the GAPIT3 package for R and the IIIVmrMLM program in Single_env mode.

GAPIT3 package found 12 associations with three phenotypic traits: duration of flowering, number of days from germination to full maturity, height of the first pod, with one association between the DF trait and the insertion of the *RTE-BovB* transposon at position Ca3_23488685 being found by two models (Table S6). The Ca3_1499163 and Ca6_24162635 insertion sites of the *Copia* and *RC Helitron* transposons associated with the height of the first pod are localized in the 5'-flanking regions of the *Ca_19414* and *Ca_11043* genes. These genes encode ribosomal protein S29 and late embryogenesis abundant protein, respectively.

² Supplementary Figures S1–S4 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Stanim_Engl_29_1.pdf



84 associations with Astrakhan station data and 114 associations with Kuban station data were found using the IIIVmrMLM program (Table S7). Three transposon insertion sites turned out to be polymorphic, in particular, Ca3_27767370, an insertion of the *PIF-Harbinger* transposon into the *Ca_08130* gene (Table S8). This insertion was associated with pod weight per plant at the Kuban station and with days from germination to full maturity at the Astrakhan station. 47 transposon insertion sites were located in the genes or in

their vicinity over a size of 1 kb. In most cases, however, these genes encoded proteins with unknown functions, and only 28 genes were functionally annotated (Table S9). An interesting example is the association of the *hAT_Charlie* transposon at position Ca6_31416746 with maturation time (Fig. 3a). This transposon is localized upstream of the *Ca_15174* gene, encoding transcription factor from the CCHC(Zn) family (Fig. 3b). In alfalfa *Medicago truncatula*, deletion of the gene encoding such a transcription factor strongly reduces seed size,

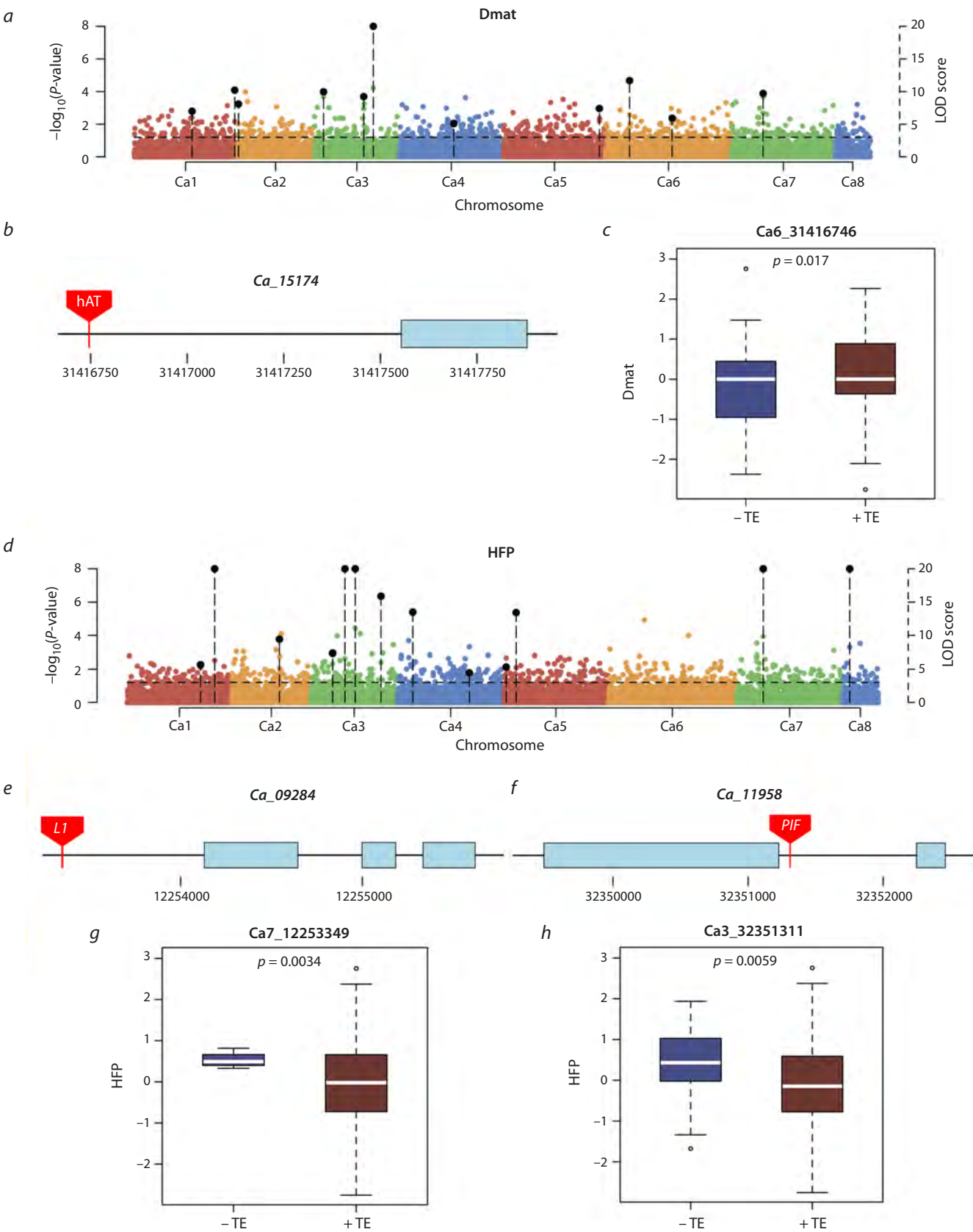


Fig. 3. TE insertion sites as a source of phenotypic variability. *a* – Manhattan plot showing associations of TE insertion sites with maturation time; *b* – *Ca_15174* gene structure; *c* – maturation time of plants from accessions with and without TE insertion; *d* – Manhattan plot showing association of TE insertion sites with the height of the first pod; *e*, *f* – *Ca_11958* (*e*) and *Ca_09284* (*f*) gene structure; *g*, *h* – height of the first pod in plants from accessions with and without TE insertion. See Figures S2–S4 for the results of transposon insertion validation.

stem length, and internode length (Radkova et al., 2019). In *Arabidopsis* plants, transcription factors of the CCHC(Zn) family are involved in RNA metabolism, transcription elongation, polyadenylation, translation, pre-mRNA splicing, RNA export and degradation, microRNA and ribosomal RNA biogenesis, and post-transcriptional gene silencing. Transposon insertion extends pod maturation time (Aceituno-Valenzuela et al., 2020) (Fig. 3c).

The height of the first pod is an important trait for reducing harvest losses. The *PIF-Harbinger* transposon at position Ca3_32351311 is associated with this trait (Fig. 3d). It is localized in the *Ca_11958* gene, which encodes the receptor for ethylene 2, a phytohormone that regulates plant growth and development (Fig. 3f) (Binder, 2020). In rice, mutations in the gene encoding the ethylene 2 receptor have been shown to affect flowering time (Wuriyangan et al., 2009). Another *non-LTR* transposon *L1* at position Ca7_12253349, also associated with this trait, is located upstream of the *Ca_09284* gene encoding chloroplast glucose-6-phosphate-1-dehydrogenase, which is involved in oxidative processes affecting germination, nitrogen metabolism, plant branching, and the response to abiotic stress (Jiang et al., 2022) (Fig. 3e). In both cases, plants with transposon insertion have a lower height of the first pod attachment, i. e. the transposon insertion has an unfavorable effect on the trait (Fig. 3g, h).

Comparison of the results of the association study using the GAPIT3 R package and the IIVmrMLM program showed that four transposon insertions are detected by both programs (Table S10).

Polymorphism of transposon insertion sites in groups of chickpea landraces from different geographical locations

ADMIXTURE analysis of plink files made from data on transposon insertion sites showed that the most preferred number of populations was five, although the CV-error for four populations was actually the same (Fig. 4a). The population structure of accessions from different geographical groups (Fig. 1) differed. Accessions from Ethiopia (ETH) were the most contrasting compared to the other samples, the admixture patterns of Indian (IND) and Central Asian (UZB) samples were similar to each other and different from the admixture pattern of Turkish (TUR) and Mediterranean (MED) accessions. It can also be noted that Lebanese (LEB) and Ethiopian (ETH) accessions were the most homogeneous in terms of admixture patterns and differed most from each other.

The number of polymorphic transposon insertion sites present in one group (unique sites) or in several, but not all groups of landraces, differed between groups (Fig. 4b, Table S11). Indian and Turkish accessions had the highest number of sites present in several groups, 650 and 705 sites, respectively. The Indian group also had the highest number of purely unique sites, namely, 44. The RUS group had the least number of unique insertion sites, which is likely due to the small number of samples in the group. The Ethiopian group stood out from all groups: it had the highest proportion (0.125) of unique sites among sites present in several groups. A more detailed analysis using the χ^2 criterion revealed 514 insertion sites, the frequency of which in the groups differed from the theoreti-

cally expected frequency calculated under the assumption of no differences. Then, to examine the groups for enrichment in insertion sites, for each site with a non-random frequency of occurrence, we calculated two differences: between the maximum frequency value and the second highest frequency in the group, and between the minimum frequency value and the frequency second from the end (Table S12). This analysis confirmed that the Ethiopian population is enriched in transposon insertion sites that occur predominantly in this population, but also contains rare sites that occur frequently in other populations.

It should be noted that the frequency of unique sites in groups, with rare exceptions, did not exceed 5 %, which indicates their relatively recent emergence. Only one *PIF-Harbinger* transposon at position Ca6_2586225 in the Ethiopian group had a very high population frequency of 0.95. This transposon is 1,979 bp away from the *Ca_10390* gene encoding the ROP-binding protein kinase RBK2 (Fig. 5a). RBK1/2 protein kinases phosphorylate small G-proteins of plant ROP and also interact with mitogen-activated protein kinase 1 (MPK1) from the auxin-responsive MPK cascade (Weiß et al., 2022). In addition, RBK1 is involved in Casparian strip formation and also plays a role in trichome branching, cytoskeleton stabilization and control of barley basal resistance to powdery mildew. Interestingly, the Ca6_2586225 position is located within a region of chromosome 6 about 100 kb long (2494265 to 2598131), which is virtually SNP-free in all *C. arietinum* samples. In addition to the *Ca_10390* gene, this region contains nine other genes encoding proteins involved in hormone-mediated control of cell elongation, plant growth, transpiration, and formation of generative organs (Table S13).

Discussion

Like all repetitive elements, transposons are characterized by extreme diversity. Each transposon family represents a continuum of more or less diverged copies, consisting of both autonomous and defective elements. This feature makes the identification and classification of transposons a challenging task, which has recently been progressively solved by high-throughput sequencing methods and the development of new programs. For example, the PoPoolationTE2 program used in this work (Kofler et al., 2016) implements an option for aggregating into one position TE insertion sites that differ by several nucleotides due to mapping inaccuracy (explained by sequence degeneracy). This allows to perform a comparative analysis of transposon insertions between different accessions.

By analyzing 190 chickpea genomes, we found 42,324 transposon insertion sites from 83 families and showed that most of the sites (70–80 %) are present in almost all accessions. The most abundant families were retrotransposons *Copia* (17,408 sites), *Gypsy* (6,813 sites), and *non-LTR* retrotransposons, represented mainly by *L1* and *RTE-BovB* (4,245 sites) (Fig. 1b, Fig. S1). The percentage of DNA transposon insertion sites in chickpea accessions is significantly lower and the most common families are the family of *Mu*-like elements *MuDR* (8 %), as well as the *PIF* (6 %), *hAT* (7 %), *CMC* (4 %), represented mainly by *CMC-EnSpm/CACTA*) and *TcMar* (2 %) families. *Copia* family insertion sites are

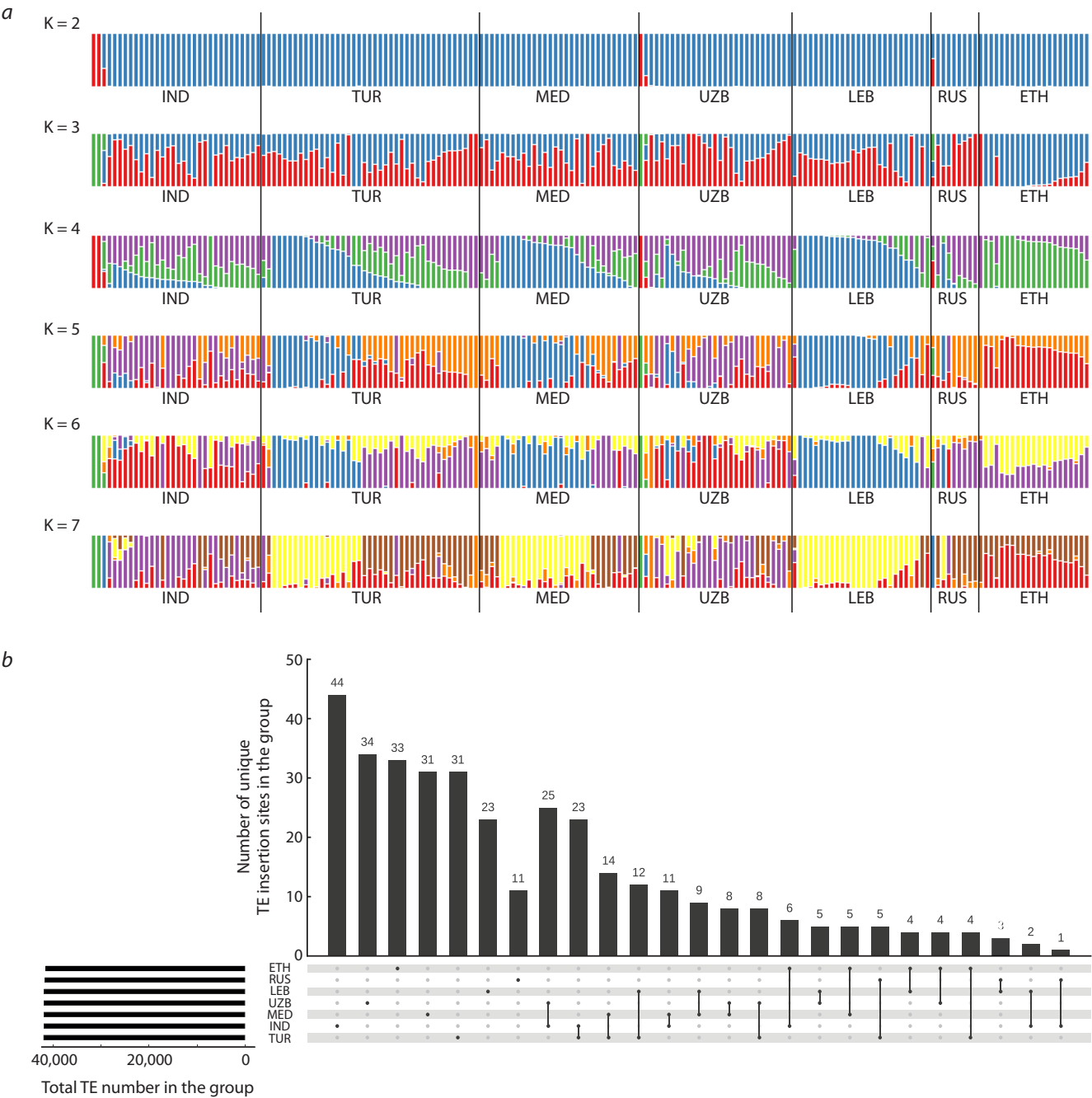


Fig. 4. Transposon insertion sites as a source of diversification of samples from different geographical regions.
a – population structure of landraces; *b* – Upset plot of transposon insertion sites. IND, MED, TUR, RUS, UZB, LEB, ETH – groups of accessions of different origin.

also prevalent in genomes of other plants (Domínguez et al., 2020; Cai et al., 2022). Our data are generally consistent with the results of the search for intact transposons in the chickpea reference genome, which also showed an excess of *Copia* family frequency over *Gypsy* and *non-LTR* frequencies, and the highest representation of the *MuDR* (*Mu*-like) family among DNA transposons (Mokhtar et al., 2021).

We found 47 transposon insertion hotspots, of which 16 contained exclusively retrotransposon insertions (Fig. 2a, Table S4). The non-random arrangement of transposon insertions of different families has also been demonstrated in other

plant genomes (Sultana et al., 2017). For example, in tomato, *Gypsy* insertion sites are predominantly located in pericentromeric regions (Domínguez et al., 2020).

Transposon insertions can influence the expression of adjacent genes (Bourque et al., 2018); therefore, the analysis of their location relative to genes and their flanking regions is of interest. In chickpea, such regions were found to be enriched in transposon family insertion sites, which was particularly evident for *non-LTR* retrotransposon insertions as well as *TcMar* and *PIF* DNA transposons (Table S5). The enrichment of gene-specific regions and their flanking regions with trans-

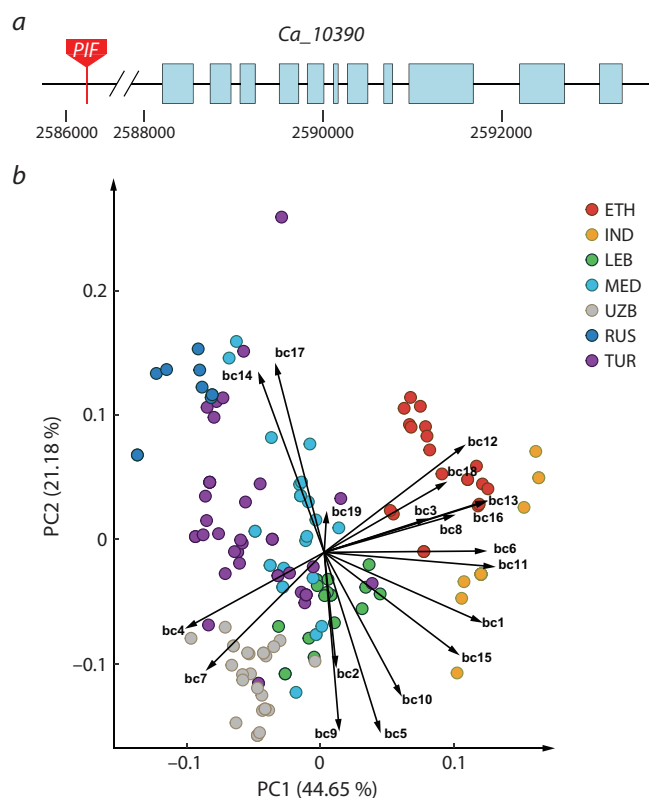


Fig. 5. *PIF-Harbinger* transposon insertion in the Ethiopian group.

a – structure of the *Ca_10390* gene with transposon insertion; *b* – principal component plot of bioclimatic variables from the collection sites. The decoding of the bioclimatic variable labels is given in Table S1.

poson insertions has been demonstrated in many plants (Qiu et al., 2021; Zhao et al., 2022). In our data, transposon insertions were least frequently recorded in exons due to their deleterious effect and the action of negative selection. The highest number of insertions in all transposon superfamilies occurred in introns, which was particularly evident in the *Copia*, *PIF*, and *TcMar* superfamilies. In *Copia* and *PIF*, the excess of insertions into introns over exon insertions was 42-fold, and in *TcMar*, almost all gene-specific insertion sites fell within introns. The highest excess of transposon insertions in flanking regions of genes over exons was observed for elements of the *Copia*, *MuDR*, and *PIF* superfamilies (Table S5).

In joint mode analysis, transposon insertion signatures are reliably identified in individual accessions, which makes it possible to analyze the contribution of transposon insertion site polymorphism to phenotypic variation. We have shown that transposon insertion sites are an important factor diversifying phenotypes and can be successfully used in genome-wide association studies as markers replacing single nucleotide polymorphisms (Tables S6, S7). In this case, the IIIVmrMLM program finds significantly more associations between insertion sites and a trait than GAPIT3 R. This is partly explained by the fact that the strict threshold for the significance of associations implemented in GAPIT3 R excludes the possibility of identifying markers with small effects. The feasibility of using transposon insertion sites in genome-wide association

studies has also been demonstrated in rice and tomato (Akakpo et al., 2020; Domínguez et al., 2020; Vourlaki et al., 2022; Yan et al., 2022).

Transposon insertion sites may have played a significant role in plant adaptation during evolution, since such changes can occur rapidly, which is critical for the organism to adapt to changing conditions (Niu et al., 2019; Schrader, Schmitz, 2019; Zhao et al., 2022; Kang et al., 2023). The primary domestication of chickpea occurred in the Fertile Crescent (modern Turkey), followed by secondary centers of diversification in India, Ethiopia, Central Asia and the Mediterranean (Igolkina et al., 2023). Due to the efforts of N.I. Vavilov, seeds of varieties from such centers are stored in the VIR collection, which makes it possible to study the polymorphism of transposon insertion sites in groups of accessions from different secondary diversification centers. It turned out that each group of accessions contained a large number of unique sites, but their frequency did not exceed 5 %, indicating their relatively recent emergence. Only one *PIF-Harbinger* transposon at position *Ca6_2586225* in the Ethiopian group of samples had a very high population frequency of 0.95. It should be noted that, in terms of the admixture pattern, the population structure of Ethiopian varieties differed most significantly from other groups (Table S12). The transposon *Ca6_2586225* is inserted into the 5'-flanking region of the *Ca_10390* gene encoding the ROP-binding protein kinase RBK2 (Fig. 5*a*), which is involved in the formation of the Casparian strip, i. e. in the regulation of the water balance of the plant (Weiß et al., 2022). As can be seen from the principal component analysis of bioclimatic variables from the collection sites (Fig. 5*b*), Ethiopian varieties are most dependent on the rainfall and humidity variables. This fact may be an indirect explanation for the spread of transposon *Ca6_2586225* in the group, since such an insertion, with the determinant role of climatic variables associated with precipitation, may be adaptive and provide plants with a selective advantage.

Transposons are a major source of genomic mutations (Bourque et al., 2018). In the case of *Ca6_2586225*, the transposon insertion appears to have resulted in a beneficial change. More often, however, transposon insertions have a deleterious effect on a trait, as we see with transposon insertions at positions *Ca7_12253349*, *Ca3_32351311* and *Ca6_31416746* (Fig. 3*b, e, f*).

Conclusion

In this work, we performed a primary analysis of transposon insertion sites in a large number of chickpea accessions, represented mainly by landraces. We found high polymorphism of such sites, characterized the representation of different transposon superfamilies, and showed uneven distribution of insertion sites along chromosomes. We also showed that transposon insertion sites, which until recently were out of the field of population genomics, are an important factor diversifying phenotypes and ensuring plant adaptation to growing conditions. The data and results obtained in this study are a valuable resource that can be used as a starting point for a more in-depth analysis of the evolutionary dynamics of transposons in the chickpea genome, their contribution to adaptation to global climate change, and the breeding of new varieties.

References

- Abbo S., Berger J., Turner N.C. Viewpoint: Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct Plant Biol.* 2003;30(10):1081-1087. doi 10.1071/fp03084
- Aceituno-Valenzuela U., Micol-Ponce R., Ponce M.R. Genome-wide analysis of CCHC-type zinc finger (ZCCHC) proteins in yeast, *Arabidopsis*, and humans. *Cell Mol Life Sci.* 2020;77(20):3991-4014. doi 10.1007/s00018-020-03518-7
- Akakpo R., Carpentier M., Hsing Y.I., Panaud O. The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytol.* 2020;226(1):44-49. doi 10.1111/nph.16356
- Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655-1664. doi 10.1101/gr.094052.109
- Alioto T., Alexiou K.G., Bardil A., Barteri F., Castanera R., Cruz F., Dhingra A., Duval H., Fernández i Martí Á., Frias L., Galán B., García J.L., Howad W., Gómez-Garrido J., Gut M., Julca I., Morata J., Puigdomènech P., Ribeca P., Rubio Cabetas M.J., Vlasova A., Wirthensohn M., Garcia-Mas J., Gabaldón T., Casacuberta J.M., Arús P. Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *Plant J.* 2020;101(2):455-472. doi 10.1111/tj.14538
- Binder B.M. Ethylene signaling in plants. *J Biol Chem.* 2020;295(22):7710-7725. doi 10.1074/jbc.rev120.010854
- Bourque G., Burns K.H., Gehring M., Gorbunova V., Seluanov A., Hammell M., Imbeault M., Izsvák Z., Levin H.L., Macfarlan T.S., Mager D.L., Feschotte C. Ten things you should know about transposable elements. *Genome Biol.* 2018;19(1):199. doi 10.1186/s13059-018-1577-z
- Cai X., Lin R., Liang J., King G.J., Wu J., Wang X. Transposable element insertion: a hidden major source of domesticated phenotypic variation in *Brassica rapa*. *Plant Biotechnol J.* 2022;20(7):1298-1310. doi 10.1111/pbi.13807
- Catlin N.S., Josephs E.B. The important contribution of transposable elements to phenotypic variation and evolution. *Curr Opin Plant Biol.* 2022;65:102140. doi 10.1016/j.pbi.2021.102140
- de la Peña T.C., Pueyo J.J. Legumes in the reclamation of marginal soils, from cultivar and inoculant selection to transgenic approaches. *Agron Sustain Dev.* 2012;32:65-91. doi 10.1007/s13593-011-0024-2
- Domínguez M., Dugas E., Benchouaia M., Leduque B., Jiménez-Gómez J.M., Colot V., Quadrana L. The impact of transposable elements on tomato diversity. *Nat Commun.* 2020;11(1):4058. doi 10.1038/s41467-020-17874-2
- Duk M.A., Kanapin A.A., Bankin M.P., Samsonova M.G. Using the IIIVmrMLM method to confirm and search new genome-wide associations in chickpea. *Biofizika = Biophysics.* 2024;69(6):1263-1278. doi 10.31857/S0006302924060126 (in Russian)
- Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Clim.* 2005;25(15):1965-1978. doi 10.1002/joc.1276
- Igolkina A.A., Noujdina N.V., Vishnyakova M., Longcore T., von Wettberg E., Nuzhdin S.V., Samsonova M.G. Historical routes for diversification of domesticated chickpea inferred from landrace genomics. *Mol Biol Evol.* 2023;40:msad110. doi 10.1093/molbev/msad110
- Jain M., Misra G., Patel R.K., Priya P., Jhanwar S., Khan A.W., Shah N., Singh V.K., Garg R., Jeena G., Yadav M., Kant C., Sharma P., Yadav G., Bhatia S., Tyagi A.K., Chattopadhyay D. A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 2013;74(5):715-729. doi 10.1111/tj.12173
- Jiang Z., Wang M., Nicolas M., Ogé L., Pérez-García M.-D., Crespel L., Li G., Ding Y., Le Gourrieric H., Grappin P., Sakr S. Glucose-6-phosphate dehydrogenases: the hidden players of plant physiology. *Int J Mol Sci.* 2022;23(24):16128. doi 10.3390/ijms232416128
- Kaloki P., Devasirvatham V., Tan D.K.Y. Chickpea abiotic stresses: combating drought, heat and cold. In: *Abiotic and Biotic Stress in Plants*. IntechOpen, 2019. doi 10.5772/intechopen.83404
- Kang M., Wu H., Liu H., Liu W., Zhu M., Han Y., Liu W., Song C.C.Y., Tan L., Yin K., Zhao Y., Yan Z., Lou S., Zan Y., Liu J. The pan-genome and local adaptation of *Arabidopsis thaliana*. *Nat Commun.* 2023;14(1):6259. doi 10.1038/s41467-023-42029-4
- Kofler R., Pandey R.V., Schlotterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* 2011;27(24):3435-3436. doi 10.1093/bioinformatics/btr589
- Kofler R., Gómez-Sánchez D., Schlötterer C. PoPoolationTE2: comparative population genomics of transposable elements using Pool-Seq. *Mol Biol Evol.* 2016;33(10):2759-2764. doi 10.1093/molbev/msw137
- 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
- Li M., Zhang Y.-W., Xiang Y., Liu M.-H., Zhang Y.-M. IIIVmrMLM: the R and C++ tools associated with 3VmrMLM, a comprehensive GWAS method for dissecting quantitative traits. *Mol Plant.* 2022a;15(8):1251-1253. doi 10.1016/j.molp.2022.06.002
- Li M., Zhang Y.-W., Zhang Z.-C., Xiang Y., Liu M.-H., Zhou Y.-H., Zuo J.-F., Zhang H.-Q., Chen Y., Zhang Y.-M. A compressed variance component mixed model for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. *Mol Plant.* 2022b;15(4):630-650. doi 10.1016/j.molp.2022.02.012
- Mann H.B., Whitney D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statistics.* 1947;18(1):50-60. doi 10.1214/aoms/1177730491
- Mhiri C., Borges F., Grandbastien M.-A. Specificities and dynamics of transposable elements in land plants. *Biology.* 2022;11(4):488. doi 10.3390/biology11040488
- Mokhtar M.M., Alsamman A.M., Abd-Elhalim H.M., Allali A.E. CicerSpTEdb: a web-based database for high-resolution genome-wide identification of transposable elements in *Cicer* species. *PLoS One.* 2021;16(11):e0259540. doi 10.1371/journal.pone.0259540
- Niu X.-M., Xu Y.-C., Li Z.-W., Bian Y.-T., Hou X.-H., Chen J.-F., Zou Y.-P., Jiang J., Wu Q., Ge S., Balasubramanian S., Guo Y.-L. Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc Natl Acad Sci USA.* 2019;116(14):6908-6913. doi 10.1073/pnas.1811498116
- Pulido M., Casacuberta J.M. Transposable element evolution in plant genome ecosystems. *Curr Opin Plant Biol.* 2023;75:102418. doi 10.1016/j.pbi.2023.102418
- Qiu Y., O'Connor C.H., Della Coletta R., Renk J.S., Monnahan P.J., Noshay J.M., Liang Z., Gilbert A., Anderson S.N., McGaugh S.E., Springer N.M., Hirsch C.N. Whole-genome variation of transposable element insertions in a maize diversity panel. *G3 (Bethesda).* 2021;11(10):jkab238. doi 10.1093/g3journal/jkab238
- Quesneville H. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. *Mob DNA.* 2020;11:28. doi 10.1186/s13100-020-00223-x
- Radkova M., Revalska M., Kertikova D., Iantcheva A. Zinc finger CCHC-type protein related with seed size in model legume species *Medicago truncatula*. *Biotechnol Biotechnol Equip.* 2019;33(1):278-285. doi 10.1080/13102818.2019.1568914
- Schrader L., Schmitz J. The impact of transposable elements in adaptive evolution. *Mol Ecol.* 2019;28(6):1537-1549. doi 10.1111/mec.14794
- Solkolkina A., Bulyntsev S.V., Chang P.L., Carrasquilla-García N., Igolkina A.A., Noujdina N.V., von Wettberg E., Vishnyakova M.A., Cook D.R., Nuzhdin S.V., Samsonova M.G. Genomic analysis of Vavilov's historic chickpea landraces reveals footprints of environ-







- mental and human selection. *Int J Mol Sci.* 2020;21(11):3952. doi 10.3390/ijms21113952
- Sultana T., Zamborlini A., Cristofari G., Lesage P. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet.* 2017;18(5):292-308. doi 10.1038/nrg.2017.7
- Varshney R., Song C., Saxena R., Azam S., Yu S., Sharpe A.G., Cannon S., ... Singh K.B., Datta S.K., Jackson S.A., Wang J., Cook D.R. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol.* 2013;31(3):240-246. doi 10.1038/nbt.2491
- Varshney R., Thudi M., Roorkiwal M., He W., Upadhyaya H.D., Yang W., Bajaj P., ... Sutton T., von Wettberg E., Vigouroux Y., Xu X., Liu X. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat Genet.* 2019;51(5):857-864. doi 10.1038/s41588-019-0401-3
- Vourlaki I.-T., Castanera R., Ramos-Onsins S.E., Casacuberta J.M., Pérez-Enciso M. Transposable element polymorphisms improve prediction of complex agronomic traits in rice. *Theor Appl Genet.* 2022;135(9):3211-3222. doi 10.1007/s00122-022-04180-2
- Weiß L., Gaelings L., Reiner T., Mergner J., Kuster B., Fehér A., Hensel G., Gahrtz M., Kumlehn J., Engelhardt S., Hückelhoven R. Posttranslational modification of the RHO of plants protein RACB by phosphorylation and cross-kingdom conserved ubiquitination. *PLoS One.* 2022;17(3):e0258924. doi 10.1371/journal.pone.0258924
- Wuriyanghan H., Zhang B., Cao W.-H., Ma B., Lei G., Liu Y.-F., Wei W., Wu H.-J., Chen L.-J., Chen H.-W., Cao Y.-R., He S.-J., Zhang W.-K., Wang X.-J., Chen S.-Y., Zhang J.-S. The ethylene receptor ETR2 delays floral transition and affects starch accumulation in rice. *Plant Cell.* 2009;21(5):1473-1494. doi 10.1105/tpc.108.065391
- Yan H., Haak D.C., Li S., Huang L., Bombarely A. Exploring transposable element-based markers to identify allelic variations underlying agronomic traits in rice. *Plant Commun.* 2022;3(3):100270. doi 10.1016/j.xplc.2021.100270
- Zhao Y., Li X., Xie J., Xu W., Chen S., Zhang X., Liu S., Wu J., El-Kassaby Y.A., Zhang D. Transposable elements: distribution, polymorphism, and climate adaptation in populus. *Front Plant Sci.* 2022;13:814718. doi 10.3389/fpls.2022.814718

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

Received July 28, 2024. Revised November 19, 2024. Accepted November 26, 2024.

doi 10.18699/vjgb-25-09

Evaluation of the biodiversity of arbuscular mycorrhizal fungi during regenerative succession in quarries

A.A. Kryukov , A.P. Yurkov , A.O. Gorbunova, T.R. Kudriashova , A.I. Gorenkova ,
Y.V. Kosulnikov , Y.V. Laktionov 

The All-Russia Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia

 aa.krukov@arriam.ru







Abstract. Arbuscular mycorrhizal fungi (AMF) play a key role in the regenerative successions of plant communities after anthropogenic disturbances, particularly in quarries. AMF help plants with water and mineral nutrition, contributing to the restoration rate of vegetation cover. The research is aimed to study the biodiversity of AMF using molecular genetic methods at different stages of overgrowth of two quarries in the Leningrad region. Molecular genetic identification of fungi was carried out using Illumina MiSeq analysis of the ITS1 and ITS2 regions as barcodes for the identification of operational taxonomic units (OTUs) with species-level identification. An adapted and error-checked AMF genetic sequence database from NCBI was used as a reference. The study applied an optimized nucleic acid isolation technique for sandy soils. The results showed maximum AMF biodiversity at the initial stages of overgrowth – pioneer and grass stages – with minimum diversity observed at the shrub stage, where it decreased by five times. At the forest stage, the biodiversity of AMF was almost restored to the level seen at the grass stage. It has been shown that the biodiversity and species composition of AMF can vary greatly between the stages of regenerative succession and probably depends primarily on the biodiversity of grasses, with which AMF most effectively enter into symbiotic relationships. The analysis showed a reliable negative correlation between the number of AMF species and the number of woody plant species. Such studies can aid in understanding how plant-fungal symbiosis develops in regenerative successions and which AMF most effectively contribute to vegetation cover restoration.

Key words: arbuscular mycorrhizal fungi; biodiversity; regenerative succession; quarry; Illumina


For citation: Kryukov A.A., Yurkov A.P., Gorbunova A.O., Kudriashova T.R., Gorenkova A.I., Kosulnikov Y.V., Laktionov Y.V. Evaluation of the biodiversity of arbuscular mycorrhizal fungi during regenerative succession in quarries. *Vavilovskii Zhurnal Genetiki i Selektcii* = *Vavilov J Genet Breed.* 2025;29(1):72-78. doi 10.18699/vjgb-25-09

Funding: This article was made with support from the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075-15-2022-320 dated 20 April 2022 on providing a grant in the form of subsidies from the Federal budget of the Russian Federation. The grant was provided as state support for the creation and development of a world-class scientific center "Agrotechnologies for the Future". The work was carried out using the equipment of the Center for Collective Use "Genomic Technologies, Proteomics and Cell Biology" at the ARRIAM.

Оценка биоразнообразия грибов арбускулярной микоризы при восстановительной сукцессии на песчаных карьерах

А.А. Крюков , А.П. Юрков , А.О. Горбунова, Т.Р. Кудряшова , А.И. Горенкова ,
Ю.В. Косильников , Ю.В. Лактионов 

Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии, Пушкин, Санкт-Петербург, Россия

 aa.krukov@arriam.ru

Аннотация. Грибы арбускулярной микоризы (АМГ) играют ключевую роль в восстановительных сукцессиях растительных сообществ после антропогенных нарушений, в частности на песчаных карьерах. АМГ помогают растениям в водном и минеральном питании, ускоряя восстановление растительного покрова. Целью исследования было изучить биоразнообразие АМГ молекулярно-генетическими методами на разных стадиях зарастания двух карьеров Ленинградской области. Молекулярно-генетическая идентификация грибов проводилась с использованием анализа Illumina MiSeq по регионам ITS1 и ITS2 в качестве баркода для поиска операционных таксономических единиц с идентификацией микроорганизмов до вида. Референсом служила адаптированная и проверенная на ошибки база генетических последовательностей АМГ из NCBI. В исследовании использовалась оптимизированная для песчаных почв методика выделения нуклеиновых кислот. Максимальное биоразнообразие АМГ наблюдалось на начальных стадиях зарастания – пионерной и злаковой. На кустарниковой стадии разнообразие снижалось в пять раз, а затем на лесной стадии восстанавливалось почти

до уровня злаковой. Биоразнообразие и видовой состав АМГ могут значительно меняться между стадиями восстановительной сукцессии, что может в первую очередь зависеть от биоразнообразия трав, с которым АМГ вступают в симбиотические отношения наиболее эффективно. Показана достоверная отрицательная корреляция между числом видов АМГ и числом видов древесных растений. Проведенное исследование может помочь в понимании того, как развивается растительно-грибной симбиоз в восстановительных сукцессиях и какие АМГ наиболее эффективно помогают в восстановлении растительного покрова.

Ключевые слова: грибы арбускулярной микоризы; биоразнообразие; восстановительная сукцессия; песчаный карьер; Illumina.

Introduction

Most Embryophytes (more than 90 % of the families) form arbuscular mycorrhiza with fungi (AMF) of the Glomeromycotina subdivision, the Mucoromycota division (Spatafora et al., 2016). AMF help the symbiotic plant with water and mineral nutrition, receiving complex organic substances in return. In regenerative successions, mycorrhizal interaction promotes the plant's ability to compete and overcome unfavorable edaphic conditions (van der Heijden et al., 1998; Lambers et al., 2008). Mycorrhiza presence may be an influential factor contributing to the successful absorption of free substrates by plants. Mycorrhizal symbionts can significantly enhance the growth conditions of pioneer plants. Mycorrhiza is a crucial adaptation for plants facing deficiencies in nitrogen, phosphorus, and other essential minerals, unfavorable water, air conditions, and a lack of organic carbon substrates (Aikio, 2000). AMF is likely to facilitate succession, but as of now, there is very little direct evidence of this in natural ecosystems (Smith, Read, 2008). The species composition of AMF communities is known to have a great influence on plant productivity, plant community structure, successional patterns, and ecosystem performance (Wu, 2017). The quantitative and species composition of micromycetes is mainly determined by the vegetation (Sumina et al., 2010). The diversity and performance of mycorrhizal fungi are crucial for biodiversity and ecosystem health, while the diversity and structure of vegetation can also influence the diversity of AMF populations (Jeffries, Barea, 2001). AMF diversity is expected to be lower in the quarry substrate than in zonal soils, as there is lower soil moisture content, which is fundamental to the very survival of these microorganisms (Ganugi et al., 2019).

Soil microbial biodiversity analysis uses high-throughput sequencing methods, primarily with Illumina MiSeq. Various genetic markers are used to identify AMF, generally by using an ITS (internal transcribed spacer) region or the SSU (Small SubUnit ribosomal gene fragments) and LSU (Large SubUnit) gene regions that flank it (Kryukov et al., 2020). In some cases, it involves other genes or even full-genome sequencing. The lack of consensus on a barcode marker in Glomeromycotina provides challenges to ecological and phylogenetic studies. The conserved regions of the SSU and LSU genes are suitable for effective AMF identification to the level of genus, however, they are poorly fit for species-level identification due to their low variability (Öpik et al., 2014). To effectively perform an AMF species identification, it is preferable to use variable ITSs, although it can often lead to the identification of virtual taxa (Kryukov et al., 2020).

This study aims to assess the AMF biodiversity using molecular genetic methods at different overgrowth stages of two quarries in the Leningrad region, Russian Federation.

Materials and methods

The materials were collected in midsummer of 2018 and 2019 at two quarries of different ages in the Vsevolozhsk district of the Leningrad region: Kuzmolovo (60.116448N, 30.545006E) and Kalelovo (60.256590N, 29.971972E). Pioneer stage communities at the quarries were sparsely closed, with total projective cover by plants (TPC) not exceeding 20 %. During the grass stage, Gramineae dominated the communities, which were predominantly composed of various grasses. During the shrub stage, the upper tier of the communities was composed of shrubs and undergrowth. The forest stage was composed of young growths of the forest vegetation type. Sample plots measuring 5 × 5 meters were established at each stage of the regenerative succession in the quarries (Gorbunova, Sumina, 2021). To conduct the genetic AMF composition analysis, 20 plots were sampled, four sites per succession stage. Soil sampling was conducted in the rhizosphere of each plot. For five plant species (*Agrostis capillaris*, *Artemisia vulgaris*, *Chamaenerion angustifolium*, *Deschampsia cespitosa*, and *Tussilago farfara*), up to 25 soil samples were collected per plot, with five samples taken for each species whenever possible. According to the published works (Wang, Qiu, 2006; Akhmetzhanova et al., 2012), these plant species form symbiosis with AMF. Botanical description of the plots and plant mycorrhization level assessment were performed earlier (Gorbunova, Sumina, 2021).

An optimized technique involving Illumina MiSeq sequencing was used for AMF molecular genetic identification (Kryukov et al., 2020; Yurkov et al., 2024). Rhizosphere soil samples containing AMF mycelium and spores were taken for identification. For DNA extraction, a 0.5 g sample of frozen soil and 1 g of garnet abrasive were taken into a 2 mL tube for mechanical grinding (Pinaev et al., 2022). After that, 700 µL of CTAB buffer (2 % CTAB; 1.4 M NaCl; 20 mM EDTA; 100 mM Tris-HCl pH = 8.0) was added to the heated tube containing soil. The tubes were shaken in a vortex mixer for 1 minute every 15 min and incubated at +65 °C for up to 2 h. After thermal, chemical, and mechanical treatment, the samples were centrifuged for 5 min, and then the supernatant was transferred to new test tubes. A second DNA washing was conducted using 500 µL of water. This process involved shaking the soil and water mixture for 5 minutes. After centrifugation, the second supernatant was combined with the first supernatant. Additional washing with water was imperative as DNA tends to adsorb on sandy soil particles. The obtained DNA was freed from impurities by double extraction with an equal volume of chloroform. After each centrifugation (10 min at 14,000 rpm, Eppendorf, Germany), the supernatant with DNA was sampled and transferred to a new tube. The DNA was precipitated with 2/3 V isopropanol with 0.4 M NaCl,

washed with 70 % ethyl alcohol, dried for 3 min, and then further dissolved in water (Maniatis et al., 1982). The DNA was purified with the AMPure XP magnetic particles (Beckman Coulter, USA).

The purified DNA was used for a separate PCR of the ITS1 and ITS2 marker regions with universal primers. The primers were synthesized in Evrogen (Russia) with the 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' adaptor for forward primers and the 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3' adaptor for reverse primers for Illumina MiSeq: ITS5 (5'-GGAAGTAAAAGTCGTAACAACAAGG-3') and our modified ITS-2RK reverse primer (5'-CGTTCAAAGATTCGATGATTCAC-3') for the ITS1 amplification; the ITS3 primers (5'-GCATCGATGAAGAACGCAGC-3') and the ITS4 reverse primer (5'-TCCTCCGCTTATTGATATGC-3') for the ITS2 amplification. After the first PCR round with 20 cycles, the PCR product was diluted 100 times and reamplified with 30 cycles. After amplification and visualization on an agarose gel, the PCR products of ITS1 and ITS2 were pooled for each sample and purified using AMPure XP magnetic particles (Beckman Coulter, USA).

Prior to sequencing, the purified amplicon libraries from each plot were combined to create a single sequencing run that reflects the AMF species composition for each plot. The amplicon libraries were sequenced on Illumina MiSeq using the MiSeq® Reagent Kit v3 (600-cycle) with paired-end reads (2 × 300 n.) (Illumina, Inc., USA). The identified sequences were then processed with the Illumina software (Illumina, Inc., USA). Illumina MiSeq sequencing resulted in FASTQ sequences from forward and reverse primers. This format covers sequence data and quality scores for each nucleotide position. The sequencing results have been submitted to the NCBI database (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997898/>, BioProject ID: PRJNA997898).

The methodology for the bioinformatics analysis is detailed in our 2020 research (Kryukov et al., 2020). In the present study, the data analysis was performed using the local database of reference sequences. As a starting point, we took the sequences from the NCBI database and filtered them for errors. Authorship, sequencing year, manual alignment analysis, and phylogenetic analysis were reviewed to filter any errors. Currently, our database contains data on 33 genera and 176 AMF species.

Results

To analyze the AMF biodiversity in four stages of regenerative succession (Table 1), 20 sequencing runs were performed with a depth of up to 100,000 reads per sample (4 sequencing runs for each succession stage for the two quarries). Bioinformatic data processing was performed with USEARCH software (Edgar, 2010). We also calculated the Margalef and Shannon diversity indexes, as well as the Williams polydominance index (Table 1).

We searched for the fungi OTUs (Operational Taxonomic Units) of different divisions (Table 2) for four succession stages in the two quarries. OTUs can denote both real individual species and virtual taxa (having no close reference in the database) of different taxonomic levels.

The lowest number of the AMF OTUs (Mucoromycota (Glomeromycotina)) has been found at the shrub stage, and the highest, at the grass stage in Kuzmolovo. Interestingly, this is different for other fungal divisions; for example, the highest biodiversity of Ascomycota and Basidiomycota fungi was observed at the pioneer stage of the regenerative succession.

About half of the identified AMF OTUs were able to be positively assigned to species (Fig. 1). Figure 1 shows both the identified species and the read count proportion during sequencing (in percentages), which can suggest their occurrence at each succession stage. 1 (pioneer stage, Kalelovo) – *Rhizophagus irregularis* (68.9 %), *Rh. sp.* (12.2 %), *Glomeraceae sp.* (7.8 %), *Glomus cerebriforme* (6.7 %), *Acaulosporaceae sp.* (3.3 %), *Paraglomus laccatum* (1.1 %); 2a (grass stage, Kalelovo) – *Entrophospora sp.* (81.2 %), *E. glacialis* (12.6 %), *Nanoglomus sp.* (3.5 %), *Glomus sp.* (2.1 %), *Paraglomus laccatum* (0.1 %), *Acaulospora brasiliensis* (0.1 %); 2b (grass stage, Kuzmolovo) – *Glomeraceae sp.* (19.8 %), *Acaulospora paulinae* (18.0 %), *Archaeosporaceae sp.* (17.1 %), *Paraglomus laccatum* (16.1 %), *Ambispora sp.* (16.0 %), *Entrophospora claroideum* (5.6 %), *Rhizophagus intraradices* (3.2 %), *Dominikia sp.* (2.8 %), *Entrophospora sp.* (1.4 %); 3 (shrub stage, Kuzmolovo) – *Entrophospora claroideum* (95.6 %), *Diversispora versiformis* (4.4 %); 4 (forest stage, Kuzmolovo) – *Archaeosporaceae sp.* (37.7 %), *Glomus sp.* (29.3 %), *Nanoglomus sp.* (27.2 %), *Diversispora versiformis* (2.6 %), *Acaulosporaceae sp.* (2.1 %), *Glomeraceae sp.* (1.1 %). Interestingly, *R. irregularis* is most prevalent at the pioneer stage; it is replaced by other AMF later on.

Table 1. AMF diversity and polydominance indexes at plots of different succession stages

No.	Succession stages	Quarry	Margalef index, D_{Mg}	Shannon diversity index, H_N	Williams polydominance index, D^{-1}
1	Pioneer	Kalelovo	1.33 ^{ab}	1.59 ^c	2.06 ^c
2a	Grass		0.89 ^c	0.92 ^d	1.46 ^d
2b	Grass	Kuzmolovo	1.18 ^b	2.97 ^a	6.95 ^a
3	Shrub		0.26 ^d	0.26 ^e	1.09 ^d
4	Forest		1.51 ^a	2.07 ^b	3.58 ^b

Note. The “a–d” indexes indicate significantly different values of the estimated parameter ($p < 0.05$). Green indicates higher values; blue indicates lower values.

Table 2. Identified fungi OTUs by division

Divisions	Succession No.				
	1	2a	2b	3	4
Ascomycota	124	78	126	86	118
Zoopagomycota	1	0	2	2	1
Basidiomycota	106	52	82	68	96
Chytridiomycota	4	0	2	0	2
Mucoromycota (Glomeromycotina)	10	7	11	2	8
Mucoromycota (Mortierellomycotina)	6	1	11	7	15
Mucoromycota (Mucoromycotina)	3	1	11	1	9
Cryptomycota	4	2	5	1	17

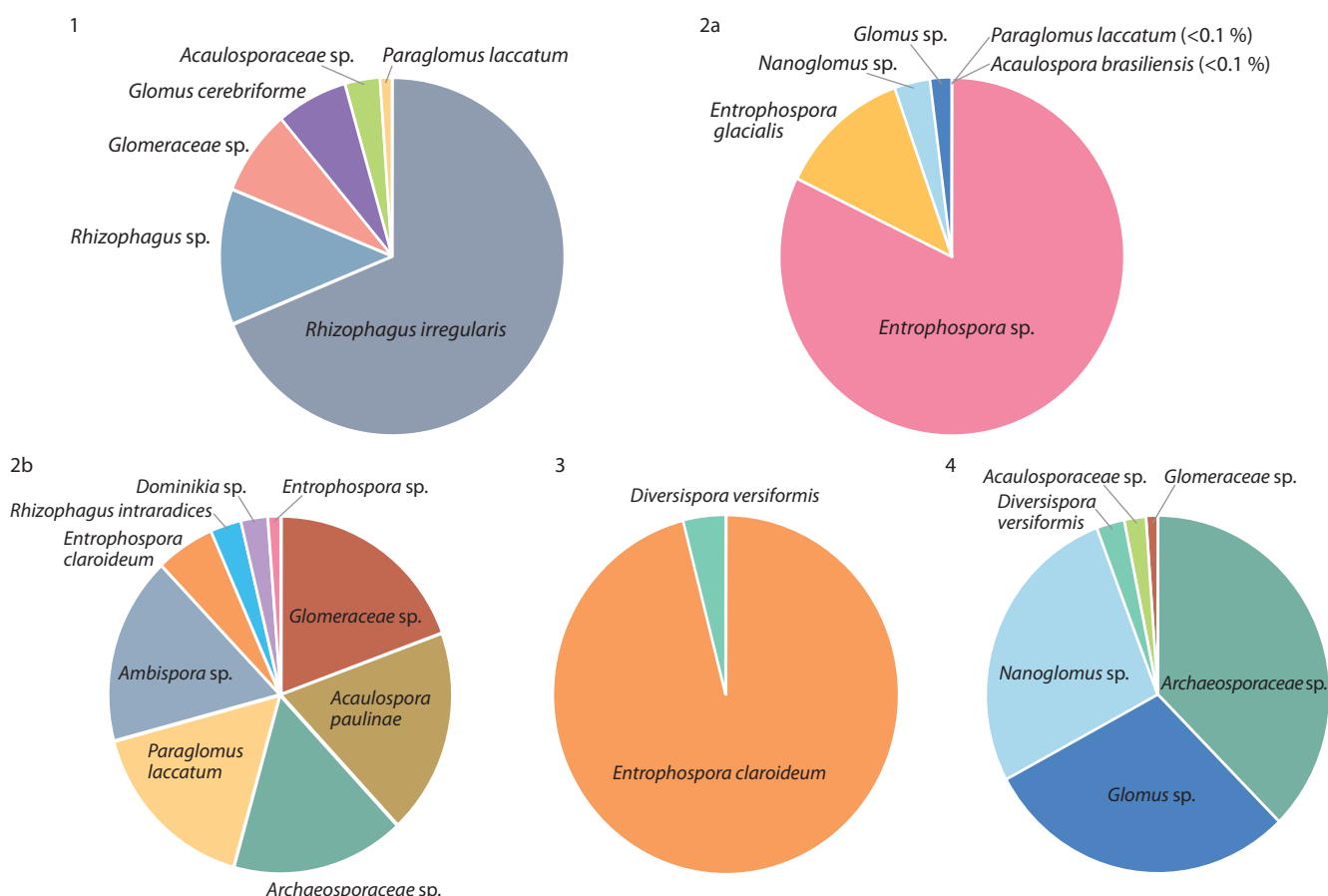


Fig. 1. Glomeromycotina species composition at different stages of regenerative succession (proportion of reads). Succession numbers correspond to Table 1.

Figure 2 shows the proportion of read counts after Illumina MiSeq sequencing for major fungal divisions. The Basidiomycota division fungi are most represented at all succession stages. Basidiomycota reads are most abundant during the pioneer stage; their proportion decreases at the grass stage, increases again at the shrub stage, and most significantly, at the forest stage. The share of Ascomycota fungi is also significant. The fungi of other divisions are scarcely represented in the DNA array at all stages of regenerative succession. At the grass stage, it is important to note that AMF species are

represented by the highest number, and they also exhibit the greatest total proportion of reads compared to other stages. In general, the proportion of AMF reads is not high if compared to other fungal divisions; this has been found previously in other studies. The maximum read percentage for AMF by ITS using universal primers amounts to up to 2 % (Senés-Guerrero, Schüßler, 2015).

Correlations were analyzed between AMF species, species from other fungal divisions, and various parameters taken from the study by A.O. Gorbunova and O.I. Sumina (2021).

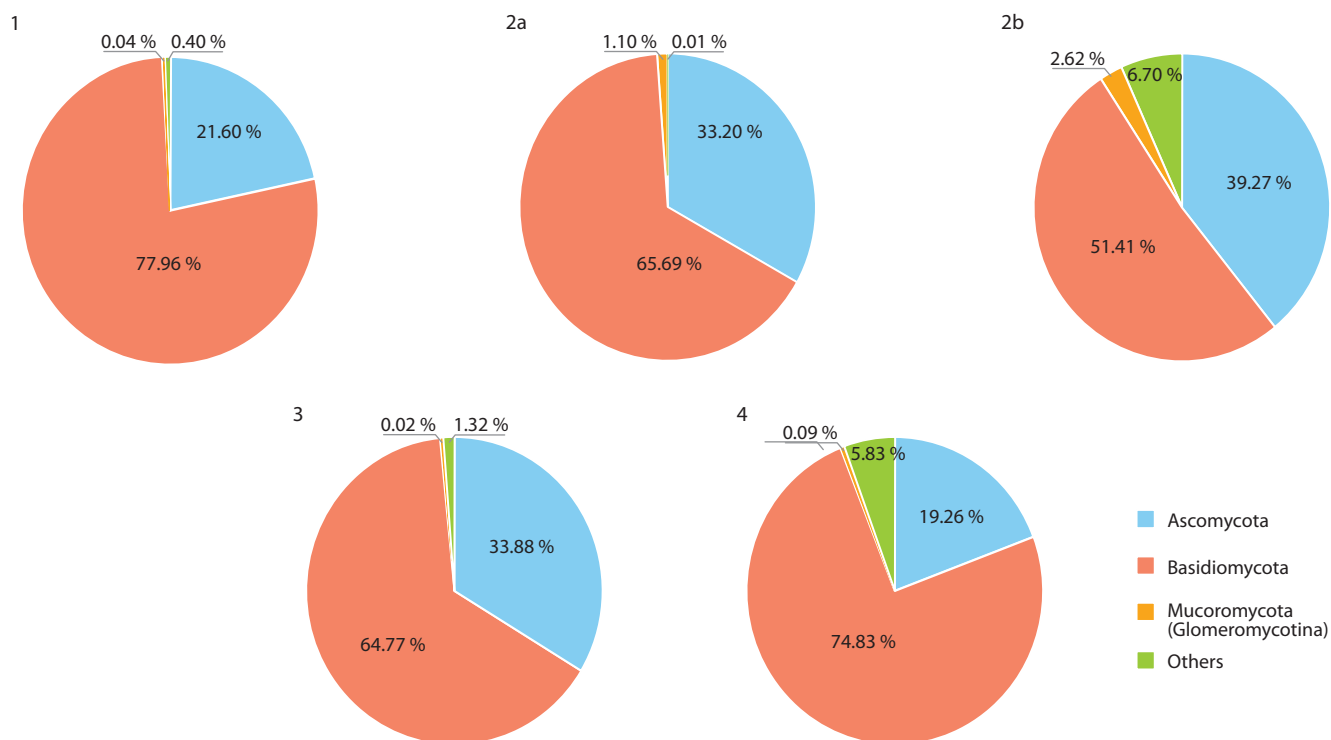


Fig. 2. Read proportion based on the Illumina MiSeq sequencing results for the fungal kingdom divisions. Succession numbers correspond to Table 1.

These parameters included the number of grass species, moss species, woody plant species, total projective cover of plants, and projective cover by plant group. The calculations were performed across different stages of regenerative succession. A strong negative correlation was found between the number of AMF species and the number of woody plant species (-0.85).

Discussion

The results are generally consistent with the global trends observed in AMF research. The highest AMF biodiversity in our work is found at the grass stage of regenerative succession, and the lowest, at the shrub stage. In our 2021 work, it was indicated that the grass stage of regenerative succession displays the highest grass diversity (up to 32 species per plot) (Gorbunova, Sumina, 2021); the forest stage shows approximately the same number (up to 26 species). At the same time, minimal grass biodiversity is present at the pioneer and shrub stages. It appears that a greater grass species diversity is more beneficial for AMF due to a wider range of nutrient providers, which may eventually benefit both fungi and plants (Kiers et al., 2011). Similar results were shown in a succession study of a fungal community in a retreating glacier in the Cascade Range (Jumpponen et al., 2012). The authors note that when the mycorrhizal fungi diversity increases, the plant species diversity and the plant community primary production increase as well; this is also in line with earlier studies (van der Heijden et al., 1998). Our research indicates a strong negative correlation (-0.85) between the number of AMF species and the number of woody plant species. The number of tree species is highest at the shrub stage with the minimum number of AMF

species. No correlation was found between herbaceous plant species and the number of AMF species. The latter is probably due to the large number of independent factors affecting vegetation and AMF (Gorbunova, Sumina, 2021).

According to earlier results, based on the microscopy study of samples (Gorbunova, Sumina, 2021), it was observed that the number of fungal propagules in soil and the diversity of mycorrhizal fungi species do not increase with succession; they even decrease at the shrub stage. This agrees in part with the data of the molecular genetic part of the study. It should be noted that the AMF count registered by microscopy is usually lower than that in molecular genetic studies (Kryukov et al., 2020).

M. Zobel and M. Öpik (2014) formulated the habitat hypothesis to distinguish the case where AMF and plant biodiversity are correlated but not in a direct cause-and-effect relationship, as opposed to the null hypothesis of no correlation (independence). For example, during primary succession, plants typically occupy the habitat before AMF and then act as a potential filter for AMF, i. e., AMF are “passengers”, as they follow the plants. However, limited distribution in a stable AMF community may result in the AMF community being a stronger determinant of which plants take root during secondary succession; in this case, the AMF community becomes the driver (Zobel, Öpik, 2014).

The biodiversity data from the two quarries in this study differ due to their different age. The Kalelovo quarry is fairly young with rather broken plots; the overgrowth process has been taking place for a shorter period; trees and shrubs are almost absent and young. The Kuzmolovo quarry has been

overgrowing for about three decades; there are some bare plots, free of trees and shrubs, but tree regeneration is quite active (Gorbunova, Sumina, 2021). At the same time, different stages at the two quarries provide an opportunity to investigate how biodiversity changes during the development of the regenerative succession. During revegetation (shrub and forest stages), the competition between soil microbiota increases, and we see the results of ecological filtering of the fungal species introduced at the first stage, which changes the species composition, and AMF, more capable of symbiosis with grasses, tend to persist (Yurkov et al., 2024). *R. irregularis* tends to vanish during the grass stage; it is more common during an unusual pioneer stage and is eventually replaced by other AMF species. Still, this species is widespread, as other studies show, including our work on AMF biodiversity assessment in the North Caucasus (Yurkov et al., 2024). Compared to the AMF diversity in the Caucasus, it is lower in quarries, which is in line with the suggestion that AMF biodiversity is lower in sandy soils than in others (Ganugi et al., 2019).

Fairly many AMF species have been observed at the pioneer stage, as the ecological niches are vacant, and the species diversity is subject to random factors, mainly fungal introductions to plots (van der Heijden et al., 2015). The efficiency of AMF-plant symbiosis reaches maximum levels during the grass stage, while the number of AMF species is also at its maximum. During the shrub stage, grasses are replaced by shrubs and undergrowth associated with ectomycorrhizal fungi (aspen, birch, etc.); grasses are suppressed by the lack of light and the presence of woody plant roots in the soil, resulting in a dramatic decrease in AMF diversity. During the forest stage, grass species become slightly more abundant due to the accession of typical forest species, but competition for light and soil resources remains high, which gives a slight increase in AMF diversity (Neuenkamp et al., 2021).

The distribution of fungal biodiversity across taxonomic groups (Fig. 2) is influenced by both random factors (mainly at the pioneer stage) and competitive selection (at the later stages). During the pioneer stage, samples predominantly contain basidio- and ascomycetes, the spores of which spread much more easily than AMF spores, brought by wind, water, humans, and animals (Janowski, Leski, 2022). At the grass stage, the AMF proportion increases, as there are more plant species with which previously introduced AMF can enter into symbiosis. The maximum AMF percentage is attributed to their complementarity with grasses, high abundance of herbaceous plant species, and reduced competition among AMF. Soils become more fertile, and conditions become favorable for other fungi (Fig. 2), the proportion of which is maximum for this succession stage. In the shrub community, the AMF share is minimal, as well as the diversity and projective coverage of mycorrhizae-forming grasses; shrubs and undergrowth dominate here; they are proactive in forming a symbiosis with ectomycorrhizal fungi, which include many basidio- and ascomycetes (to a lesser extent). The picture is similar in plots with the forest stage, except that the trees are even more developed, which may influence the increased share of basidiomycetes; the shares of other groups are also significant, as the soils are more developed, and there is a lot of forest litter (Gorbunova, Sumina, 2021).

Conclusion

Regenerative succession in quarries is represented by four consecutive overgrowth stages of the free substrate and woody vegetation regeneration: pioneer, grass, shrub, and forest stages. AMF biodiversity is high at the first stages of regenerative succession. The highest AMF diversity is observed at the grass stage of the succession development. This is primarily due to the significant number of herbaceous species with which AMF form better symbiotic relationships. The lowest AMF biodiversity at the shrub stage is caused by the fact that herbaceous plants (including grasses) do not grow well in the shrub shade which reduces AMF species and their abundance. As has been shown, the most widespread AMF at the pioneer stage is *R. irregularis*. Other AMF types then take advantage. The results show that the biodiversity and species composition of AMF can vary widely among stages of regenerative succession and are likely to depend primarily on grass biodiversity. Although there are different hypotheses as to whether it is the fungus or the plant that is in charge, our work shows that it is likely that plants determine which AMF will be associated with them, and not vice versa.

References

- Aikio S. Plant Adaptive Strategies in Relation to Variable Resource Availability, Soil Microbial Processes and Ecosystem Development. Acad. diss. Oulu, 2000
- Akhmetzhanova A.A., Soudzilovskaia N.A., Onipchenko V.G., Cornwell W.K., Agafonov V.A., Selivanov I.A., Cornelissen J.H.C. A rediscovered treasure: mycorrhizal intensity database for 3000 vascular plant species across the former Soviet Union. *Ecology*. 2012; 93(2):689-690. doi 10.1890/11-1749.1
- Edgar R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460-2461. doi 10.1093/bioinformatics/btq461
- Ganugi P., Masoni A., Pietramellara G., Benedettelli S. A review of studies from the last twenty years on plant-arbuscular mycorrhizal fungi associations and their uses for wheat crops. *Agronomy*. 2019; 9(12):840. doi 10.3390/agronomy9120840
- Gorbunova A.O., Sumina O.I. Dynamics of mycorrhization in some plant species during progressive succession on sand quarries (Leningrad region). *Botanicheskii Zhurnal = Botanical Journal*. 2021; 106(1):22-42. doi 10.31857/S0006813621010051 (in Russian)
- Janowski D., Leski T. Factors in the distribution of mycorrhizal and soil fungi. *Diversity*. 2022;14(12):1122. doi 10.3390/d14121122
- Jeffries P., Barea J.M. Arbuscular mycorrhiza – a key component of sustainable plant-soil ecosystems. In: Hock B. (Ed.) *The Mycota*. Vol. 9. Berlin; Heidelberg: Springer, 2001;95-113. doi 10.1007/978-3-662-07334-6_6
- Jumpponen A., Trappe J.M., Cazares E. Occurrence of ectomycorrhizal fungi on the forefront of retreating Lyman Glacier (Washington, USA) in relation to time since deglaciation. *Mycorrhiza*. 2012;12: 43-49. doi 10.1007/s00572-001-0152-7
- Kiers E.T., Duhamel M., Beesetty Y., Mensah J.A., Franken O., Verbruggen E., Fellbaum C.R., Kowalchuk G.A., Hart M.M., Bago A., Palmer T.M., West S.A., Vandenkoornhuyse P., Jansa J., Bucking H. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science*. 2011;333(6044):880-882. doi 10.1126/science.1208473
- Kryukov A.A., Gorbunova A.O., Machs E.M., Mikhaylova Y.V., Rodionov A.V., Zhurbenko P.M., Yurkov A.P. Perspectives of using Illumina MiSeq for identification of arbuscular mycorrhizal fungi. *Vavilov J Genet Breed*. 2020;24(2):158-167. doi 10.18699/VJ19.38-o
- Lambers H., Raven J.A., Shaver G.R., Smith S.E. Plant nutrient-acquisition strategies change with soil age. *Trends Ecol Evol*. 2008;23(2): 95-103. doi 10.1016/j.tree.2007.10.008

- Maniatis T., Fritsch E.F., Sambrook J. Molecular Cloning. A Laboratory Manual. New York: Cold Spring Harbor Laboratory, 1982
- Neuenkamp L., Zobel M., Koorem K., Jairus T., Davison J., Öpik M., Vasar M., Moora M. Light availability and light demand of plants shape the arbuscular mycorrhizal fungal communities in their roots. *Ecol Lett.* 2021;24(3):426-437. doi 10.1111/ele.13656
- Öpik M., Davison J., Moora M., Zobel M. DNA based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences. *Botany.* 2014;92(2):135-147. doi 10.1139/cjb-20130110
- Pinaev A.G., Kichko A.A., Aksenova T.S., Safronova V.I., Kozhenkova E.V., Andronov E.E. RIAM: a universal accessible protocol for the isolation of high purity DNA from various soils and other humic substances. *Methods Protoc.* 2022;5(6):99. doi 10.3390/mps5060099
- Senés-Guerrero C., Schüßler A. A conserved arbuscular mycorrhizal fungal core-species community colonizes potato roots in the Andes. *Fungal Divers.* 2015;77(1):317-333. doi 10.1007/s13225-015-0328-7
- Smith S.E., Read D.J. Mycorrhizal Symbiosis. Cambridge: Academic Press, 2008
- Spatafora J.W., Chang Y., Benny G.L., Lazarus K., Smith M.E., Berbee M.L., Bonito G., Corradi N., Grigoriev I., Gryganskyi A., James T.Y., O'Donnell K., Roberson R.W., Taylor T.N., Uehling J., Vilgalys R., White M.M., Stajich J.E. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia.* 2017;108(5):1028-1046. doi 10.3852/16-042
- Sumina O.I., Vlasov D.Y., Dolgova L.L., Safronova E.V. Formation features of the micromycete communities in overgrown sand quarries of the North of Western Siberia. *Vestnik SPbGU. Seriya 3. Biologiya = Vestnik of Saint Petersburg University.* 2010;2:84-90 (in Russian)
- van der Heijden M.G.A., Klironomos J.N., Ursic M., Moutoglis P., Streitwolf-Engel R., Boller T., Wiemken A., Sanders I.R. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature.* 1998;396:72-75. doi 10.1038/23932
- van der Heijden M.G.A., Martin F.M., Selosse M.-A., Sanders I.R. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol.* 2015;205:1406-1423. doi 10.1111/nph.13288
- Wang B., Qiu Y.L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza.* 2006;16:299-363. doi 10.1007/s00572-005-0033-6
- Wu Q.-S. (Ed.) Arbuscular Mycorrhizas and Stress Tolerance of Plants. Singapore: Springer, 2017. doi 10.1007/978-981-10-4115-0
- Yurkov A.P., Kryukov A.A., Gorbunova A.O., Kudriashova T.R., Kovalchuk A.I., Gorenkova A.I., Bogdanova E.M., Laktionov Y.V., Zhurbenko P.M., Mikhaylova Y.V., Puzanskiy R.K., Bagrova T.N., Yakhin O.I., Rodionov A.V., Shishova M.F. Diversity of arbuscular mycorrhizal fungi in distinct ecosystems of the North Caucasus, a temperate biodiversity hotspot. *J Fungi.* 2024;10(1):11. doi 10.3390/jof10010011
- Zobel M., Öpik M. Plant and arbuscular mycorrhizal fungal (AMF) communities – which drives which? *J Veg Sci.* 2014;25(5):1133-1140. doi 10.1111/jvs.12191

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


Received August 30, 2024. Revised October 22, 2024. Accepted October 30, 2024.

doi 10.18699/vjgb-25-10

Symbiosis of intracellular bacteria *Wolbachia* with insects: a hundred years of study summarized

O.D. Shishkina , N.E. Gruntenko 

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

 shishkinaod@bionet.nsc.ru

Abstract. *Wolbachia pipientis* is an α -proteobacterium, which is a widespread intracellular symbiont in a number of Arthropoda and some Nematoda species. With insects, *W. pipientis* forms a symbiont-host system characterized by very close interactions between its components. The mutual effects of *Wolbachia* on the host and the host on *Wolbachia* are important biotic factors for both components of this symbiotic system. *Wolbachia* is able to affect both host reproduction and somatic organ function. Due to its prevalence among insects and a wide variety of both negative (cytoplasmic incompatibility and androcydosis are among the most well-known examples) and positive (increasing resistance to biotic and abiotic factors, providing vitamins and metabolites) effects on the host organism, *Wolbachia* is of great interest for both entomologists and microbiologists. The diversity of host phenotypes induced by *Wolbachia* provides a broad choice of evolutionary strategies (such as reproductive parasitism or mutually beneficial symbiont-host relationships) that it utilizes. The influence of *Wolbachia* is to be considered in the design of any experiment conducted on insects. The application of sequencing technologies has led to new approaches being created to study the existing relationships within the *Wolbachia*-insect system, but interpretation of the data obtained is challenging. Nevertheless, the prospects for the use of the whole-genome analysis data to study *Wolbachia*-host coevolution are beyond doubt. Ongoing projects to introduce *Wolbachia* strains, which provide antiviral host defense, into insect populations to control the spread of RNA-viruses are actively pursued, which could result in saving many human lives. The aim of this brief review is to summarize the data collected by scientists over the past hundred years of *Wolbachia* studies and the current understanding of its genetic diversity and mechanisms of interaction with the host, including those based on transcriptome analysis.

Key words: *Wolbachia*; insects; *Drosophila melanogaster*.

For citation: Shishkina O.D., Gruntenko N.E. Symbiosis of intracellular bacteria *Wolbachia* with insects: a hundred years of study summarized. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025;29(1):79-91. doi 10.18699/vjgb-25-10


Funding. The work was supported by BP No. FWNR-2022-0019.

Author contribution. O.D.S. – writing, original draft preparation, visualization; N.E.G. – writing, review and editing.

Симбиоз внутриклеточных бактерий *Wolbachia* с насекомыми: некоторые итоги ста лет изучения

О.Д. Шишкина , Н.Е. Груntenko 

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

 shishkinaod@bionet.nsc.ru

Аннотация. *Wolbachia pipientis* – α -протеобактерия, широко распространенный внутриклеточный симбионт у ряда видов Arthropoda и некоторых видов Nematoda. Вместе с насекомыми *W. pipientis* образует систему «симбионт–хозяин», характеризующуюся очень тесными взаимодействиями между ее компонентами. Влияния в обоих направлениях, которые оказывает как вольбахия на хозяина, так и хозяин на вольбахию, являются важными биотическими факторами для обеих составляющих этой симбиотической системы. Вольбахия способна оказывать воздействие как на размножение хозяина, так и на работу соматических органов. Благодаря своей распространенности среди насекомых и большому разнообразию как отрицательных (среди самых известных примеров – цитоплазматическая несовместимость и андроцид), так и положительных эффектов (повышение устойчивости к биотическим и абиотическим факторам, обеспечение витаминами и метаболитами), оказываемых на организм хозяина, вольбахия вызывает огромный интерес у энтомологов и микробиологов. Разнообразие вызываемых вольбахией фенотипов хозяина обеспечивает широкий выбор эволюционных стратегий, таких как репродуктивный паразитизм или взаимовыгодные отношения между симбионтом и хозяином, которыми она пользуется. Влияние вольбахии необходимо учитывать при постановке любого эксперимента, проводимого на насекомых. Применение технологий секвенирования привело к появлению новых подходов для изучения существующих связей внутри системы «*Wolbachia*–насекомое»,

однако интерпретация полученных данных представляет определенную сложность. Тем не менее перспективы использования данных полногеномного анализа для изучения коэволюции *Wolbachia* и хозяина не вызывают сомнений. Активно осуществляются проекты по внедрению в популяции насекомых штаммов вольбахии, обеспечивающих противовирусную защиту хозяина, для контроля распространения РНК-вирусов, что может способствовать спасению многих человеческих жизней. Целью этого обзора стало обобщение данных, полученных учеными за прошедшие сто лет изучения *Wolbachia*, и современных представлений о ее генетическом разнообразии и механизмах взаимодействия с хозяином, в том числе основанных на данных транскриптомного анализа.

Ключевые слова: *Wolbachia*; насекомые; *Drosophila melanogaster*

Introduction

Relations within the endosymbiont-host system deserve considerable attention from an evolutionary perspective because mutual adaptations of the symbiont to the host and the host to the symbiont guide the advancement of both species. Despite that, numerous surprising effects of symbiont influence on the host were not immediately linked to infection status. The observed effects of the intracellular α -proteobacterium *Wolbachia* on host insects are particularly well documented, but even in this symbiotic system the relationships remain poorly understood. At present, numerous studies of specific *Wolbachia* strains and their impact on completely different aspects of host species are being conducted using whole-genome sequencing and transcriptomic analysis. The purpose of this brief review is to highlight the progress that has been made in the field of studying the *Wolbachia*-host symbiotic system.

The establishment and development of an interest in *Wolbachia*

The genus *Wolbachia* belongs to the family Anaplasmataceae, a member of the order Rickettsiales, class α -proteobacteria (Hertig, Wolbach, 1924). *Wolbachia* is a widespread intracellular symbiont bacterium of a number of Arthropoda species and some Nematoda species. Approximately 50 % of all insect species on our planet are infected with this bacterium (Hilgenboecker et al., 2008; Zug, Hammerstein, 2012). The estimations of different groups of researchers vary due to the difficulty of conducting such large-scale studies and limitations in sample sizes. There is variation in the frequency of infection in different geographical locations, and in some, infection occurs at very low frequencies, which increases the likelihood of false negatives when testing for *Wolbachia* (since there is an increased probability of randomly selecting a sample without *Wolbachia*, even though it occurs in the host population).

Although the discovery of this bacterium took place a century ago (Hertig, Wolbach, 1924), even the specification of the number of species in the genus *Wolbachia* has been a matter of debate for many decades. The reason is that there is no clear concept of species boundaries applicable to endosymbiotic bacteria. At the moment, it is accepted that all discovered variants of *Wolbachia* belong to one species, *Wolbachia pipientis*. In this paper, according to tradition, this bacterium will be referred to as *Wolbachia* (genus name only) or *Wolbachia pipientis* (name of the species that has

remained traditionally). However, it should be noted that there is still no established consensus in the research community on the vagueness of the taxonomy of this genus (Lo et al., 2007).

It is believed that M. Hertig and S.B. Wolbach (1924) were driven to the discovery of the bacterium, which has been defined as “rickettsia-like,” by a deadly typhoid epidemic (Porter, Sullivan, 2023). Typhus is a disease, the source of which is the bacterium *Rickettsia prowazekii*, a bacterium carried by the body louse *Pediculus humanus corporis* (Linnaeus, 1758). As a result of the search for potential agents of typhus, other intracellular organisms have been discovered that later acquired the name *Wolbachia pipientis* (Porter, Sullivan, 2023). Although *Wolbachia* is not a threat to humans, as with many discoveries in biology, the initial stimulus for the development of the study of this genus came from medicine.

After the first discovery of this bacterium and several years of dormancy, the next discovery that revitalized interest in *Wolbachia* was the conditional sterility of some insects caused by certain *Wolbachia* strains. To this day, this effect is the most well-known when it comes to this bacterium (Burdina, Gruntenko, 2022). The underlying mechanism behind this phenomenon is called cytoplasmic incompatibility (CI) (Laven, 1967). The way cytoplasmic incompatibility is realized in the first mitotic division of the zygote was later studied cytologically (Ryan, Saul, 1968). But only a relatively short time ago the elements that cause CI have been elucidated (Beckmann et al., 2017; LePage et al., 2017; Chen et al., 2019).

The current understanding of the prevalence of *Wolbachia* in insects would not be possible without the PCR identification of *Wolbachia*-specific DNA-markers. Even with the latest light and fluorescence microscopes, it is difficult to repeat M. Hertig and S.B. Wolbach’s achievement (Hertig, Wolbach, 1924) for other insects because *Wolbachia* are often inferior in size (diameter 0.25 to 1.8 μm) even to mitochondria (Yu, Walker, 2006). Screening as many insect species as possible by analyzing cytological specimens, which for each host species requires several specimens isolated from populations (single isolates), is an almost impossible task, while the same volume of isolates examined by the more sensitive PCR method requires less time and effort. With the help of this key molecular technique, modern biology has been able to discover that *Wolbachia* lives in almost all insects on the planet (Hilgenboecker et al., 2008).

Influence of *Wolbachia* on the host

Wolbachia are vertically transmitted from the mother to the offspring through the cytoplasm of oocytes. The transmission mechanism may not always run flawlessly and sometimes spontaneous loss of infection occurs (Werren, 1997). Nevertheless, *Wolbachia* is consistently found in natural and laboratory insect populations. *Wolbachia* has no free-living analogues; all representatives of the Rickettsiales order, to which it belongs, are intracellular organisms (Yu, Walker, 2006). The ecological niche occupied by *Wolbachia* is the internal environment of its animal host. It grows in the cytoplasm of its host cell in the membrane-bound vacuole (Yu, Walker, 2006). For that reason the interactions occurring between *Wolbachia* and the host are very intimate and are both mutualistic and parasitic in nature (Burdina, Gruntenko, 2022).

Maternal inheritance is a common feature for mitochondria and *Wolbachia*. In addition, they are inherited in a linked manner rather than independently, forming a certain “cytotype” (Ilinsky, 2013). Due to the intracellular nature of this bacterium, its study is complicated; for example, it is difficult to explore the specifics of its metabolism. The establishment of passaged cell cultures of the insect hosts containing *Wolbachia* is possible, but also complicated. The first such cell line was created from cells of the mosquito *Aedes albopictus* (O’Neill et al., 1997). Stable cell cultures might become an invaluable tool for studying this genus, but this is made challenging by the spontaneous loss of infection that often occurs in them.

Wolbachia can only live in symbiosis with its host, but most host species are able to live and reproduce while uninfected. The influence *Wolbachia* has on the host is demonstrated in a number of different traits that can be observed when comparing infected host individuals with uninfected ones, as well as comparing individuals infected with different strains (Burdina, Gruntenko, 2022). The successful expansion of *Wolbachia* is partially explained by the ability of this organism to interfere with sex determination mechanisms, alter the development and reproductive patterns of the host for its own benefit.

Among the numerous effects of *Wolbachia*, the most well-studied ones are those it has on the reproductive function of the host:

- androcyde – selective death of males on the embryonic or larval developmental stage,
- feminization of genetic males – acquisition of phenotypic traits of females by infected males,
- stimulation of parthenogenesis,
- cytoplasmic incompatibility.

The most attention was always paid to the phenomenon of cytoplasmic incompatibility caused by *Wolbachia*. CI in insects is defined as follows: infected females can reproduce by being fertilized by both uninfected males and infected males, while uninfected females cannot reproduce with infected males (Kaur et al., 2021). Thus, infected females do not experience the negative consequences of CI and have reproductive advantage. Since *Wolbachia* is inherited

through the maternal lineage along with the cytoplasm, this mechanism ensures that *Wolbachia* is effectively spread in insect populations (Lassya, Karrb, 1996).

Molecular mechanisms responsible for causing CI are connected to the disruption of the first mitotic division of the zygote (Poinso et al., 2003). It has been shown, that the deubiquitylase CidA, which initiates CI in males, and protein CidB, which allows to overcome it, when expressed in females, are involved in the formation of CI (Beckmann et al., 2017). This confirms the previously formulated hypothesis of the “modification–rescue” pair put forward to explain the phenomenon of CI (Werren, 1997). Genes of the CI factors, called *cifA–cifB* pairs, were found to be integrated from the prophage WO into the genomes of *Wolbachia* that cause CI in the host (LePage et al., 2017). An alternative mechanism utilizes the nuclease CinA and its binding protein CinB (Chen et al., 2019), which also operate as the same “modifier–rescuer”. No other mechanism of *Wolbachia*’s influence on the host has been described in such detail.

Wolbachia-induced reproductive effects have been found in various insects, but there are exceptions, for example, they are completely not characteristic or weakly manifested in the most of the studied *Drosophila melanogaster* lines (Fry et al., 2004; Ilinsky, Zakharov, 2011).

Besides *Wolbachia*’s influence on reproduction, numerous effects it causes on the somatic cells of the host have been discovered. This is possible due to the fact that *Wolbachia* is found not only in the reproductive organs of the females, where it is the most expected based on the mechanism of transmission of this symbiont to offspring, but also in the fat body, Malpighian vessels, muscle and nerve tissues (Fig. 1a) (Pietri et al., 2016). A variant of the pathogenic *Wolbachia* strain wMelPop, infecting *D. melanogaster* and known for causing premature death of the flies (Min, Benzer, 1997), wMelPop-CLA, changes male behavior, reducing male aggression by decreasing octopamine production in the brain (Rohrscheib et al., 2015). Individuals of *D. melanogaster* with the same nuclear genotype but infected with different strains of *Wolbachia* have different optimal temperature ranges (Truitt et al., 2019). This may affect the prevalence of certain strains at different latitudes.

Maintaining an endosymbiont is usually associated with costs to the host in terms of resources that both it and the bacterium require. Often more successfully selected by evolution are those symbionts that can provide greater benefit to the host. This minimizes the effect of its costs in maintaining the symbiont, and a mutually beneficial relationship is established in the system. The strategy of providing benefit by positively affecting aspects of the host’s life may explain why *Wolbachia* is so common among species in which its manipulation of the host’s reproductive system is not pronounced.

Increased lifespan has been shown in infected *Drosophila* (Maistrenko et al., 2016); *Wolbachia* can also supply their hosts with vitamins and essential amino acids. For example, the wCLe strain, infecting the bedbug *Cimex lectularius*, provides the host with vitamin B7 (biotin) (Newton,

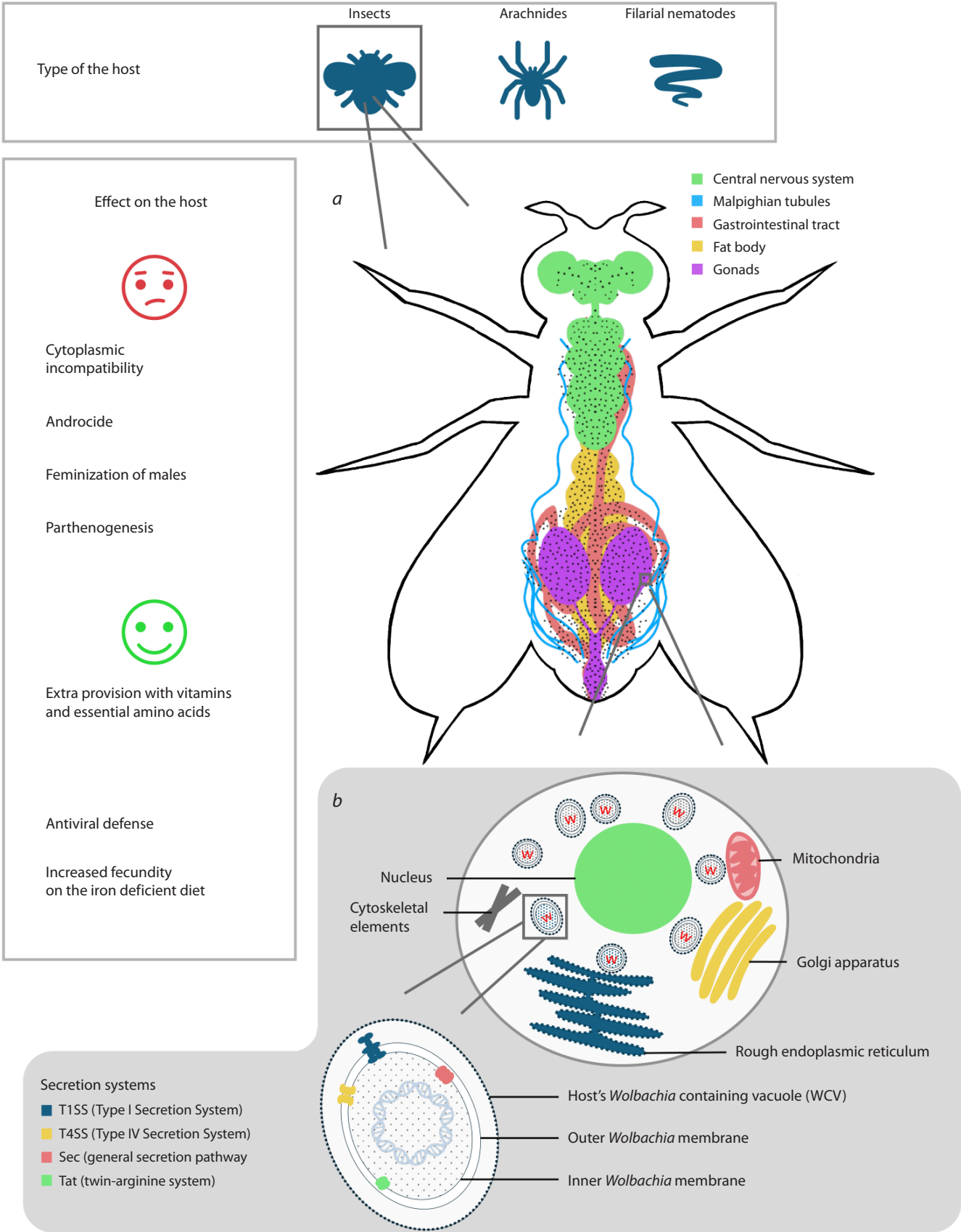


Fig. 1. Scheme of *Wolbachia*–insect symbiotic system at different levels of organization.
a – localization of *Wolbachia* in different organs of *Drosophila*; *b* – localization of *Wolbachia* in the insect host cell and organelles that interact with *Wolbachia*. The *Wolbachia*-containing vacuole is shown at the bottom of the illustration.

Rice, 2020). *Wolbachia* protects the cells of parasitoid wasps *Asobara tabida* from iron excess through expression of bacterioferritin (Kremer et al., 2009), and in *D. melanogaster* infected with *Wolbachia*, fecundity is increased when maintained on iron-deficient diet (Brownlie et al., 2009).

Bacterioferritin binds free divalent iron and promotes absorption in the gut of the fly larvae (Brownlie et al., 2009). Evidence has been received that host insulin/insulin-like growth factor signaling cascade is suppressed and the hypoxia-inducible factor (HIF) signaling cascade is activated upon *Wolbachia* infection (Currin-Ross et al., 2021). *Wolbachia* have been shown to require iron acquisition from the host, based on which the authors suggest that iron is a fundamental aspect of *Wolbachia*–host interactions (Currin-Ross et al., 2021).

Because reproductive anomalies induced by *Wolbachia* are not characteristic of *D. melanogaster*, other physiological effects observed in this symbiotic pair have received more attention. There have been numerous studies carried out on the effects of specific strains of this bacterium on different lines of this host species. This makes *Wolbachia*–*D. melanogaster* symbiotic system one of the most studied in terms of the genetic diversity of both the host and the bacterium, as well as the effect of the combination of their genotypes on host fitness.

For instance, three lines of *D. simulans* of different origin, but infected with the same *Wolbachia* strain, showed different effects of the symbiont on host adaptability: in one of the lines under study, the introduction of *Wolbachia* by microinjection increased the fitness estimated in population cage assays, meanwhile, in two other lines, the fitness was not influenced by the bacterium (Dean, 2006). Different effects of *Wolbachia* on the lifespan, fecundity and developmental rate of different *D. melanogaster* lines were also found by A.J. Fry and D.M. Rand (2002) and N.V. Adonyeva et al. (2023).

On the other hand, infecting a single line of *D. melanogaster* with different *Wolbachia* variants resulted in changes in dopamine metabolism in flies infected with *Wolbachia* of the wMelCS genotype, but not in those infected with *Wolbachia* of the wMel genotype (Gruntenko et al., 2017; Burdina et al., 2021). Similar differences in the influence of *Wolbachia* genotype on its effects on host physiology have been shown for juvenile hormone metabolism (Gruntenko et al., 2019).

At the same time, several effects on *D. melanogaster* attributed to *Wolbachia*, as far as is currently known, do not depend on symbiont genotype. So, infection of one line of *D. melanogaster* with seven different *Wolbachia* variants promoted an increase in the host fly's lipid stores (Karpova et al., 2023). An increase in glucose and triglyceride (TAG) content in the host was also common to different bacterial variants (Zhang et al., 2021; Karpova et al., 2023), but trehalose levels remained unchanged in all lines compared to uninfected flies (Karpova et al., 2023). These lines differed from uninfected lines in their increased survival under nutritional deficiency. Increased glucose-6-phosphate levels were

also observed in *Wolbachia*-infected mosquitos *Aedes fluviatilis* (da Rocha Fernandes et al., 2014).

A study carried out on a transgenic line of *D. melanogaster* with impaired function of insulin receptor showed that the presence of *Wolbachia* increases the adaptability of such mutants (Ikeya et al., 2009). Removal of *Wolbachia* by antibiotics in such flies resulted in an enhanced mutant phenotype (which is manifested by reduced growth and fecundity). The authors hypothesized that *Wolbachia* activates insulin/insulin-like growth factor (I/IGF) signaling cascade (Ikeya et al., 2009). However, a more recent study suggests otherwise. In the work (Currin-Ross et al., 2021), they examined the metabolic response of *D. melanogaster* to infection status and showed that the I/IGF-mediated signaling pathway is suppressed by *Wolbachia*.

Some strains of *Wolbachia* are known to improve the host's defense against a number of pathogens, as they are able to inhibit the replication of RNA viruses (Hedges et al., 2008; Teixeira et al., 2008; Moreira et al., 2009). Due to their antiviral defense properties, *Wolbachia* are used for biological control purposes (Hoffmann et al., 2011; LePage, Bordenstein, 2013). A number of *Wolbachia* strains, the native host of which is *D. melanogaster*, have been introduced by microinjection into individuals of the mosquito *Aedes aegypti*, which is a vector of dengue virus (dengue virus – DENV) (Hoffmann et al., 2011, Gu et al., 2022). Introduction of *Wolbachia*-infected individuals into natural populations resulted in their successful spread due to CI (Hoffmann et al., 2011), which may reduce the efficiency of dengue virus transmission, since blocking of the latter by *Wolbachia* in *Ae. aegypti* has been demonstrated in laboratory conditions (Gu et al., 2022).

There are several hypotheses as to how different properties of *Wolbachia* strains may influence antiviral defense, and selection of the most effective strains is the goal of many studies. Since CI promotes the predominant spread of a particular strain (the one that causes this abnormality in the host) in the population, the joint inheritance of antiviral defense and the ability to induce CI makes such strains more effective when using a substitution strategy. It is noted that strains characterized by increased *Wolbachia* content in host cells (such as wMelPop) contribute more to the host's ability to successfully fight the virus. Based on this fact, it is hypothesized that there is a correlation between the effectiveness of antiviral defense and high *Wolbachia* content in cells (Chrostek et al., 2013; Gu et al., 2022). However, the optimal temperature range for *Wolbachia* strains in the habitat of the insects, into the population of which a new *Wolbachia* strain is introduced, is also worth considering. Attempts have been made to use *Wolbachia* to control other arboviruses that pose a threat to humans (Kametchum-Tatuene et al., 2017).

The adaptive or deleterious nature of some *Wolbachia* effects is difficult to determine unequivocally, but it is generally clear that some of them (for example, manipulation of host reproduction) can be attributed to parasitic effects, whereas other effects, such as increased resistance to viral infection and starvation, provide an adaptive advantage not

only to endosymbionts in this system but also to host insects. The full range of effects of *Wolbachia* on the host cannot be considered without addressing the diversity of strains of this bacterium, as many of the effects it exerts are specific to a particular strain of *Wolbachia*. Although *Wolbachia* genomes share a common core set of genes, different strains differ significantly from each other.

Genetic diversity of *Wolbachia*

Since it is generally accepted that there is only one species of *Wolbachia* – *W. pipientis* (Hertig, Wolbach, 1924), the entire diversity of these insect endosymbionts is described by different strains divided into supergroups. The division into supergroups is based on phylogenetic analysis of the sequences of several genes used for multilocus typing. Several groups of genes for multilocus typing of *Wolbachia* strains have been proposed: *dnaA*, *16SrRNA*, *wsp*, *gltA* and *groEL*, *ftsZ* (Lo, Evans, 2007), *gatB*, *hcpA*, *fbpA*, *coxA* (Baldo et al., 2006b). According to different sources, from 10 to 13 supergroups are distinguished, designated by Latin letters A–F, N–M and S (Kaur et al., 2021); the classical classification includes seven supergroups (A–F and H) (Ros et al., 2009; Augustinos et al., 2011).

The most universal genotyping system – the process of identifying genetic differences and similarities between different groups of organisms – for *Wolbachia* strains is currently multilocus sequence typing (MLST), which uses five protein-coding genes: *ftsZ*, *gatB*, *coxA*, *hcpA* and *fbpA* (Baldo et al., 2006b). Based on analysis of the combination of 5 or more polymorphic markers, ST (sequence type) profiles are compiled. The utilization of several alleles as markers provides more accurate and complete information than the utilization of a single allele.

The genomes of *Wolbachia* are characterized by a wide diversity, which is also formed by strain isolation due to maternal inheritance along with the cytoplasm. Although the hosts may be closely related species, their associated *Wolbachia* strains can differ greatly at the genetic level. Sequences from hypervariable loci can be used to separate recently diverged strains, although the possibility of recombination of *Wolbachia* strains, which has been demonstrated experimentally (Baldo et al., 2006a), and the presence of a large number of repeats and mobile elements in the *Wolbachia* genome (Wu et al., 2004) must be taken into account.

In the vast array of host–*Wolbachia* combinations, each is characterized by its own unique set of adaptations of the symbiont to the host and vice versa, which affects the type of symbiotic relationship. In addition, new strains of this bacterium are discovered and described almost every year, and, as a general rule, researchers focus their work on the effects of specific *Wolbachia* strains on their objects of interest (Burdina et al., 2021; Duarte et al., 2021; Ilinsky et al., 2022).

We will examine in more detail the diversity of *Wolbachia* strains found in the classical model object *D. melanogaster*. *Wolbachia* infection in *D. melanogaster* was first detected in 1988 (Hoffman, 1988), but the wMel strain was described

only ten years later (Zhou et al., 1998). In 2005, M. Riegler et al. (2005) identified five different *Wolbachia* genotypes in *D. melanogaster* based on polymorphic markers. Several different lineages were assumed to have originated from a single ancestral *Wolbachia* infection (Riegler et al., 2005; Hilgenboecker et al., 2008). In the literature, new and first described *Wolbachia* in *D. melanogaster* are usually referred to as strains (Lo et al., 2007). Often there is insufficient information in a study presenting a new strain to assign it to one of the known genotypes.

To date, six genotypes of *W. pipientis* found in *D. melanogaster* have been described (Fig. 2). They are divided into two groups: wMel (which includes genotypes wMel, wMel2, wMel3, wMel4) and wMelCS (which includes wMelCS and wMelCS2) (Riegler et al., 2005; Ilinsky, 2013). Sequencing of *Wolbachia* genomes revealed the presence of a large number of repeats, including insertion sequences (IS) and variable number tandem repeats (VNTR). Genotypes are distinguished by polymorphisms of five genome markers: the presence of inversion in the locus WD0394–WD0541 (in Figure 2, the direction of the fragment is indicated by an arrow); variable number tandem repeat markers VNTR-105, VNTR-141 (in Figure 2, the number of repeats is indicated by numbers under them); IS5 WD1310, IS5 WD0516/7 – IS element insertion loci. These markers are used for genotyping *Wolbachia* from isolates of natural and laboratory populations of *D. melanogaster* (Riegler et al., 2005; Ilinsky, 2013).

It should be noted that two strains have also been described for the wMelCS genotype that differ in their effect on the host and in their genetic composition, although these differences are not detected by Riegler genotyping (Riegler et al., 2005). The first of these strains is the pathogenic strain wMelPop (from the word “popcorn”), which causes premature death of flies infected with it through its unrestricted proliferation leading to overcrowding and rupture of host cells (Min, Benzer, 1997) and has an increased copy number of a region of eight *Octomom* genes that has been associated with the pathology caused by the wMelPop strain (Chrostek et al., 2013; Chrostek, Teixeira, 2015). The second strain, wMelPlus (from “plus”, meaning a “positive sign”), not defined by M. Riegler et al. but distinguished by a large (approximately 1/6 of the genome) inversion from other representatives of the wMelCS genotype (Korenskaia et al., 2022), on the contrary, has a positive effect on host fitness, increasing its resistance to heat stress (Burdina et al., 2021). The discoveries of these strains were a great surprise when investigating the phenotypic differences between *D. melanogaster* lines carrying them and lines with “normal” characteristics. A strain named wMelM that increases host resistance of *D. melanogaster* to heat stress, but does not differ in markers (according to M. Riegler et al.) from the wMel genotype was also discovered (Gu et al., 2022). These three examples demonstrate that great genetic diversity can be hidden from researchers behind identical genotype labels.

Whole-genome sequencing is suitable for detecting such differences in the genome of strains. It should be taken into

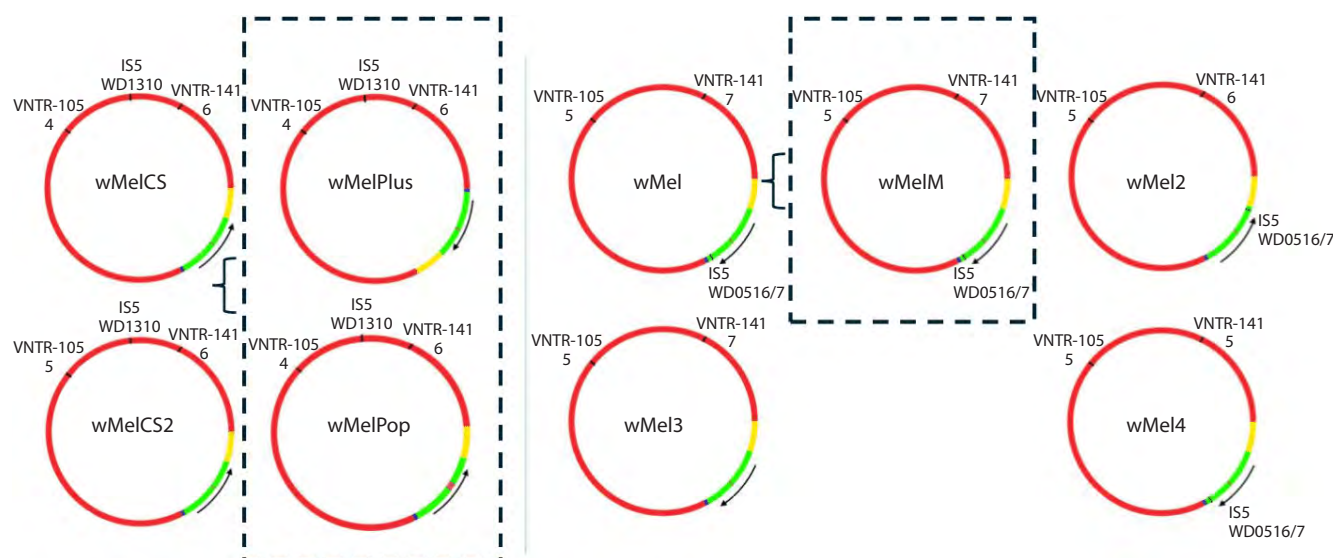


Fig. 2. Chromosome maps of six different *W. pipientis* genotypes isolated from *D. melanogaster*, as well as three unique strains (wMelPlus and wMelPop, belonging to the wMelCS genotype, and wMelM, belonging to the wMel genotype).

The green color indicates the inversion that distinguishes the wMel genotype from wMelCS. The yellow and blue regions denote sequences included in the inversion in wMelPlus but unaffected by the inversion in the wMel group. The magenta region denotes the Octomom sequence (Chrostek, Teixeira, 2018).

account that when assembling the genome using a reference genome, it is possible to miss the presence of inversions (there are difficulties due to the presence of repeats in the genome and short lengths of reads while sequencing).

A number of issues related to *Wolbachia* genotypes that infect *D. melanogaster* deserve special attention. In natural populations of *D. melanogaster*, genotypes wMel and wMelCS are the most commonly found, with wMel significantly predominating (Riegler et al., 2005; Nunes et al., 2008; Ilinsky, 2013). It is hypothesized that this genotype gradually displaced the previously predominant wMelCS (Riegler et al., 2005). It has been shown that the effect of shifting thermal preference toward lower temperatures in *Drosophila* infected with *Wolbachia* compared to uninfected flies is strongest in *D. melanogaster* lines infected with *Wolbachia* strains of the wMelCS group (Truitt et al., 2019). On the other hand, there is evidence of low genetic polymorphism of wMelCS group genotypes in the Palaearctic, contradicting the hypothesis that the global replacement of *Wolbachia* genotypes occurred recently and indicating that there is still much that remains unknown in this field (Bykov et al., 2019).

The strain wMel is the first strain of the *Wolbachia* bacterium with a completely sequenced and annotated genome (Wu et al., 2004). The genome size of this strain is 1,267,782 bp; it includes about 1,270 protein-coding genes (Porter, Sullivan, 2023). No significant differences in size and gene composition from strains of the wMelCS genotype (excluding the unique wMelPop strain, which has a special genome region formed by *Octomom* sequence repeats (Duarte et al., 2021)) have been shown (Chrostek et al., 2013; Korenskaia et al., 2022).

Studies dedicated to the mechanisms of interactions in the *Wolbachia*–host system

Large-scale searches for possible effector molecules, which *Wolbachia* can utilize to have an influence on the host's organism, have been conducted (Ote et al., 2016; Sheehan et al., 2016; Rice et al., 2017). For a bacterium to influence processes within eukaryotic host's cells, the effector molecules presumably must have homology with some molecules synthesized in the host organism.

The bacterial genome often acquires foreign genetic material from eukaryotic cells that retains at least some of its original activity, and the products of these domains are released into the cytoplasm of the eukaryotic cell (De Felipe et al., 2005). A study was conducted in which 163 gene candidates from the genome of the wMel strain were selected via bioinformatics methods, and then 84 transcription products of these genes were analyzed for their effects on the yeast *Saccharomyces cerevisiae* (Rice et al., 2017). In this analysis, yeast growth defects and 14 possible effector genes were identified (Rice et al., 2017), three of which contain ankyrin repeats, which may indicate their involvement in protein-protein interactions with their arthropod hosts.

Since there is a barrier between endosymbiont and host organisms, specialized secretion systems are required to release effector molecules outside the bacterium. Bacterial secretion systems consist of protein complexes and are responsible for the passage of macromolecules through membranes. In bacteria, secretion is necessary for adaptation to environmental conditions and to enable pathogenicity in some bacteria. Due to *Wolbachia* being an endosymbiont, secretion system is an important tool for interactions with the host's cells. *Wolbachia* utilizes two types of secretion sys-

tems (Fig. 1b): T1SS – type I secretion system and T4SS – type IV secretion system (Lindsey, 2020). The first type of secretion system consists of three proteins: ABC-transporter, which is ATP-dependent, MFP – membrane fusion protein and OMP – outer membrane protein. The forth type of secretion system usually consists of 12 protein components: VirB1–VirB11 and VirD4 (Fronzes et al., 2009). Genes of this secretion system are located in two clusters in the *Wolbachia* genome: tandem genes of five proteins (VirB8, VirB9, VirB10, VirB11, VirD4) and those of three proteins (VirB3, VirB4, VirB6); meanwhile, genes VirB1, VirB2, VirB5 and VirB7 have been eliminated. The sequence and organization of these genes have been shown to be conserved in 37 *Wolbachia* strains under study (Pichon et al., 2009). These two secretory systems allow the secretion of a wide range of substrates, from single proteins to protein-protein and protein-DNA complexes (Backert, Meyer, 2006).

In the *Wolbachia* genomes, there are genes coding the channels of the Sec (general secretion system) and Tat (twin-arginine translocation) systems. These systems are involved in the protein transport through the *Wolbachia*'s cell membrane into the periplasmic space (Sec transports unfolded proteins, while Tat transports proteins folded to the tertiary structure) (Lindsey, 2020).

Wolbachia-containing vacuoles share a common origin with the Golgi apparatus and the endoplasmic reticulum of insects (Fig. 1b, the location of vesicles in the cell, the lower part of the Figure depicts one such vacuole) (Cho et al., 2011). It is suggested that the polar proteins Van Gogh/Strabismus and Scribble can be responsible for positioning of such vacuoles close to the site of membrane synthesis (Cho et al., 2011). *Wolbachia* interacts with the cytoskeleton of the host's cell to maintain the integrity and stability of vacuoles, similar to how the bacterial pathogens utilize such vacuoles to defend themselves against the host's immune system (Ferree et al., 2005; Kumar, Valdivia, 2009; Creasey, Isberg, 2014).

Wolbachia requires a supply of many metabolites from the host (Jiménez et al., 2019; Newton, Rice, 2020). It has been hypothesized that the wMel strain native to *D. melanogaster* is dependent on the host for alanine, glycine, and serine metabolism, as well as lipopolysaccharide and biotin production (Jiménez et al., 2019; Newton, Rice, 2020). *Wolbachia* is completely dependent on the host for iron supply (Gill et al., 2014; Jiménez et al., 2019). On the other hand, dependence on substances supplied by *Wolbachia* has been shown for some insect species. For example, the bedbug *Cimex lectularius* utilizes riboflavin (Moriyama et al., 2015) and biotin (Nikoh et al., 2014) provided by the bacterium.

Among the key mechanisms of *Wolbachia*-host interaction is its impact on the cytoskeleton of host cells. Interaction with dynein and kinesin of host cell microtubules ensures *Wolbachia*'s passage into oocytes and hence its spread to the next generation (Ferree et al., 2005). *Wolbachia* is also reliant on clathrin/dynein-dependent capture by host cells for transport from somatic cell to germ cell (White et al., 2017).

Spontaneous loss of *Wolbachia* is sometimes reported, which can be explained by the response of the host's immune system to the bacterium. Damaged organelles (for example, mitochondria) pose a threat to the cell. When such damage is detected, the organelle is eliminated by selective autophagy. This mechanism has recently been shown to be applicable to *Wolbachia* (Hargitai et al., 2022). Lysosome-mediated degradation of vacuoles containing *Wolbachia* may be a major cause of the host curing itself. Aging has been shown to decrease the efficiency of *Wolbachia* removal from the cells, resulting in *Wolbachia* actively proliferating and increasing its density in the host cells (Hargitai et al., 2022). Based on the obtained data, the authors conclude that autophagy may be a mechanism for controlling *Wolbachia* virulence.

It is logical to assume that if endosymbionts are observed in many generations of the same hosts, the host immune response to that organism is reduced. Since *Wolbachia* is the most common symbiont of invertebrates, it is likely that these bacteria have evolved an effective mechanism of protection against the host's immunity, which only occasionally fails. It has been hypothesized that a new acquisition of *Wolbachia* infection triggers an immune response and oxidative stress in the host, whereas if there is evidence of a long time of symbiosis with a particular strain (a stable association of a strain of bacterium and a particular insect population), infection is not associated with oxidative stress (Zug, Hammerstein, 2015).

Transcriptome analysis studies dedicated to the interactions in the *Wolbachia*-host system

Current approaches to determining the links between *Wolbachia* and the host rely on sequencing analysis. It is important to interpret the data from the studies of *Wolbachia* strain genomes in tandem with the results of host transcriptome studies.

Transcriptome analysis of the *D. melanogaster* lines infected with *Wolbachia*, equally with genomic studies, may shed light on the molecular mechanisms of interaction between these parts of the system. However, this method has drawbacks that have been repeatedly emphasized in the conducted studies. The host's material is always in a larger quantity than material from the endosymbiont. To get around this limitation, it would make sense to use not the whole insect, but only the organs that have a higher density of this bacterium. The reproductive organs of the insect are suitable for this requirement, and appropriate studies have been made: on the ovaries (He et al., 2019; Frantz et al., 2023) and on the testes (He et al., 2019; Detcharoen et al., 2021). However, differences in gene expression levels between independent samples of the same type (one line infected with one strain) are often as significant as differences in gene expression levels between different types of samples (Detcharoen et al., 2021). This is most likely due to the contribution of other factors, such as unstable external conditions at the time of RNA extraction.

The transcriptome in *Wolbachia*-infected *D. melanogaster* has also been analyzed using virgin and fertilized females

(Detcharoen et al., 2021; Lindsey et al., 2021; Gruntenko et al., 2023), embryos (Mateos et al., 2019). However, the latter work found no significant differences in mRNA makeup between *Wolbachia*-infected and uninfected embryos (Mateos et al., 2019), which can probably also be explained by the contribution of other factors.

Despite these drawbacks of using transcriptome analysis to study the influence of *Wolbachia*, it has been able to provide meaningful results concerning different aspects of the *Wolbachia*–*Drosophila* interaction. Further on we review several studies conducted over the last five years.

In a study investigating the phenomenon of CI and its mechanisms, first the ovarian transcriptome and then the testes transcriptome of adult *D. melanogaster* were analyzed (He et al., 2019). Comparisons were made between the transcriptomes of uninfected insects and those infected with the wMel strain. The authors identified the following functional groups of genes that are potentially susceptible to *Wolbachia*: “metabolism”, “transport”, “oxidation-reduction processes”, “immunity” and “individual development”. The authors hypothesize that *Wolbachia* is responsible for the regulation of the transcription in the opposite directions of a number of genes in female and male *Drosophila*. According to this hypothesis, when infected males mate with uninfected females, the resulting embryos have an imbalance in the levels of fertility restoration components, causing a cytoplasmic incompatibility effect (He et al., 2019). This popular hypothesis of the origin of CI is called titration-restitution model (Poinso et al., 2003).

Another group of researchers also obtained transcriptome data on the topic of cytoplasmic incompatibility. A study was conducted to investigate the effect of various endosymbiotic bacteria on the transcriptome of early *D. melanogaster* embryos, but the authors found no effect of the *Wolbachia* wMel strain used in the study on the host transcriptome (Mateos et al., 2019). The authors concluded that the wMel strain does not alter maternal transcripts and does not lead to their degradation (Mateos et al., 2019).

There was a study of *Wolbachia*’s influence on *D. melanogaster* lines with different genotypes (Frantz et al., 2023). The authors studied ovarian transcriptomes of eight lines of *D. melanogaster*: four genetically diverse lines carrying one genotype of *Wolbachia* and derivatives of these lines that were cured of *Wolbachia* by tetracycline treatment. The host’s line genotype turned out to be a more significant factor affecting the transcriptome of the lines studied than the presence or absence of *Wolbachia* in them. However, the authors were still able to detect *Wolbachia*-induced differences in the expression of host genes involved in pathways related to cell cycle checkpoints, translation and metabolism, as well as cell division and recombination processes (Frantz et al., 2023).

The study conducted on the testes of two *Drosophila* species was aimed at investigating differences in the effect of the wMel strain on the native host species (*D. melanogaster*) and on a novel host species (*D. nigrosparsa*) to which the

indicated strain was introduced by artificial transinfection of *Wolbachia* (Detcharoen et al., 2021). The detected differences affected such groups of orthologous genes as “oxidation-reduction processes”, “iron ions binding”, “activity of voltage-gated potassium channels” (Detcharoen et al., 2021).

In order to investigate the mechanisms of antiviral protection of host insects provided by *Wolbachia*, the transcriptomes of *D. melanogaster* flies infected with the *Wolbachia* wMel2 strain were analyzed (Lindsey et al., 2021). Two factors were simultaneously taken into account in the experimental design: *Wolbachia* infection or its absence, and Sindbis virus (SINV) infection or its absence. Four groups of insects (all possible combinations of these two factors) were acquired.

As a result of this analysis, the authors identified the following functional groups of genes that are potentially susceptible to *Wolbachia*: “stress response”, “RNA binding and processing”, “metabolism”, “ubiquitination”, and “transcription and translation”. The authors were unable to identify specific genes, the expression level of which would change as a result of the interaction between *Wolbachia* and virus. However, they constructed one core gene network linking genes responding to *Wolbachia*, genes responding to viruses, and genes, the response of which was induced by the combined effect of *Wolbachia* and the virus. Only genes attributed to the “metabolism” group (mainly amino acid metabolism and purine biosynthesis) got included in this network. The authors suggested that the discovered effect of *Wolbachia* on the synthesis of host nucleotides may be the reason for the suppression of virus replication (Lindsey et al., 2021).

In the study of the positive effect of the *Wolbachia* wMel-Plus strain on stress resistance of *D. melanogaster* flies, the transcriptomes of adult females of three lines of flies with the same nuclear genotype but differing in infection status (uninfected, infected with the wMelPlus strain, infected with the wMelCS¹¹² strain) were compared (Gruntenko et al., 2023). Both *Wolbachia* strains induced changes in the expression levels of genes that belong to the functional groups “transmembrane transport”, “proteolysis”, “carbohydrate transport and metabolism”, “oxidation-reduction processes”, “regulation of alkaline phosphatase activity”, “embryogenesis”, and “stress response”. Nevertheless, the groups’ composition of differentially expressed genes partially differed between fly lines infected with different strains of *Wolbachia* (a pairwise comparison of the transcriptomes of infected fly lines against the transcriptomes of uninfected ones was conducted). The main difference in the expression of stress response genes was an increase in the level of transcription of the corazonin receptor (*CrzR*) gene in flies infected with the wMelPlus strain. Differences were also found between fly lines infected with different *Wolbachia* strains in the expression of different genes of alkaline phosphatases (which play a role in the stress response, participate in the dopamine synthesis cascade) (Gruntenko et al., 2023).

To summarize the above-mentioned studies, it can be concluded that *Wolbachia* affects the expression of hundreds of genes in flies of the genus *Drosophila*. These changes affect a multitude of processes that are combined into functional groups of the genes involved, the list of which differs in a number of studies. In turn, the functional groups can be matched to known *Wolbachia* effects that influence the observed host phenotype. The results of these large-scale transcriptome studies of *Wolbachia*-infected insects may help to guide more pinpointed experiments to specify the mechanisms in the *Wolbachia*–host system in the future.

With the development of sequencing technologies, new tools have become available. CappableSeq has been used to assemble the *Wolbachia* transcriptome of nematodes (Luck et al., 2017). This method could also be very promising for the study of insect *Wolbachia* transcriptomes, but no results of such an analysis have been published yet.

However, it is difficult to move from the results of specific studies to more global conclusions. Comprehensive analysis of data compiled from several experiments is known as meta-analysis. This direction of scientific search may in the future prove to be the most promising in studying the influence of *Wolbachia* on the host transcriptome.

Conclusion

The *Wolbachia*–host system is very stable. *Wolbachia* evolved together with host species, and was also one of the factors directing their evolution. This mutualistic relationship is so deep and ancient that *Wolbachia* is compared to cell organelles located in the cytoplasm, such as mitochondria and chloroplasts. And even though a huge amount of information has been accumulated in this area, much is still unknown concerning the mechanisms maintaining this system.

This area of biology still lacks a systematization of knowledge that would not give rise to contradictions, beginning with the systematics of the genus and ending with the schematization of the molecular mechanisms of its effects. *Wolbachia* has acquired a controversial reputation, acting as a parasitic organism in some cases and as a mutualistic symbiont in others. A hundred years of studying this object does not provide a complete picture.

Since *Wolbachia* has become famous for manipulating the host's reproductive strategy, most studies are devoted to this topic, and not enough attention is paid to another important area – the *Wolbachia* influence on the processes occurring in somatic cells. *Wolbachia* not only affects reproduction but other vital signs in the host as well. It is necessary to continue investigation of less popular and well-studied aspects of the *Wolbachia*–host interactions using new bioinformatics methods and technologies that allow for fundamentally new experiments. The application of these approaches has already contributed to significant progress in the area, but the development of ideas concerning the relationship between insects and the endosymbiotic bacterium *W. pipientis* is not yet complete.

References

- Adonyeva N.V., Efimov V.M., Gruntenko N.E. The effect of genotype combinations of *Wolbachia* and its *Drosophila melanogaster* host on fertility, developmental rate and heat stress resistance of flies. *Insects*. 2023;14(12):928. doi 10.3390/insects14120928
- Augustinos A.A., Santos-Garcia D., Dionyssopoulou E., Moreira M., Papapanagiotou A., Scarvelakis M., Doudoumis V., Ramos S., Aguiar A.F., Borges P.A.V., Khadem M., Latorre A., Tsiamis G., Bourtzis K. Detection and characterization of *Wolbachia* infections in natural populations of Aphids: is the hidden diversity fully unraveled? *PLoS One*. 2011;6(12):e28695. doi 10.1371/journal.pone.0028695
- Backert S., Meyer T.F. Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol*. 2006;9(2):207-217. doi 10.1016/j.mib.2006.02.008
- Baldo L., Bordenstein S.R., Wernegreen J.J., Werren J.H. Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol*. 2006a;23:437-449. doi 10.1093/molbev/msj049
- Baldo L., Hotopp J.C.D., Jolley K.A., Bordenstein S.R., Biber S.A., Choudhury R.R., Hayashi C., Maiden M.C.J., Tettelin H., Werren J.H. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol*. 2006b;72(11):7098-7110. doi 10.1128/AEM.00731-06
- Beckmann J.F., Ronau J.A., Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol*. 2017;2:17007. doi 10.1038/nmicrobiol.2017.7
- Brownlie J.C., Cass B.N., Riegler M., Witsenburg J.J., Iturbe-Ormaetxe I., McGraw E.A., O'Neill S.L. Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. *PLoS Pathog*. 2009;5(4):e1000368. doi 10.1371/journal.ppat.1000368
- Burdina E.V., Gruntenko N.E. Physiological aspects of *Wolbachia pipientis*–*Drosophila melanogaster* relationship. *J Evol Biochem Phys*. 2022;58(2):303-317. doi 10.1134/s0022093022020016
- Burdina E.V., Bykov R.A., Menshanov P.N., Ilinsky Y.Y., Gruntenko N. Unique *Wolbachia* strain wMelPlus increases heat stress resistance in *Drosophila melanogaster*. *Arch Insect Biochem Physiol*. 2021;106(4):e21776. doi 10.1002/arch.21776
- Bykov R.A., Yudina M.A., Gruntenko N.E., Zakharov I.K., Voloshina M.A., Melashchenko E.S., Danilova M.V., Mazunin I.O., Ilinsky Y.Y. Prevalence and genetic diversity of *Wolbachia* endosymbiont and mtDNA in Palearctic populations of *Drosophila melanogaster*. *BMC Evol Biol*. 2019;19(Suppl. 1):48. doi 10.1186/s12862-019-1372-9
- Chen H., Ronau J.A., Beckmann J.F., Hochstrasser M. A *Wolbachia* nuclease and its binding partner provide a distinct mechanism for cytoplasmic incompatibility. *Proc Natl Acad Sci USA*. 2019;116(44):22314-22321. doi 10.1073/pnas.1914571116
- Cho K.O., Kim G.W., Lee O.K. *Wolbachia* bacteria reside in host Golgi-related vesicles whose position is regulated by polarity proteins. *PLoS One*. 2011;6(7):e22703. doi 10.1371/journal.pone.0022703
- Chrostek E., Teixeira L. Mutualism breakdown by amplification of *Wolbachia* genes. *PLoS Biol*. 2015;13(2):e1002065. doi 10.1371/journal.pbio.1002065
- Chrostek E., Teixeira L. Within host selection for faster replicating bacterial symbionts. *PLoS One*. 2018;13(1):e0191530. doi 10.1371/journal.pone.0191530
- Chrostek E., Marialva M.S.P., Esteves S.S., Weinert L.A., Martinez J., Jiggins F.M., Teixeira L. *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: A phenotypic and phylogenomic analysis. *PLoS Genet*. 2013;9(12):e1003896. doi 10.1371/journal.pgen.1003896
- Creasey E.A., Isberg R.R. Maintenance of vacuole integrity by bacterial pathogens. *Curr Opin Microbiol*. 2014;17(1):46-52. doi 10.1016/j.mib.2013.11.005

- Currin-Ross D., Husdell L., Pierens G.K., Mok N.E., O'Neill S.L., Schirra H.J., Brownlie J.C. The metabolic response to infection with *Wolbachia* implicates the insulin/insulin-like-growth factor and hypoxia signaling pathways in *Drosophila melanogaster*. *Front Ecol Evol.* 2021;9:623561. doi 10.3389/fevo.2021.623561
- da Rocha Fernandes M., Martins R., Pessoa Costa E., Pacidônio E.C., Araujo de Abreu L., da Silva Vaz I. Jr., Moreira L.A., da Fonseca R.N., Logullo C. The modulation of the symbiont/host interaction between *Wolbachia pipientis* and *Aedes fluviatilis* embryos by glycogen metabolism. *PLoS One.* 2014;9(6):e98966. doi 10.1371/journal.pone.0098966
- De Felipe K.S., Pampou S., Jovanovic O.S., Pericone C.D., Ye S.F., Kalachikov S., Shuman H.A. Evidence for acquisition of *Legionella* type IV secretion substrates via interdomain horizontal gene transfer. *J Bacteriol.* 2005;187(22):7716-7726. doi 10.1128/JB.187.22.7716-7726.2005
- Dean M.D. A *Wolbachia*-associated fitness benefit depends on genetic background in *Drosophila simulans*. *Proc Biol Sci.* 2006;273(1592):1415-1420. doi 10.1098/rspb.2005.3453
- Detcharoen M., Schilling M.P., Arthofer W., Schlick-Steiner B.C., Steiner F.M. Differential gene expression in *Drosophila melanogaster* and *D. nigrosparsa* infected with the same *Wolbachia* strain. *Sci Rep.* 2021;11(1):11336. doi 10.1038/s41598-021-90857-5
- Duarte E.H., Carvalho A., López-Madrigal S., Costa J., Teixeira L. Forward genetics in *Wolbachia*: regulation of *Wolbachia* proliferation by the amplification and deletion of an additive genomic island. *PLoS Genet.* 2021;17(6):e1009612. doi 10.1371/journal.pgen.1009612
- Ferree P.M., Frydman H.M., Li J.M., Cao J., Wieschaus E., Sullivan W. *Wolbachia* utilizes host microtubules and Dynein for anterior localization in the *Drosophila* oocyte. *PLoS Pathog.* 2005;1(2):0111-0124. doi 10.1371/journal.ppat.0010014
- Frantz S.I., Small C.M., Cresko W.A., Singh N.D. Ovarian transcriptional response to *Wolbachia* infection in *D. melanogaster* in the context of between-genotype variation in gene expression. *G3.* 2023;13(5):jkad047. doi 10.1093/g3journal/jkad047
- Fronzes R., Christie P.J., Waksman G. The structural biology of type IV secretion systems. *Nat Rev Microbiol.* 2009;7(10):703-714. doi 10.1038/nrmicro2218
- Fry A.J., Rand D.M. *Wolbachia* interactions that determine *Drosophila melanogaster* survival. *Evolution.* 2002;56(10):1976-1981. doi 10.1111/j.0014-3820.2002.tb00123.x
- Fry A.J., Palmer M.R., Rand D.M. Variable fitness effects of *Wolbachia* infection in *Drosophila melanogaster*. *Heredity.* 2004;93(4):379-389. doi 10.1038/sj.hdy.6800514
- Gill A.C., Darby A.C., Makepeace B.L. Iron necessity: the secret of *Wolbachia*'s success? *PLoS Negl Trop Dis.* 2014;8(10):e3224. doi 10.1371/journal.pntd.0003224
- Gruntenko N.E., Ilinsky Y.Y., Adonyeva N.V., Burdina E.V., Bykov R.A., Menshanov P.N., Rauschenbach I.Y. Various *Wolbachia* genotypes differently influence host *Drosophila* dopamine metabolism and survival under heat stress conditions. *BMC Evol Biol.* 2017;17:252. doi 10.1186/s12862-017-1104-y
- Gruntenko N.E., Karpova E.K., Adonyeva N.V., Andreenkova O.V., Burdina E.V., Ilinsky Y.Y., Bykov R.A., Menshanov P.N., Rauschenbach I.Y. *Drosophila* female fertility and juvenile hormone metabolism depends on the type of *Wolbachia* infection. *J Exp Biol.* 2019;222(Pt. 4):jeb195347. doi 10.1242/jeb.195347
- Gruntenko N.E., Deryuzhenko M.A., Andreenkova O.V., Shishkina O.D., Bobrovskikh M.A., Shatskaya N.V., Vasiliev G.V. *Drosophila melanogaster* transcriptome response to different *Wolbachia* strains. *Int J Mol Sci.* 2023;24(24):17411. doi 10.3390/ijms242417411
- Gu X., Ross P.A., Rodriguez-Andres J., Robinson K.L., Yang Q., Lau M.J., Hoffmann A.A. A wMel *Wolbachia* variant in *Aedes aegypti* from field-collected *Drosophila melanogaster* with increased phenotypic stability under heat stress. *Environ Microbiol.* 2022;24(4):2119-2135. doi 10.1111/1462-2920.15966
- Hargitai D., Kenéz L., Al-Lami M., Szenczi G., Lőrincz P., Juhász G. Autophagy controls *Wolbachia* infection upon bacterial damage and in aging *Drosophila*. *Front Cell Dev Biol.* 2022;10:976882. doi 10.3389/fcell.2022.976882
- He Z., Zheng Y., Yu W.J., Fang Y., Mao B., Wang Y.F. How do *Wolbachia* modify the *Drosophila* ovary? New evidences support the "titration-restitution" model for the mechanisms of *Wolbachia*-induced CI. *BMC Genomics.* 2019;20(1):608. doi 10.1186/s12864-019-5977-6
- Hedges L.M., Brownlie J.C., O'Neill S.L., Johnson K.N. *Wolbachia* and virus protection in insects. *Science.* 2008;322(5902):702. doi 10.1126/science.1162418
- Hertig M., Wolbach S.B. Studies on rickettsia-like micro-organisms in insects. *J Med Res.* 1924;44(3):329-374.7
- Hilgenboecker K., Hammerstein P., Schlattmann P., Telschow A., Werren J.H. How many species are infected with *Wolbachia*? A statistical analysis of current data. *FEMS Microbiol Lett.* 2008;281(2):215-220. doi 10.1111/j.1574-6968.2008.01110.x
- Hoffmann A.A. Partial cytoplasmic incompatibility between two Australian populations of *Drosophila melanogaster*. *Entomol Exp Appl.* 1988;48(1):61-67. doi 10.1111/j.1570-7458.1988.tb02299.x
- Hoffmann A.A., Montgomery B.L., Popovici J., Iturbe-Ormaetxe I., Johnson P.H., Muzzi F., Greenfield M., Durkan M., Leong Y.S., Dong Y., Cook H., Axford J., Callahan A.G., Kenny N., Omodei C., McGraw E.A., Ryan P.A., Ritchie S.A., Turelli M., O'Neill S.L. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature.* 2011;476(7361):454-459. doi 10.1038/nature10356
- Ikeya T., Broughton S., Alic N., Grandison R., Partridge L. The endosymbiont *Wolbachia* increases insulin/IGF-like signalling in *Drosophila*. *Proc Biol Sci.* 2009;276(1674):3799-3807. doi 10.1098/rspb.2009.0778
- Ilinsky Y. Coevolution of *Drosophila melanogaster* mtDNA and *Wolbachia* genotypes. *PLoS One.* 2013;8(1):e54373. doi 10.1371/journal.pone.0054373
- Ilinsky Y.Y., Zakharov I.K. Cytoplasmic incompatibility in *Drosophila melanogaster* is caused by different *Wolbachia* genotypes. *Russ J Genet Appl Res.* 2011;1(5):458-462. doi 10.1134/S2079059711020031
- Ilinsky Y., Demenkova M., Bykov R., Bugrov A. Narrow genetic diversity of *Wolbachia* symbionts in acrididae grasshopper hosts (Insecta, Orthoptera). *Int J Mol Sci.* 2022;23(2):853. doi 10.3390/ijms23020853
- Jiménez N.E., Gerdtsen Z.P., Olivera-Nappa Á., Salgado J.C., Conca C. A systems biology approach for studying *Wolbachia* metabolism reveals points of interaction with its host in the context of arboviral infection. *PLoS Negl Trop Dis.* 2019;13(8):e0007678. doi 10.1371/journal.pntd.0007678
- Kamtchum-Tatuene J., Makepeace B.L., Benjamin L., Baylis M., Solomon T. The potential role of *Wolbachia* in controlling the transmission of emerging human arboviral infections. *Curr Opin Infect Dis.* 2017;30(1):108-116. doi 10.1097/QCO.0000000000000342
- Karpova E.K., Bobrovskikh M.A., Deryuzhenko M.A., Shishkina O.D., Gruntenko N.E. *Wolbachia* effect on *Drosophila melanogaster* lipid and carbohydrate metabolism. *Insects.* 2023;14(4):357. doi 10.3390/insects14040357
- Kaur R., Shropshire J.D., Cross K.L., Leigh B., Mansueto A.J., Stewart V., Bordenstein S.R., Bordenstein S.R. Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host Microbe.* 2021;29(6):879-893. doi 10.1016/j.chom.2021.03.006
- Korenskaia A.E., Shishkina O.D., Klimenko A.I., Andreenkova O.V., Bobrovskikh M.A., Shatskaya N.V., Vasiliev G.V., Gruntenko N.E.

- New *Wolbachia pipientis* genotype increasing heat stress resistance of *Drosophila melanogaster* host is characterized by a large chromosomal inversion. *Int J Mol Sci.* 2022;23(24):16212. doi 10.3390/ijms232416212
- Kremer N., Voronin D., Charif D., Mavingui P., Mollereau B., Vavre F. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. *PLoS Pathog.* 2009;5(10):e1000630. doi 10.1371/journal.ppat.1000630
- Kumar Y., Valdivia R.H. Leading a sheltered life: intracellular pathogens and maintenance of vacuolar compartments. *Cell Host Microbe.* 2009;5(6):593-601. doi 10.1016/j.chom.2009.05.014
- Lassy C.W., Karr T.L. Cytological analysis of fertilization and early embryonic development in incompatible crosses of *Drosophila simulans*. *Mech Dev.* 1996;57(1):47-58. doi 10.1016/0925-4773(96)00527-8
- Laven H. Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility. *Nature.* 1967;216:383-384. doi 10.1038/216383a0
- LePage D., Bordenstein S.R. *Wolbachia*: can we save lives with a great pandemic? *Trends Parasitol.* 2013;29(8):385-393. doi 10.1016/j.pt.2013.06.003
- LePage D.P., Metcalf J.A., Bordenstein S.R., On J., Perlmuter J.I., Shropshire J.D., Layton E.M., Funkhouser-Jones L.J., Beckmann J.F., Bordenstein S.R. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature.* 2017;543(7644):243-247. doi 10.1038/nature21391
- Lindsey A.R.I. Sensing, signaling, and secretion: a review and analysis of systems for regulating host interaction in *Wolbachia*. *Genes (Basel).* 2020;11(7):813. doi 10.3390/genes11070813
- Lindsey A.R.I., Bhattacharya T., Hardy R.W., Newton I.L.G. *Wolbachia* and virus alter the host transcriptome at the interface of nucleotide metabolism pathways. *mBio.* 2021;12(1):e03472-20. doi 10.1128/mBio.03472-20
- Lo N., Evans T.A. Phylogenetic diversity of the intracellular symbiont *Wolbachia* in termites. *Mol Phylogenet Evol.* 2007;44(1):461-466. doi 10.1016/j.ympev.2006.10.028
- Lo N., Paraskevopoulos C., Bourtzis K., O'Neill S.L., Werren J.H., Bordenstein S.R., Bandi C. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. *Int J Syst Evol Microbiol.* 2007;57(3):654-657. doi 10.1099/ijs.0.64515-0
- Luck A.N., Slatko B.E., Foster J.M. Removing the needle from the haystack: enrichment of *Wolbachia* endosymbiont transcripts from host nematode RNA by Cappable-seq™. *PLoS One.* 2017;12(3):e0173186. doi 10.1371/journal.pone.0173186
- Maistrenko O.M., Serga S.V., Vaiserman A.M., Kozeretska I.A. Longevity-modulating effects of symbiosis: insights from *Drosophila-Wolbachia* interaction. *Biogerontology.* 2016;17(5-6):785-803. doi 10.1007/s10522-016-9653-9
- Mateos M., Silva N.O., Ramirez P., Higareda-Alvear V.M., Aramayo R., Erickson J.W. Effect of heritable symbionts on maternally-derived embryo transcripts. *Sci Rep.* 2019;9(1):8847. doi 10.1038/s41598-019-45371-0
- Min K.-T., Benzer S. *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proc Natl Acad Sci USA.* 1997;94(20):10792-10796. doi 10.1073/pnas.94.20.10792
- Moreira L.A., Iturbe-Ormaetxe I., Jeffery J.A., Lu G., Pyke A.T., Hedges L.M., Rocha B.C., Hall-Mendelin S., Day A., Riegler M., Hugo L.E., Johnson K.N., Kay B.H., McGraw E.A., van den Hurk A.F., Ryan P.A., O'Neill S.L. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and *Plasmodium*. *Cell.* 2009;139(7):1268-1278. doi 10.1016/j.cell.2009.11.042
- Moriyama M., Nikoh N., Hosokawa T., Fukatsu T. Riboflavin provisioning underlies *Wolbachia*'s fitness contribution to its insect host. *mBio.* 2015;6(6):e01732-15. doi 10.1128/mBio.01732-15
- Newton I.L.G., Rice D.W. The Jekyll and Hyde symbiont: could *Wolbachia* be a nutritional mutualist? *J Bacteriol.* 2020;202(4):e00589-19. doi 10.1128/JB.00589-19
- Nikoh N., Hosokawa T., Moriyama M., Oshima K., Hattori M., Fukatsu T. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci USA.* 2014;111(28):10257-10262. doi 10.1073/pnas.1409284111
- Nunes M.D.S., Nolte V., Schlötterer C. Nonrandom *Wolbachia* infection status of *Drosophila melanogaster* strains with different mtDNA haplotypes. *Mol Biol Evol.* 2008;25(11):2493-2498. doi 10.1093/molbev/msn199
- O'Neill S.L., Pettigrew M.M., Sinkins S.P., Braig H.R., Andreadis T.G., Tesh R.B. *In vitro* cultivation of *Wolbachia pipientis* in an *Aedes albopictus* cell line. *Insect Mol Biol.* 1997;6(1):33-39. doi 10.1046/j.1365-2583.1997.00157.x
- Ote M., Ueyama M., Yamamoto D. *Wolbachia* protein TomO targets *nanos* mRNA and restores germ stem cells in *Drosophila sex-lethal* mutants. *Curr Biol.* 2016;26(17):2223-2232. doi 10.1016/j.cub.2016.06.054
- Pichon S., Bouchon D., Cordaux R., Chen L., Garrett R.A., Grève P. Conservation of the Type IV secretion system throughout *Wolbachia* evolution. *Biochem Biophys Res Commun.* 2009;385(4):557-562. doi 10.1016/j.bbrc.2009.05.118
- Pietri J.E., DeBruhl H., Sullivan W. The rich somatic life of *Wolbachia*. *Microbiologyopen.* 2016;5(6):923-936. doi 10.1002/mbo.3.390
- Poinsot D., Charlat S., Merçot H. On the mechanism of *Wolbachia*-induced cytoplasmic incompatibility: confronting the models with the facts. *BioEssays.* 2003;25(3):259-265. doi 10.1002/bies.10234
- Porter J., Sullivan W. The cellular lives of *Wolbachia*. *Nat Rev Microbiol.* 2023;21(11):750-766. doi 10.1038/s41579-023-00918-x
- Rice D.W., Sheehan K.B., Newton I.L.G. Large-scale identification of *Wolbachia pipientis* effectors. *Genome Biol Evol.* 2017;9(7):1925-1937. doi 10.1093/gbe/evx139
- Riegler M., Sidhu M., Miller W.J., O'Neill S.L. Evidence for a global *Wolbachia* replacement in *Drosophila melanogaster*. *Curr Biol.* 2005;15(15):1428-1433. doi 10.1016/j.cub.2005.06.069
- Rohrscheib C.E., Bondy E., Josh P., Riegler M., Eyles D., van Swinderen B., Weible M.W., Brownlie J.C. *Wolbachia* influences the production of octopamine and affects *Drosophila* male aggression. *Appl Environ Microbiol.* 2015;81(14):4573-4580. doi 10.1128/AEM.00573-15
- Ros V.I.D., Fleming V.M., Feil E.J., Breeuwer J.A.J. How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: Tetranychidae). *Appl Environ Microbiol.* 2009;75(4):1036-1043. doi 10.1128/AEM.01109-08
- Ryan S.L., Saul G.B. 2nd. Post-fertilization effect of incompatibility factors in *Mormoniella*. *Mol Gen Genet.* 1968;103(1):29-36. doi 10.1007/BF00271154
- Sheehan K.B., Martin M., Lesser C.F., Isberg R.R., Newton I.L.G. Identification and characterization of a candidate *Wolbachia pipientis* type IV effector that interacts with the actin cytoskeleton. *mBio.* 2016;7(4):e00622-16. doi 10.1128/mBio.00622-16
- Teixeira L., Ferreira Á., Ashburner M. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* 2008;6(12):2753-2763. doi 10.1371/journal.pbio.1000002
- Truitt A.M., Kapun M., Kaur R., Miller W.J. *Wolbachia* modifies thermal preference in *Drosophila melanogaster*. *Environ Microbiol.* 2019;21(9):3259-3268. doi 10.1111/1462-2920.14347
- Werren J.H. Biology of *Wolbachia*. *Annu Rev Entomol.* 1997;42:587-609. doi 10.1146/annurev.ento.42.1.587

- White P.M., Pietri J.E., Debec A., Russell S., Patel B., Sullivan W. Mechanisms of horizontal cell-to-cell transfer of *Wolbachia* spp. in *Drosophila melanogaster*. *Appl Environ Microbiol.* 2017;83(7): e03425-16. doi 10.1128/AEM.03425-16
- Wu M., Sun L.V., Vamathevan J., Riegler M., Deboy R., Brownlie J.C., McGraw E.A., Martin W., Esser C., Ahmadinejad N., Wiegand C., Madupu R., Beanan M.J., Brinkac L.M., Daugherty S.C., Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P., Berry K., Young M.B., Utterback T., Weidman J., Niernan W.C., Paulsen I.T., Nelson K.E., Tettelin H., O'Neill S.L., Eisen J.A. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* 2004;2(3):E69. doi 10.1371/journal.pbio.0020069
- Yu X.-J., Walker D.H. The Order Rickettsiales. In: The Prokaryotes. Springer, 2006;493-528. doi 10.1007/0-387-30745-1_20
- Zhang H., Id Z.C., Qiao J., Zhong Z., Pan C. Metabolomics provide new insights into mechanisms of *Wolbachia*-induced paternal defects in *Drosophila melanogaster*. *PLoS Pathog.* 2021;17(8):e1009859. doi 10.1371/journal.ppat.1009859
- Zhou W., Rousset F., O'Neil S. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proc Biol Sci.* 1998;265(1395):509-515. doi 10.1098/rspb.1998.0324
- Zug R., Hammerstein P. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40 % of terrestrial arthropod species are infected. *PLoS One.* 2012;7(6):e38544. doi 10.1371/journal.pone.0038544
- Zug R., Hammerstein P. *Wolbachia* and the insect immune system: what reactive oxygen species can tell us about the mechanisms of *Wolbachia*-host interactions. *Front Microbiol.* 2015;6:1201. doi 10.3389/fmicb.2015.01201

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

Received November 20, 2024. Revised December 25, 2024. Accepted December 25, 2024.

doi 10.18699/vjgb-25-11

The liver-brain axis under the influence of chronic *Opisthorchis felinus* infection combined with prolonged alcoholization in mice

D.F. Avgustinovich ^{1,2}✉, I.V. Chadaeva ¹, A.V. Kizimenko ¹, A.V. Kovner ¹, D.V. Bazovkina ¹,
D.V. Ponomarev ¹, V.I. Evseenko ², V.A. Naprimerov^{1,3}, M.N. Lvova ¹

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

² Institute of Solid State Chemistry and Mechanochemistry of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

³ Novosibirsk State Agrarian University, Novosibirsk, Russia

✉ avgust@bionet.nsc.ru

Abstract. Our purpose was to model a combination of a prolonged consumption of ethanol with *Opisthorchis felinus* infection in mice. Four groups of C57BL/6 mice were compiled: OF, mice infected with *O. felinus* for 6 months; Eth, mice consuming 20 % ethanol; Eth+OF, mice subjected to both adverse factors; and CON, control mice not exposed to these factors. In the experimental mice, especially in Eth+OF, each treatment caused well-pronounced periductal and cholangiofibrosis, proliferation of bile ducts, and enlargement of areas of inflammatory infiltration in the liver parenchyma. Simultaneously with liver disintegration, the infectious factor caused – in the frontal cerebral cortex – the growth of pericellular edema (OF mice), which was attenuated by the administration of ethanol (Eth+OF mice). Changes in the levels of some proteins (Iba1, IL-1 β , IL-6, and TNF) and in mRNA expression of genes *Aif1*, *Il1b*, *Il6*, and *Tnf* were found in the hippocampus and especially in the frontal cortex, implying region-specific neuroinflammation. Behavioral testing of mice showed that ethanol consumption influenced the behavior of Eth and Eth+OF mice in the forced swimming test and their startle reflex. In the open field test, more pronounced changes were observed in OF mice. In mice of all three experimental groups, especially in OF mice, a disturbance in the sense of smell was detected (fresh peppermint leaves). The results may reflect an abnormality of regulatory mechanisms of the central nervous system as a consequence of systemic inflammation under the combined action of prolonged alcohol consumption and helminth infection.








Key words: mice; *Opisthorchis felinus* infection; chronic ethanol consumption; liver; brain; microglia; proinflammatory cytokine; behavior.

For citation: Avgustinovich D.F., Chadaeva I.V., Kizimenko A.V., Kovner A.V., Bazovkina D.V., Ponomarev D.V., Evseenko V.I., Naprimerov V.A., Lvova M.N. The liver-brain axis under the influence of chronic *Opisthorchis felinus* infection combined with prolonged alcoholization in mice. *Vavilovskii Zhurnal Genetiki i Selektzii* = *Vavilov J Genet Breed.* 2025;29(1):92-107. doi 10.18699/vjgb-25-11

Funding. The work was supported by the Russian Foundation for Basic Research (grant No. 20-04-00139); by a government-funded project of the ICG SB RAS (grant No. FWN-2022-0021); and in part by a state assignment for the Institute of Solid State Chemistry and Mechanochemistry SB RAS (project No. 121032500061-7).

Acknowledgements. The authors are grateful to the Multi-Access Center for Microscopic Analysis of Biological Objects at the ICG SB RAS (<http://www.bionet.nsc.ru/microscopy/>) for the equipment as well as to the Center for Genetic Resources of Laboratory Animals (ICG SB RAS), which is supported by the Russian Ministry of Science and Higher Education (unique identifier of the project: RFMEFI62119X0023). The English language was corrected and certified by shevchuk-editing.com (language certificate of July 3, 2024).

Ось «печень-мозг» при хронической инфекции *Opisthorchis felinus* в сочетании с длительной алкоголизацией мышей

Д.Ф. Августинович ^{1,2}✉, И.В. Чадаева ¹, А.В. Кизименко ¹, А.В. Ковнер ¹, Д.В. Базовкина ¹,
Д.В. Пономарёв ¹, В.И. Евсеенко ², В.А. Напримеров^{1,3}, М.Н. Львова ¹

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

² Институт химии твёрдого тела и механохимии Сибирского отделения Российской академии наук, Новосибирск, Россия

³ Новосибирский государственный аграрный университет, Новосибирск, Россия

✉ avgust@bionet.nsc.ru

Аннотация. Описторхоз – паразитарное заболевание, вызываемое трематодами семейства Opisthorchiidae, к которому относится вид *Opisthorchis felinus*, максимально представленный у жителей Обь-Иртышского речного бассейна Западной Сибири. Хроническое употребление алкоголя влечет за собой неизлечимые заболевания пе-

чени, которые могут быть усугублены привнесением гельминтной инфекции. Хроническое воспаление на фоне двух воздействий может вызывать нарушения в других органах и системах, в том числе в центральной нервной системе. Сочетание длительного потребления 20 % этанола и инфекции *O. felineus* моделировали у мышей C57BL/6 с целью исследования изменений в мозге таких животных. Были сформированы четыре группы мышей: OF – мыши, инфицированные *O. felineus* в течение 6 мес; Eth – мыши, потребляющие 20 % этанол; Eth+OF – мыши с сочетанным действием двух факторов; CON – контрольные мыши, не испытывающие эти воздействия. Оба фактора вызывали у мышей, особенно в группе Eth+OF, выраженные перидуктальный фиброз и холангиофиброз, пролиферацию желчных протоков и увеличение участков воспалительной инфильтрации в паренхиме печени. Одновременно с нарушениями в печени, во фронтальной коре мозга инфекционный фактор обуславливал усиление перичеллюлярного отека (мыши OF), сдерживаемого введением этанола (мыши Eth+OF). Также обнаружены колебания уровней белка (Iba1, IL-1 β , IL-6, TNF) и экспрессии генов *Aif1*, *Il1b*, *Il6* и *Tnf* в гиппокампе и особенно во фронтальной коре, предполагающие регион-специфическое нейровоспаление. Поведенческое тестирование мышей показало, что потребление этанола влияло на поведение мышей Eth and Eth+OF в тесте принудительного плавания и на Startle-рефлекс. В тесте открытого поля более выраженные изменения происходили у мышей OF. У животных всех трех экспериментальных групп, но особенно у OF, выявлена аномалия обоняния (на свежие листья мяты перечной). Полученные данные могут отражать нарушение регуляторных механизмов в головном мозге вследствие системного воспаления при сочетанном действии длительного употребления алкоголя и гельминтной инфекции.

Ключевые слова: мыши; инфекция *Opisthorchis felineus*; хроническое потребление этанола; печень; мозг; микроглия; провоспалительные цитокины; поведение.

Introduction

The family Opisthorchiidae includes three species of trematodes: *Clonorchis sinensis*, most common in China, Korea, and the Far East of the Russian Federation (35 million people); *Opisthorchis viverrini*, widespread in Thailand and Laos (10 million people); and *Opisthorchis felineus*, manifesting the highest prevalence of infection among fish in water bodies of the Ob-Irtysh basin in Russia (1.2 million people) (Mordvinov, Furman, 2010; Petney et al., 2013; Saijuntha et al., 2021). The last two species cause opisthorchiasis when fish of the family Cyprinidae are eaten raw or undercooked. During prolonged helminth infection, mature worms parasitize not only bile ducts of the liver and the gall bladder of piscivorous mammals but also pancreatic ducts, as shown for *O. felineus* and *C. sinensis* (Bernstein et al., 1994; Mordvinov, Furman, 2010; Lvova et al., 2023), thereby causing complications such as various forms of pancreatitis (Gundamaraju, Vemuri, 2014). Given that opisthorchiasis upregulates proinflammatory cytokines and leukocytes in the blood (Sripa et al., 2012; Avgustinovich et al., 2022a), this disease is considered a systemic illness that provokes pathologies in other organs and systems of the body (Boonpucknavig et al., 1992; Akhmedov, Kritevich, 2009; Bychkov et al., 2011), including the central nervous system (CNS) (Lvova et al., 2020; Avgustinovich et al., 2016, 2022b). One can expect the development of neuroinflammation during opisthorchiasis, judging by tenets of the “liver–brain axis” paradigm (D’Mello, Swain, 2011). According to those authors, three proinflammatory cytokines – interleukin 1 beta (IL-1 β), interleukin 6 (IL-6), and tumor necrosis factor (TNF) – in the blood are key promoters of central neural changes in chronic liver diseases. In addition, in the context of liver inflammation, activation of microglia takes place with subsequent recruitment of blood monocytes into the brain. All this strongly drives hepatic inflammation-associated sickness behavior.

The daily alcohol (ethanol) abuse/misuse is a major cause of inevitable damage to the liver (Collins, Neafsey, 2012). The progression of alcoholic liver disease induces cirrhosis, liver cancer, and acute and chronic liver failure and can be fatal (Stickel et al., 2017). As determined by the World Health Or-

ganization (WHO. Alcohol. <https://www.who.int/news-room/fact-sheets/detail/alcohol>), worldwide, 3 million deaths every year result from overconsumption of alcohol, and this figure represents 5.3 % of all deaths.

As with opisthorchiasis, excessive alcohol consumption entails pathological changes in other organs and systems of organs, for example, in the gut microbiota (Saltykova et al., 2016; Bishehsari et al., 2017; Bajaj, 2019; Ketpueak et al., 2020; Pakharukova et al., 2023; Yao et al., 2023). Aside from the direct negative effect on the intestines, alcohol disrupts bile acid synthesis in the liver during inflammation and impairs bile acid entry into the gallbladder for subsequent secretion into the small intestine. In this context, the reabsorption of bile by the liver is impeded, which under normal conditions is 95 %. Accordingly, in such patients, in addition to alcoholic liver diseases, microbial composition and functions of the intestine change, leading to functional alterations in the “gut–liver–brain axis” (Bishehsari et al., 2017). As a consequence, symptoms of hepatic encephalopathy are aggravated, which are associated with microglial activation and subsequent cognitive deficits (Bajaj, 2019).

On the other hand, alcohol can also have a direct impact on the CNS by inducing cerebral cortical edema and electrolyte (Na⁺ or K⁺) accumulation (Collins et al., 1998), and during chronic consumption, alcohol can cause neuronal loss in some brain structures (the cerebral cortex, hypothalamus, hippocampus, and cerebellum) (Harper, 1998; de la Monte, Kril, 2014; Fowler et al., 2014). These effects are attributable to activation of resident microglia and peripheral-macrophage infiltration of the CNS, particularly in the hippocampus, and the two processes together contribute to overexpression of proinflammatory markers in various regions of the brain, including the cortex and hippocampus (Yang et al., 2014; Henriques et al., 2018; Lowe et al., 2020). These phenomena negatively affect cognitive abilities, learning, and memory (Geil et al., 2014).

In humans, these two adverse factors (alcohol consumption and *O. felineus* infection) can often occur simultaneously. Unfortunately, a combination of ethanol consumption with chronic *O. felineus* infection can have irreversible consequences for humans, as previously shown in a model of such

a combination (Avgustinovich et al., 2022a). In that work, a more pronounced liver pathology was accompanied by splenomegaly due to structural changes in the spleen as well as elevated levels of IL-6 and higher leukocyte and monocyte counts in the blood. Taken together, they meant substantial whole-body inflammation. Considering these data as well as a known statement about possible brain disturbances during severe hepatic inflammation (D’Mello, Swain, 2011), the purpose of our current study was to investigate – by histological, immunohistochemical, and molecular methods – changes in inflammatory markers in the cerebral cortex and hippocampus, as assessed by determination of microglial activation and of expression of proinflammatory cytokines IL-1 β , IL-6, and TNF. Because any disturbances in the brain manifest themselves in behavior in mammals, behavioral patterns of mice were assessed here by the open field test (Ramos, Mormède, 1998) and forced swimming test (Porsolt et al., 1977), along with estimation of the startle reflex in response to acoustic signals (Paylor, Crawley, 1997).

Materials and methods

Animals. Male mice of inbred strain C57BL/6 were obtained from the multi-access center Vivarium of Conventional Animals of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (ICG SB RAS). All animals were maintained in standard cages (36×22×12 cm) at 2–5 individuals per cage, on a 12/12 h light/dark cycle, at 24±1 °C with free access to pelleted feed and liquid. The study was conducted in accordance with Directives of the Council of the European Union (86/609/EEC) of November 24, 1986 and according to a decision of the Bioethics Commission of the ICG SB RAS (protocol No. 113 of December 9, 2021).

Obtaining of *O. felineus* metacercariae. Metacercariae of *O. felineus* were isolated from naturally infected ides caught

in the Ob River (Novosibirsk Oblast) by a method described previously (Avgustinovich et al., 2016, 2021). Metacercariae of *O. felineus* were administered to mice intragastrically (100 larvae per mouse) using specialized probes (Braintree Scientific, Inc.).

The design of the experiment. As a result, four groups of mice were set up: CON ($n = 15$), not subjected to any pathogenic procedures; OF ($n = 15$), infected with *O. felineus* (duration of infection: 6 months); Eth ($n = 15$), consuming 20 % ethanol for 6 months; and Eth+OF ($n = 13$), subjected to both procedures (Fig. 1). Mice were trained to consume ethanol according to a protocol described before (Avgustinovich et al., 2022a). Five months after the infection initiation, the behavior of all mice was recorded in the open field test. By the end of 6 months, the startle response of mice to an acoustic signal (startle reflex) was evaluated. The animals were then housed individually in 28×14×10 cm cages to evaluate behavior in the forced swimming test. At the end of the experiment, the animals were killed by decapitation, and brain samples were collected for subsequent analyses. The hippocampus and frontal cortex were isolated on ice, placed in liquid nitrogen and then in a freezer at –70 °C until the expression of genes of interest was assayed by real-time polymerase chain reaction (qPCR). Brains from five animals in each group were put in 10 % formalin for subsequent histological and immunohistochemical examination. The liver of these animals was also placed in formalin for subsequent determination of pathomorphological structural alterations in this organ under the influence of the two tested adverse factors.

The open field test. This is one of the most popular tests in behavioral studies, which assesses effects of external factors on rodents (Ramos, Mormède, 1998). When animals are first exposed to an unfamiliar open space, behavioral responses are believed to be mediated by anxiety, whereas repeated

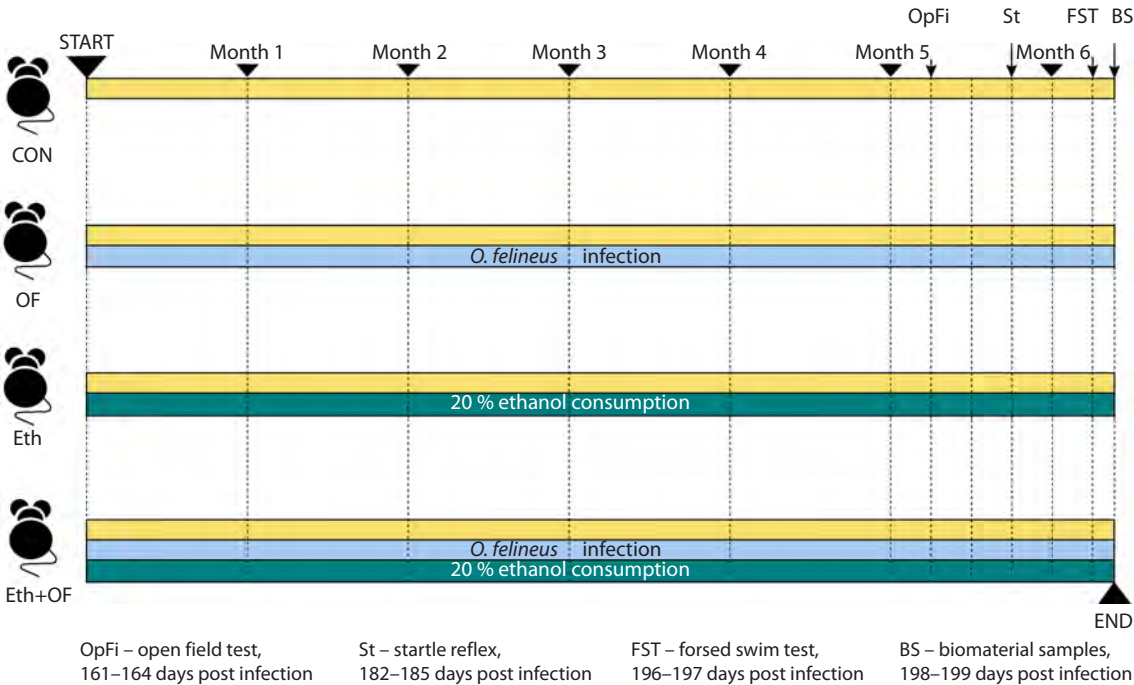


Fig. 1. Design of the experiment.

exposure can lead to exploration of new objects (or odors) by the animals in a nonthreatening familiar environment (Belzung, 1992; Choleris et al., 2001). For testing, a 40 × 40 cm field made of white plastic material was utilized, which was visibly subdivided into squares (4 × 4) with a side of 10 cm. The illumination at the field level was 320 lx. In the middle of one side of the field, an inverted office metallic tumbler for pencils (hereafter: tumbler) was placed. Five minutes before the testing, the animals were brought into the test room for activation, after which they were always placed in the middle of the wall opposite the wall with the tumbler. Each animal's behavior was recorded twice (for 3 min each time) using a video camera in the session with an empty tumbler and next with a tumbler containing fresh peppermint leaves. Further analysis of animal behavior was carried out in the Behavioral Observation Research Interactive Software (BORIS) (Friard, Gamba, 2016), in which we determined the number, time (duration), and latency of approaches and/or turns to the tumbler, the number of square crossings, and the number, time (duration), and latency of rearings near a wall.

Startle reflex testing. The startle reflex of humans and animals toward a sharp sound signal is an innate reflex and characterizes the CNS's ability to filter sensory information (Paylor, Crawley, 1997). The behavioral response of mice to an acoustic stimulus was measured using an SR-Pilot device (San-Diego Instruments, USA). Background white noise was set to 65 dB. Each animal was placed in the chamber, and after 3-min adaptation, 10 impulses (P) of 115 dB intensity and 40 ms duration were presented, alternating with 10 impulses preceded by weak (85 dB, 40 ms) prepulses (PP), which were presented 100 ms before the main impulse. The interval between stand-alone P impulses and the PP combination was 15 s. The size of the response of a mouse to the stimuli (amplitude) was displayed by the device's screen in relative units. Prepulse inhibition (PPI, %) was calculated (Paylor, Crawley, 1997) by means of the formula: $100 - [(PP/P) \times 100]$.

The forced swim test. This test is the most popular in research on depressive-like behavior in rodents because it is sensitive to the action of antidepressants (Porsolt et al., 1977). During the testing, transparent plastic cylinders 30 cm high and 9.5 cm in diameter were used, which were filled with water ($t = 25 \pm 1$ °C) to a level of 19 cm. Five minutes before the test, the mice were sequentially brought into the test room and then placed for 5 min into the cylinder filled with water; their behavior was recorded on a video camera. Further analysis was performed in the BORIS software (Friard, Gamba, 2016). The duration of active and passive swimming was assessed; the latter included drift (when the mouse made weak movements with one or two hind legs to keep the head above water) and a state of complete immobility.

RNA isolation for qPCR. To obtain RNA from mouse brain structures, the TRI reagent (Sigma-Aldrich, USA) was employed according to the protocol for samples with a high fat content, then the samples were purified on magnetic particles (Agencourt RNAClean XP Kit, Beckman). The concentration of the resulting RNA was determined using a Qubit™ 2.0 fluorimeter (Invitrogen/Life Technologies) with a kit (RNA High Sensitivity, Invitrogen). Next, the isolated RNA was treated with DNase using the DNase I, RNase-free kit (Thermo Fisher Scientific, USA). Complementary DNA (cDNA) was

synthesized by means of a reverse-transcription kit (#OT-1, Syntol, Russia). All stages of RNA isolation and analysis and those of cDNA preparation were carried out according to the protocols of the manufacturers of the corresponding kits.

Using the PrimerBLAST web service, we designed oligonucleotide primers for qPCR. qPCR was conducted with the EVA Green I kit (#R-441, Syntol, Russia) according to the manufacturer's instructions, and amplification efficiency for each primer pair was 90–110 %. qPCR was carried out in three technical replicates on a CFX-96 thermal cycler (Bio-Rad, USA). qPCR efficiency was determined by means of serial dilutions of cDNA (standards); after completion of qPCR, melting curves of the products were constructed to monitor the specificity of the reaction. Expression levels of each analyzed gene were normalized to two reference genes, the expression stability of which was checked in both brain regions under study and in each of the four groups of mice. Based on literature data (Stephens et al., 2011), three reference genes were chosen: *Actb* (beta actin: a highly conserved protein that participates in cell motility, structure, and integrity), *B2m* (beta-2-microglobulin: a light-chain component of major histocompatibility complex class I), and *Hprt1* (hypoxanthine phosphoribosyltransferase 1: a eukaryotic enzyme of purine metabolism).

The genes of interest were selected taking into account their functional characteristics listed in the Human Protein Atlas (<https://www.proteinatlas.org/>). The following genes were investigated in the experiment: *Aif1* (a marker of microglial activity) and *Il1b*, *Il6*, and *Tnf* (markers of inflammation).

Sequences of the chosen primers were as follows:

- 1 *Actb* (F) 5'-TATTGGCAACGAGCGGTTCC
(R) 5'-TGGCATAGAGGTCTTTACGG
- 2 *Aif1* (F) 5'-GGATTTCAGGGAGGAAAA
(R) 5'-TGGGATCATCGAGGAATTG
- 3 *B2m* (F) 5'-CTGCTACGTAACACAGTTCACCC
(R) 5'-CATGATGCTTGATCACATGTCTCG
- 4 *Hprt1* (F) 5'-GAGGAGTCCTGTTGATGTTGCCAG
(R) 5'-GGCTGGCCTATAGGCTCATAGTGC
- 5 *Il1b* (F) 5'-ACACTCCTTAGTCCCTCGGCCA
(R) 5'-CCATCAGAGGCAAGGAGGAA
- 6 *Il6* (F) 5'-ACAAAGCCAGAGTCCTTCAGAG
(R) 5'-ACGCACTAGGTTTGCCGAG
- 7 *Tnf* (F) 5'-AGCCGATGGGTTGTACCTTG
(R) 5'-GGTTGACTTCTCCTGGTATGAGA

Histological examination of sections of the cerebral cortex and liver. After 10 days of fixation in a 10 % formaldehyde solution, the brain and liver of mice were sectioned for subsequent processing in an STP 120 carousel-type apparatus for automatic incubation in a graded series of ethanol and xylene (Thermo Fisher Scientific, USA). The frontal section separated the middle part of the brain, containing the hippocampus, at level 60–64 according to the Allen Mouse Brain Atlas (<http://mouse.brain-map.org/static/atlas>). Samples containing bile ducts and parenchyma and measuring 100 × 150 × 70 mm were dissected from the large lobe of the liver. After dehydration, the tissue samples were embedded in the HISTOMIX synthetic paraffin medium (Russia) by means of an EC-350 embedding

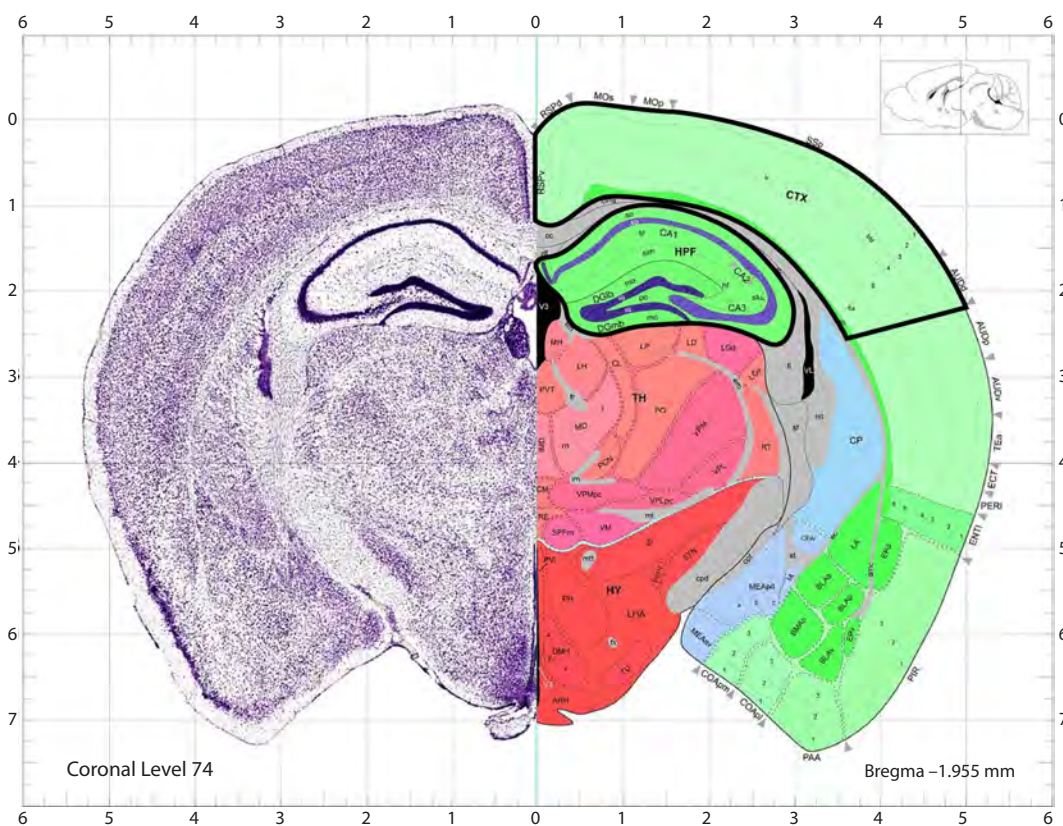


Fig. 2. Areas of the cerebral cortex and hippocampal formation (highlighted with a black line on the right) from the brain of C57BL/6 mice, after immunohistochemical analysis.

The image of the frontal brain section at levels 72–74 was borrowed from the Allen Mouse Brain Atlas (Allen Institute Publications for Brain Science; <http://mouse.brain-map.org/static/atlas>).

station (Thermo Scientific, USA). Sections with a thickness of 3.5–4.0 μm were prepared on a Microm HM 355S rotary microtome (Thermo Fisher Scientific, USA). To study the cortex and hippocampus, brain samples were cut to levels 72–74 according to the brain atlas (Fig. 2). The obtained samples of the cortex and liver were subsequently stained by standard techniques: basic survey staining with hematoxylin and eosin and the Masson method (staining of connective tissue). After that, sections mounted in BioMount (BIO-OPTICA, Italy) were visualized under an Axioskop 2 plus microscope equipped with an AxioCam MRc camera (Carl Zeiss, Germany) and AxioVision software (release 4.12). Morphometry of structural changes in the cerebral cortex was performed using a closed test system targeting 100 points with an area of $3.64 \times 10^5 \mu\text{m}^2$. Meanwhile, numerical density of perivascular and pericellular edema and the total number of blood vessels were assessed. The scoring method for these measurement data had been described in detail earlier (Pakharukova et al., 2019). In the liver, the absence/presence of periductal fibrosis, cholangiofibrosis, and inflammatory infiltration and proliferation of bile ducts was recorded.

Immunohistochemical analysis of sections of the hippocampus and cerebral cortex. The hippocampus and cortex at levels 72–74 of frontal brain sections were examined on glass slides with an adhesive poly-L-lysine coating (BVS, Russia). To count activated brain microglia cells and cells synthesizing proinflammatory cytokines, we utilized an indirect biotin-

free peroxidase immunohistochemical technique for staining paraffin sections using a kit (SpringBioScience Kit HRP-125, Pleasanton, CA, USA) and primary antibodies specific to ionized calcium-binding adapter protein 1 (Iba1) (Abcam, cat. # ab5076, dilution 1:300), IL-6 (Abcam, ab6672, 1:100), TNF (Abcam, ab6671, 1:100), and IL-1 β (Abcam, ab9722, 1:100). According to this technique, after standard deparaffinization and dehydration, antigens were retrieved on the sections in citrate buffer (pH 6.0) in a microwave oven at 700 W for 5 min. After washing four times in phosphate-buffered saline (PBS, pH 7.6), endogenous peroxidase was blocked for 30 min with fresh 3 % H_2O_2 . Next, blocking with horse serum was carried out for 1.0–1.5 h with pre-washing in PBS. The duration and conditions of probing with primary antibodies were chosen according to the manufacturer's instructions (sections with antibodies were incubated overnight at 4 °C). Next, after washing four times (PBS, pH 7.6), the tissue sections were incubated with secondary antibodies for 45 min, and then, after washing (PBS, pH 7.6), the tissue sections were incubated with diaminobenzidine as a substrate until a brown color appeared upon visual inspection. After that, the sections were counterstained with Mayer's hematoxylin for 1 min, placed in tap water for 5 min, passed through a graded series of ethanol and xylenes, covered with the synthetic BioMount medium, and placed under a cover glass. Cells positive for staining with the above antibodies were counted in all subfields of view in the cortical and hippocampal areas highlighted with

a black line in Figure 2, by a previously described approach (Pakharukova et al., 2019).

Statistics. To compare groups, two-way ANOVA and three-way ANOVA were performed in the STATISTICA 6.0 software, followed by a *post hoc* comparison of groups by the least significant difference (LSD) test. The following factors were analyzed during the intergroup comparison: “infection” (*O. felineus* or no *O. felineus*) and “ethanol” (ethanol or no ethanol). In the open field test, an additional within-group factor was “peppermint”: the first or second 3 min of observation, i. e., in the absence or presence of peppermint in the tumbler, respectively. The difference in the startle amplitude between the first and 10th stimulus in each group of mice was assessed by the Wilcoxon matched-pairs test. In all analyses, data with a p -value ≤ 0.05 were considered statistically significant, and those with $0.05 < p < 0.1$ were regarded as an insignificant tendency. All data are given as means \pm SEM.

Results

Pathomorphological changes in the liver of mice of experimental groups

Control animals had normal liver architecture with well-defined portal triads (Fig. S1)¹. On liver sections of animals from groups OF and Eth+OF, bile ducts appeared dilated, helminths were present in some of them, and there was a noticeable number of proliferating bile ducts and considerable lymphocytic–monocytic inflammatory infiltration. The bile duct epithelium became stratified in OF and Eth+OF mice. In the liver of animals from group Eth, fatty dystrophy of hepatocytes was noted.

Masson staining of liver sections revealed expansion of connective tissue in OF and Eth+OF mice, both in the region of large bile ducts (periductal fibrosis) and in the region of proliferation of small bile ducts (cholangiofibrosis) (Fig. S2). It is noteworthy that the combination of the two adverse factors (infection and ethanol) featured a significantly aggravated course of opisthorchiasis, namely the severity of fibrosis and the size of infiltration loci; these signs were absent with each adverse factor applied alone. Additionally, in OF and Eth+OF animals, there was a change of epithelial cells consistent with intestinal metaplasia. Hemozoin granules were found in some mice, in agreement with what has been previously observed in *O. felineus*-infected Syrian hamsters (Lvova et al., 2016).

Histological analyses of the cerebral cortex

Histological examination of brain sections uncovered differences in the number of perivascular zones of edema (around the vessels) and pericellular zones of edema (around the cells) in the cerebral cortex. There was a somewhat greater number of perivascular edema zones in OF mice compared to Eth and Eth+OF mice (Fig. 3A); an effect of the ethanol factor on this parameter was detected. OF mice differed significantly from all other animals by having an increased number of pericellular edema zones (Fig. 3B); the effect of the infection factor was significant. Ethanol contributed to a decrease in this parameter, especially in mice from group Eth. Moreover, the ethanol factor had a >4 times more pronounced impact than the infection

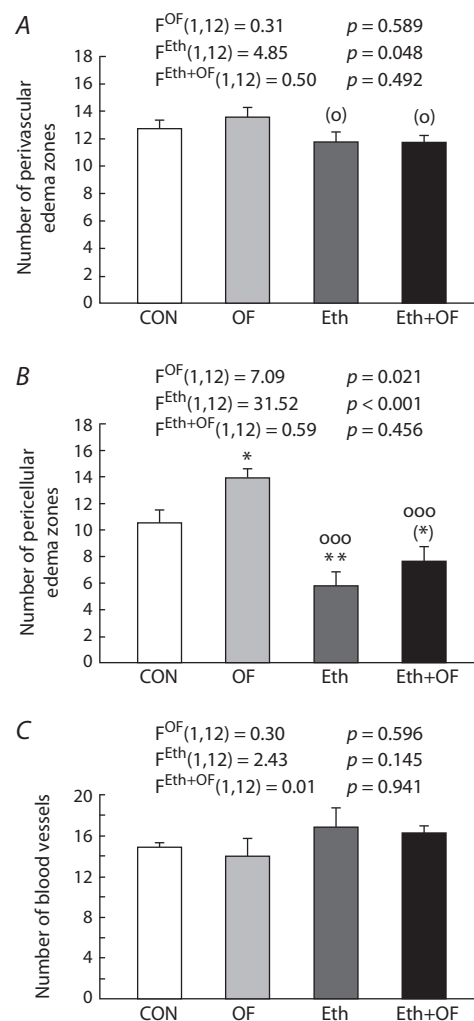


Fig. 3. Changes in the number of perivascular (A) and pericellular edema foci (B) and blood vessels (C) in the cerebral cortex of control mice (CON), *O. felineus*-infected mice (OF), mice consuming 20 % ethanol (Eth), and mice subjected to both procedures (Eth+OF).

* $p < 0.05$; ** $p < 0.01$; (*) $0.05 < p < 0.1$ as compared with group CON; ooo $p < 0.001$; (o) $0.05 < p < 0.1$ as compared with OF mice. The values are presented as mean \pm SEM and were processed by two-way ANOVA followed by the LSD test.

factor did; in the absence of their interaction, this outcome contributed to a lower value of this parameter in Eth+OF mice. The mouse groups did not differ in the number of vessels in the cerebral cortex (Fig. 3C). The assessed parameters in mice of all the experimental groups are presented in photographs of histological sections of the cerebral cortex (Fig. S3).

Immunohistochemical analysis of hippocampus and cortex sections

Analysis of Iba1-positive cells in the hippocampus revealed no difference among the groups (Fig. 4A). In the cerebral cortex (Fig. 4B), this parameter was higher in mice consuming ethanol (group Eth). Nonetheless, with a statistically significant interaction between the two adverse factors, this parameter decreased to the control level in Eth+OF mice. The number of IL-1 β -positive cells was higher in both the hippocampus (Fig. 4C) and cortex (Fig. 4D) during combined treatments (Eth+OF). Although in the hippocampus, this phenomenon

¹ Supplementary Figures S1–S5 and Table S1 are available at: <https://vavilovj-icg.ru/download/pict-2025-29/appx5.pdf>

was determined by the influence of the infection factor, in the cortex, in addition to infection, this outcome was facilitated by an even more significant influence of ethanol, and there was a synergistic interaction of the two factors.

A similar but even more pronounced pattern was documented for proinflammatory cytokine IL-6. In the hippocampus (Fig. 4E), infection to a greater extent and ethanol to a

lesser extent contributed to a higher value of this parameter in Eth+OF mice. In the cortex (Fig. 4F), on the contrary, ethanol exerted a greater effect on an increase in the number of IL-6-positive cells, and a statistically significant interaction of the adverse factors caused a significantly higher value of this parameter in Eth+OF animals compared to the other three groups of mice (CON, OF, and Eth). There was a pronounced

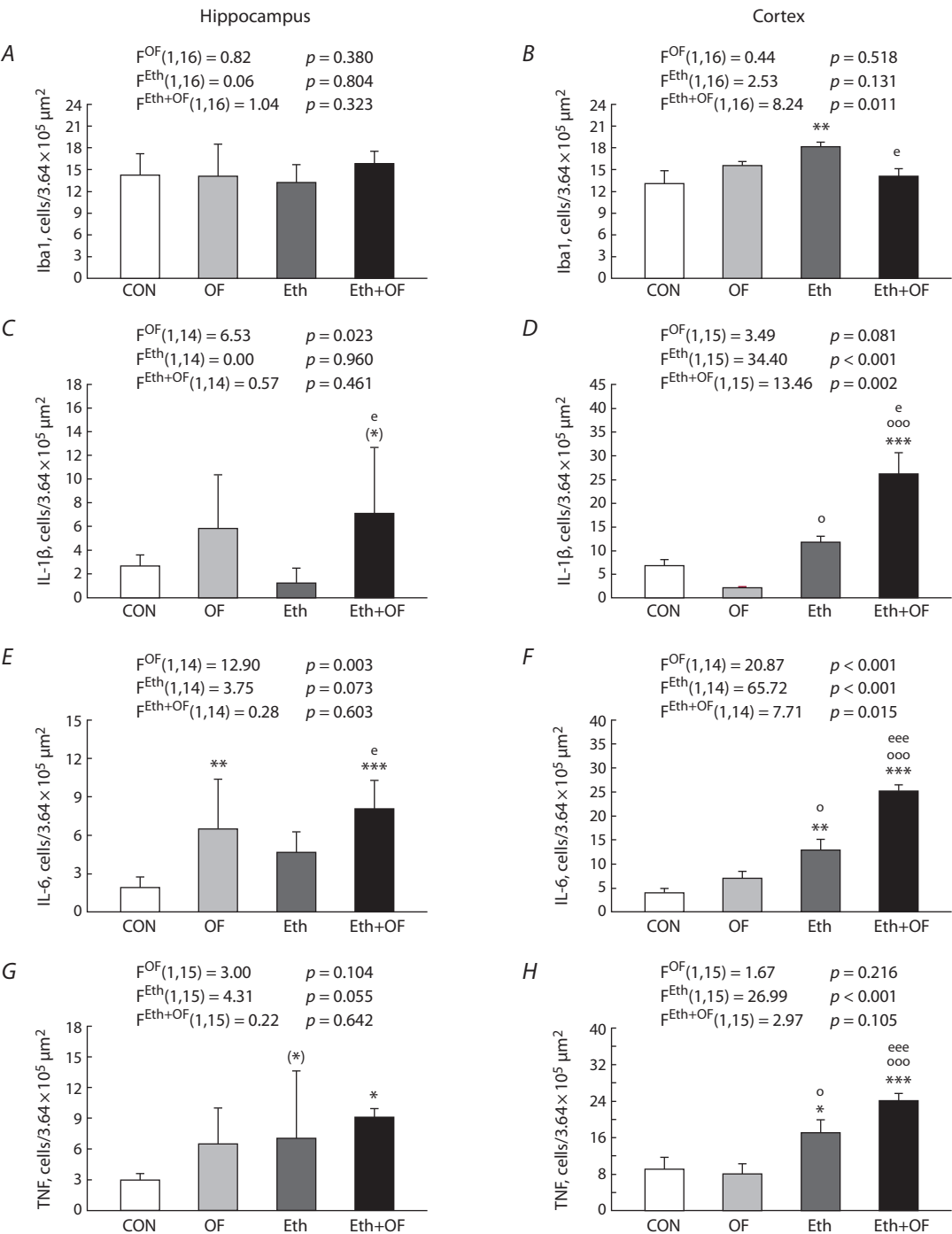


Fig. 4. Numerical density of brain cells in the hippocampus (A, C, E, G) and cortex (B, D, F, H) stained for Iba1, IL-1 β , IL-6, and TNF in the four mouse groups: control (CON), *O. felineus*-infected (OF), consuming 20 % ethanol (Eth), and subjected to both procedures (Eth+OF).

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; (*) $0.05 < p < 0.1$ in comparison with group CON; ° $p < 0.05$ and °°° $p < 0.001$ in comparison with group OF; e $p < 0.05$ and eee $p < 0.001$ in comparison with group Eth. The values are presented as mean \pm SEM and were processed by two-way ANOVA followed by the LSD test.

effect of ethanol, but not infection, on the increase in the number of TNF-positive cells in the hippocampus (Fig. 4G) and especially in the cortex (Fig. 4H). As in the analyses of the other proinflammatory cytokines, a statistically significantly higher value of this parameter was found in the cortex of Eth+OF mice relative to the other groups. Examples of immunohistochemical staining of the analyzed brain sections from the four mouse groups are presented in Figures S4 and S5.

Gene expression analysis in the hippocampus and cortex

Our quantitative analysis (qPCR) detected a significant effect of both adverse factors on the expression of the *Aif1* gene in the two brain structures under study (Fig. 5A, B). In the hippocampus, ethanol consumption had a more pronounced effect, contributing to a 28-fold higher value of this parameter in Eth mice compared to CON mice and a 7–8-fold higher value compared to groups OF and Eth+OF. An interaction of

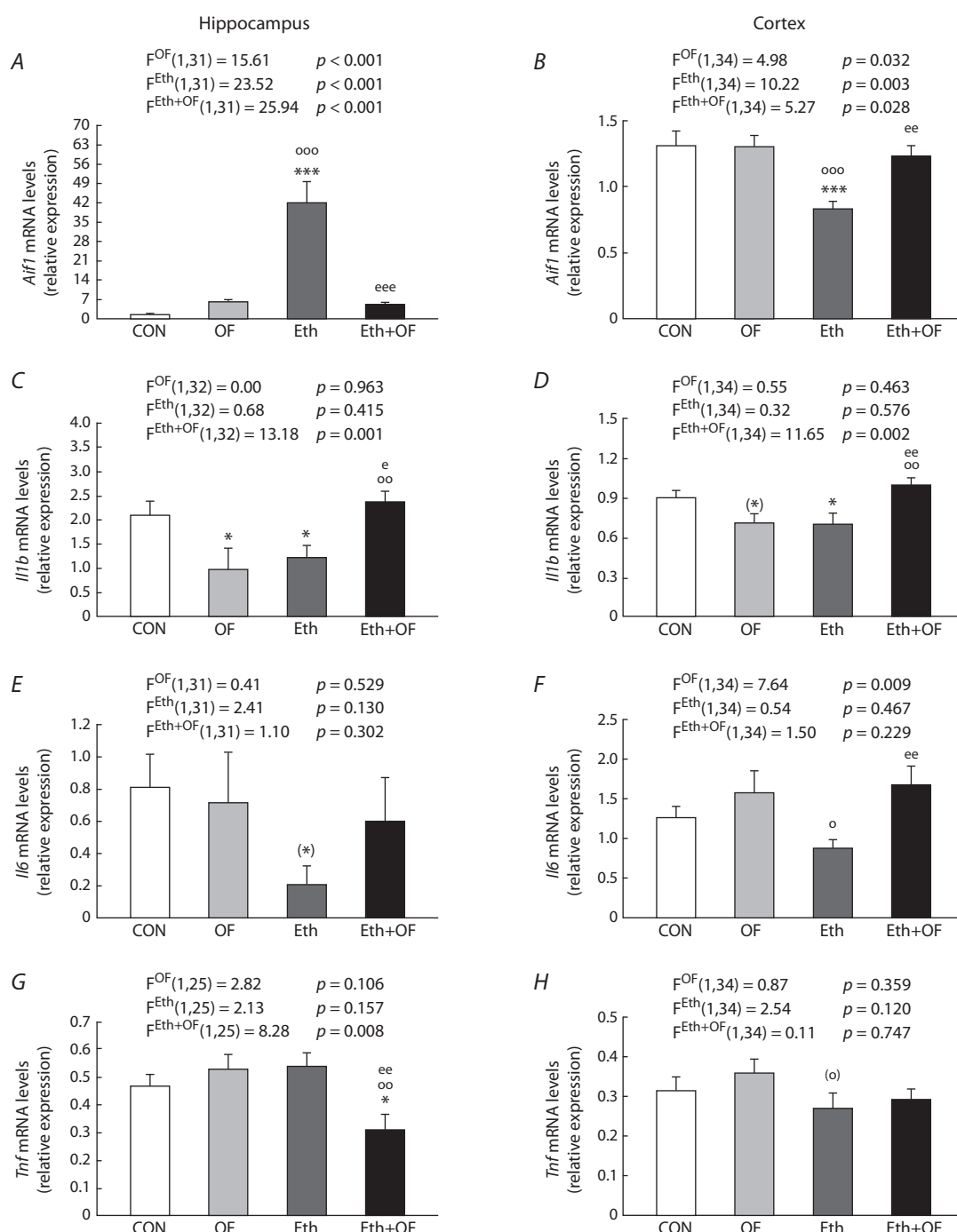


Fig. 5. Relative levels of mRNA expression of genes *Aif1*, *Il1b*, *Il6*, and *Tnf* in the hippocampus (A, C, E, G) and frontal cortex (B, D, F, H) of mice from the four groups: control (CON), *O. felineus*-infected (OF), consuming 20 % ethanol (Eth), and subjected to both procedures (Eth+OF).

* $p < 0.05$, *** $p < 0.001$, (*) $0.05 < p < 0.1$ in comparison with group CON; ° $p < 0.05$, °° $p < 0.01$, °°° $p < 0.001$, (°) $0.05 < p < 0.1$ in comparison with group OF; ° $p < 0.05$, °° $p < 0.01$, °°° $p < 0.001$ in comparison with group Eth.

the factors was noted to cause underexpression of the *Aif1* gene in Eth+OF mice. In the frontal cortex, on the contrary, ethanol consumption contributed to *Aif1* underexpression (the Eth group), and when the two adverse factors were combined, this parameter rose to the level of control mice.

When the expression of the *Il1b* gene was assessed in the hippocampus, no statistically significant effect of either factor alone was revealed, although a subsequent *post hoc* comparison of the groups detected a lower value of this parameter in mice from groups OF and Eth compared to CON mice. Nevertheless, a significant interaction of the two factors (Eth+OF mice; Fig. 5C) caused an increase in this parameter to the control level. A similar pattern was registered in the frontal cortex (Fig. 5D).

There was no statistically significant effect of each factor or of their interaction on the expression of the *Il6* gene in the hippocampus (Fig. 5E). The infection factor had a significant effect on the *Il6* expression in the frontal cortex (Fig. 5F), and this outcome contributed to an increase in the parameter in Eth+OF mice to the control level; this parameter was lower in Eth mice than in OF mice.

The combination of the two adverse factors caused down-regulation of the *Tnf* gene in the hippocampus of Eth+OF mice compared with the other groups of animals (CON, OF, and Eth) (Fig. 5G). In the frontal cortex of the brain, the expression of this gene was approximately the same in all four groups of animals (Fig. 5H); no statistically significant influence of each factor or of their interaction was detectable.

The startle reflex

Neither the prolonged consumption of ethanol nor chronic infection significantly affected PPI according to the evaluation of the startle response to acoustic stimuli (Fig. 6A). Nevertheless, the ethanol-drinking animals showed no habituation to the repetitive sound: in Eth mice, the startle response was the same for the 1st and 10th pulse with an intensity of 115 dB; the same was true for Eth+OF mice (Fig. 6B). Furthermore, in CON mice (statistically significantly) and in OF mice (insignificantly), the startle response to the sharp sound applied 10 times was found to decrease. It is noteworthy that in Eth+OF mice, the response to the first sound was the lowest when compared with the other groups, and there was a significant influence of ethanol [$F^{\text{Eth}}(1,51) = 8.10, p = 0.006$] and an insignificant influence of infection [$F^{\text{OF}}(1,51) = 3.09, p = 0.085$], but there was no interaction of the two factors [$F^{\text{Eth+OF}}(1,51) = 1.81, p = 0.185$]. The reaction of the mice to the 10th sound signal was the same among the four groups.

The open field test

In this test, intergroup differences were detected in the response of the mice to the odor of peppermint placed in the tumbler (Fig. 7A–C, Table S1). Even though the duration of staying near the tumbler with peppermint was approximately the same among all the groups, the number of approaches to it was lower in the second 3 min (i.e., when peppermint was present), especially in the CON group; the impact of the peppermint factor on this parameter was significant (Fig. 7A). An interaction of the ethanol and infection factors was detectable too (Table S1). Furthermore, CON and Eth mice sensed the peppermint smell faster than the other animals did because

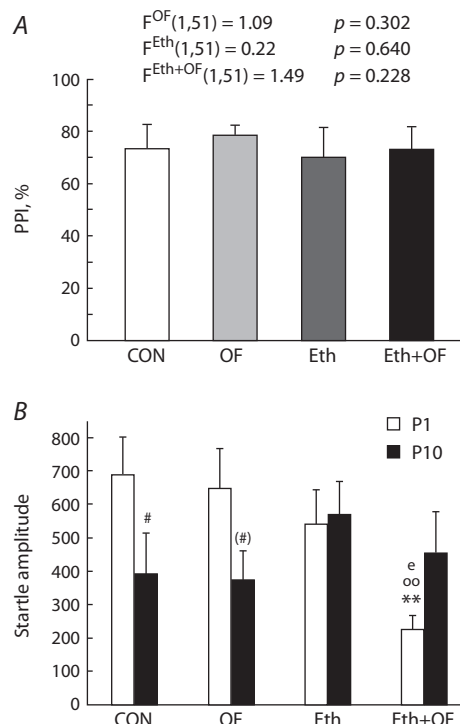


Fig. 6. Prepulse inhibition (PPI) (A) of the startle response to an acoustic stimulus and startle amplitude (B) in response to the first (P1) and 10th (P10) stimulus in control mice (CON), *O. felineus*-infected mice (OF), mice consuming 20 % ethanol (Eth), and mice subjected to both procedures (Eth+OF).

** $p < 0.01$ in comparison with group CON; °° $p < 0.01$ in comparison with OF mice; ° $p < 0.05$ in comparison with group Eth; # $p < 0.05$ and (#) $0.05 < p < 0.1$ in comparison with the P1 data. The values are presented as mean \pm SEM. Two-way ANOVA was used followed by the LSD test (A) and the Wilcoxon matched-pairs test (B).

the latency period of the first approach to the tumbler with peppermint was significantly shorter in groups CON and Eth (Fig. 7C). In OF mice, this parameter did not differ from the latency period of approaching the empty tumbler, whereas in group Eth+OF, there was an insignificant decrease from the first 3 min to the second 3 min. This finding may reflect impairment of olfactory sensitivity in the course of prolonged parasitosis.

In the first 3 min of the test, the number of crossed squares (reflecting locomotor activity of the animals) was lower in the OF group than in groups CON and Eth (Fig. 7D); the effect of infection was statistically significant (Table S1). With a change in the motivation of mice after peppermint was placed in the tumbler, this parameter markedly diminished in all four groups of mice (Fig. 7D), and a significant impact of the peppermint factor was registered (Table S1).

Exploratory activity, assessed by means of the number and duration of rearings, was influenced by the infection and peppermint factors (Fig. 7E, F; Table S1). In the first 3 min of the test, these parameters were lower in OF mice than in CON mice. As compared to the first 3 min, the number of rearings in the second 3 min was significantly lower in CON mice and insignificantly so in group Eth+OF, whereas the duration of rearing stayed approximately the same in all

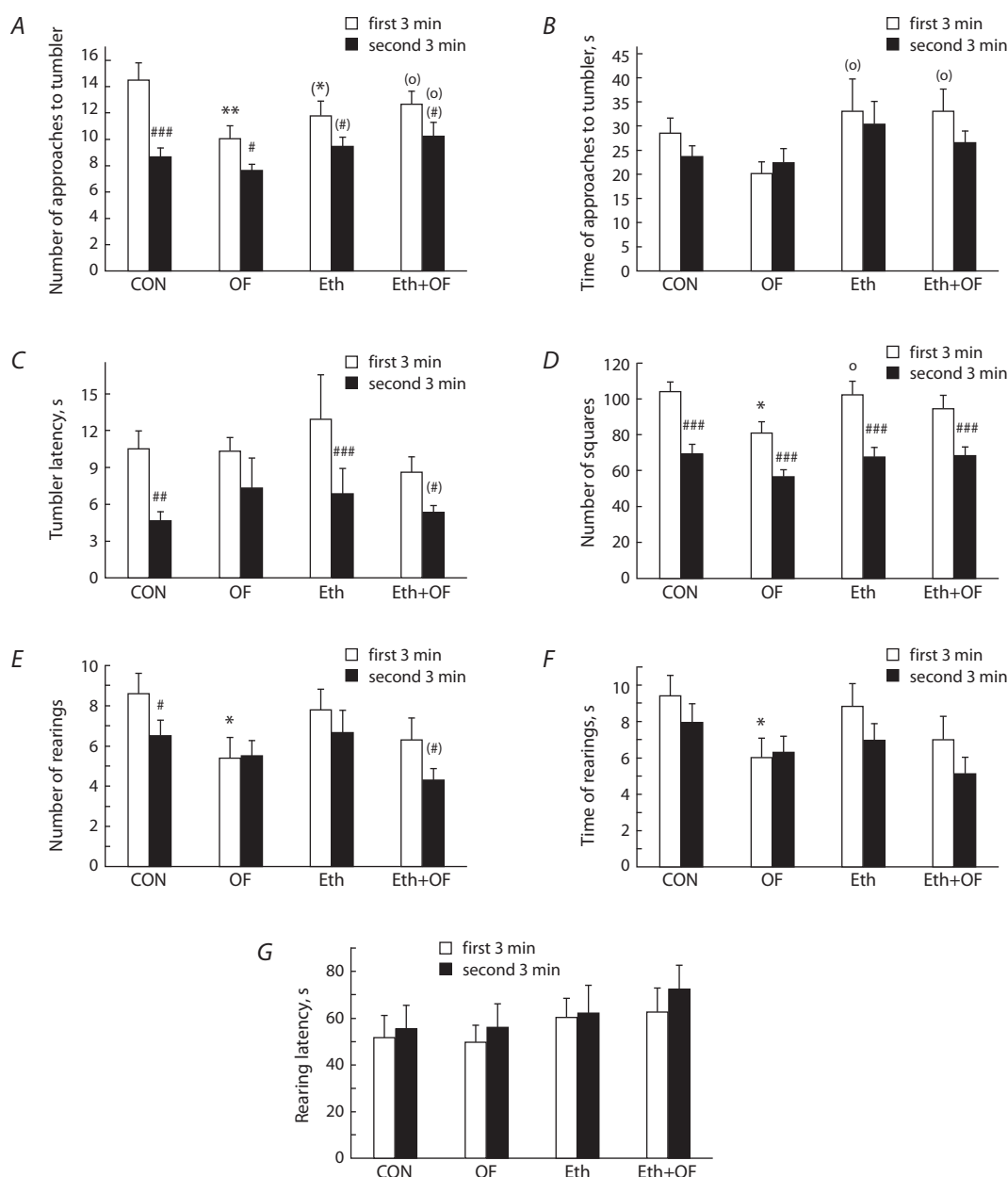


Fig. 7. Behavioral parameters of control mice (CON), *O. felineus*-infected mice (OF), mice consuming 20 % ethanol (Eth), and mice subjected to both procedures (Eth+OF) in the open field test with an empty tumbler (the first 3 min) and a tumbler filled with peppermint leaves (the second 3 min). The number (A), time (duration) (B), and latency (C) of approaches to the tumbler, the number of square crossings (D), and the number (E), time (duration) (F), and latency (G) of rearings near a wall were determined.

$p < 0.05$, ## $p < 0.01$, ### $p < 0.001$, and (#) $0.05 < p < 0.1$ as compared with the first 3 min of the test. * $p < 0.05$, ** $p < 0.01$, and (*) $0.05 < p < 0.1$ as compared with the corresponding time slot in group CON. ° $p < 0.05$ and (°) $0.05 < p < 0.1$ as compared with the corresponding time slot in group OF. The values are presented as mean \pm SEM. Three-way ANOVA followed by the LSD test.

groups. There was no influence of factors or of their interaction and no differences among the four groups in the latency of rearings either in the first or in the second 3 min of the test (Fig. 7G; Table S1).

The forced swimming test

Prolonged consumption of ethanol had a significant effect on the behavior of mice in the forced swimming test. Active-swimming duration was longer (Fig. 8A), and passive-

swimming duration was shorter (Fig. 8B), especially in mice of the Eth group and, to a lesser extent, in Eth+OF mice, both in comparison with CON and with OF mice.

Discussion

This is the first study on the states of the liver and brain during exposure of mice to a combination of two adverse factors: infection (*O. felineus* helminths) and a toxic chemical (ethanol). Influence of each factor and of their combined action on the

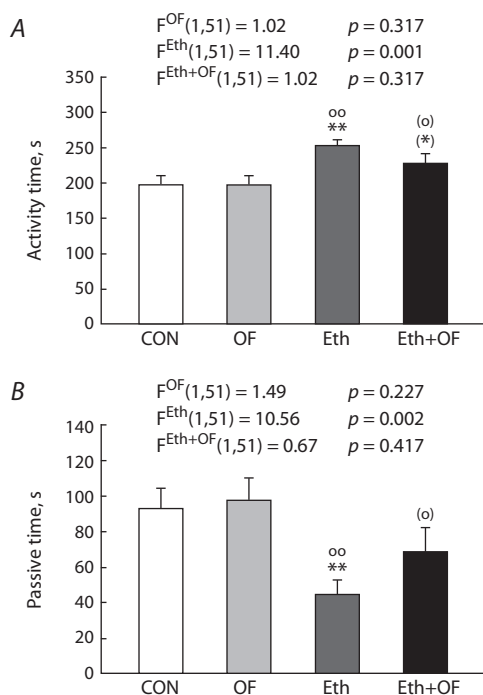


Fig. 8. The duration of active (A) and passive swimming (B) in control mice (CON), *O. felineus*-infected mice (OF), mice consuming 20 % ethanol (Eth), and mice subjected to both procedures (Eth+OF) in the forced swimming test.

** $p < 0.01$ and (*) $0.05 < p < 0.1$ as compared with group CON. oo $p < 0.01$ and (o) $0.05 < p < 0.1$ as compared with group OF. The values are presented as mean \pm SEM. Two-way ANOVA followed by the LSD test.

cortex and hippocampus was found in the brain of mice, as determined not only by molecular and histological methods but also at the level of behavior, which is considered a marker of functional changes in the brain.

First of all, as previously (Avgustinovich et al., 2022a), it was shown here that both adverse factors, especially their combination, cause hepatic inflammation. In mice subjected to the combination of the factors, there was not only a bile duct pathology caused by the infection (cholangio- and periductal fibrosis and inflammatory infiltration) but also hepatocyte dystrophy induced by ethanol. These data – just as previously detected elevated blood levels of leukocytes (especially stab neutrophils and monocytes) and of IL-6 (Avgustinovich et al., 2022a) – point to substantial inflammation in the body. It is known that virus- and alcohol-induced liver diseases are accompanied by inflammation and fibrosis associated with the suppression of activation of NF- κ B in hepatocytes: a transcription factor that is a key regulator of inflammation and cell death in the liver (Seki, Schwabe, 2015). In this context, Kupffer cells increase their production of proinflammatory cytokines, including TNF, IL-1 β , and IL-6 (Bilzer et al., 2006), with their subsequent traffic to the brain, where monocytes are also attracted from the bloodstream (D’Mello, Swain, 2011). As other researchers believe (Yang et al., 2014; Simon et al., 2019), these events are followed by activation of resident and recruited microglial cells, contributing to the development of neuroinflammation due to the production of inflammation mediators.

One of important indicators of inflammation in the brain is edema, the pathogenesis of which involves many factors, including liver failure (Adeva et al., 2012). In acute liver failure, edema is initially localized to the perivascular space and a zone of large swollen astrocytes. In our study, the experimental groups of mice did not differ from controls in perivascular edema, although in the animals consuming ethanol (groups Eth and Eth+OF), this parameter was somewhat less pronounced than that in OF mice. It is thought that when a pathological process in the liver is prolonged, structural changes occur in blood–brain barrier components (capillary endothelium, basal membrane, and astrocyte vascular peduncles), and these alterations reduce its protective function (Mishchenko et al., 1993). Blood–brain barrier destabilization facilitates penetration of “noxious agents” [e. g., ammonia, cytokines, or bacterial cell wall endotoxins: lipopolysaccharides (LPSs)] (Jayakumar et al., 2012) into the brain, with consequent pericellular edema due to the accumulation of fluid in the intracellular space, as demonstrated by magnetic resonance imaging in humans (Chavarria et al., 2011). We noticed significant predominance of pericellular edema in OF animals. At the same time, in the Eth+OF group of mice, ethanol “corrected” the number of edema foci, thereby reducing this parameter below the control value. Therefore, the actions of the two factors had opposite directions.

We assume that the increase in the number of pericellular edema zones during the helminth infection may proceed according to the hepatic-encephalopathy scenario, associated not only with upregulation of inflammatory cytokines (TNF, IL-1 β , IL-6, and IFN- γ) in the blood but also with elevated blood and brain ammonia levels, as shown in acute and chronic liver failure (Butterworth, 2003; Rama Rao et al., 2014; Upadhyay, 2017). Ammonia is extremely toxic to the brain and leads to hepatic coma. Considering the evidence from other researchers that alcohol promotes edema in the brain (Collins et al., 1998; Collins, Neafsey, 2012; de la Monte, Kril, 2014), the decrease in the number of pericellular edema zones in mice consuming ethanol in our experiment remains unexplained. It can be hypothesized that aquaporins (AQPs) are involved, which are water-selective plasma membrane channels that enhance water permeability of cells (Huber et al., 2007), because there are reports confirming that ethanol can diminish swelling in the cortex after brain injury and that this phenomenon is associated with underexpression of AQP4 and AQP9 simultaneously with an improvement of cognitive and motor functions in animals (Wang et al., 2013). Because it is known that aquaporins are enriched within brain astroglia (Huber et al., 2007; Collins, Neafsey, 2012), the decrease in pericellular edema during prolonged ethanol consumption in our experiment may be related to its degenerative effect on astrocytes (de la Monte, Kril, 2014). As demonstrated by the optical fractionation technique, in severely ill alcoholics, there is a loss of 37 % of glial cells in the hippocampus, primarily astrocytes and oligodendrocytes and to a lesser extent microglial cells, without a loss of neurons (Korbo, 1999). Apparently, our data require further research.

It is known that brain microglia, when responding to any pathological stimulus coming from the periphery (an LPS challenge, vaccination, or alcohol), begin to produce proinflammatory cytokines with subsequent stimulation of a release

of cytokines and chemokines from neurons and astrocytes (Miller et al., 2009; Yang et al., 2014; Norden et al., 2016; Henriques et al., 2018).

According to our findings, there are brain region-specific changes in the expression of the *Aif1* gene, reflecting the activity of microglia, during the exposure to the two adverse factors. In the hippocampus, where microglial density is reported to be the highest (Silvin, Ginhoux, 2018; Tan et al., 2020), a statistically significant influence of both factors and of their interaction was documented in our work. Chronic consumption of 20 % ethanol contributed to a significant increase in this parameter; the impact of the infectious factor was less pronounced but statistically significant, as found previously (Avgustinovich et al., 2022a). It is important that when the two factors acted together, *Aif1* expression in the hippocampus was low. Apparently, in this situation, activation of microglia cells in the hippocampus is so substantial that the activated microglia can be driven into apoptosis, in order to prevent the brain from entering a state of chronic inflammation. Currently, the mechanisms controlling microglial apoptosis are characterized incompletely, but the possibility of a chain of events (in the brain) proceeding from the stage of activated microglia to their apoptosis has been examined by other researchers (Fu et al., 2015). Further microscopic research on phenotypic alterations in microglial cells is necessary to identify their transformation from a resting state (ramified morphology) to active status (amoeboid morphology) (Tan et al., 2020).

In the frontal cortex, a significant influence of each adverse factor and of their statistically significant interaction on the expression of the *Aif1* gene was noted too. In contrast to the hippocampus, in this brain structure, ethanol reduced this parameter, and the combination of the two factors contributed to its increase to the level of control individuals. Evidently, the infectious factor has an effect opposite to that of ethanol, thus increasing the expression of this gene. Immunohistochemistry revealed a statistically significantly greater number of Iba1-positive cells in the cortex in a more distant frontal section (level 74) as compared to the prefrontal region of the brain. Therefore, we can theorize that the observed downregulation of *Aif1* in Eth mice is a compensatory reaction of genes to the large protein amount.

Thus, in our paper, region-specific changes in microglial activity were established in terms of the expression of the *Aif1* gene and of a protein (Iba1) in response to the infectious factor (*O. felineus*), and especially to the toxic factor (ethanol). This result can be explained in accordance with the current understanding of the heterogeneity of microglia in the cell number, morphomolecular signatures, and homeostatic functions in different anatomical structures of the healthy CNS (Tan et al., 2020) and during alcoholic pathologies (He, Crews, 2008). In this regard, our data indicate a possible modifying effect of helminth infection on the expression of the *Aif1* gene, but not its protein, in the hippocampus and cortex. Ethanol had a more pronounced effect on this parameter in both brain structures. Taking into account the interaction of the two factors, we can say that their effects have different directions. Activation of microglia in the cortex under the influence of ethanol was also recorded in terms of the level of the Iba1 protein.

It has been reported that peripheral insults (an LPS or bacterial challenge or ethanol) that cause microglial activa-

tion induce upregulation of proinflammatory cytokines (TNF, IL-1 β , and IL-6) at the protein and/or mRNA level along with engagement of Toll-like receptors (TLR-2 and TLR-4) (Collins, Neafsey, 2012; Fernandez-Lizarbe et al., 2013; Yang et al., 2014; Hoogland et al., 2015). As reported by C. D'Mello and M.G. Swain (2011), cytokines TNF, IL-1 β , and IL-6 are likely to be key promoters of central neural alterations in chronic liver diseases. Accordingly, we next examined changes in expression of three proinflammatory cytokines (IL-1 β , IL-6, and TNF) in response to the adverse factors.

The expression of the *Il1b* gene in the two brain structures under study did not reflect a statistically significant effect of each factor alone, but there was a significant effect when they were combined: the weak expression of this gene in groups OF and Eth increased in Eth+OF mice to the control level. Immunohistochemical analysis of sections of the hippocampus, and especially that of the cerebral cortex, pointed to high levels of this proinflammatory cytokine in mice during the exposure to the two factors, thereby implying neuroinflammation.

We believe that our results on protein and mRNA levels of *Il6* in the hippocampus reflect a variable process associated with the duration of the adverse factors: a change in the amount of the proinflammatory protein was followed by a change in mRNA expression, and these processes seem to be somewhat separated in time. In the cerebral cortex of Eth+OF mice, there was the highest level of IL-6 among all groups of mice, which is explained by the synergy of the two adverse factors. Nonetheless, there were no pronounced differences between the groups in *Il6* expression within this brain structure, and this outcome may be ascribed to regulation by a negative feedback mechanism. Similar dynamics of expression (from unchanged to increased/decreased) of genes responsible for levels of IL-6, IL-1 β , and TNF in the three structures of the rat brain were recorded by other researchers after forced consumption of ethanol by rats (for 6 months) (Nunes et al., 2019). Those authors proposed that the upregulation of some cytokines may mean infiltration of immune cells (T cells in particular) into the brain, and these phenomena are indicative of a severe impairment of the blood-brain barrier especially during the synergism of the two adverse factors. Under such conditions, the delayed changes in the genes' expression may have a protective/compensatory effect against expansion of neuroinflammatory-cytokine expression in the brain.

In contrast to the other two genes, the combination of the helminth infection and the prolonged ethanol consumption caused underexpression of *Tnf* in the hippocampus of Eth+OF mice; this phenomenon may be a compensatory reaction to the elevated amount of the TNF protein in this brain structure. An even higher level of this protein was detected in the cortex at level 74 of the frontal brain section during ethanol administration, but the *Tnf* mRNA level in the frontal cortex was the same among all the groups of mice.

Thus, three proinflammatory cytokines were found to differ in mRNA or protein expression depending not only on the nature of the adverse factor(s) but also on localization in the brain. The most pronounced variations of the parameters were noticed in the frontal cortex, especially during prolonged ethanol consumption, and to a lesser extent in the hippocampus. It is known that both brain structures in question, especially frontal lobes of the cerebral cortex, are sensitive to alcohol-induced

damage (Fowler et al., 2014). In this regard, the pathogenic effect of alcohol is associated with white matter atrophy, neuroinflammation, and synaptogenesis disturbances, leading to emotional instability and cognitive impairment (Harper, 2009; de la Monte, Kril, 2014). The outcomes observed here – just as previously obtained evidence that inflammation in the liver during the action of the two adverse factors is accompanied by an increase in blood concentrations of monocytes and proinflammatory cytokines (Avgustinovich et al., 2022a) – are thought to contribute to neuroinflammation and may induce changes in central neurotransmission that are manifested in the behavior of animals (D’Mello, Swain, 2011). That is why we performed an extensive analysis of murine behavior, which reflects disturbances in the brain.

Even though PPI was the same among our groups of mice, the animals in the groups consuming 20 % ethanol for a long time (Eth and Eth+OF) did not get accustomed to the administered signals: startle amplitudes of the 1st and 10th signals were the same. In addition, the mice subjected to both adverse factors (Eth+OF) showed the weakest reaction to the first sharp sound signal, but the response increased by the 10th signal. These results mean that the brain of mice consuming ethanol is always ready to respond strongly to repeated harsh sound signals. Considering that the startle reflex is regarded as a behavioral indicator of CNS excitability (Blendov et al., 2019), we can assume permanent high excitability of brain neurons in mice consuming 20 % ethanol.

Throughout almost the whole forced swimming test (on average 252 out of 300 s), mice of the Eth group tried to actively get out of the water, and passive swimming was shorter. In Eth+OF mice, the changes in these parameters were smaller: the differences from groups CON and OF were insignificant. The duration of immobility in this test is regarded as an indicator of depressive-like behavior in rodents and is reduced by known antidepressants (Lucki, 2001). Nonetheless, we believe that the prolonged consumption of 20 % ethanol did not have an antidepressant effect but rather promoted CNS hyperexcitability, which involves an imbalance in the activities of the glutamatergic and GABAergic systems of the brain. It is known that chronic alcohol consumption leads to hyperexcitation of neurons because of downregulation of GABAergic functions as a consequence of pseudo-immaturity in the hippocampus and prefrontal cortex (Murano et al., 2017). A blockade of NMDA receptors and of the nitric oxide/cyclic-guanosine monophosphate pathway may be involved in the antidepressant-like effect of ethanol in mice (Khan et al., 2021).

The open field test in its various modifications is utilized by researchers to assess many behavioral parameters in rodents: locomotor and exploratory activities, emotionality and anxiety, and a reaction to an unfamiliar object when re-tested (Choleris et al., 2001). Because our aim was, first of all, to assess the sense of smell in the four groups of mice, the unfamiliar smell of peppermint was presented to the mice during the second part of the test, after they were familiarized with the test arena in the first 3 min. New odors are often aversive to rodents, for example, rats avoid the peppermint smell at first exposure (Brown, Willner, 1983). On the other hand, in some studies on mice, investigators have described repeated use of peppermint for treating olfactory impairment (Kim et al., 2019).

In our experiments, mice subjected to prolonged ethanol consumption (group Eth), just as CON mice, quickly identified the unfamiliar odor because their latency period for approaching the tumbler with peppermint diminished. The infectious factor did not affect this parameter in OF mice and had a weak effect on Eth+OF mice, implying disturbances of the sense of smell in these mice. Besides, although the control mice exhibited a pronounced avoidance reaction toward the tumbler with peppermint (as evidenced by the number of approaches or turns to the tumbler), in the other groups of mice, this parameter was less pronounced. This result also points to some anomalies in the sense of smell in mice of the three experimental groups [subjected to an adverse factor(s)]. We believe that the changes in the sense of smell resulting from liver fluke infection and consumption of 20 % ethanol may be associated with aberrations in the CNS. In any case, hyposmia is considered an early symptom of Parkinson’s disease (Chen et al., 2012), which is also associated with neuroinflammation.

During the evaluation of other patterns of behavior in the open field test, it was found here that the 6-month consumption of 20 % ethanol did not have a significant effect on the locomotor and exploratory activities of mice, as evaluated via the number of squares crossed and rearing parameters in the first 3 min of the test. By contrast, these parameters were lower in OF mice. A decrease in these parameters, according to other articles (Henderson et al., 2004; Seibenhener, Wooten, 2015), can be viewed as a manifestation of anxious behavior. Taking this into account, we could say that there is a likely pro-anxiety effect of prolonged *O. felineus* infection on animals.

Conclusion

The presented data are an experimental model of situations often occurring in human society: people with chronic opisthorchiasis – perhaps being unaware of the infection – abuse alcohol or, conversely, by relying on disinfecting properties of ethanol, begin to drink it, sometimes in large amounts, when there is a threat of this infection. Nonetheless, our experimental data indicate that under such circumstances, the liver is not the only organ that receives a double “blow” (the toxic injury plus the infectious one). When chronic alcoholization is combined with prolonged *O. felineus* infection, the brain also receives a double impact: aside from direct entry of ethanol into the CNS through the blood–brain barrier, according to D’Mello and Swain (2011), peripheral proinflammatory signals begin to arrive with the blood, primarily IL-1 β , IL-6, TNF, and monocytes.

Under these conditions, as revealed by two-way ANOVA, there are statistically significant effects of interaction of the two adverse factors on histological and molecular characteristics of microglia and on proinflammatory cytokines, and these effects are brain region-specific. For instance, in the hippocampus, the infectious factor attenuated ethanol-induced *Aif1* overexpression, which reflects the activation of microglia. By contrast, in the frontal cortex, the expression of this gene was low during the prolonged alcoholization and increased to control values in the mice subjected to both factors. During a statistically significant interaction of the factors, this finding indicates that directions of the two impacts are different. It is possible that helminths exert a “corrective” effect here that is designed to preserve the health of the host (at whose expense

they live and reproduce) because excessive activation of microglia can have irreversible neurodegenerative consequences and may ultimately kill the host. Identical directions of the effects of the two factors were noted during quantification of the expression of the *Il1b* gene (in the cortex and hippocampus) and of the *Tnf* gene (in the hippocampus): these effects promoted an increase in the former parameter and a decrease in the latter and may be attributed to the proteins' levels at this stage of the pathology.

In the cortex, high concentrations of cytokines IL-1 β , IL-6, and TNF were found at levels 72–74 of the frontal brain slices in mice subjected to both factors; this finding implies that the alterations induced in the prefrontal region of the cerebral cortex might be similar. Together with an increase in these parameters in the hippocampus, this finding indicates the development of neuroinflammation.

We believe that the obtained results indicate a variable process that is largely explained by the duration of the stimuli: a change in the amounts of proinflammatory proteins is followed by a change in mRNA expression. Furthermore, these processes are brain region-specific and seem to be somewhat separated temporally. This is because these processes not only are regulated by the activity of brain cells but also depend on the arrival of peripheral proinflammatory signals into the brain. This phenomenon in turn affects the behavior of the animals. Behavioral testing of our mice revealed that ethanol has a stimulatory effect, which manifested itself in two tests reflecting alterations in the regulatory mechanisms of the CNS. In our mice, the behavioral pathology associated with *O. felineus* infection is suggestive of the development of anxiety. It should be pointed out that both factors altered the mice's sense of smell (the infection did so to a greater extent). Mechanisms that counteract these adverse effects remain to be explored. At this stage of the project, we can state only that the combination of the two factors – in addition to the destructive effects on the liver – can affect the brain too, by contributing to the development of neuroinflammation.

References

- Adeva M.M., Souto G., Donapetry C., Portals M., Rodriguez A., Lamas D. Brain edema in diseases of different etiology. *Neurochem Int.* 2012;61(2):166-174. doi 10.1016/j.neuint.2012.05.007
- Akhmedov V.A., Kritevich M.A. Chronic opisthorchiasis is a multiple-organ pathology. *Vestnik NGU. Seriya: Biologiya, Klinicheskaya Meditsina = Vestnik NSU. Series: Biology and Clinical Medicine.* 2009;7(1):118-121. [https://nsu.ru/xmlui/bitstream/handle/nsu/4344/21.pdf] (in Russian)
- Avgustinovich D.F., Marenina M.K., Zhanaeva S.Y., Tenditnik M.V., Katokhin A.V., Pavlov K.S., Sivkov A.Y., Vishnivetskaya G.B., Lvova M.N., Tolstikova T.G., Mordvinov V.A. Combined effects of social stress and liver fluke infection in a mouse model. *Brain Behav Immun.* 2016;53:262-272. doi 10.1016/j.bbi.2016.01.012
- Avgustinovich D., Kovner A., Kashina E., Shatskaya N., Vishnivetskaya G., Bondar N., Lvova M. The pathogenic potential of the combined action of chronic *Opisthorchis felineus* infection and repeated social defeat stress in C57BL/6 mice. *Int J Parasit.* 2021;51:353-363. doi 10.1016/j.ijpara.2020.10.003
- Avgustinovich D., Kizimenko A., Marenina M., Lvova M., Kovner A., Orlovskaya I., Toporkova L., Goiman E., Tsyganov M., Ponomarev D. Prolonged liver fluke infection combined with alcoholization: an experimental mouse model. *Exp Parasitol.* 2022a;242:108399. doi 10.1016/j.exppara.2022.108399
- Avgustinovich D.F., Tenditnik M.V., Bondar N.P., Marenina M.K., Zhanaeva S.Y., Lvova M.N., Katokhin A.V., Pavlov K.S., Evseenko V.I., Tolstikova T.G. Behavioral effects and inflammatory markers in the brain and periphery after repeated social defeat stress burdened by *Opisthorchis felineus* infection in mice. *Physiol Behav.* 2022b;252:113846. doi 10.1016/j.physbeh.2022.113846
- Bajaj J.S. Alcohol, liver disease and the gut microbiota. *Nat Rev Gastroenterol Hepatol.* 2019;16(4):235-246. doi 10.1038/s41575-018-0099-1
- Belzung C. Hippocampal mossy fibres: implication in novelty reactions or in anxiety behaviours? *Behav Brain Res.* 1992;51(2):149-155. doi 10.1016/s0166-4328(05)80208-6
- Bernstein D.E., Dickinson G.M., Kim K.-J., Al Karawi M.A., Bar-kin J.S. Parasitic causes of pancreatic and biliary tract disease: a growing concern in a highly mobile population. *Mil Med.* 1994;159(4):331-338. doi 10.1093/milmed/159.4.331
- Bilzer M., Roggel F., Gerbes A.L. Role of Kupffer cells in host defense and liver disease. *Liver Int.* 2006;26(10):1175-1186. doi 10.1111/j.1478-3231.2006.01342.x
- Bishehsari F., Magno E., Swanson G., Desai V., Voigt R.M., Forsyth C.B., Keshavarzian A. Alcohol and gut-derived inflammation. *Alcohol Res.* 2017;38(2):163-171
- Blednov Y.A., Bajo M., Roberts A.J., Da Costa A.J., Black M., Edmunds S., Mayfield J., Roberto M., Homanics G.E., Lasek A.W., Hitzemann R.J., Harris R.A. *Scn4b* regulates the hypnotic effects of ethanol and other sedative drugs. *Genes Brain Behav.* 2019;18(6):e12562. doi 10.1111/gbb.12562
- Boonpucknavig S., Boonpucknavig V., Tanvanich S., Doungchawee G., Thamavit W. Development of immune-complex glomerulonephritis and amyloidosis in Syrian golden hamsters infected with *Opisthorchis viverrini*. *J Med Assoc Thai.* 1992;75(Suppl.1):7-19
- Brown R.E., Willner J.A. Establishing an "affective scale" for odor preferences of infant rats. *Behav Neural Biol.* 1983;38(2):251-260. doi 10.1016/s0163-1047(83)90254-6
- Butterworth R.F. Role of circulating neurotoxins in the pathogenesis of hepatic encephalopathy: potential for improvement following their removal by liver assist devices. *Liver Int.* 2003;23(Suppl.3):5-9. doi 10.1034/j.1478-3231.23.s.3.1.x
- Bychkov V.G., Solov'eva O.G., Khadieva E.D., Ivanova L.A., Kulikova S.V., Garchuk I.V., Orlov S.A., Sabirov A.Kh., Shilin K.O., Beliaeva M.I. Morphogenesis of the structural changes of the internal organs in superinvasive opisthorchiasis. *Morfologiya = Morphology.* 2011;140(5):22-27 (in Russian)
- Chavarria L., Alonso J., Rovira A., Córdoba J. Neuroimaging in acute liver failure. *Neurochem Int.* 2011;59(8):1175-1180. doi 10.1016/j.neuint.2011.09.003
- Chen W., Chen S., Kang W.Y., Li B., Xu Z.M., Xiao Q., Liu J., Wang Y., Wang G., Chen S.D. Application of odor identification test in Parkinson's disease in China: a matched case-control study. *J Neurol Sci.* 2012;316(1-2):47-50. doi 10.1016/j.jns.2012.01.033
- Choleris E., Thomas A.W., Kavaliers M., Prato F.S. A detailed ethological analysis of the mouse open field test: effects of diazepam, chlordiazepoxide and an extremely low frequency pulsed magnetic field. *Neurosci Biobeh Rev.* 2001;25(3):235-260. doi 10.1016/s0149-7634(01)00011-2
- Collins M.A., Neafsey E.J. Neuroinflammatory pathways in binge alcohol-induced neuronal degeneration: oxidative stress cascade involving aquaporin, brain edema, and phospholipase A2 activation. *Neurotox Res.* 2012;21:70-78. doi 10.1007/s12640-011-9276-5
- Collins M.A., Zou J.-Y., Neafsey E.J. Brain damage due to episodic alcohol exposure *in vivo* and *in vitro*: furosemide neuroprotection implicates edema-based mechanism. *FASEB J.* 1998;12(2):221-230. doi 10.1096/fasebj.12.2.221
- de la Monte S.M., Kril J.J. Human alcohol-related neuropathology. *Acta Neuropathol.* 2014;127(1):71-90. doi 10.1007/s00401-013-1233-3
- D'Mello C., Swain M.G. Liver-brain inflammation axis. *Am J Physiol Gastrointest Liver Physiol.* 2011;301(5):G749-G761. doi 10.1152/ajpgi.00184.2011

- Fernandez-Lizarbe S., Montesinos J., Guerri C. Ethanol induces TLR4/TLR2 association, triggering an inflammatory response in microglial cells. *J Neurochem*. 2013;126(2):261-273. doi 10.1111/jnc.12276
- Fowler A.K., Thompson J., Chen L., Dagda M., Dertien J., Dossou K.S., Moaddel R., Bergeson S.E., Kruman I.I. Differential sensitivity of prefrontal cortex and hippocampus to alcohol-induced toxicity. *PLoS One*. 2014;9(9):e106945. doi 10.1371/journal.pone.0106945
- Friard O., Gamba M. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol Evol*. 2016;7:1325-1330. doi 10.1111/2041-210X.12584
- Fu Y., Yang M.-S., Jiang J., Ganesh T., Joe E., Dingleline R. EP2 receptor signaling regulates microglia death. *Mol Pharmacol*. 2015;88(1):161-170. doi 10.1124/mol.115.098202
- Geil C.R., Hayes D.M., McClain J.A., Liput D.J., Marshall S.A., Chen K.Y., Nixon K. Alcohol and adult hippocampal neurogenesis: promiscuous drug, wanton effects. *Prog Neuropsychopharmacol Biol Psychiatry*. 2014;54:103-113. doi 10.1016/j.pnpbp.2014.05.003
- Gundamaraju R., Vemuri R.C. *Opisthorchis viverrini* (liver fluke) as the lot of baleful parasite of tropical region-A replete synopsis. *Asian Pac J Trop Dis*. 2014;4(1):61-66. doi 10.1016/S2222-1808(14)60316-9
- Harper C. The neuropathology of alcohol-specific brain damage, or does alcohol damage the brain? *J Neuropathol Exp Neurol*. 1998;57(2):101-110. doi 10.1097/00005072-199802000-00001
- Harper C. The neuropathology of alcohol-related brain damage. *Alcohol Alcohol*. 2009;44(2):136-140. doi 10.1093/alcal/agn102
- He J., Crews F.T. Increased MCP-1 and microglia in various regions of the human alcoholic brain. *Exp Neurol*. 2008;210(2):349-358. doi 10.1016/j.expneurol.2007.11.017
- Henderson N.D., Turri M.G., DeFries J.C., Flint J. QTL analysis of multiple behavioral measures of anxiety in mice. *Behav Genet*. 2004;34(3):267-293. doi 10.1023/B:BEGE.0000017872.25069.44
- Henriques J.F., Portugal C.C., Canedo T., Relvas J.B., Summavielle T., Socodato R. Microglia and alcohol meet at the crossroads: microglia as critical modulators of alcohol neurotoxicity. *Toxicol Lett*. 2018;283:21-31. doi 10.1016/j.toxlet.2017.11.002
- Hoogland I.C., Houbolt C., van Westerloo D.J., van Gool W.A., van de Beek D. Systemic inflammation and microglial activation: systematic review of animal experiments. *J Neuroinflammation*. 2015;12:114. doi 10.1186/s12974-015-0332-6
- Huber V.J., Tsujita M., Yamazaki M., Sakimura K., Nakada T. Identification of arylsulfonamides as Aquaporin 4 inhibitors. *Bioorg Med Chem Lett*. 2007;17(5):1270-1273. doi 10.1016/j.bmcl.2006.12.010
- Jayakumar A.R., Tong X.Y., Ospel J., Norenberg M.D. Role of cerebral endothelial cells in the astrocyte swelling and brain edema associated with acute hepatic encephalopathy. *Neuroscience*. 2012;218:305-316. doi 10.1016/j.neuroscience.2012.05.006
- Ketpueak T., Thienmimit P., Apaijai N., Chattipakorn S.C., Chattipakorn N. Association of chronic opisthorchis infestation and microbiota alteration on tumorigenesis in cholangiocarcinoma. *Clin Transl Gastroenterol*. 2020;12(1):e00292. doi 10.14309/ctg.0000000000000292
- Khan M.I., Nikoui V., Naveed A., Mumtaz F., Zaman H., Haider A., Aman W., Wahab A., Khan S.N., Ullah N., Dehpour A.R. Antidepressant-like effect of ethanol in mice forced swimming test is mediated via inhibition of NMDA/nitric oxide/cGMP signaling pathway. *Alcohol*. 2021;92:53-63. doi 10.1016/j.alcohol.2021.01.005
- Kim B.Y., Park J.Y., Kim E.J., Kim B.G., Kim S.W., Kim S.W. The neuroplastic effect of olfactory training to the recovery of olfactory system in mouse model. *Int Forum Allergy Rhinol*. 2019;9(7):715-723. doi 10.1002/alr.22320
- Korbo L. Glial cell loss in the hippocampus of alcoholics. *Alcohol Clin Exp Res*. 1999;23(1):164-168. doi 10.1111/j.1530-0277.1999.tb04039.x
- Lowe P.P., Morel C., Ambade A., Iracheta-Vellve A., Kwiatkowski E., Satishchandran A., Furi I., Cho Y., Gyongyosi B., Catalano D., Lefebvre E., Fischer L., Seyedkazemi S., Schafer D.P., Szabo G. Chronic alcohol-induced neuroinflammation involves CCR2/5-dependent peripheral macrophage infiltration and microglia alterations. *J Neuroinflammation*. 2020;17(1):296. doi 10.1186/s12974-020-01972-5
- Lucki I. A prescription to resist proscriptions for murine models of depression. *Psychopharmacology (Berl)*. 2001;153(3):395-398. doi 10.1007/s002130000561
- Lvova M., Zhukova M., Kiseleva E., Mayboroda O., Hensbergen P., Kizilova E., Ogienko A., Besprozvannykh V., Sripa B., Katokhin A., Mordvinov V. Hemozoin is a product of heme detoxification in the gut of the most medically important species of the family Opisthorchiidae. *Int J Parasitol*. 2016;46(3):147-156. doi 10.1016/j.ijpara.2015.12.003
- Lvova M.N., Shevelev O.B., Serdobintseva V.V., Kalinin D.V., Starostenko D.A., Zavjalov E.L., Krivoschapkin A.L., Logachev P.V., Mordvinov V.A., Avgustinovich D.F. Effect of silicon dioxide nanoparticles on Syrian hamsters infected by *Opisthorchis felinus*: ¹H MRS study of the brain. *Dokl Biochem Biophys*. 2020;495(1):319-324. doi 10.1134/S1607672920060095
- Lvova M.N., Ponomarev D.V., Tarasenko A.A., Kovner A.V., Minkova G.A., Tsyganov M.A., Li M., Lou Y., Evseenko V.I., Dushkin A.V., Sorokina I.V., Tolstikova T.G., Mordvinov V.A., Avgustinovich D.F. Curcumin and its supramolecular complex with disodium glycylrhizinate as potential drugs for the liver fluke infection caused by *Opisthorchis felinus*. *Pathogens*. 2023;12(6):819. doi 10.3390/pathogens12060819
- Miller A.H., Maletic V., Raison C.L. Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. *Biol Psychiatry*. 2009;65(9):732-741. doi 10.1016/j.biopsych.2008.11.029
- Mishchenko V.A., Goryukhina O.A., Ilyuk R.D. Changes in the blood-brain barrier during experimental cirrhosis of the liver. *Bull Exp Biol Med*. 1993;116:1547-1550. doi 10.1007/BF00785497
- Mordvinov V.A., Furman D.P. The digenea parasite *Opisthorchis felinus*: a target for the discovery and development of novel drugs. *Infect Disord Drug Targets*. 2010;10(5):385-401. doi 10.2174/187152610793180858
- Murano T., Koshimizu H., Hagihara H., Miyakawa T. Transcriptomic immaturity of the hippocampus and prefrontal cortex in patients with alcoholism. *Sci Rep*. 2017;7:44531. doi 10.1038/srep44531
- Norden D.M., Trojanowski P.J., Villanueva E., Navarro E., Godbout J.P. Sequential activation of microglia and astrocyte cytokine expression precedes increased Iba-1 or GFAP immunoreactivity following systemic immune challenge. *Glia*. 2016;64(2):300-316. doi 10.1002/glia.22930
- Nunes P.T., Vedder L.C., Deak T., Savage L.M. A pivotal role for thiamine deficiency in the expression of neuroinflammation markers in models of alcohol-related brain damage. *Alcohol Clin Exp Res*. 2019;43(3):425-438. doi 10.1111/acer.13946
- Pakharukova M.Y., Zapparina O.G., Kovner A.V., Mordvinov V.A. Inhibition of *Opisthorchis felinus* glutathione-dependent prostaglandin synthase by resveratrol correlates with attenuation of cholangiocyte neoplasia in a hamster model of opisthorchiasis. *Int J Parasitol*. 2019;49(12):963-973. doi 10.1016/j.ijpara.2019.07.002
- Pakharukova M.Y., Lishai E.A., Zapparina O., Baginskaya N.V., Hong S.J., Sripa B., Mordvinov V.A. *Opisthorchis viverrini*, *Clonorchis sinensis* and *Opisthorchis felinus* liver flukes affect mammalian host microbiome in a species-specific manner. *Cell Mol Life Sci*. 2023;17(2):e0011111. doi 10.1371/journal.pntd.0011111
- Paylor R., Crawley J.N. Inbred strain differences in prepulse inhibition of the mouse startle response. *Psychopharmacology*. 1997;132:169-180. doi 10.1007/s002130050333
- Petney T.N., Andrews R.H., Saijuntha W., Wenz-Mücke A., Sithithaworn P. The zoonotic, fish-borne liver flukes *Clonorchis sinensis*, *Opisthorchis felinus* and *Opisthorchis viverrini*. *Int J Parasitol*. 2013;43(12-13):1031-1046. doi 10.1016/j.ijpara.2013.07.007
- Porsolt R.D., Bertin A., Jalfre M. Behavioral despair in mice: a primary screening test for antidepressants. *Arch Int Pharmacodyn Ther*. 1977;229:327-336

- Rama Rao K.V., Jayakumar A.R., Norenberg M.D. Brain edema in acute liver failure: mechanisms and concepts. *Metab Brain Dis.* 2014;29(4):927-936. doi 10.1007/s11011-014-9502-y
- Ramos A., Mormède P. Stress and emotionality: a multidimensional and genetic approach. *Neurosci Biobeh Rev.* 1998;22(1):33-57. doi 10.1016/s0149-7634(97)00001-8
- Saijuntha W., Sithithaworn P., Petney T.N., Andrews R.H. Foodborne zoonotic parasites of the family Opisthorchiidae. *Res Vet Sci.* 2021; 135:404-411. doi 10.1016/j.rvsc.2020.10.024
- Saltykova I.V., Petrov V.A., Logacheva M.D., Ivanova P.G., Merzlikin N.V., Sazonov A.E., Ogorodova L.M., Brindley P.J. Biliary microbiota, gallstone disease and infection with *Opisthorchis felineus*. *PLoS Negl Trop Dis.* 2016;10(7):e0004809. doi 10.1371/journal.pntd.0004809
- Seibenhener M.L., Wooten M.C. Use of the open field maze to measure locomotor and anxiety-like behavior in mice. *J Vis Exp.* 2015; (96):e52434. doi 10.3791/52434
- Seki E., Schwabe R.F. Hepatic inflammation and fibrosis: functional links and key pathways. *Hepatology.* 2015;61(3):1066-1079. doi 10.1002/hep.27332
- Silvin A., Ginhoux F. Microglia heterogeneity along a spatio-temporal axis: more questions than answers. *Glia.* 2018;66(10):2045-2057. doi 10.1002/glia.23458
- Simon E., Obst J., Gomez-Nicola D. The evolving dialogue of microglia and neurons in Alzheimer's disease: microglia as necessary transducers of pathology. *Neuroscience.* 2019;405:24-34. doi 10.1016/j.neuroscience.2018.01.059
- Sripa B., Thinkhamrop B., Mairiang E., Laha T., Kaewkes S., Sithithaworn P., Periago M.V., Bhudhisawasdi V., Yonglithipagon P., Mulvenna J., Brindley P.J., Loukas A., Bethony J.M. Elevated plasma IL-6 associates with increased risk of advanced fibrosis and cholangiocarcinoma in individuals infected by *Opisthorchis viverrini*. *PLoS Negl Trop Dis.* 2012;6(5):e1654. doi 10.1371/journal.pntd.0001654
- Stephens A.S., Stephens S.R., Morrison N.A. Internal control genes for quantitative RT-PCR expression analysis in mouse osteoblasts, osteoclasts and macrophages. *BMC Res Notes.* 2011;4:410. doi 10.1186/1756-0500-4-410
- Stickel F., Datz C., Hampe J., Bataller R. Pathophysiology and management of alcoholic liver disease: update 2016. *Gut Liver.* 2017;11(2): 173-188. doi 10.5009/gnl16477
- Tan Y.L., Yuan Y., Tian L. Microglial regional heterogeneity and its role in the brain. *Mol Psychiatry.* 2020;25:351-367. doi 10.1038/s41380-019-0609-8
- Upadhyay R.K. Stem cell therapeutics of acute liver diseases, transplantation, and regeneration. *J Cell Sci Ther.* 2017;8(4):1000275. doi 10.4172/2157-7013
- Wang T., Chou D.Y., Ding J.Y., Fredrickson V., Peng C., Schafer S., Guthikonda M., Kreipke C., Rafols J.A., Ding Y. Reduction of brain edema and expression of aquaporins with acute ethanol treatment after traumatic brain injury. *J Neurosurg.* 2013;118(2):390-396. doi 10.3171/2012.8.JNS12736
- Yang J.Y., Xue X., Tian H., Wang X.X., Dong Y.X., Wang F., Zhao Y.N., Yao X.C., Cui W., Wu C.F. Role of microglia in ethanol-induced neurodegenerative disease: pathological and behavioral dysfunction at different developmental stages. *Pharmacol Ther.* 2014;144(3): 321-337. doi 10.1016/j.pharmthera.2014.07.002
- Yao H., Zhang D., Yu H., Yuan H., Shen H., Lan X., Liu H., Chen X., Meng F., Wu X., Zhang G., Wang X. Gut microbiota regulates chronic ethanol exposure-induced depressive-like behavior through hippocampal NLRP3-mediated neuroinflammation. *Mol Psychiatry.* 2023;28(2):919-930. doi 10.1038/s41380-022-01841-y

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

Received March 22, 2024. Revised July 22, 2024. Accepted September 2, 2024.

doi 10.18699/vjgb-25-12


The lowest chromosome number in the family Pteromalidae (Hymenoptera: Chalcidoidea): the karyotype and other genetic features of *Pachycrepoideus vindemmiae* (Rondani, 1875)

V.E. Gokhman ¹, A.S. Ryabinin ², R.A. Bykov ², Yu.Yu. Ilinsky ^{2, 3}

¹ Russian Entomological Society, Moscow, Russia

² Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

³ Center for Immunology and Cell Biology, Immanuel Kant Baltic Federal University, Kaliningrad, Russia

 vegokhman@hotmail.com

Abstract. Various genetic features of the *hitman* strain of the widespread parasitoid of Drosophilidae (Diptera), *Pachycrepoideus vindemmiae* (Rondani, 1875) (Pteromalidae, Pachyneurinae) were studied. This strain was established and is maintained at the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia). An analysis of air-dried chromosome preparations from prepupae of this parasitoid showed that it has $n = 4$ and $2n = 8$ in males and females, respectively, which is the lowest known chromosome number in the family Pteromalidae. All chromosomes in the karyotype of this species are metacentric. The first and second chromosomes are of similar size, the remaining ones are substantially shorter. The same results were obtained for an additional strain of this species kept at the Moscow State University (Moscow, Russia). A comparison of the DNA sequence of the barcoding region of the mitochondrial cytochrome c oxidase (COI) gene of the *hitman* strain of *P. vindemmiae* with those available from the GenBank and BoLD databases demonstrated that this strain clustered together with conspecifics originating from China, Turkey and Italy. Despite certain endosymbionts being previously reported for the genus *Pachycrepoideus* Ashmead, 1904 as well as for *P. vindemmiae* itself, the *hitman* strain turned out to be free of endosymbiotic bacteria in the genera *Arsenophonus* Gherna et al., 1991, *Cardinium* Zchori-Fein et al., 2004, *Rickettsia* da Rocha-Lima, 1916, *Spiroplasma* Saglio et al., 1973 and *Wolbachia* Hertig, 1936. The above-mentioned results improve our knowledge of various genetic features of parasitoids of the family Pteromalidae and those of *P. vindemmiae* in particular.

Key words: Hymenoptera; Pteromalidae; parasitoids; chromosomes; karyotype; DNA barcoding; endosymbionts.

For citation: Gokhman V.E., Ryabinin A.S., Bykov R.A., Ilinsky Yu.Yu. The lowest chromosome number in the family Pteromalidae (Hymenoptera: Chalcidoidea): the karyotype and other genetic features of *Pachycrepoideus vindemmiae* (Rondani, 1875). *Vavilovskii Zhurnal Genetiki i Selektzii* = *Vavilov J Genet Breed*. 2025;29(1):108-112. doi 10.18699/vjgb-25-12

Acknowledgements. The authors are grateful to Ivan V. Zenin (Department of Evolutionary Theory, Moscow State University) for providing additional living material for the present study, as well as to Ekaterina V. Tselikh (Zoological Institute, Russian Academy of Sciences, St. Petersburg, Russia) for identifying the parasitoids. We thank the Insect Collection of the Federal Research Center "Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences" (supported by the Russian Ministry of Science and Higher Education, project FWNR-2022-0019), which hosts the *hitman* strain of *P. vindemmiae*.


Наименьшее число хромосом в семействе Pteromalidae (Hymenoptera: Chalcidoidea): кариотип и другие генетические особенности *Pachycrepoideus vindemmiae* (Rondani, 1875)

В.Е. Гохман ¹, А.С. Рябинин ², Р.А. Быков ², Ю.Ю. Илинский ^{2, 3}

¹ Русское энтомологическое общество, Москва, Россия

² Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

³ Центр иммунологии и клеточной биологии, Балтийский федеральный университет им. Иммануила Канта, Калининград, Россия

 vegokhman@hotmail.com

Аннотация. Изучены различные генетические особенности культуры *hitman* широко распространенного паразитоида Drosophilidae (Diptera) *Pachycrepoideus vindemmiae* (Rondani, 1875) (Pteromalidae, Pachyneurinae), созданной и поддерживаемой в Институте цитологии и генетики Сибирского отделения Российской академии наук (Новосибирск, Россия). Анализ высушенных на воздухе хромосомных препаратов, полученных из предкуколок, показывает, что число хромосом для самцов и самок этого наездника составляет $n = 4$ и $2n = 8$ соответственно; это наименьшее из известных для семейства Pteromalidae. Все хромосомы в кариотипе данного вида являются метацентриками. Первая и вторая хромосомы близки по размерам, остальные существенно короче. Такие же ре-

зультаты получены еще для одной культуры этого вида, содержащейся в Московском государственном университете (Россия). Сравнение последовательности ДНК баркодинг-участка, т.е. митохондриального гена цитохром с-оксидазы (*COI*) культуры *hitman* *P. vindemmiae*, с информацией, доступной в базах данных GenBank и BoLD, показало, что эта культура кластеризуется с конспецифичными особями, происходящими из Китая, Турции и Италии. Хотя некоторые эндосимбионты ранее указывались как для рода *Pachycrepoideus* Ashmead, 1904, так и для самого *P. vindemmiae*, оказалось, что культура *hitman* свободна от эндосимбиотических бактерий, принадлежащих к родам *Arsenophonus* Gherna et al., 1991, *Cardinium* Zchori-Fein et al., 2004, *Rickettsia* da Rocha-Lima, 1916, *Spiroplasma* Saglio et al., 1973 и *Wolbachia* Hertig, 1936. Вышеприведенные данные пополняют наши знания о различных генетических особенностях наездников семейства Pteromalidae, и в частности *P. vindemmiae*.

Ключевые слова: Hymenoptera; Pteromalidae; наездники; хромосомы; кариотип; ДНК-баркодинг; эндосимбионты.

Introduction

Parasitoid Hymenoptera are one of the most species-rich, taxonomically complicated and economically important groups of insects (Bebber et al., 2014; Forbes et al., 2018). In particular, the superfamily Chalcidoidea, with its exceptionally high morphological and ecological diversity, contains more than 27 thousand known species (Cruaud et al., 2024). Until recently, Pteromalidae represented the second largest family of Chalcidoidea with about four thousand members, but now it is subdivided into several smaller families (Huber, 2017; Burks et al., 2022). Nevertheless, karyotypes of less than twenty species of Pteromalidae s. l. have been studied up to now (Gokhman, 2024), as opposed to about 230 members of other Chalcidoidea (Gokhman, 2009, 2020). Among other Pteromalidae s. str. (hence Pteromalidae), we have recently studied the karyotype of a widespread parasitoid of *Drosophilidae* (Diptera), *Pachycrepoideus vindemmiae* (Rondani, 1875), using routine staining and morphometric analysis of chromosomes.

To ensure the precise identification of this species, which is of considerable interest as an effective agent of biological control (see, e. g., Bezerra Da Silva et al., 2019), we sequenced the barcoding region of the mitochondrial cytochrome c oxidase (*COI*) gene of the same strain. In addition, many chalcids harbor maternally inherited bacterial endosymbionts that can cause various cytogenetic effects, for example, diploid thelytoky (Werren et al., 2008; Gokhman, Kuznetsova, 2018), which, in turn, can promote rapid fixation of chromosomal mutations. Specifically, these endosymbionts belong to the genera *Arsenophonus* Gherna et al., 1991, *Cardinium* Zchori-Fein et al., 2004, *Rickettsia* da Rocha-Lima, 1916, *Spiroplasma* Saglio et al., 1973 and *Wolbachia* Hertig, 1936 (Gavotte et al., 2007; Werren et al., 2008; Duron et al., 2010; Pilgrim et al., 2021; Nadal-Jimenez et al., 2023). Since the karyotype discovered in *P. vindemmiae* turned to be fairly aberrant for Pteromalidae (Gokhman, 2024) (see below), we have therefore conducted an additional study aimed at testing for the presence of various endosymbionts in the above-mentioned strain.

Material and methods

Origin of insects. The *hitman* strain of *P. vindemmiae* has been maintained in the Laboratory of Molecular Genetics of Insects (Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia) since 2018. It is reared on *Drosophila melanogaster* Meigen, 1830 (Diptera, Drosophilidae) under 19–22 °C and 60 ± 10 % humidity. The founder specimens of the strain were

isolated by Dr. Sophia N. Panteleeva (Institute of Systematics and Ecology of Animals of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia) from *D. melanogaster* pupae that were exposed at the Novosibirsk Arboretum in 2018. This strain can also be developed in the laboratory on *Drosophila virilis* Sturtevant, 1916 and *D. mercatorum* Patterson et Wheeler, 1942. For the karyotypic study, a few additional individuals were used from the laboratory stock kept at the Department of Evolutionary Theory (Moscow State University, Russia).

Karyotypic study. Chromosomal preparations were obtained from cerebral ganglia of seven male and four female parasitoid prepupae generally following the protocol developed by Imai et al. (1988) with certain modifications. Ganglia were extracted from insects dissected in 0.5 % hypotonic sodium citrate solution containing 0.005 % colchicine. The extracted ganglia were then transferred to a fresh portion of hypotonic solution and incubated for 30 min at room temperature. The material was transferred onto a pre-cleaned microscope slide using a Pasteur pipette and then gently flushed with Fixative I (glacial acetic acid : absolute ethanol : distilled water 3:3:4). The tissues were disrupted using dissecting needles in an additional drop of Fixative I. A drop of Fixative II (glacial acetic acid : absolute ethanol 1:1) was applied to the center of the area, and the more aqueous phase was blotted off the edges of the slide. The same procedure was performed with Fixative III (glacial acetic acid). The slides were then dried for approximately half an hour and stored at room temperature. Chromosome preparations were stained overnight with freshly prepared 3 % Giemsa solution.

Metaphase plates were analyzed under a Zeiss Axioskop 40 FL epifluorescence microscope (Carl Zeiss, Germany). Images of chromosomes from 21 haploid and 31 diploid mitotic divisions were taken with Zeiss AxioCam 208 digital camera using ZEN software version 3.0. To prepare illustrations, the resulting images were arranged and enhanced with GIMP 2.10. KaryoType software version 2.0 was also used for taking chromosome measurements from five diploid metaphase plates of good quality. The chromosomes were classified following the guidelines provided by Levan et al. (1964).

Molecular study. For barcoding and screening for endosymbionts, DNA was extracted from at least five pooled specimens. During 2018–2024, the *hitman* strain was tested four times for maternally inherited endosymbionts in conjunction with replacements of the host strains.

Parasitoid specimens were homogenized in 200 µl extraction buffer (0.1M NaCl, 10 mM Tris HCl (pH 8.0), 25 mM

Table 1. Primers used in this study

Primers	Target genes	Sequences (5'–3')	References
LCO1490	<i>COI</i>	GGTCAACAATCATAAAGATATTGG	Folmer et al., 1994
HCO2198		TAAACTTCAGGGTGACCAAAAAATCA	
ArsF1	<i>Arsenophonus</i> , 16S rDNA	GGGTGTAAAGTACTTTTCAGTCGT	Duron et al., 2008
ArsR2		GTAGCCCTRCTCGTAAGGGCC	
Car281F	<i>Cardinium</i> , 16S rDNA	GGTAGGGGTTCTTAGTGGAAG	Brown et al., 2018
Car269R		TGCTCCCCACGCTTTCGTG	
gyrB859F	<i>Cardinium</i> , <i>gyrB</i>	ATGCAYGTAACGGGDTTTAAAAG	Tarlachkov et al., 2023
gyrB1498R		CATAATYACAATTTTATGGTAMCG	
glt1	<i>Rickettsia</i> , <i>gltA</i> (1st round)	GATTGCTTTACTTACGACCC	Ilgolkina et al., 2015
glt2		TGCATTTCTTCCATTGTGC	
glt3	<i>Rickettsia</i> , <i>gltA</i> (2nd round)	TATAGACGGTGATAAAGGAATC	
glt4		CAGAACTACCGATTCTTTAAGC	
SpiF1	<i>Spiroplasma</i> , 16S rDNA	GGGTGAGTAACACGTATCT	Sanada-Morimura et al., 2013
SpiR3		CCTTCCTCTAGCTTACACTA	
ftsZunif1	<i>Wolbachia</i> , <i>ftsZ</i> (1st round)	GGYAARGGTGCRGCAGAAGA	Lo et al., 2002
ftsZunif2		ATCRATRCCAGTTGCAAG	
ftsZ_F1	<i>Wolbachia</i> , <i>ftsZ</i> (2nd round)	TYATGGARCATATAAARGATAG	Baldo et al., 2006
ftsZ_R1		TCRAGYAATGGATTTRGATAT	

EDTA, 0.5 % SDS) and incubated at 56 °C for an hour. DNA was then salted out with 100 µl 5M potassium acetate/3M acetic acid with further precipitation and dissolution in 100 µl double-distilled water. All PCRs were carried out in 20 µl mix containing chemicals from the Biomaster HS-Taq PCR kit (Biolabmix, Russia), together with a specific primer set and genomic DNA, with the following cycling conditions: an initial denaturation at 95 °C for 5 min, 35 cycles at 95 °C – 15 sec, 53 °C – 1 min for *COI* and *Spiroplasma*, 55 °C – 30 sec for *Cardinium* and 40 sec for *Arsenophonus* and nested PCRs followed by elongation at 72 °C – 40 sec, and a final elongation at 72 °C for 2 min. The presence of *Rickettsia* and *Wolbachia* was checked by nested PCR, with 15 cycles for the first round and 25 cycles for the second round; 1 µl of the reaction volume from the first round was used in the second one. Primers used in this study are listed in Table 1. The amplicon was purified with exonuclease (ExoI) (New England Biolabs, USA), and sequenced using the BrilliantDye™ Terminator Cycle Sequencing Kit (NimaGen, The Netherlands). The sequence of the *COI* gene was deposited in GenBank under accession number PP727399.

We retrieved all sequences of the barcoding fragment of the *COI* gene deposited under the name of *P. vindemmiae* in the Barcode of Life Database (BoLD) (Ratnasingham, Hebert, 2007) and GenBank. Using BLAST nucleotide search (<https://blast.ncbi.nlm.nih.gov>), we also found *COI* gene sequences for a few other species with the highest similarity with those of *P. vindemmiae*. These sequences were also included into the analysis. The maximum likelihood (ML) phylogenetic tree of the *COI* gene was reconstructed using MEGA6 software (Tamura et al., 2013) under the General Time Reversible model as the best fit and bootstrapping at 1,000 iterations.

Results

The haploid karyotype of *P. vindemmiae* harbors four meta-centric chromosomes ($n = 4$), although the first chromosome is close to a submetacentric one (Fig. 1a, Table 2). The second metacentric chromosome is similar in length to the first one, the remaining chromosomes are distinctly shorter. Conse-



Fig. 1. Karyograms of *P. vindemmiae*: a – male (haploid), b – female (diploid).
Bar = 10 µm.

Table 2. Relative lengths (RLs) and centromeric indices (CIs) of *P. vindemmiae* chromosomes (mean ± SD)

Chromosome no.	RL	CI
1	29.77 ± 0.48	41.07 ± 3.94
2	28.52 ± 0.87	41.76 ± 3.40
3	22.86 ± 0.37	40.94 ± 1.37
4	18.85 ± 0.77	44.11 ± 2.81

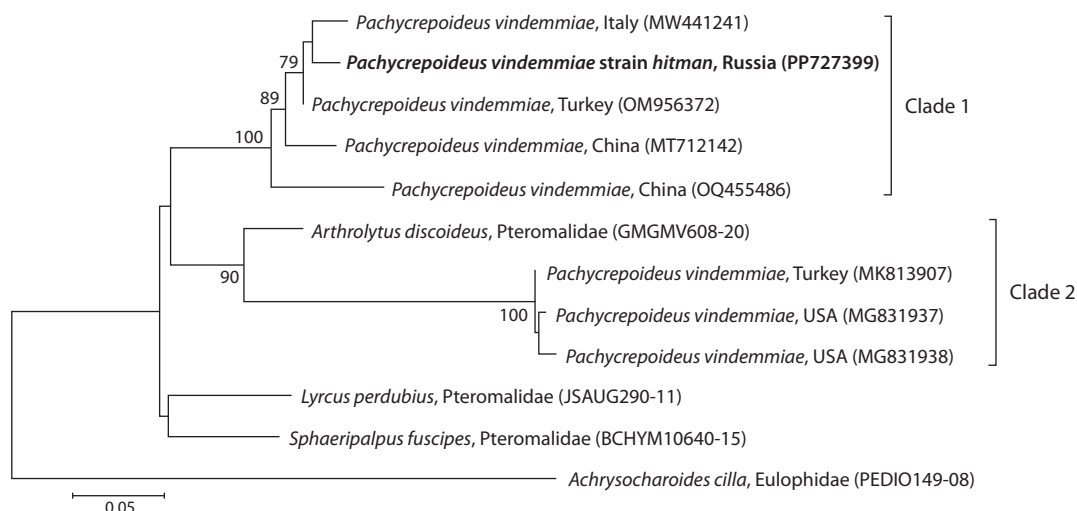


Fig. 2. The maximum likelihood (ML) phylogenetic tree of *Pachycrepoideus* COI mitochondrial DNA sequences (577 bp region) reconstructed with the GTR+G model.

GenBank and/or BoLD accession numbers as well as origins of samples are indicated. The *hitman* strain indicated in bold. COI sequences of *Arthrolytus discoideus* (Nees, 1834) (Pteromalidae), *Achrysocharoides cilla* (Walker, 1839) (Eulophidae), as well as *Lycus perdubius* (Girault, 1916) and *Sphaeripalpus fuscipes* (Walker, 1833) (Pteromalidae) were used as outgroups. Bootstrap values higher than 75 (1,000 iterations) are indicated. The scale bar denotes the number of substitutions per site.

quently, the diploid karyotype of this species contains eight chromosomes ($2n = 8$) (Fig. 1b). No obvious chromosomal difference was detected between the strains from Novosibirsk and Moscow.

We sequenced 652 bp of the mitochondrial COI gene of the *hitman* strain and reconstructed the ML phylogenetic tree, which included all annotated sequences available for *P. vindemmiae* (Fig. 2). There are two clades on the ML phylogenetic tree, in which the *hitman* strain is clustered with conspecifics from China, Turkey and Italy (Clade 1), while another cluster is formed by two strains from the USA as well as by another one from Turkey (Clade 2). However, the latter clade also turned out to include *Arthrolytus discoideus* (Nees, 1834) (Pteromalidae, Pteromalinae). We did not find any molecular evidence for the presence of any checked endosymbiont, i.e., *Arsenophonus*, *Cardinium*, *Rickettsia*, *Spiroplasma* and *Wolbachia*.

Discussion

Among other members of Pteromalidae s. l., $n = 4$ was reported only for *Spalangia endius* Walker, 1839 (Spalangidae) from Thailand (Kitthawee, Vasinpiyamongkol, 2002). However, the same chromosome number found in *P. vindemmiae* represents the lowest n value known for the family Pteromalidae (Gokhman, 2024), with other members of the family having $n = 5-7$. The most frequent chromosome number, which is characteristic of most species of Pteromalidae, is $n = 5$ (Gokhman, 2024). Although it is unclear at present which n value can be considered ancestral for the family, this is almost certainly not $n = 4$, i.e., the chromosome number found in *P. vindemmiae*. Moreover, the latter species and all other Pteromalidae with known karyotypes belong to the subfamilies Pachyneurinae and Pteromalinae respectively (Burks et al., 2022). It is therefore not surprising that *P. vindemmiae* demonstrates deviating chromosomal characters. Moreover, strong behavioral and molecular differences between this

species and many other Pteromalidae were already noted by previous authors (van den Assem, 1974; Huang et al., 2023).

Taking into account the large genetic distance between Clades 1 and 2, we suggest that the latter clade does not actually belong to *P. vindemmiae*. Indeed, according to the available data, studied samples of *P. vindemmiae* that belong to the second clade appear to be more closely related to *Arthrolytus discoideus* than to the strains of Clade 1 of *P. vindemmiae* (Fig. 2). However, the pteromalid genera *Arthrolytus* Thomson, 1878 and *Pachycrepoideus* Ashmead, 1904 belong to different subfamilies (see above), and therefore identifications of these samples of *P. vindemmiae* may well be wrong.

Maternally inherited endosymbionts are constantly transferred to the offspring, and therefore they can be effectively considered facultative components of the host genome. Currently, *Arsenophonus* remains the only endosymbiont genus detected in *P. vindemmiae* (Duron et al., 2010; Nadal-Jimenez et al., 2023). Moreover, *Wolbachia* and *Rickettsia* were also previously reported for the genus *Pachycrepoideus* (Gavotte et al., 2007; Pilgrim et al., 2021). However, the *hitman* strain appears to be free from all these endosymbionts. Since we have recurrently checked this strain for any endosymbiotic microorganisms for a few years starting from 2018 (see above), we can assume that the *hitman* strain was neither initially infected with these endosymbionts nor obtained them via different hosts in the process of rearing.

Conclusions

Chromosome preparations from prepupae of *Pachycrepoideus vindemmiae* showed that it has $n = 4$ and $2n = 8$ in males and females, respectively, which are the lowest known chromosome numbers in the family Pteromalidae. A comparison of the DNA sequence of the COI barcoding region of the *hitman* strain of *P. vindemmiae* demonstrated that this strain clustered together with conspecifics originating from China, Turkey and Italy. The *hitman* strain turned out to be free of endosymbiotic

bacteria in the genera *Arsenophonus*, *Cardinium*, *Rickettsia*, *Spiroplasma* and *Wolbachia*. These results improve our knowledge of various genetic features of parasitoids of the family Pteromalidae and those of *P. vindemmiae* in particular.

References

- Baldo L., Dunning Hotopp J.C., Jolley K.A., Bordenstein S.R., Biber S.A., Choudhury R.R., Hayashi C., Martin M.C.J., Tettelin H., Werren J.H. Multilocus sequence typing system for the endosymbiont. *Appl Environ Microbiol.* 2006;72(11):7098-7110. doi 10.1128/aem.00731-06
- Bebber D.P., Polaszek A., Wood J.R.I., Barker C., Scotland R.W. Taxonomic capacity and author inflation. *New Phytol.* 2014;202:741-742. doi 10.1111/nph.12745
- Bezerra Da Silva C.S., Price B.E., Soohoo-Hui A., Walton V.M. Factors affecting the biology of *Pachycrepoideus vindemmiae* (Hymenoptera: Pteromalidae), a parasitoid of spotted-wing drosophila (*Drosophila suzukii*). *PLoS One.* 2019;14(7). doi 10.1371/journal.pone.0218301
- Brown A.M.V., Wasala S.K., Howe D.K., Peetz A.B., Zasada I.A., Denver D.R. Comparative genomics of *Wolbachia-Cardinium* dual endosymbiosis in a plant-parasitic nematode. *Front Microbiol.* 2018;9. doi 10.3389/fmicb.2018.02482
- Burks R., Mitroiu M.-D., Fusu L., Heraty J.M., Janšta P., Heydon S., Dale-Skey Papilloud N., Peters R.S., Tselikh E.V., Woolley J.B., van Noort S., Baur H., Cruaud A., Darling C., Haas M., Hanson P., Krogmann L., Rasplus J.Y. From hell's heart I stab at thee! A determined approach towards a monophyletic Pteromalidae and reclassification of Chalcidoidea (Hymenoptera). *J Hymenopt Res.* 2022;94:13-88. doi 10.3897/jhr.94.94263
- Cruaud A., Rasplus J.-Y., Zhang J., Burks R., Delvare G., Fusu L., Gurnovsky A., Huber J.T., Janšta P., Mitroiu M.-D., Noyes J.S., van Noort S., Baker A., Böhmová J., Baur H., Blaimer B.B., Brady S.G., Bubeníková K., Chartois M., Copeland R.S., Dale-Skey Papilloud N., Dal Molin A., Dominguez C., Gebiola M., Guerrieri E., Kresslein R.L., Krogmann L., Lemmon E., Murray E.A., Nidelet S., Nieves-Aldrey J.L., Perry R.K., Peters R.S., Polaszek A., Sauné L., Torrén J., Triapitsyn S., Tselikh E.V., Yoder M., Lemmon A.R., Woolley J.B., Heraty J.M. The Chalcidoidea bush of life: evolutionary history of a massive radiation of minute wasps. *Cladistics.* 2024;40:34-63. doi 10.1111/cla.12561
- Duron O., Bouchon D., Boutin S., Bellamy L., Zhou L., Engelstädter J., Hurst G.D. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol.* 2008;6:27. doi 10.1186/1741-7007-6-27
- Duron O., Wilkes T.E., Hurst G.D. Interspecific transmission of a male-killing bacterium on an ecological timescale. *Ecol Lett.* 2010;13(9): 1139-1148. doi 10.1111/j.1461-0248.2010.01502.x
- 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.
- Forbes A.A., Bagley R.K., Beer M.A., Hippee A.C., Widmayer H.A. Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecol.* 2018;18:21. doi 10.1186/s12898-018-0176-x
- Gavotte L., Henr H., Stouthamer R., Charif D., Charlat S., Boulétreau M., Vavre F. A survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol.* 2007;24(2):427-435. doi 10.1093/molbev/msl171
- Gokhman V.E. Karyotypes of Parasitic Hymenoptera. Dordrecht: Springer, 2009. doi 10.1007/978-1-4020-9807-9
- Gokhman V.E. Chromosomes of parasitic wasps of the superfamily Chalcidoidea (Hymenoptera): an overview. *Comp Cytogenet.* 2020; 14:399-416. doi 10.3897/compcytogen.v14i3.56535
- Gokhman V.E. Comparative cytogenetics of the families Pteromalidae and Spalangidae – a review. *Entomol Exp Appl.* 2024;172:467-471. doi 10.1111/eea.13406
- Gokhman V.E., Kuznetsova V.G. Parthenogenesis in Hexapoda: holometabolous insects. *J Zool Syst Evol Res.* 2018;56(1):23-34. doi 10.1111/jzs.12183
- Huang Y., Yang Y., Qi L., Hu H., Rasplus J.-Y., Wang X. Novel gene rearrangement pattern in *Pachycrepoideus vindemmiae* mitochondrial genome: new gene order in Pteromalidae (Hymenoptera: Chalcidoidea). *Animals.* 2023;13(12):1985. doi 10.3390/ani13121985
- Huber J.T. Biodiversity of Hymenoptera. Footitt R.G., Adler P.H. (Eds) Insect Biodiversity: Science and Society. 2nd ed. Oxford: Wiley Blackwell, 2017;419-461
- Igolkina Y.P., Rar V.A., Yakimenko V.V., Malkova M.G., Tancev A.K., Tikunov A.Y., Epikhina T.I., Tikunova N.V. Genetic variability of *Rickettsia* spp. in *Ixodes persulcatus*/*Ixodes trianguliceps* sympatric areas from Western Siberia, Russia: identification of a new *Candidatus Rickettsia* species. *Infect Genet Evol.* 2015;34:88-93. doi 10.1016/j.meegid.2015.07.015
- Imai H.T., Taylor R.W., Crosland M.W.J., Crozier R.H. Modes of spontaneous chromosomal mutation and karyotype evolution in ants with reference to the minimum interaction hypothesis. *Jpn J Genet.* 1988;63:159-185. doi 10.1266/jgg.63.159
- Kitthawee S., Vasinpiyamongkol L. Mitotic karyotype of *Spalangia endius* Walker (Hymenoptera: Pteromalidae), a pupal parasitoid of tephritid flies (Diptera: Tephritidae) in Thailand. *Cytologia.* 2002; 67:435-438. doi 10.1508/cytologia.67.435
- Levan A., Fredga K., Sandberg A.A. Nomenclature for centromeric position on chromosomes. *Hereditas.* 1964;52:201-220. doi 10.1111/j.1601-5223.1964.tb01953.x
- Lo N., Casiraghi M., Salati E., Bazzocchi C., Bandi C. How many *Wolbachia* supergroups exist? *Mol Biol Evol.* 2002;19(3):341-346. doi 10.1093/oxfordjournals.molbev.a004087
- Nadal-Jimenez P., Parratt S.R., Siozios S., Hurst G.D. Isolation, culture and characterization of *Arsenophonus* symbionts from two insect species reveal loss of infectious transmission and extended host range. *Front Microbiol.* 2023;14:1089143. doi 10.3389/fmicb.2023.1089143
- Pilgrim J., Thongpremp P., Davison H.R., Siozios S., Baylis M., Zakharov E.V., Ratnasingham S., deWaard J.R., Macadam C.R., Smith M.A., Hurst G.D. Torix *Rickettsia* are widespread in arthropods and reflect a neglected symbiosis. *GigaScience.* 2021;10(3):giab021. doi 10.1093/gigascience/giab021
- Ratnasingham S., Hebert P.D. BoLD: the barcode of life data system (<http://www.barcodinglife.org>). *Mol Ecol Notes.* 2007;7(3):355-364. doi 10.1111/j.1471-8286.2007.01678.x
- Sanada-Morimura S., Matsumura M., Noda H. Male killing caused by a *Spiroplasma* symbiont in the small brown planthopper, *Laodelphax striatellus*. *J Hered.* 2013;104(6):821-829. doi 10.1093/jhered/est052
- Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725-2729. doi 10.1093/molbev/mst197
- Tarlachkov S.V., Efeykin B.D., Castillo P., Evtushenko L.I., Subbotin S.A. Distribution of bacterial endosymbionts of the *Cardinium* clade in plant-parasitic nematodes. *Int J Mol Sci.* 2023;24(3):ijms24032905. doi 10.3390/ijms24032905
- van den Assem J. Male courtship patterns and female receptivity signal of Pteromalinae (Hym., Pteromalidae), with a consideration of some evolutionary trends and a comment on the taxonomic position of *Pachycrepoideus vindemmiae*. *Neth J Zool.* 1974;24:253-278. doi 10.1163/002829674x00066
- Werren J.H., Baldo L., Clark M.E. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol.* 2008;6(10):741-751. doi 10.1038/nrmicro1969

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


Received June 28, 2024. Revised October 2, 2024. Accepted October 3, 2024.

doi 10.18699/vjgb-25-13

Genetic characteristics of local horse breeds by microsatellite DNA loci

N.V. Blohina , L.A. Khrabrova 

All-Russian Research Institute of Horse Breeding, Divovo, Rybnovsky district, Ryazan region, Russia

 nbloh16@yandex.ru

Abstract. Russia has a significant pedigree diversity of horse breeds with unique gene pools that are well adapted to a wide variety of harsh natural and climatic conditions, are characterized by universal performance and high productive qualities, and are of significant interest to the world horse breeding. Genetic studies of population diversity in horse breeding are very relevant, since many domestic horse breeds are under threat of extinction. Biomaterials (hair, blood, semen) from horses of 15 local breeds bred in the Russian Federation and neighboring countries (CIS) were selected for the research. The sample included 2,193 horses, including: Altaiskaya ($n = 48$), Bashkirsкая ($n = 130$), Buryatskaya ($n = 30$), Vyatskaya ($n = 220$), Zabaikalskaya ($n = 34$), Kyrgyzskaya ($n = 100$), Mezenskaya ($n = 148$), Mugalzarskaya ($n = 109$), Novoaltaiskaya ($n = 514$), Pechorskaya ($n = 31$), Shetland pony ($n = 47$), Priobskaya ($n = 85$), Tuvinskaya ($n = 600$), Khakas-skaya ($n = 47$) and Yakutskaya ($n = 50$) breeds. The following indicators were used in the genetic and population analysis: the total number of allele variants (N_a) in 17 microsatellite loci, the level of polymorphism (A_e), the average number of alleles per locus (N_v), observed (H_o) and expected (H_e) heterozygosity, coefficients of genetic similarity and genetic distances, as well as the coefficient of intrapopulation inbreeding (F_{is}). Modern local horse breeds, even relatively small in number, have a high level of biodiversity and a peculiar genetic structure, often with the presence of private alleles, which persists despite periodic crossing with stud breeds of different specializations. It was found that horses of local breeds possess a number of unique alleles, including *ASB2T*, *HMS7S*, *HMS6J*, *HMS6H*, *HMS2T*, *HMS1O*, *HTG7L*, *HTG6L*, *HTG6H*, *VHL20S*, *ASB17Z*, *ASB17X*, *ASB17U*, *LEX3S*, *LEX3R* and *CA425E*, which were not detected in representatives of stud breeds in the studied European populations. The majority of the studied breeds were characterized by a negative F_{is} value and the absence of inbreeding. The coefficients of genetic similarity of local breeds varied in a relatively wide range (0.828–0.973) and testified to the uniqueness of the gene pools of most local horse breeds of the Russian Federation, as well as confirmed the common origin of the Kyrgyzskaya horse with the horse populations of Southern Siberia.

Key words: *Equus caballus*; genetic diversity; DNA microsatellites; local breeds; horse.

For citation: Blohina N.V., Khrabrova L.A. Genetic characteristics of local horse breeds by microsatellite DNA loci. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025;29(1):113-121. doi 10.18699/vjgb-25-13

Funding. The work was financially supported by the Ministry of Science and Higher Education of the Russian Federation, grant No. 075-15-2021-1037 (internal No. 15.BRK.21.0001).

Генетическая характеристика лошадей местных пород по локусам микросателлитов ДНК

Н.В. Блохина , Л.А. Храброва 

Всероссийский научно-исследовательский институт коневодства, пос. Дивово, Рыбновский район, Рязанская область, Россия

 nbloh16@yandex.ru

Аннотация. Россия обладает значительным породным разнообразием конских пород с уникальными генофондами, которые хорошо адаптированы к самым разнообразным суровым природно-климатическим условиям, характеризуются универсальной работоспособностью, высокими продуктивными качествами и представляют существенный интерес для мирового коневодства. Генетические исследования популяционного разнообразия в коневодстве являются весьма актуальными, так как многие отечественные породы лошадей находятся под угрозой исчезновения. Для исследований были отобраны биоматериалы (волосы, кровь, сперма) от лошадей 15 местных пород, разводимых в России и сопредельных странах. Выборка включала 2193 лошади алтайской ($n = 48$), башкирской ($n = 130$), бурятской ($n = 30$), вятской ($n = 220$), забайкальской ($n = 34$), кыргызской ($n = 100$), мезенской ($n = 148$), мугалжарской ($n = 109$), новоалтайской ($n = 514$), печорской ($n = 31$), шетлендских пони

($n = 47$), приобской ($n = 85$), тувинской ($n = 600$), хакасской ($n = 47$) и якутской ($n = 50$) пород. При проведении генетико-популяционного анализа использовали следующие показатели: общее количество вариантов аллелей (N_a) в 17 микросателлитных локусах, уровень полиморфности (A_e), среднее число аллелей на локус (N_v), наблюдаемая (H_o) и ожидаемая (H_e) гетерозиготность, коэффициенты генетического сходства и генетических дистанций, коэффициент внутрипопуляционного инбридинга (F_{is}). Современные местные породы лошадей даже при сравнительно небольшой численности имеют высокий уровень биоразнообразия и своеобразную генетическую структуру, часто с наличием приватных аллелей, которая сохраняется несмотря на периодическое скрещивание с заводскими породами разной специализации. Установлено, что лошади местных пород обладают рядом уникальных аллелей, включая *ASB2T*, *HMS7S*, *HMS6J*, *HMS6H*, *HMS2T*, *HMS1O*, *HTG7L*, *HTG6L*, *HTG6H*, *VHL20S*, *ASB17Z*, *ASB17X*, *ASB17U*, *LEX3S*, *LEX3R* и *CA425E*, которые не выявлены у представителей заводских пород в изученных европейских популяциях. Для большинства изученных пород было характерно отрицательное значение F_{is} и отсутствие внутривидового инбридинга. Коэффициенты генетического сходства местных пород менялись в сравнительно широком диапазоне (0.828–0.973) и свидетельствовали об уникальности генофондов большинства местных конских пород в России, а также подтверждали общность происхождения кыргызской лошади с популяциями лошадей Южной Сибири.

Ключевые слова: *Equus caballus*; генетическое разнообразие; микросателлиты ДНК; местные породы; лошадь.

Introduction

Until the beginning of the last century, horse breeding occupied a special place in the economy and agricultural production not only in Russia, but also in many other countries. The horse was not only a symbol of the power and prestige of the country, but also saved peoples in difficult times of history. However, in the middle of the 20th century, with the development of mechanization in agriculture, transport, army and industry, the approach to the use of horses changed, and the number of horses decreased sharply, but then stabilized and even began to grow in many countries. Now horses play an important role in tourism, sports, racing business, as well as in food production (milk, koumiss, meat). The observed global trend of increasing horse meat production (Askarov et al., 2020) is explained by the high dietary properties of horse meat. Horse meat is easily digested and contains practically no allergenic amino acids; therefore, it is considered a valuable dietary product (Stanislawczyk et al., 2020).

World practice shows that if the expenses on breeding work are reduced, the productivity of livestock decreases, and ultimately, animal breeding becomes unprofitable. The assessment of genetic diversity within individual breed populations and entire breeds makes it possible to create and improve both breeding plans and programs for the conservation of the gene pool of these breeds (Marzanov et al., 2010; Kalashnikova et al., 2022). In our country, local horse breeds represent more than 50 % of the total horse population, and their breeding provides employment, livestock production, preservation of cultural traditions and the development of new territories.

Currently, 47 horse breeds are included in the State Register of Breeding Achievements of the Russian Federation, including 20 local and productive breeds created on their basis, which are mainly distributed in regions with harsh climatic conditions (Fig. 1).

The evolution of local horse breeds took place mainly under the influence of natural selection using methods of native breeding, which resulted in high adaptability to environmental conditions. To date, a significant part of the native breeds have been systematically improved by crossing with various stud breeds due to the need to produce a larger horse for ag-

riculture and increase horse meat production. Nowadays, the value of local breeds is increasing, as they are the basis of a productive horse breeding industry (Askarov et al., 2020). As a result of crossing horses of local breeds with stud, mainly draft breeds, it was possible to increase the size of native horses while maintaining their type and economically useful qualities. However, since the end of the last century to the present, there has been no purposeful breeding work with many local breeds, breeding records have been established almost only in the Mezenskaya, Vyatskaya, Bashkirskaia and Kalmykskaya breeds.

Uncontrolled crossing leads to a change in the type and pure original aboriginal forms of horses of most local breeds, which in the future may lead to the disappearance of the breeds themselves. Out of the 40 native horse breeds described by hippologists at the turn of the 19–20th centuries, only 16 have actually survived to the present day. At the same time, there is virtually no information about the current state of the Pechorskaya, Chumyshskaya, Kuznetskaya and Verkhneyniseiskaya breeds included in the State Register of Breeding Achievements of the Russian Federation (Belousova, 2018).

Table 1 provides information on the number of horses tested and the number of mares of the studied local breeds. It follows from the above data that critically low numbers of livestock are observed in Vyatskaya, Mezenskaya, Pechorskaya and Priobskaya breeds, which are in urgent need of conservation measures. Therefore, comprehensive programs are needed to preserve small-numbered horse breeds, which are a national achievement of Russia, including genetic research for study and evaluation of the valuable domestic gene pool.

Today, DNA technologies are widely used in the control of animal origin, the study of phylogenetic relationships and microevolution of breeds, in order to improve the genotypic assessment of animals at individual and population levels, as well as in diverse and genomic studies (Marzanov et al., 2010; Roh et al., 2020; Nwachukwu et al., 2022; Pozharskiy et al., 2023).

In recent years, microsatellites have been the most popular markers in studies of the genetic characteristics of farm animals (Ernst, Zinovieva, 2008; Glazko et al., 2023), they are



Fig. 1. Distribution of horses of local breeds in the territory of the Russian Federation.

Table 1. The total number of livestock and the number of samples taken from horses of local breeds

Breed	Number of samples taken	Total number of mares
Altayskaya	48	500*
Bashkirskaya	130	2,400*
Buryatskaya	30	1,000*
Vyatskaya	220	230*
Zabaikalskaya	34	1,500*
Kyrgyzskaya	100	64,500
Mezenskaya	148	171*
Mugalzharskaya	109	5,200
Novoaltaiskaya	514	3,245
Pechorskaya	31	100*
Shetland pony	47	396
Priobskaya	85	200*
Tuvinskaya	600	1,200*
Khakasskaya	47	5,500*
Yakutskaya	50	51,800*

* Data: Belousova N.F. (2018).

easy to use and have a high degree of information. Microsatellites are interesting because they are subject to a higher level of mutation than the rest of the genome (Glazko et al., 2023). Studies by many scientists researching the genetic structure of horses of different specializations using DNA markers

(Kalashnikov et al., 2011; Blohina et al., 2018; Gavrilicheva, 2019; Khaudov et al., 2019) have shown a high level of allelic variability in the studied populations and confirmed the presence of genetic specificity of the allelofund of most horse breeds, often with a limited breeding area (Vdovina, Yurieva, 2021; Khrabrova et al., 2022). Microsatellites are an effective tool for studying the features of the gene pool, genetic polymorphism, phylogeny, and obtaining data on the formation and evolution of animals (R2D2 Consortium et al., 2021).

The purpose of our study was to investigate the allele pool and genetic diversity of STR loci in the genomes of horses of 15 native breeds bred in the territory of the Russian Federation and CIS countries, as well as to study their phylogenetic relationships.

Materials and methods

The materials for these studies were selected from representatives of local breeds living on the territory of the Russian Federation. The study included 2,193 horses, including: Altayskaya ($n = 48$), Bashkirskaya ($n = 130$), Buryatskaya ($n = 30$), Vyatskaya ($n = 220$), Zabaikalskaya ($n = 34$), Kyrgyzskaya ($n = 100$), Mezenskaya ($n = 148$), Mugalzharskaya ($n = 109$), Novoaltaiskaya ($n = 514$), Pechorskaya ($n = 31$), Shetland pony ($n = 47$), Priobskaya ($n = 85$), Tuvinskaya ($n = 600$), Khakasskaya ($n = 47$) and Yakutskaya ($n = 50$) breeds.

The studies were conducted in the certified laboratory of genetics of the Federal State Budgetary Scientific Institution All-Russian Research Institute for Horse Breeding for 17 STR DNA loci: *HMS2*, *HMS3*, *HMS1*, *AHT4*, *VHL20*, *AHT5*, *HTG7*, *HTG6*, *HTG4*, *HTG10*, *HMS7*, *HMS6*, *ASB23*, *ASB2*, *ASB17*, *LEX3* and *CA425*, using standardized techniques recommended by ISAG.

DNA isolation from biomaterials (hair, blood, sperm, etc.) was carried out using CoRDIS SPRINT reagents (Russia). Amplification of the obtained DNA was performed using a 17-plex set of primers for genotyping horses of domestic production CoRDIS Horse (Russia). The separation and detection of amplification products were carried out by capillary electrophoresis on an automatic 4-capillary genetic analyzer NANA FOR 05 (Russia). After recording the electrophoresis data using the GeneMapper™ V.4.0 program, the sizes of the amplified DNA fragments were calculated. The interpretation of the results was carried out using a control DNA profile with a known genotype and data from international comparative tests (Horse Comparison Tests) conducted by ISAG in 2008–2020. An international alphabetic code was used to designate alleles. The analysis of the genetic and population parameters of the breeds was carried out and graphically visualized in the program R Studio 1.3.1093 (Francis, 2017), R package “diveRcity”, using the package “POPHELPER”. The expected (*Ho*) and observed (*He*) heterozygosity values were calculated using the PLINK 1.9 software packages (Chang et al., 2015); MS Excel 2010, Statistics 12 (<https://statsoft-statistica.ru/>) and GenAlEx (ver.6.5.1) (<https://biology-assets.anu.edu.au/GenAlEx/Download.html>) were also used in the calculations. The phylogenetic tree was constructed using the Neighbor-Net algorithm using the SplitsTree4 4.14.5

program (https://www.advanceduninstaller.com/SplitsTree4-4_14_5-72c0418345e4a971ba5b353bfae970d6-application.htm).

When characterizing the breeds, the following indicators were calculated: the total number of alleles in 17 STR loci (*Na*), the average number of alleles per locus (*Nv*), the level of polymorphism (*Ae*), observed (*Ho*) and expected (*He*) heterozygosity, the coefficient of intrapopulation inbreeding (*Fis*), genetic kinship and genetic distances.

Results

Genotyping of 2,193 horses of 15 local breeds at 17 STR loci revealed 521 alleles with large fluctuations in loci from 3 for *HTG6* (Pechorskaya) to 21 for *ASB17* (Tuvinskaya).

A comparative analysis of the generalizing indicators shows that the richest spectrum of alleles was recorded in horses of the Tuvinskaya (170), Novoaltaiskaya (158) and Mugalzarskaya (154) breeds, while horses of the Buryatskaya breed had the minimum number of alleles, which was 117 (Table 2).

The Tuvinskaya breed was the best in terms of polymorphism (*Ae* = 5.186). Horses of the Kyrgyzskaya (*Ae* = 5.080) and Mugalzarskaya (*Ae* = 5.051) breeds had similar indicators of this value. The lowest level of polymorphism was recorded in horses of the Khakasskaya (*Ae* = 4.299) and

Table 2. Characteristics of horses of native breeds (*n* = 2,193) according to 17 STR DNA markers

Breed	<i>n</i>	<i>Na</i>	<i>Nv</i>	<i>Ae</i>	<i>Ho</i>	<i>He</i>	<i>Fis</i>	<i>HWE</i>
Altaiskaya	48	134	7.880	4.761	0.784±0.025	0.759±0.024	−0.038	***
Bashkirskaya	130	153	9.001	4.837	0.778±0.023	0.774±0.018	−0.005	ns
Buryatskaya	30	117	6.883	4.420	0.740±0.025	0.750±0.017	0.012	***
Vyatskaya	220	129	7.588	4.101	0.738±0.028	0.723±0.027	−0.022	ns
Zabaikalskaya	34	129	7.588	4.579	0.747±0.024	0.763±0.020	0.015	**
Kyrgyzskaya	100	153	9.000	5.080	0.766±0.020	0.781±0.019	0.016	**
Mezenskaya	148	132	7.764	4.197	0.736±0.033	0.723±0.032	−0.021	ns
Mugalzarskaya	109	154	9.060	5.051	0.801±0.022	0.781±0.019	−0.025	ns
Novoaltaiskaya	514	158	9.294	4.923	0.752±0.031	0.757±0.029	0.006	ns
Pechorskaya	31	121	7.117	4.580	0.755±0.035	0.755±0.023	−0.006	**
Shetland pony	47	131	7.705	4.337	0.714±0.024	0.748±0.020	0.046	***
Priobskaya	85	147	8.647	4.661	0.702±0.026	0.752±0.027	0.058	**
Tuvinskaya	600	170	10.000	5.186	0.754±0.025	0.782±0.021	0.028	ns
Khakasskaya	47	123	7.235	4.299	0.729±0.031	0.745±0.020	0.025	ns
Yakutskaya	50	128	7.529	4.307	0.693±0.032	0.740±0.025	0.055	**
Russian Heavy Draft*	615	138	8.117	3.667	0.674±0.030	0.689±0.030	0.015	ns
Soviet Heavy Draft*	286	129	7.587	3.871	0.694±0.043	0.697±0.042	0.004	ns

Note. *n* – number of horses; *Na* – total number of alleles in 17 microsatellite loci; *Ae* – level of polymorphism; *Ho* – observed heterozygosity; *He* – expected heterozygosity; *Fis* – coefficient of intrapopulation inbreeding; *Nv* – average number of alleles per locus; *HWE* – deviation from the Hardy–Weinberg equilibrium (ns – not significant, ** – significant at *p* < 0.001, *** – significant at *p* < 0.0001), * – horse breeds participating as improvers of local horses.

Yakutskaya ($A_e = 4.307$) breeds. The highest indicators of actual heterozygosity were determined in horses of the Mugalzarskaya ($H_o = 0.801$), Altaiskaya ($H_o = 0.784$), Bashkirskaya ($H_o = 0.778$) and Kyrgyzskaya ($H_o = 0.766$) breeds, the lowest degree was found in Yakutskaya ($H_o = 0.693$) horses.

The predominance of heterozygous genotypes and the absence of intrapopulation inbreeding were observed in horses of the Altaiskaya, Bashkirskaya, Vyatskaya, Mezenskaya and Mugalzarskaya breeds. A slight shift in the genetic balance towards an excess of homozygotes according to the *Fis* coefficient was noted in horses of the Zabaikalskaya, Kyrgyzskaya, Novoaltaiskaya, Priobskaya, Tuvinskaya, Khakasskaya and Yakutskaya breeds. In most subpopulations and breeds of horses, we observed deviations from the Hardy–Weinberg equilibrium, significant at $p < 0.001$ or $p < 0.0001$ for all studied loci. The Hardy–Weinberg equilibrium was observed in horses of Bashkirskaya, Vyatskaya, Mezenskaya, Mugalzarskaya, Novoaltaiskaya, Tuvinskaya, and Khakasskaya breeds ($p > 0.05$).

The results of the analysis of the genetic structure of horses of native breeds demonstrate that each of the analyzed groups differs in the spectrum, frequency of occurrence and set of alleles. It should be noted that a comparative analysis of 17 STR loci in horses of local breeds revealed 16 new alleles that were missing from the standardized ISAG nomenclature (Van de Goor et al., 2010), namely alleles: *HMS7S*, *HMS6J*, *HMS6H*, *HMS2T*, *HMS1O*, *ASB2T*, *HTG7L*, *HTG6L*, *HTG6H*, *ASB17Z*, *ASB17X*, *ASB17U*, *VHL20S*, *LEX3S*, *LEX3R* and *CA425E*.

In horses of the Mezenskaya breed bred in the Arkhangelsk region, five unique alleles were found at once: *HMS6J* (0.003), *ASB17Y* (0.019), *ASB17X* (0.010), *LEX3S* (0.039) and *LEX3R* (0.016). The alleles *HMS7L* (0.685), *HMS3M* (0.432), *AHT4O* (0.417), *HTG7O* (0.437), *HTG7K* (0.425), *HTG6O* (0.799), *HTG4M* (0.419) and *LEX3M* (0.535) had the highest concentration in this northern forest breed.

In the genetic structure of the Bashkirskaya breed, there is a high frequency of occurrence of certain alleles (*HTG10O* – 0.447, *HTG6O* – 0.508, *HTG4M* – 0.589 and *HMS7L* – 0.487) and the presence of a rare allele *ASB17U* (0.041), found in horses of Tuvinskaya breed.

Two alleles turned out to be unique alleles for horses of the Vyatskaya breed *HTG6L* (0.004) and *AHT5P* (0.009), and the typical ones were *HMS7L* (0.470), *HMS2H* (0.457), *HMS1M* (0.468), *AHT5J* (0.427), *HTG7* (0.576), *HTG6O* (0.712) and *HTG4M* (0.689).

The local horse breeds of Siberia differed markedly from European populations in all genetic parameters. High values of all basic population parameters were determined in horses of the Kyrgyzskaya and Tuvinskaya breeds that previously had contiguous ranges. High allele frequencies were also identified in horses of these breeds: *HTG6O* (0.548; 0.530), *HTG4M* (0.632; 0.597), *HMS7L* (0.420; 0.400). The Tuvinskaya breed rarely had alleles *HMS6H* (0.001), *HMS3L* (0.004), *VHL20K* (0.002), *ASB23N* (0.001), *ASB17Z* (0.002), *LEX3J* (0.054) and *LEX3I* (0.024). Two unique alleles were identified in the Kyrgyzskaya horse breed, *HTG4J* (0.005) and *HTG7Q* (0.009), absent from other studied breeds within the studied groups.

The *HTG6R* allele was detected only in horses of four breeds: Kyrgyzskaya (0.005), Mugalzarskaya (0.007), Pri-

obskaya (0.023) and Tuvinskaya (0.006). At the *VHL20* locus, the *S* allele was found in horses of the Mugalzarskaya (0.046), Novoaltaiskaya (0.005) and Tuvinskaya (0.002) breeds.

In the *LEX3* locus localized on the X chromosome, 12 alleles were identified in local horses, three of which (*F*, *L*, *M*) were found in all the studied breeds (Fig. 2).

Representatives of the Tuvinskaya and Altaiskaya breeds are characterized by the presence of a rare allele *LEX3J* (0.019; 0.056), which is absent in other groups studied in this work. Only in horses of local breeds bred in Southern and Western Siberia, such as Kyrgyzskaya, Novoaltaiskaya, Mugalzarskaya and Tuvinskaya, the *LEX3G* allele was found, which occurred with low frequency.

Rare alleles have been identified in the genotypes of horses of the Mezenskaya breed, *LEX3S* (0.045) and *LEX3R* (0.017), missing from other populations (Fig. 2).

Horses of local breeds were characterized by the presence of separate alleles at the *CA425* locus (*I*, *J*, *L*, *M* and *N*). The maximum frequency of occurrence of the *CA425M* allele was detected in the Tuvinskaya and Khakasskaya populations, and the *CA425N* allele was determined in horses of the Buryatskaya (0.511), Altaiskaya (0.463), Mugalzarskaya (0.408) and Novoaltaiskaya (0.435) breeds. The new allele *CA425E* was found only in horses of the Mugalzarskaya breed bred in Kazakhstan (0.009). The unique *CA425P* allele was identified in Shetland ponies, Bashkirskaya and Khakasskaya horses, and was absent in other breeds we studied.

The indicators of the level of polymorphism and the degree of heterozygosity in local breeds were high at low *Fis* values, which indicates a genetic balance in the studied populations. Based on the results obtained, it can be noted that the highest indicators of genetic diversity were found in horses of the Tuvinskaya breed, in comparison with other breeds analyzed in the framework of the presented study. And the lowest level of genetic diversity is observed in horses of the Vyatskaya breed.

The coefficients of genetic kinship between local horse breeds varied in the range of 0.828–0.973 (Table 3). The highest coefficients of genetic relationship were determined in horses of the Kyrgyzskaya breed with the Tuvinskaya (0.973), Bashkirskaya (0.939), Altaiskaya (0.938), Zabaikalskaya (0.934) and Khakasskaya (0.926) breeds. The lowest level of genetic relationship was found with Shetland ponies. Genetic differences between the studied horse breeds by microsatellite markers confirm the values of genetic distances, which varied in the range of 0.027–0.331. Horses of the Tuvinskaya and Kyrgyzskaya breeds have the closest genetic distances (0.027).

On the phylogenetic tree, the studied horse breeds form three independent clusters (Fig. 3). The first cluster includes horses of the Kyrgyzskaya and Tuvinskaya breeds, characterized by a common origin, which is adequately consistent with history. Ethnic groups actively roamed on horseback throughout Eastern, Western and Central Siberia, hence the genetic relationship of the populations.

The second cluster is formed by horses of the Bashkirskaya, Khakasskaya, Mugalzarskaya and Orlov Trotters breeds. At first glance, the inclusion of the Orlov Trotters in the group of local steppe breeds looks somewhat unexpected, but most likely it is due to the periodic use of this trotter breed to improve the economically useful qualities of horses of local

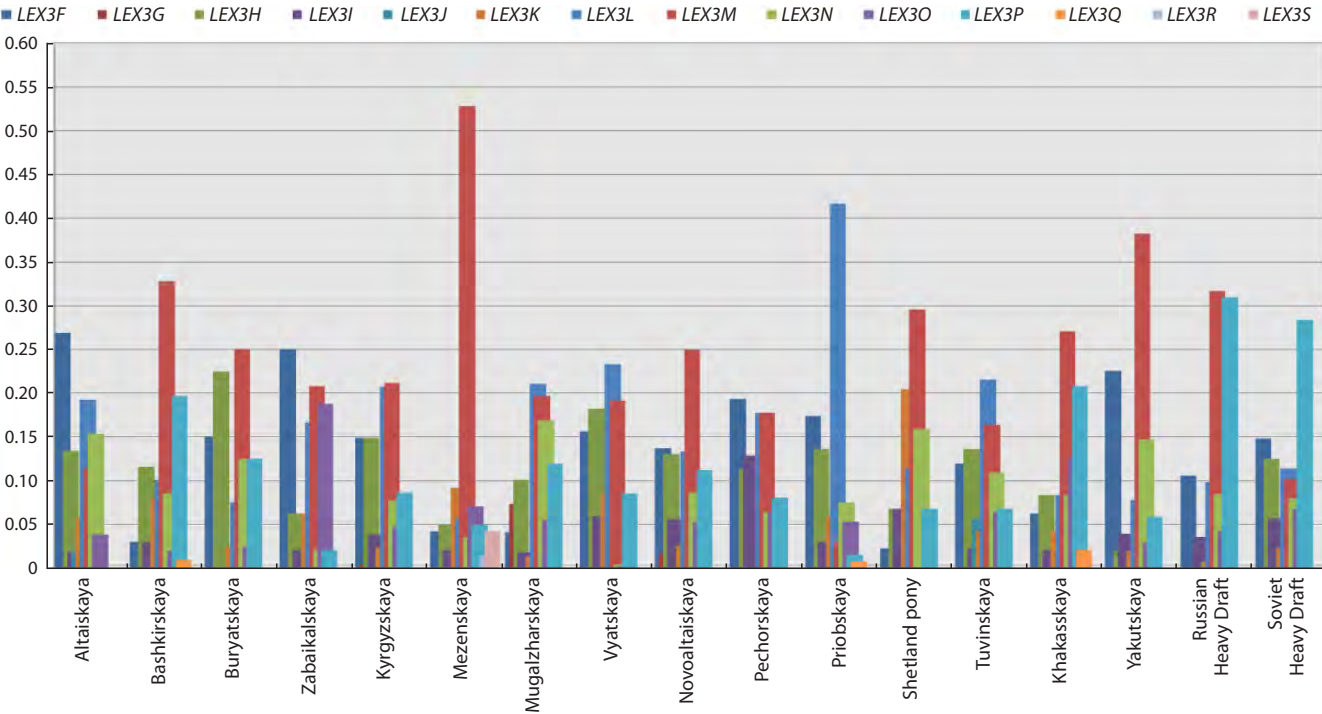


Fig. 2. Histogram of allele frequencies at the *LEX3* locus in horses of local breeds.

Table 3. Coefficients of genetic kinship and genetic distances in horses of local breeds

Genetic distance	Genetic kinship															
	Altaiskaya	Bashkirskaya	Buryatskaya	Zabaikalskaya	Kyrgyzskaya	Mezenskaya	Mugalzarskaya	Vyatskaya	Novolaitkaya	Pechorskaya	Priobskaya	Shetland pony	Tuvinskaya	Khakasskaya	Yakutskaya	Orlov Trotter
Altaiskaya	x	0.886	0.862	0.911	0.938	0.805	0.878	0.872	0.901	0.866	0.852	0.777	0.924	0.870	0.844	0.829
Bashkirskaya	0.114	x	0.869	0.903	0.939	0.840	0.890	0.817	0.886	0.856	0.841	0.814	0.945	0.915	0.864	0.800
Buryatskaya	0.138	0.131	x	0.872	0.914	0.836	0.851	0.821	0.883	0.831	0.810	0.779	0.887	0.865	0.833	0.744
Zabaikalskaya	0.089	0.097	0.128	x	0.934	0.816	0.876	0.842	0.899	0.858	0.874	0.789	0.929	0.874	0.865	0.779
Kyrgyzskaya	0.062	0.061	0.086	0.066	x	0.850	0.909	0.864	0.916	0.888	0.871	0.828	0.973	0.926	0.871	0.833
Mezenskaya	0.195	0.160	0.164	0.184	0.150	x	0.760	0.781	0.838	0.831	0.777	0.805	0.856	0.823	0.799	0.701
Mugalzarskaya	0.122	0.110	0.149	0.124	0.091	0.240	x	0.782	0.885	0.787	0.819	0.768	0.913	0.867	0.855	0.788
Vyatskaya	0.128	0.183	0.179	0.158	0.136	0.219	0.218	x	0.873	0.816	0.829	0.711	0.844	0.822	0.775	0.732
Novolaitkaya	0.099	0.114	0.117	0.101	0.084	0.162	0.115	0.127	x	0.879	0.888	0.775	0.911	0.914	0.846	0.820
Pechorskaya	0.134	0.144	0.169	0.142	0.112	0.169	0.213	0.184	0.121	x	0.820	0.811	0.889	0.844	0.833	0.748
Priobskaya	0.148	0.159	0.190	0.126	0.129	0.223	0.181	0.171	0.112	0.180	x	0.752	0.881	0.829	0.789	0.768
Shetland pony	0.223	0.186	0.221	0.211	0.172	0.195	0.232	0.289	0.225	0.189	0.248	x	0.851	0.767	0.806	0.669
Tuvinskaya	0.076	0.055	0.113	0.071	0.027	0.144	0.087	0.156	0.089	0.111	0.119	0.149	x	0.915	0.894	0.818
Khakasskaya	0.130	0.085	0.135	0.126	0.074	0.177	0.133	0.178	0.086	0.156	0.171	0.233	0.085	x	0.820	0.847
Yakutskaya	0.156	0.136	0.167	0.135	0.129	0.201	0.145	0.225	0.154	0.167	0.211	0.194	0.106	0.180	x	0.701
Orlov Trotter	0.171	0.200	0.256	0.221	0.167	0.299	0.212	0.268	0.180	0.252	0.232	0.331	0.182	0.153	0.299	x

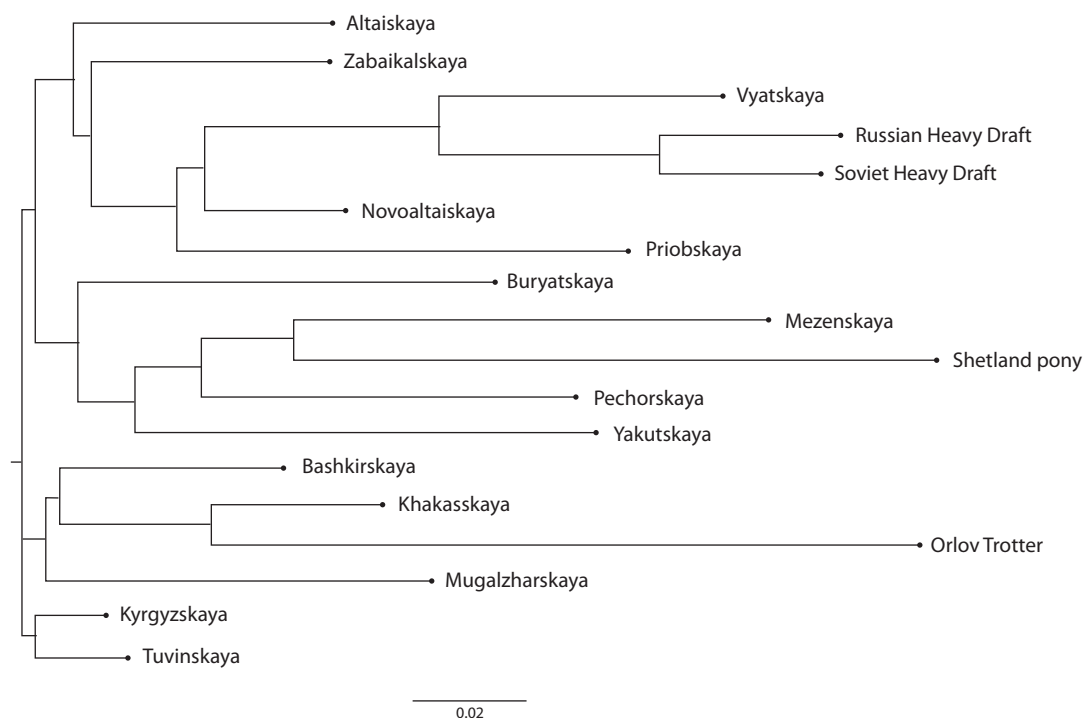


Fig. 3. Dendrogram of genetic distances between native horse breeds, constructed using the Neighbor-Net method.

populations. In this subcluster, a branch of the geographically isolated Mugalzarskaya breed stands out somewhat, possessing a number of unique alleles.

The most extensive third cluster includes most of the studied aboriginal horse breeds of the forest and forest-steppe zone, as well as breeds of domestic draft horses previously used as improvers of local livestock. The dendrogram of genetic distances clearly reflects the noticeable influence of stallions of the Russian draft breed on the formation of the Vyatskaya breed over the past decades of its development.

Thus, based on the results obtained, it can be concluded that the formation of genetic profiles of aboriginal breeds, their levels of polymorphism, differences in the structure of loci, genetic and population characteristics, as well as commonalities, are influenced by two groups of significant factors of different vectors: the natural geographical isolation of animals and origin from common ancestors. Our molecular genetic analysis of 15 native breeds showed that all animal groups were characterized by high genetic diversity.

Discussion

Russia has unique genetic resources of horse breeding, the study of biological characteristics of which is the basis for programs for their conservation and improvement.

In the course of the conducted studies in horses of native breeds, high values of the polymorphism level of STR markers were determined in almost all breeds ($A_e = 4.101-5.186$) analyzed in the framework of the presented study. The data obtained indicate a complex and diverse system of crossing and breeding in the studied groups, as well as the presence of genetic diversity associated with the adaptive qualities of

horses and their ability to adapt to extreme environmental conditions in an evolutionary context.

In addition to the standardized nomenclature (Van de Goor et al., 2010), 16 new alleles were found in horses of local breeds, which could have remained in the centers of domestication of ancient horses in the territory, as well as appeared as a result of genomic mutations or the introduction of genes with horses of nomads from different regions of Asia. Scientists from China (Ling et al., 2011) confirm this fact with their own studies of local Chinese horses, in which a wide range of alleles of microsatellite markers of oriental origin has been identified.

The analysis revealed significant differences in the main genetic parameters (A_e , N_v , H_o , H_e , F_{is}). In addition to the high degree of genetic variability, a characteristic feature of local breeds was the presence of a number of alleles unique for domestic breeds (*ASB2T*, *HMS7S*, *HMS6J*, *HMS6H*, *HMS2T*, *HMS1O*, *HTG7L*, *HTG6L*, *HTG6H*, *VHL20S*, *ASB17Z*, *ASB17X*, *ASB17U*, *LEX3S*, *LEX3R* and *CA425E*), which were not found in horses of stud breeds and in the studied European populations (Seo et al., 2016; Baena et al., 2020).

Horses of native breeds have unique alleles: Tuvinskaya – *HTG6L*, *VHL20S*, *HMS6H*, *ASB17X*, *ASB17U*, *ASB17Z*; Bashkirskaya – *ASB17U*; Altaiskaya – *HMS2T*; Buryatskaya – *HTG6L*, *HTG6H*; Vyatskaya – *AHT5P*, *HTG6L*; Mezenskaya – *ASB17Y*, *ASB17X*, *HMS6J*, *LEX3R*, *LEX3S*. Modern aboriginal horse breeds, even with a common origin from Mongolian roots (Yun et al., 2022), have their own characteristic genetic structure with the presence of private alleles, despite periodic crossing with stud breeds of riding, trotting and draft directions.

Our results confirm the published data of foreign scientists (Lippold et al., 2011; Librado et al., 2021) regarding the area of horse domestication having occupied a significant part of modern Russia, which, due to its geographical location, was a historical crossroads of the routes of many nomadic peoples of Eurasia, which contributed to the intensive process of forming horses of new breeds.

Conclusion

Thus, the conducted studies have shown that domestic horse breeds have an original genetic structure, an inherent allele pool and are characterized by a high level of genetic diversity. Private alleles have been identified in horses of native breeds, which must be taken into account when controlling the origin and assessing population diversity, as well as when conducting genetic monitoring and planning programs for the conservation and breeding of horses of local breeds.

The results of the constructed phylogenetic tree show that local horse breeds bred in the territories of neighboring regions have the highest degree of genetic similarity. Cluster analysis combined horse breeds into three groups according to the genetic structure of DNA microsatellite loci, which confirmed their suitability as markers of phylogenetic relationship of populations. The obtained coefficients of genetic similarity adequately reflect the relationships of local horse breeds in accordance with the history of their formation.

The study of the features of the genetic structure and phylogenetic relationships of domestic aboriginal horse breeds by 17 STR markers is of undoubted interest both from a theoretical and a practical point of view. Genetic breeding methods make it possible not only to assess the degree of genetic diversity of breeds, but also to control the level of inbreeding, and based on this to form a strategy for breeding programs.

References

- Askarov A., Kuznetsova A., Gusmanov R., Askarova A., Kovshov V. Cost-effective horse breeding in the Republic of Bashkortostan, Russia. *Vet World*. 2020;13(10):2039-2045. doi 10.14202/vetworld.2020.2039-2045
- Baena M.M., Diaz S., Moura R.S., Meirelles S.L.C. Genetic characterization of Mangalarga Marchador breed horses based on microsatellite molecular markers. *J Equine Vet Sci*. 2020;95:103231. doi 10.1016/j.jevs.2020.103231
- Belousova N.F. Local (aboriginal) Horse Breeds in Russia. Divovo: Publishing House of the All-Russia Research Institute of Horse Breeding, 2018 (in Russian)
- Blohina N.V., Khrabrova L.A., Zaitsev A.M., Gavrilicheva I.S. Assessment of the genetic diversity of microsatellite loci in horses of heavydraft breeds. *Genetika i Razvedenie Zhivotnykh = Animal Genetics and Breeding*. 2018;2:39-44. doi 10.31043/2410-2733-2018-2-39-44 (in Russian)
- Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7. doi 10.1186/s13742-015-0047-8
- Ernst L.K., Zinovieva N.A. Biological Issues of Animal Husbandry in the XXI Century. Moscow, 2008 (in Russian)
- Francis R.M. POPHELPER: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour*. 2017;17(1):27-32. doi 10.1111/1755-0998
- Gavrilicheva I.S. Genetic and population characteristics of Russian trotting horse breed by DNA microsatellite loci. *AgroZooTekhnika = Agricultural and Livestock Technology*. 2019;2(3):2. doi 10.15838/alt.2019.2.3.2 (in Russian)
- Glazko V.I., Kosovsky G.Yu., Glazko T.T., Fedorova L.M. DNA markers and "microsatellite code" (review). *Sel'skokhozyaystvennaya Biologiya = Agric Biol*. 2023;58(2):223-248. doi 10.15389/agrobiology.2023.2.223eng
- Kalashnikov V.V., Khrabrova L.A., Zaitsev A.M., Zaitseva M.A., Kalinkova L.V. Polymorphism of microsatellite DNA in horses of stud and local breeds. *Sel'skokhozyaystvennaya Biologiya = Agric Biol*. 2011;46(2):41-45 (in Russian)
- Kalashnikova L.A., Novikov A.A., Semak M.S. Development of the genetic testing of breeding products in animal husbandry. *Zootekhniya = Zootechniya*. 2022;11:25-28. doi 10.25708/ZT.2022.19.42.008 (in Russian)
- Khaudov A.D., Duduev A.S., Kokov Z.A. Diversity of Kabardian horses and their genetic relationships with selected breeds in the Russian Federation based on 17 microsatellite loci. *IOP Conf Ser: Earth Environ Sci*. 2019;341:012072. doi 10.1088/1755-1315/341/1/012072
- Khrabrova L.A., Blohina N.V., Belousova N.F., Kotran E.G. Estimation of the genealogical structure of Vyatka horse breed (*Equus ferus caballus*) using DNA analysis. *Russ J Genet*. 2022;58(4):462-466. doi 10.1134/S1022795422040068
- Librado P., Khan N., Fages A. The origins and spread of domestic horses from the Western Eurasian steppes. *Nature*. 2021;598:634-640. doi 10.1038/s41586-021-04018-9
- Ling Y.H., Ma Y.H., Guan W.J., Cheng Y.J., Wang Y.P., Han J.L., Mang L., Zhao Q.J., He X.H., Pu Y.B., Fu B.L. Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers. *Anim Genet*. 2011;42(1):56-65. doi 10.1111/j.1365-2052.2010.02067.x
- Lippold S., Matzke N.J., Reissmann M., Hofreiter M. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol Biol*. 2011;11:328. doi 10.1186/1471-2148-11-328
- Marzanov N.S., Nasibov M.G., Marzanova L.K., Ozerov M.Yu., Kantanen Yu., Lobkov V.Yu. Genetic Markers in the Theory and Practice of Sheep Breeding. Moscow: Pioneer Publ., 2010 (in Russian)
- Nwachukwu E.N., Kalla D.J.U., Ukwu H.O., Ogbu C.C., Ezea J., Udoh U.H., Ekumankama O.O. Genetic diversity and population structure of four Nigerian indigenous cattle breeds. *Trop Anim Health Prod*. 2022;54(2):132. doi 10.1007/s11250-022-03132-8
- Pozharskiy A., Abdrakhmanova A., Beishova I., Shamshidin A., Nameitov A., Ulyanova T., Bekova G., Kikebayev N., Kovalchuk A., Ulyanov V., Turabayev A., Khusnitdinova M., Zhambakin K., Sapakhova Z., Shamekova M., Gritsenko D. Genetic structure and genome-wide association study of the traditional Kazakh horses. *Animal*. 2023;17(9):100926. doi 10.1016/j.animal.2023.100926
- R2D2 Consortium; Fugerey-Scarbel A., Bastien C., Dupont-Nivet M., Lemarié S. Why and how to switch to genomic selection: lessons from plant and animal breeding experience. *Front Genet*. 2021;12:629737. doi 10.3389/fgene.2021.629737
- Roh H.J., Kim S.C., Cho C.Y., Lee J., Jeon D., Kim D.K., Kim K.W., Afrin F., Ko Y.G., Lee J.H., Batsaikhan S., Susanti T., Hegay S., Kongvongxay S., Gorkhali N.A., Thi L.A.N., Thao T.T.T., Manikku L. Estimating genetic diversity and population structure of 22 chicken breeds in Asia using microsatellite markers. *Asian-Australas J Anim Sci*. 2020;33(12):1896-1904. doi 10.5713/ajas.19.0958
- Seo J.H., Park K.D., Lee H.K., Kong H.S. Genetic diversity of Halla horses using microsatellite markers. *J Anim Sci Technol*. 2016;58:40. doi 10.1186/s40781-016-0120-6

- Stanislawczyk R., Rudy M., Gil M. Quality characteristics of horse meat as influence by the age of horse. *Int J Food Prop.* 2020;23(1): 864-877. doi 10.1080/10942912.2020.1764579
- Van de Goor L.H., Panneman H., Van Haeringen W.A. A proposal for standardization in forensic equine DNA typing: allele nomenclature for 17 equine-specific STR loci. *Anim Genet.* 2010;41(2):122-127. doi 10.1111/j.1365-2052.2009.01975.x
- Vdovina N.V., Yuryeva I.B. Monitoring for the genetic structure of Mezen breed of horses in terms of DNA microsatellites. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed.* 2021;25(2): 202-207. doi 10.18699/VJ21.024
- Yun J., Oyungerel B., Kong H.S. Genetic diversity and population structure of Mongolian regional horses with 14 microsatellite markers. *Anim Biosci.* 2022;35(8):1121-1128. doi 10.5713/ab.21.0497

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

Received March 14, 2024. Revised August 22, 2024. Accepted September 27, 2024.

doi 10.18699/vjgb-25-14

Association of two missense mutations in the *MSS51* and *KAT6B* genes with body weight at different ages in cows of the Yaroslavl breed

A.V. Igoshin ¹ , N.S. Yudin ¹, D.M. Larkin ² 

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

² Royal Veterinary College, University of London, London, United Kingdom

 igoshin@bionet.nsc.ru; dmlarkin@gmail.com

Abstract. The Yaroslavl cattle is a native Russian dairy breed developed in the 19th century from the Northern Great Russian cattle, which were adapted to withstand harsh climates and poor forage conditions. Previous studies identified two breed-specific missense mutations in the *MSS51* (Ala415Glu) and *KAT6B* (Val105Met) genes that negatively impact the body weight of the animals. This study aimed to confirm the association of these missense mutations in the *MSS51* and *KAT6B* genes, along with the mutant haplotype containing both mutations, with live weight at various ages in the Yaroslavl breed using an expanded sample set. We genotyped 113 cows for these missense variants and analyzed their associations with live weight at birth, as well as at 6, 10, 12, 15, and 18 months in a combined sample of 143 animals, which includes earlier data. We employed linear regression and one-way ANOVA for statistical analysis. The results from linear regression indicated significant associations with live weight at 6, 12, and 18 months for the mutation in the *KAT6B* gene. The *MSS51* gene mutation was associated with live weight at 6, 12, 15, and 18 months. Notably, the mutant haplotype was linked to live weight across all ages from 6 to 18 months. One-way ANOVA revealed significant associations of live weight with *KAT6B* genotypes only at 6 months. For the *MSS51* gene mutation and the mutant haplotype, significant associations were found at 6, 12, 15, and 18 months. In both statistical tests, the most significant association was observed for the mutant haplotype rather than for the individual variants. These findings could be instrumental in enhancing the live weight of beef hybrids utilising the Yaroslavl cattle breed.

Key words: Yaroslavl breed; live weight; age; *KAT6B* gene; *MSS51* gene; missense mutation; haplotype; selection.

For citation: Igoshin A.V., Yudin N.S., Larkin D.M. Association of two missense mutations in the *MSS51* and *KAT6B* genes with body weight at different ages in cows of the Yaroslavl breed. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2025;29(1):122-127. doi 10.18699/vjgb-25-14

Funding. This work was supported by State Budget Project No. FWNR-2024-0012.

Ассоциация двух миссенс-мутаций в генах *MSS51* и *KAT6B* с массой тела в разном возрасте у коров ярославской породы

А.В. Игошин ¹ , Н.С. Юдин ¹, Д.М. Ларкин ² 

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

² Королевский ветеринарный колледж, Лондон, Великобритания

 igoshin@bionet.nsc.ru; dmlarkin@gmail.com

Аннотация. Ярославская порода крупного рогатого скота – отечественная порода молочного направления продуктивности, выведенная в XIX в. на основе северного великорусского скота, адаптированного к суровому климату и скудному рациону. Ранее у животных этой породы мы обнаружили две высокочастотные почти породоспецифичные миссенс-мутации в генах *MSS51* (Ala415Glu) и *KAT6B* (Val105Met), которые имели отрицательную связь с массой тела на выборке из 30 животных. Целью работы было подтверждение ассоциации миссенс-мутаций в генах *MSS51* и *KAT6B*, а также мутантного гаплотипа, содержащего обе мутации, с живой массой в разном возрасте у коров ярославской породы на расширенной выборке животных. Мы генотипировали 113 коров по вышеупомянутым миссенс-вариантам и на объединенной с предыдущими данными выборке в 143 животных провели анализ ассоциаций с живой массой в возрасте 0, 6, 10, 12, 15 и 18 мес. с использованием линейной регрессии и однофакторного дисперсионного анализа. С помощью линейной регрессии для мутации в гене *KAT6B* были выявлены значимые ассоциации с живой массой в возрасте 6, 12 и 18 мес. Мутация в гене *MSS51* была ассоциирована с живой массой в возрасте 6, 12, 15 и 18 мес. Гаплотип с обеими мутациями был ассоциирован с

живой массой во всех возрастах, от 6 до 18 мес. включительно. По результатам однофакторного дисперсионного анализа значимые ассоциации живой массы с генотипами по мутации в гене *KAT6B* были обнаружены только в возрасте 6 мес. Для мутации в гене *MSS51*, как и для мутантного гаплотипа, ассоциации имелись в возрасте 6, 12, 15 и 18 мес. В обоих статистических тестах наибольшей значимости достигла ассоциация не с отдельными вариантами, а с мутантным гаплотипом. Полученные данные могут быть использованы при селекции для производства говядины за счет откормочного контингента молодняка, а также помесей ярославской породы с быками мясных пород.

Ключевые слова: ярославская порода; живая масса; возраст; ген *KAT6B*; ген *MSS51*; миссенс-мутация; гаплотип; селекция.

Introduction

There are currently more than a thousand officially recognized cattle breeds in the world (FAO, 2024). A significant part of them can be attributed to the so-called local (indigenous, native) breeds. Local breeds usually have lower productivity compared to commercial breeds with a large breeding area, but are well adapted to local climatic factors, pathogens and farming conditions (Curone et al., 2019). Local breeds are a valuable reservoir of genetic diversity that can be used to improve the adaptive and productive traits of cattle in the face of climate change around the world (Yudin, Larkin, 2019; Colombi et al., 2024).

The Yaroslavl cattle is a native Russian dairy breed developed in the 19th century on the territory of the former Yaroslavl province as the result of “folk selection”, by pure breeding of the Northern Great Russian cattle, which were short and had low productivity, but were adapted to withstand harsh climates and poor forage conditions (Dmitriev, Ernst, 1989).

Animals of the Yaroslavl breed are mainly black in color. The head is white, with characteristic black markings (“glasses”) around the eyes. The belly and the lower part of the limbs, as well as the tip of the tail, are white (Monoenkov, 1974). Until the beginning of the 1880s, animals of the Dutch, Tyrolean, Angeln, Simmental, Allgau, and Kholmogory breeds were imported into the Yaroslavl province in small numbers. However, it appears they did not significantly affect the Yaroslavl breed, as it retained its specific exterior (Dmitriev, Ernst, 1989). In the USSR, the Yaroslavl breed was crossed with the Friesian and Dutch cattle (since 1937), as well as with the Holstein breed (since 1978), in order to increase milk productivity (Monoenkov, 1974; Tamarova, 2009). Nevertheless, studies based on genome-wide SNP genotyping arrays (Iso-Touru et al., 2016; Yurchenko et al., 2018) and microsatellite analysis (Abdelmanova et al., 2020) have shown that the Yaroslavl breed has mostly retained its unique genetics, which differs from other Russian native and foreign breeds.

Previously, we conducted a study to search for signatures of selection in the genomes of animals of the Yaroslavl breed, in which two almost breed-specific high-frequency missense mutations were identified on chromosome 28 in the *MSS51* (Ala415Glu) and *KAT6B* (Val105Met) genes, forming a single haplotype (Ruvinskiy et al., 2022). Genotyping of these mutations and subsequent association analysis carried out on a sample of 30 cows showed a negative relationship between the mutant haplotype and the live weight of animals, as well as withers height and heart girth. We hypothesized that the mutant haplotype, being associated with lower body weight of animals, had advantages under cold climate conditions and

poor food supply. Therefore, it has undergone selection in the ancestral populations of the Yaroslavl breed.

The aim of this study was to confirm the association of the missense mutations in the *MSS51* and *KAT6B* genes, as well as the mutant haplotype containing both mutations, with live weight at various ages in the Yaroslavl cows on an expanded sample set.

Materials and methods

Blood samples from 113 Yaroslavl cows from two farms of the Yaroslavl region were used in the study. Information on live weight at the age of 0, 6, 10, 12, 15 and 18 months was obtained from breeding records. DNA isolation was performed using the standard phenol-chloroform extraction method with preliminary proteolytic digestion (Sambrook, Russell, 2006). Genotyping of the missense mutations in the *MSS51* and *KAT6B* genes was carried out by restriction fragment length polymorphism (RFLP) analysis after polymerase chain reaction (PCR) (Ota et al., 2007). Primers were designed using the Vector NTI software package (Lu, Moriyama, 2004). The specificity of each primer pair was evaluated using the primer-BLAST web tool (Ye et al., 2012). The primers, PCR reaction conditions and restriction enzymes are given in Table 1. Information on 30 previously studied individuals (Ruvinskiy et al., 2022) was added to the genotyping data of 113 animals. Thus, the total sample consisted of 143 animals. The test for deviation from the Hardy–Weinberg equilibrium (--hardy option) and estimation of linkage disequilibrium between the studied SNPs (--ld option) were carried out in PLINK v1.9 (Purcell et al., 2007). Preliminarily, the genotyping data were converted to PED format recognized by the program.

Statistical analysis was performed using the linear regression and one-way analysis of variance (ANOVA) implemented in the “lm” and “aov” R functions, respectively. When using linear regression, the genotypes for each mutation were coded as 0, 1, and 2 according to the dose of the mutant allele. In addition to associations with genotypes, we also tested the association of live weight with the dose of the haplotype containing both mutations. Double homozygotes for mutant alleles were considered as carriers of two doses of the mutant haplotype. Animals homozygous for one gene mutation and heterozygous for the other were considered carriers of one dose. Double heterozygotes were also considered as carriers of one dose of the haplotype. We believe this assumption is justified, since, according to the genotyping results, mutations in both genes were in strong linkage disequilibrium. This means that the vast majority of double heterozygote carriers have mutant alleles in *cis* position, that is, on the same homologous chromosome.

Table 1. The primers, PCR reaction conditions and restriction enzymes for genotyping of missense mutations in the *MSS51* and *KAT6B* genes

Gene, substitution	Forward, reverse primer	PCR conditions	Restriction enzyme	Fragment size
<i>KAT6B</i> , G>A (Val105Met)	ACTTGCAAACCCACTTTATACAGAGTGG, CTGATCTTTCTCGTGGGGTAGAAGG	1 cycle: 95 °C – 3 min 35 cycles: 95 °C – 1 min; 60 °C – 1 min; 72 °C – 55 s	HpySE526 I, cleavage at the presence of ancestral (G) allele	613 bp, upon digestion: 303 and 310 bp
<i>MSS51</i> , G>T (Ala415Glu)	CTTGGCTTTCTTATCCCTTCAAAGTGC, ATCCAGTCATGATCTGGCTCAGC	1 cycle: 95 °C – 3 min 35 cycles: 95 °C – 1 min; 58 °C – 1 min; 72 °C – 45 s	HinfI, cleavage at the presence of mutant (T) allele	390 bp, upon digestion: 224 and 166 bp

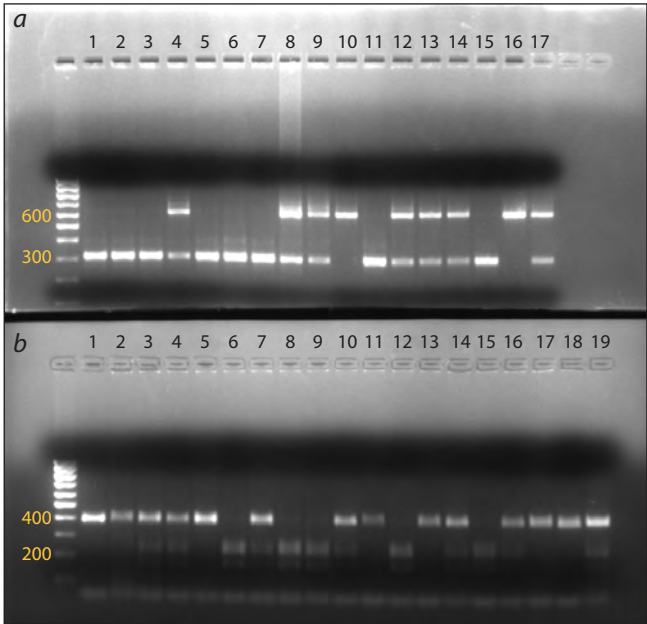
Results

The target fragments were amplified for both mutations and all the studied DNA samples were successfully genotyped (see the Figure). The distributions of genotype frequencies for both mutations did not deviate significantly from those expected under the Hardy–Weinberg equilibrium. The mutant allele frequencies for the *KAT6B* and *MSS51* genes were 0.455 and 0.434, respectively (Table 2). The mutant allele carrier rates were 0.72 and 0.699, respectively. The coefficient of linkage disequilibrium between the two loci was $r^2 = 0.891$.

When using linear regression, significant ($p < 0.05$) associations of the *KAT6B* gene mutation with live weight were identified at 6, 12, and 18 months (Table 3). The *MSS51* gene mutation was associated with live weight at 6, 12, 15, and 18 months. Notably, the dose of the mutant haplotype was associated with live weight across all ages from 6 to 18 months, inclusively. One-way ANOVA revealed significant associations of live weight with *KAT6B* genotypes only at 6 months. For both the *MSS51* gene mutation and the mutant haplotype, significant associations were found at 6, 12, 15, and 18 months. None of the statistical tests revealed an association with the live weight at birth.

Discussion

The results obtained in this work confirm the previously identified associations of mutations in the *MSS51* and *KAT6B* genes, as well as the mutant haplotype containing both variants, with the live weight of cows at different ages (Ruvinskiy et al., 2022). More significant association at most ages was achieved when using linear regression for both individual mutations and haplotype, compared to one-way ANOVA. This seems to indicate the additive effect of mutant alleles/



Examples of electropherograms of PCR-RFLP analysis for missense mutations in the studied genes.

a – genotypes for the *KAT6B* gene: GG – lanes 1, 2, 3, 5, 6, 7, 11, and 15; GA – 4, 8, 9, 12, 13, 14, and 17; AA – 10, and 16; *b* – genotypes for the *MSS51* gene: GG – lanes 1, 2, 5, 11, 13, 17, and 18; GT – 3, 4, 7, 10, 14, 16, and 19; TT – 6, 8, 9, 12, and 15.

haplotype. Reduced live weight in carriers of two copies of the mutant allele/haplotype compared to carriers of one copy was indeed observed in animals aged from 6 to 18 months, inclusively. At the same time, since both mutations are in

Table 2. Characteristics of genotyped missense mutations in the sample of Yaroslavl cows

Gene	Position (ARS-UCD1.2)	SNP	Genotype	Genotype frequency	Allele	Allele frequency
<i>KAT6B</i>	BTA28:30646253	G>A (Val105Met)	GG	0.28	G	0.545
			GA	0.531	A	0.455
			AA	0.189		
<i>MSS51</i>	BTA28:29414270	G>T (Ala415Glu)	GG	0.301	G	0.566
			GT	0.531	T	0.434
			TT	0.168		

Table 3. Associations of the studied mutations and the mutant haplotype with the live weight of cows at different ages

Gene / haplotype	Genotype / dose of haplotype (number of carriers)	Live weight at the age of (kg):					
		0 months	6 months	10 months	12 months	15 months	18 months
<i>KAT6B</i>	GG (40)	30.5 ± 4.2	146.2 ± 22.1	214.8 ± 28.3	252.1 ± 31.7	302.1 ± 37.7	352.7 ± 34.3
	GA (76)	29.4 ± 4.4	136 ± 18.3	209.3 ± 25.8	246.3 ± 28.9	300.7 ± 36.9	344.3 ± 42.8
	AA (27)	30.2 ± 3.9	134.4 ± 13.9	205.4 ± 18.3	237 ± 29.3	284.4 ± 31.7	329.5 ± 34.9
	<i>p</i> -value (lm)	0.6081	0.0074	0.1241	0.0457	0.0734	0.021
	<i>p</i> -value (aov)	0.3583	0.0113	0.302	0.1289	0.0972	0.0627
<i>MSS51</i>	GG (43)	30.6 ± 4.4	146 ± 20.9	215.8 ± 28.6	252.8 ± 31.6	302.7 ± 38.1	351.7 ± 34
	GT (76)	29.3 ± 4.3	135.8 ± 18.5	208.8 ± 24.9	246.2 ± 28.4	300.9 ± 36.1	345.1 ± 42.6
	TT (24)	30.2 ± 4	134 ± 14.5	203.9 ± 18.7	234 ± 29.5	280.3 ± 31.1	325.7 ± 34.9
	<i>p</i> -value (lm)	0.5283	0.0051	0.0519	0.0161	0.0322	0.0145
	<i>p</i> -value (aov)	0.3016	0.0084	0.1479	0.0482	0.0325	0.0324
Haplotype	0 (44)	30.7 ± 4.5	146.8 ± 21.4	217.1 ± 29.5	254.3 ± 32.8	304.2 ± 38.9	352.9 ± 34.5
	1 (75)	29.2 ± 4.2	135.1 ± 17.9	208 ± 24	245.2 ± 27.3	300 ± 35.5	344.3 ± 42.4
	2 (24)	30.2 ± 4	134 ± 14.5	203.9 ± 18.7	234 ± 29.5	280.3 ± 31.1	325.7 ± 34.9
	<i>p</i> -value (lm)	0.4048	0.0022	0.0251	0.007	0.0177	0.0089
	<i>p</i> -value (aov)	0.1797	0.0024	0.0691	0.0261	0.0278	0.0248

Note. Data are presented as mean ± standard deviation. lm – linear regression, aov – one-way ANOVA. *p*-values that reach statistical significance are highlighted in bold.

strong linkage disequilibrium, it is difficult to determine which of them is causative, that is, directly affects the phenotype. *In silico* analysis of the effect of an amino acid substitution in the previous study predicted a significant impairment of function specifically for the mutation in *KAT6B* (Ruvinskiy et al., 2022). However, associations with live weight for the missense mutation in the *MSS51* gene were more significant than those for the *KAT6B* gene. Probably, the simultaneous presence of both mutations is important for the manifestation of their effect on live weight. This assumption is supported by the fact that associations were most significant with the dose of the mutant haplotype.

The *MSS51* gene encodes a mitochondrial translation activator predominantly expressed in muscle tissue and involved in various metabolic processes, such as fatty acid oxidation, oxidative phosphorylation, and glycolysis (Moyer, Wagner, 2015). *MSS51* knockout mice have been shown to have reduced body weight compared to normal animals. However, their weight loss was due to fat, not muscle tissue (Gonzalez et al., 2019). Other authors have shown the involvement of *MSS51* in age-related muscle loss in mice. Moreover, adding betaine, which suppresses the expression of *MSS51* mRNA, to the diet of animals slowed down the decline in muscle mass and other functional parameters of skeletal muscles with age (Chen et al., 2024).

The *KAT6B* gene encodes lysine acetyltransferase 6B involved in histone modification, particularly the acetylation of H3K9 and H3K23, which increases the accessibility of chromatin in the regions of the target genes and, accordingly,

increases their expression (Bergamasco et al., 2024a). In this regard, it can be assumed that the mutation in *KAT6B* has a modifying effect on the activity of *MSS51*. Mutations in the *KAT6B* gene cause growth and developmental delay in humans (Zhang et al., 2020; Zhu et al., 2020). Of note is a study showing that mice heterozygous for a deletion in the *KAT6B* gene exhibit a significant reduction in body weight, compared to normal homozygotes. In this case, homozygotes for the deletion were not viable (Bergamasco et al., 2024b). Taken together, the biological functions of the two genes suggest that both missense variants can be causative and, probably, their effect on the live weight of animals is realized only when they are combined in a haplotype.

A limitation of this work is the fact that the study sample is represented by the animals of one sex. However, it can be assumed that the association we identified between body weight and mutations in the *MSS51* and *KAT6B* genes will be valid for bulls as well. For example, E.M.M. van der Heide et al. showed for the Aberdeen-Angus cattle that the heritability coefficients of body weight at different ages do not differ considerably between the sexes (van der Heide et al., 2016).

Also, it should be noted that the live weight of animals of the Yaroslavl breed has increased significantly over the history of its breeding. For example, in 1973, in the breeding farms of the Yaroslavl region, the average weight of heifers at the age of 0, 6, 12, and 18 months was 28, 134, 224, and 294 kg, respectively (Monoenkov, 1974). These values in our sample were 30, 139, 246, and 344 kg. Live weight was an important

selection trait of Yaroslavl cattle in the USSR, along with milk yield, since large animals capable of consuming more feed and producing more products from one stall are more efficient under industrial technology conditions (Monoenkov, 1974).

Selection to increase live weight was continued in the post-Soviet period. Thus, in most farms in the Yaroslavl region, during the period from 2000 to 2012, a significant increase in the live weight of Yaroslavl cows was recorded (Korenev et al., 2013). This can explain the fact that the frequency of mutant alleles in the populations of Yaroslavl cattle is far from fixation. It can be assumed that the selection in favor of the mutant haplotype took place in the early period of the formation of the Yaroslavl breed during “folk selection”. However, later, the frequency of this haplotype in the breed began to decrease in the course of selection aimed, among other traits, at increasing the live weight of animals.

The Yaroslavl is a dairy breed. However, as mentioned above, live weight is also an important selection trait. In addition, beef production in Russia is mainly based on fattening of young stock of dairy breeds, as well as their crosses with beef breeds (Kochetkov, 2011). In particular, there is a successful experience in creating hybrids of the Yaroslavl breed with the Limousin (Kochetkov, 2011) and Galloway (Burmistrov, 2013) breeds. Our results can be used in marker-assisted and genomic selection to increase the weight of animals of the Yaroslavl breed and its hybrids.

Conclusion

In this study, we confirmed the previously identified associations of mutations in the *MSS51* and *KAT6B* genes, as well as the mutant haplotype, with live weight in Yaroslavl cows at different ages. The obtained data can be used for selection to increase the live weight of animals in cattle breeding for beef production.

References

- Abdelmanova A.S., Kharzinova V.R., Volkova V.V., Mishina A.I., Dotsev A.V., Sermyagin A.A., Boronetskaya O.I., Petrikeeva L.V., Chinarov R.Y., Brem G., Zinovieva N.A. Genetic diversity of historical and modern populations of Russian cattle breeds revealed by microsatellite analysis. *Genes (Basel)*. 2020;11(8):940. doi 10.3390/genes11080940
- Bergamasco M.I., Abeysekera W., Garnham A.L., Hu Y., Li-Wai-Suen C.S.N., Sheikh B.N., Smyth G.K., Thomas T., Voss A.K. *KAT6B* is required for histone 3 lysine 9 acetylation and *SOX* gene expression in the developing brain. *Life Sci Alliance*. 2024a;8(2):e202402969. doi 10.26508/lsa.202402969
- Bergamasco M.I., Vanyai H.K., Garnham A.L., Geoghegan N.D., Vogel A.P., Eccles S., Rogers K.L., Smyth G.K., Blewitt M.E., Hannan A.J., Thomas T., Voss A.K. Increasing histone acetylation improves sociability and restores learning and memory in *KAT6B*-haploinsufficient mice. *J Clin Invest*. 2024b;134(7):e167672. doi 10.1172/JCI167672
- Burmistrov V.A., Pogosyan G.A., Asianin V., Golubeva A.I. Cross-breeding to thorough-bred Galloway bulls improves beef-making qualities of cattle bred in the Yaroslavl region. *Molochnoye i Myasnoye Skotovodstvo = Dairy and Meat Cattle Breed*. 2013;7:17-19 (in Russian)
- Chen S., He T., Chen J., Wen D., Wang C., Huang W., Yang Z., Yang M., Li M., Huang S., Huang Z., Zhu H. Betaine delays age-related muscle loss by mitigating *Mss51*-induced impairment in mitochondrial respiration via Yin Yang1. *J Cachexia Sarcopenia Muscle*. 2024;15(5):2104-2117. doi 10.1002/jcsm.13558
- Colombi D., Perini F., Bettini S., Mastrangelo S., Abeni F., Conte G., Marletta D., Cassandro M., Bernabucci U., Ciampolini R., Lasagna E. Genomic responses to climatic challenges in beef cattle: a review. *Anim Genet*. 2024;55(6):854-870. doi 10.1111/age.13474
- Curone G., Filipe J., Cremonesi P., Piccioli-Cappelli F., Trevisi E., Amadori M. Relevance of the dairy cow biodiversity in the development of a profitable and environmentally sustainable livestock. *CABI Rev*. 2019;14:1-11. doi 10.1079/PAVSNNR201914024
- Dmitriev N.G., Ernst L.K. Animal Genetics Resources of the USSR. Rome: Food and Agriculture Organization of the United Nations, 1989
- FAO. Status and Trends of Animal Genetic Resources. Rome: Commission on Genetic Resources for Food and Agriculture, 2024
- Gonzalez Y.I.R., Moyer A.L., LeTexier N.J., Bratti A.D., Feng S., Sun C., Liu T., Mula J., Jha P., Iyer S.R., Lovering R.M., O'Rourke B., Noh H.L., Suk S., Kim J.K., Essien Umanah G.K., Wagner K.R. *Mss51* deletion enhances muscle metabolism and glucose homeostasis in mice. *JCI Insight*. 2019;4(20):e122247. doi 10.1172/jci.insight.122247
- Iso-Touru T., Tapio M., Vilkki J., Kiseleva T., Ammosov I., Ivanova Z., Popov R., Ozerov M., Kantanen J. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim Genet*. 2016;47(6):647-657. doi 10.1111/age.12473
- Kochetkov A.A. Use of world and Russian breed resources of beef cattle to increase the production of high-quality meat in the North Caucasus and the central zone of Russia: Dr. (Agricultural Sci.) Dissertation. 2011 (in Russian)
- Korenev M.M., Furaeva N.S., Khrustaleva V.I., Ursol A.Yu., Vorobyova S.S., Konovalov A.V., Kosyachenko N.M., Ilyina A.V., Muratova N.S., Gvazava D.G., Tarasenkova N.A., Malyukova M.A. Breeding Measures for the Preservation and Improvement of the Yaroslavl Cattle Breed for 2013–2020. Yaroslavl: Kantsler Publ., 2013 (in Russian)
- Lu G., Moriyama E.N. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform*. 2004;5(4):378-388. doi 10.1093/bib/5.4.378
- Monoenkov M.I. Yaroslavl Breed of Cattle. Yaroslavl, 1974 (in Russian)
- Moyer A.L., Wagner K.R. Mammalian *Mss51* is a skeletal muscle-specific gene modulating cellular metabolism. *J Neuromuscul Dis*. 2015;2(4):371-385. doi 10.3233/JND-150119
- Ota M., Fukushima H., Kulski J.K., Inoko H. Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. *Nat Protoc*. 2007;2(11):2857-2864. doi 10.1038/nprot.2007.407
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi 10.1086/519795
- Ruvinskiy D., Igoshin A., Yurchenko A., Ilina A.V., Larkin D.M. Resequencing the Yaroslavl cattle genomes reveals signatures of selection and a rare haplotype on BTA28 likely to be related to breed phenotypes. *Anim Genet*. 2022;53(5):680-684. doi 10.1111/age.13230
- Sambrook J., Russell D.W. The Condensed Protocols from Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, 2006
- Tamarova R.V. Will we save our Yaroslavl wet nurse? On the 140th anniversary of the creation of the Yaroslavl breed. *Vestnik APK Verkhnevolzh'ya = Agroindustrial Complex of Upper Volga Region Herald*. 2009;3(7):20-23 (in Russian)
- van der Heide E.M.M., Lourenco D.A.L., Chen C.Y., Herring W.O., Sapp R.L., Moser D.W., Tsuruta S., Masuda Y., Ducro B.J., Misztal I. Sexual dimorphism in livestock species selected for economically important traits. *J Anim Sci*. 2016;94(9):3684-3692. doi 10.2527/jas.2016-0393

- 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
- Yudin N.S., Larkin D.M. Whole genome studies of origin, selection and adaptation of the Russian cattle breeds. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2019;23(5):559-568. doi 10.18699/VJ19.525 (in Russian)
- Yurchenko A., Yudin N., Aitnazarov R., Plyusnina A., Brukhin V., Soloshenko V., Lhasaranov B., Popov R., Paronyan I.A., Plemyshev K.V., Larkin D.M. Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds. *Heredity (Edinb)*. 2018;120(2):125-137. doi 10.1038/s41437-017-0024-3
- Zhang L.X., Lemire G., Gonzaga-Jauregui C., Molidpere S., Galaz-Montoya C., Liu D.S., Verloes A., ... Bamshad M.J., Lee B.H., Yang X.-J., Lupski J.R., Campeau P.M. Further delineation of the clinical spectrum of *KAT6B* disorders and allelic series of pathogenic variants. *Genet Med*. 2020;22(8):1338-1347. doi 10.1038/s41436-020-0811-8
- Zhu L., Lv L., Wu D., Shao J. *KAT6B* genetic variant identified in a short stature chinese infant: a report of physical growth in clinical spectrum of *KAT6B*-related disorders. *Front Pediatr*. 2020;8:124. doi 10.3389/fped.2020.00124

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

Received December 2, 2024. Revised December 25, 2024. Accepted December 27, 2024.

doi 10.18699/vjgb-25-15

Features of toll-like receptor genes (*TLR-2*, *TLR-3*, *TLR-4* and *TLR-6*) polymorphism in open-angle glaucoma patients

A.V. Shevchenko ¹, V.F. Prokofiev ¹, V.I. Konenkov ¹, V.V. Chernykh ², A.N. Trunov ²

¹ Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk Branch of the S.N. Fedorov National Medical Research Center “MNTK “Eye Microsurgery” of the Ministry of Health of the Russian Federation, Novosibirsk, Russia

 shalla64@mail.ru

Abstract. Modern research shows that innate immunity plays an important role in the pathogenesis of primary open-angle glaucoma (POAG). An increase in the content of toll-like receptors (TLR) in the glaucomatous retina of the human eye was revealed. TLRs can modulate the immune response in glaucoma; provide early recognition of damaging agents, activation of signaling pathways and effector mechanisms of the nonspecific immune defense system aimed at restoring homeostasis. The *TLR*-encoding genes' polymorphism alters the amino acid structure of the receptors, which leads to changes in their immune functions: expression level, ligand-binding and coreceptor functions, transport and signal transmission. The aim was to analyze the association of the *TLR2* (rs5743708), *TLR3* (rs3775291), *TLR4* (rs4986790, rs4986791) and *TLR6* (rs5743810) polymorphisms with primary open-angle glaucoma in patients of Western Siberia. Methods: 99 patients (52 men and 47 women) with a diagnosis of primary open-angle glaucoma were examined. The comparison group consisted of 100 people (81 women and 19 men). *TLR2* (rs5743708), *TLR3* (rs3775291), *TLR4* (rs4986790, rs4986791) and *TLR6* (rs5743810) polymorphisms were analyzed by RT-PCR using test systems with Syber Green (Lytx, Russia). Statistical analysis was performed using the software package SPSS 23.0 and Arlequin 3.5.2.2. Results: the distribution of genotypes in the patient group and in the control group corresponded to the Hardy–Weinberg equilibrium. The genotype frequencies did not significantly differ between the two analyzed groups. The frequency of *TLR2*-753 ArgArg:*TLR6*-249 ProPro was increased in the group of patients with POAG. The linkage disequilibrium between two polymorphic positions of the *TLR4* gene was revealed. In addition, the linkage disequilibrium between *TLR2*-*TLR6* gene for the glaucoma group and the control group was revealed. Conclusion: an increase in certain genotypes in the patient group relative to the control group may indirectly indicate the involvement of infectious factors in the initiation of POAG. However, despite the proven importance of the participation of their protein products in the pathogenesis of glaucoma, the relationship of *TLR* polymorphism requires additional research taking into account the ethnic characteristics of patients and intergenic interactions for a better understanding of the complex mechanisms of disease development. This will help carry out early diagnosis and develop the necessary therapeutic strategy.

Key words: primary open-angle glaucoma; POAG; polymorphism of toll-like receptor genes; TLR; linkage disequilibrium.


For citation: Shevchenko A.V., Prokofiev V.F., Konenkov V.I., Chernykh V.V., Trunov A.N. Features of toll-like receptor genes (*TLR-2*, *TLR-3*, *TLR-4* and *TLR-6*) polymorphism in open-angle glaucoma patients. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2025;29(1):128-134. doi 10.18699/vjgb-25-15

Особенности полиморфизма генов толл-лайк рецепторов (*TLR-2*, *TLR-3*, *TLR-4* и *TLR-6*) при открытоугольной глаукоме

А.В. Шевченко ¹, В.Ф. Прокофьев ¹, В.И. Коненков ¹, В.В. Черных ², А.Н. Трунов ²

¹ Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал Федерального исследовательского центра Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский филиал Национального медицинского исследовательского центра «МНТК «Микрохирургия глаза» им. академика С.Н. Федорова Министерства здравоохранения Российской Федерации, Новосибирск, Россия

 shalla64@mail.ru

Аннотация. Современные исследования показывают, что в патогенезе первичной открытоугольной глаукомы (ПОУГ) важную роль играет врожденный иммунитет. Выявлено повышение содержания толл-лайк рецепторов (TLR) в глаукоматозной сетчатке глаза человека. TLR могут модулировать иммунный ответ при глаукоме, обеспечивают раннее распознавание повреждающих агентов, активацию сигнальных путей и эффекторных механизмов системы неспецифической иммунной защиты, направленных на восстановление гомеостаза. По-

лиморфизм кодирующих *TLR* генов влияет на аминокислотную структуру рецепторов, приводя к изменению лигандсвязывающей и корецепторной функции, транспортировку и передачу сигналов. Целью работы был анализ ассоциированности полиморфизма генов *TLR2* (rs5743708), *TLR3* (rs3775291), *TLR4* (rs4986790, rs4986791), *TLR6* (rs5743810) с первичной открытоугольной глаукомой у пациентов Западной Сибири. Обследовано 99 пациентов (52 мужчины и 47 женщин) с диагнозом первичной открытоугольной глаукомы. Группу сравнения составили 100 человек (81 женщина и 19 мужчин). Полиморфизм генов *TLR2* (rs5743708), *TLR3* (rs3775291), *TLR4* (rs4986790, rs4986791), *TLR6* (rs5743810) анализировали методом РТ-ПЦР с использованием коммерческих тест-систем с интеркалирующим красителем Syber Green (Lytech, Россия). Статистический анализ проводился с использованием программного пакета SPSS 23.0 и Arlequin 3.5.2.2. Показано, что распределение полиморфных маркеров в группе пациентов и в контрольной группе соответствовало равновесию Харди–Вайнберга. Их частоты между двумя анализируемыми группами достоверно не различались. Частота *TLR2-753 ArgArg:TLR6-249 ProPro* была повышена в группе пациентов с ПОУГ. Выявлено неравновесное сцепление между двумя полиморфными позициям гена *TLR4*. Кроме того, выявлено нарушение равновесия между парами генов *TLR2-TLR6* для группы с глаукомой и контрольной группы. Повышение определенных генотипов в группе пациентов относительно контрольной группы может косвенно свидетельствовать об участии инфекционных факторов в инициации ПОУГ. Однако связь полиморфизма *TLR* генов, несмотря на доказанную значимость участия их белковых продуктов в патогенезе глаукомы, требует дополнительных исследований с учетом этнических особенностей пациентов и межгенных взаимодействий для лучшего понимания сложных механизмов развития заболевания. Это поможет проводить раннюю диагностику и разрабатывать необходимую терапевтическую стратегию.

Ключевые слова: первичная открытоугольная глаукома; ПОУГ; полиморфизм генов толл-лайк рецепторов; *TLR*; неравновесное сцепление.

Introduction

Primary open-angle glaucoma (POAG) is the multifactorial disease leading to progressive and irreversible vision loss, is currently a serious medical problem, including due to insufficiently studied mechanisms of damage to the optic nerve and death of retinal ganglion cells (Baudouin et al., 2021; Tezel, 2022). Modern research shows that innate immunity plays an important role in the pathogenesis of POAG. It has been established that the inducers of inflammation at the cellular level in glaucoma are the molecular structures of DAMPs (damage associated molecular patterns) released from the tissue membranes of the eye when they are damaged, including those formed as a result of an increase in intraocular pressure level (IOP) (Tezel, 2022). Excessive accumulation of DAMPs is identified by cellular pattern-associated receptors (PRRs), which are located on endosomal membranes and in the cytoplasm. It has been shown that with the development of POAG, PRRs provide early recognition of damaging agents, activation of signaling pathways and effector mechanisms of the nonspecific immune defense system aimed at restoring homeostasis (Luo et al., 2010).

The most well-studied family of PRRs are Toll-like receptors (TLRs), the expression of which has been detected in all membranes of the human eye (Stewart et al., 2015). Proteomic and immunohistochemical studies have shown an increase in TLR expression in the human glaucomatous retina, indicating that TLRs can modulate the immune response in glaucoma (Luo et al., 2010; Titi-Lartey et al., 2022). To date, two groups of functionally different TLRs have been identified in humans: transmembrane, which include *TLR1*, *TLR2*, *TLR4*, *TLR5*, *TLR6* and *TLR11*, and intracellular – *TLR3*, *TLR7*, *TLR8*, *TLR9*. It has been shown that polymorphism of TLR-encoding genes affects the amino acid structure of receptors, leading to changes in their expression level, ligand-binding and coreceptor functions, and signal transport and transmission. In addition, the features of the functions are related to the location of the polymorphic *TLR* site. Polymorphism of loci encoding the extracellular domain of the receptor may additionally lead to a

change in binding affinity and subsequent immune response, whereas mutations in the cytoplasmic domain of TLR may lead to a change in downstream signaling, despite normal binding (Törmänen et al., 2017; Macedo et al., 2019; Zhang et al., 2021). The aim of our research is to analyze the association of *TLR2* (rs5743708), *TLR3* (rs3775291), *TLR4* (rs4986790, rs4986791), *TLR6* (rs5743810) gene polymorphisms with primary open-angle glaucoma in patients of Western Siberia.

Materials and methods

Patients. 99 patients with diagnosed stage II primary open-angle glaucoma were examined – 52 (52.53 %) men and 47 (47.47 %) women. The average age of the patients was 62.8 ± 4.3 years. The diagnosis was established on the basis ophthalmological examination (determination of visual acuity, binocular ophthalmoscopy, spheroperimetry, echophthalmography, optical coherence tomography, measurement of intraocular pressure). The criteria for diagnosis were: a pronounced change in the field of vision in the paracentral region, a narrowing of the field of vision from the nose in the upper or lower nasal segment by more than 10 degrees relative to normal values, but not less than 15 degrees from the fixation point; the marginal nature of the deepening of the optic nerve. Patients of the main group had compensated (<22 mmHg (against the background of drug therapy)) or moderately elevated (<33 mmHg) intraocular pressure. The comparison group consisted of 100 people – 81 women and 19 men. The average age was 63.5 ± 0.4 years. The criterion for inclusion in the comparison group was the absence of a diagnosis of glaucoma in the subjects.

Both groups of patients did not significantly differ in age characteristics. The patients of both groups were representatives of the phenotypically Caucasian population of Russia, who were born in this territory, identifying themselves and their forebearers as “Russians”. The exclusion criteria for both groups were: acute chronic inflammatory diseases of the visual organ and their exacerbations, the presence of diabetic retinopathy, neovascular glaucoma, uveitis of various

etiologies and localization, hemophthalmos, autoimmune and tumor processes of any localization, diabetes mellitus without ophthalmological manifestations. The study was approved by the Committees on Biomedical Ethics of the Scientific Research Institute of Clinical and Experimental Lymphology, a branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Protocol No. 177 dated 02.02.2003) and the Novosibirsk Branch of FSAI “The academician S.N. Fyodorov Federal State Institution Intersectoral Research and Technology Complex ‘Eye Microsurgery’” of the Ministry of Health of the Russian Federation (Protocol No. 2 dated 02.09.2018). Informed consent was obtained from all patients for blood collection, as well as for the use of research data for scientific purposes.

DNA isolation and genotyping. Genomic DNA was obtained from whole blood samples taken for EDTA using the phenol chloroform method. Single nucleotide polymorphism (SNP) of the *TLR2* (rs5743708), *TLR3* (rs3775291), *TLR4* (rs4986790, rs4986791), *TLR6* (rs5743810) genes was detected by real-time polymerase chain reaction (RT-PCR) using commercial test systems with intercalating dye Syber Green (Lytx, Russia) in accordance with the manufacturer’s instructions.

Statistical analysis. The “case-control” scheme was used in the study. The distribution of polymorphic markers in the patient group and the control group was checked for compliance with the Hardy–Weinberg equilibrium (HWE) using the chi-square criterion. The frequency differences were determined using a two-way Fisher precision test. A $p < 0.05$ was considered statistically significant. If the null hypotheses

were not confirmed at a given level of significance $\alpha = 0.05$, then in cases of multiple comparisons, the adjusted value of p was determined using the Bonferroni correction calculated by the one-step method (Narkevich et al., 2020). Odds ratios (OR) were calculated with a 95 % confidence interval (CI). The analysis of nonequilibrium coupling was carried out by the maximum likelihood analysis method. All statistics were carried out using the software package SPSS 23.0 and Arlequin 3.5.2.2.

Results

We analyzed polymorphic variants of the coding regions of the *TLR2* (rs5743708), *TLR3* (rs3775291), *TLR4* (rs4986790, rs4986791), *TLR6* (rs5743810) genes in a group of patients with primary open-angle glaucoma II (advanced) stage relative to the control group. The distribution of polymorphic markers in the patient group and in the control group corresponded to the Hardy–Weinberg equilibrium (Table 1).

The frequency distribution in the positions analyzed by us did not significantly differ between the two groups (Table 2). Assuming that the presence of features of complex network interactions of protein products of the genes we study is the reflection of their genetic structure, we analyzed the differences in the complexes of genotypes in two groups. We identified a single *TLR2*-753 ArgArg:*TLR6*-249 ProPro complex, the frequency of which was increased in the group of patients with POAG (OR = 1.84, $p = 0.0425$, $p_{\text{cor}} = 0.297$).

Since the *TLR4* polymorphic positions analyzed by us are located in one exon of the gene, and the polymorphic loci of the *TLR2*, *TLR3*, *TLR6* genes are on the same chromosome,

Table 1. Correspondence of the frequencies of polymorphic markers to the Hardy–Weinberg equilibrium in the group of glaucoma patients and the control group

Polymorphic position	Amino acid (genotype)	Patients with glaucoma				The control group			
		Frequencies		χ^2	p	Frequencies		χ^2	p
		observed	expected			observed	expected		
<i>TLR2</i> -753 (rs5743708)	ArgArg (GG)	91.92	90.20	3.39	0.18	86.0	86.49	0.57	0.75
	ArgGln (GA)	7.07	8.59			14.0	13.02		
	GlnGln (AA)	1.01	0.20			0.0	0.49		
<i>TLR3</i> -412 (rs3775291)	LeuLeu (AA)	52.53	48.79	2.38	0.31	49.0	46.24	1.61	0.45
	LeuPhe (AG)	35.35	41.42			38.0	43.52		
	PhePhe (GG)	12.12	8.79			13.0	10.24		
<i>TLR4</i> -299 (rs4986790)	AspAsp (AA)	80.81	80.01	0.00	1.00	82.0	81.00	1.25	0.54
	AspGly (AG)	18.18	17.98			16.0	18.00		
	GlyGly (GG)	1.01	1.01			2.0	1.00		
<i>TLR4</i> -399 (rs4986791)	ThrThr (CC)	82.83	81.82	0.05	0.98	88.0	88.36	0.41	0.82
	ThrIle (CT)	16.16	16.36			12.0	11.28		
	IleIle (CC)	1.01	0.82			0.0	0.36		
<i>TLR6</i> -249 (rs5743810)	ProPro(CC)	47.47	42.68	3.71	0.16	36.0	36.60	0.06	0.97
	ProSer (CT)	36.36	44.65			49.0	47.80		
	SerSer (TT)	16.17	11.68			15.0	15.60		

Table 2. Analysis of polymorphic markers in the group of patients with primary open-angle glaucoma and in the control group

Polymorphic marker	Amino acid (genotype)	Patients with glaucoma <i>n</i> (%)	The control group <i>n</i> (%)	OR	OR_CI95	<i>p</i> *
<i>TLR2</i> -753 (rs5743708)	ArgArg (GG)	91 (91.9)	86 (86.0)	1.85	0.74–4.63	0.258
	ArgGln (GA)	7 (7.1)	14 (14.0)	0.47	0.18–1.21	0.165
	GlnGln (AA)	1 (1.0)	0 (0.0)	2.04	0.18–22.86	0.497
<i>TLR3</i> -412 (rs3775291)	LeuLeu (AA)	52 (52.5)	48 (48.5)	1.18	0.67–2.05	0.670
	LeuPhe (AG)	35 (35.3)	38 (38.4)	0.88	0.49–1.56	0.768
	PhePhe (GG)	12 (12.1)	13 (13.1)	0.91	0.39–2.11	1.000
<i>TLR4</i> -299 (rs4986790)	AspAsp (AA)	80 (80.8)	82 (82.0)	0.92	0.45–1.89	0.857
	AspGly (AG)	18 (18.2)	16 (16.0)	1.17	0.56–2.44	0.710
	GlyGly (GG)	1 (1.0)	2 (2.0)	0.50	0.04–5.60	1.000
<i>TLR4</i> -399 (rs4986791)	ThrThr (CC)	82 (82.8)	88 (88.0)	0.66	0.30–1.46	0.322
	ThrIle (CT)	16 (16.2)	12 (12.0)	1.41	0.63–3.17	0.422
	IleIle (CC)	1 (1.0)	0 (0.0)	2.04	0.18–22.86	0.497
<i>TLR6</i> -249 (rs5743810)	ProPro(CC)	47 (47.5)	36 (36.0)	1.61	0.91–2.83	0.115
	ProSer (CT)	36 (36.4)	49 (49.0)	0.59	0.34–1.05	0.086
	SerSer (TT)	16 (16.2)	15 (15.0)	1.09	0.51–2.35	0.847
<i>TLR2</i> -753: <i>TLR6</i> -249	ArgArg:ProPro (GG:CC)	46 (46.5)	32 (32.0)	1.84	1.04–3.28	0.042

Note. OR_CI95 is the 95 % confidence interval for OR, *p** is the level of statistical significance of differences according to the exact Fisher method (twosided).

Table 3. Characteristics of single nucleotide positions

SNP	Alleles of the analyzed gene (main/minor)	Position on the chromosome, bp	The frequency of the minor allele		<i>p</i>
			Patients with glaucoma	Comparison group	
rs5743708	<i>TLR2</i> (G/A)	Chr4:153705165	0.045	0.071	0.39
rs3775291	<i>TLR3</i> (A/G)	Chr4:186082920	0.298	0.323	0.66
rs4986790	<i>TLR4</i> (A/G)	Chr9:117713024	0.101	0.100	1.00
rs4986791	<i>TLR4</i> (C/T)	Chr9:117713324	0.091	0.060	0.26
rs5743810	<i>TLR6</i> (C/T)	Chr4:38828729	0.343	0.395	0.30

Note. Position is the distance from the telomeres of the short arm of the chromosome, bp – base pair.

we analyzed the linkage disequilibrium (LD) of these positions. The characteristics of the analyzed single nucleotide positions of *TLR* genes are given in Table 3. The frequency of the minor allele in most of the loci analyzed by us was more than 5 %, with the exception of rs5743708 of the *TLR2* gene.

We have revealed the linkage disequilibrium between two polymorphic positions of the *TLR4* gene (Table 4). The analysis of multiple SNPs showed that the most common haplotype for the *TLR4* rs4986790 and rs4986791 SNPs is A/C for both groups, the A/T haplotype is completely absent in the comparison group. The Lewontin's *D'* coefficient between SNP rs4986790 and rs4986791 are 0.8146 in the patient and 1.0000 in the comparison group. In addition, we found

the linkage disequilibrium between the *TLR2*-*TLR6* genes (*D'* = 0.6615 and *D'* = 0.5277 for the glaucoma group and the control group, respectively). For the *TLR3*-*TLR6* genes, *D'* = 0.1997 and *D'* = 0.2008 in the glaucoma group and the control group, respectively. At the same time, the analysis of haplotype frequencies between the groups did not reveal any significant differences.

Discussion

Open-angle glaucoma is considered as a multifactorial disease with convincing evidence of the involvement of the genetic component in its development. To date, studies of genetic associations have identified many loci that contribute to the genetic risk of developing POAG. TLRs are important fac-

Table 4. Haplotype frequencies and parameters of the linkage disequilibrium between the analyzed polymorphic loci

Analyzed positions	Haplotypes	Patients with POAG, <i>n</i> = 99		The comparison group, <i>n</i> = 100		<i>p</i>
		frequency	parameters of linkage disequilibrium of polymorphic positions	frequency	parameters of linkage disequilibrium of polymorphic positions	
TLR4 (rs4986790), TLR4 (rs4986791)	A/C	0.884	$\chi^2 = 116.94$	0.900	$\chi^2 = 114.89$	1.00
	A/T	0.015	$p = 0.000$	0.000	$p = 0.000$	0.12
	G/C	0.025	$df = 1$	0.040	$df = 1$	0.57
	G/T	0.076	$r^2 = 0.591$	0.060	$r^2 = 0.574$	0.56
TLR2 (rs5743708)/ TLR3 (rs3775291)	G/A	0.677	$\chi^2 = 0.97$	0.641	$\chi^2 = 2.15$	0.46
	G/G	0.278	$p = 0.325$	0.288	$p = 0.142$	0.82
	A/A	0.025	$df = 1$	0.035	$df = 1$	0.58
	A/G	0.020	$r^2 = 0.005$	0.035	$r^2 = 0.011$	0.38
TLR2 (rs5743708)/ TLR6 (rs5743810)	A/C	0.646	$\chi^2 = 7.89$	0.585	$\chi^2 = 6.42$	0.22
	A/T	0.308	$p = 0.005$	0.345	$p = 0.011$	0.46
	G/C	0.010	$df = 1$	0.020	$df = 1$	0.69
	G/T	0.035	$r^2 = 0.040$	0.050	$r^2 = 0.032$	0.62
TLR3 (rs3775291)/ TLR6 (rs5743810)	A/C	0.500	$\chi^2 = 6.41$	0.449	$\chi^2 = 5.87$	0.32
	A/T	0.202	$p = 0.011$	0.227	$p = 0.015$	0.54
	G/C	0.157	$df = 1$	0.157	$df = 1$	1.00
	G/T	0.141	$r^2 = 0.032$	0.167	$r^2 = 0.030$	0.49

Note. *df* – degree of freedom, r^2 – correlation coefficient.

tors of the innate immune system; however, the results of the study concerning the association of *TLR* polymorphism with the disease are quite contradictory.

It is known that *TLR2* is a mediator of retinal degeneration in response to oxidative stress, functions as a “bridge” between oxidative damage and complement-mediated retinal pathology and is associated with the development of a number of ophthalmopathologies (Mulfaul at al 2020; Titi-Lartey et al., 2022). It has been shown that the p.Arg753Gln missense mutation leads to a deficiency in *TLR2* signaling due to impaired *TLR2*-*TLR6* heterodimerization, tyrosine phosphorylation and further cascade, without affecting *TLR2* expression (Xiong at al., 2012). However, we did not identify the association of *TLR2* and *TLR6* polymorphisms in the analyzed positions with the development of POAG. Similar results have been shown by Japanese researchers for *TLR2* (Nakamura et al., 2009). We did not find any data on the polymorphism of the *TLR6* gene in glaucoma in the literature. At the same time, the analysis of the Pro249Ser marker and the construction of a three-dimensional model for *TLR6* revealed conformational changes in the structure of the mutant protein, presumably affecting the binding of ligands and receptors: in the wild type, binding pockets near proline (Pro) are larger in volume, whereas in the mutant one, the walls of the pockets are located close to each other. This significantly affects the ability of the mutant protein to enter into significant interactions, since it is known that most binding regions and active sites are located in the largest pocket

cavity. In addition, the wild-type protein, being more flexible, has more possibilities for ligand-induced movements, whereas in the mutant ligand, induced movement is limited only by side chain rearrangements. In addition, the mutant protein is less stable. All this confirms that *TLR6* polymorphism affects the structure and functionality of the protein (Hamann et al., 2013; Senglali et al., 2018). Considering that *TLR2* and *TLR6* function during the formation of a heterodimer, we analyzed their complex polymorphism during the development of POAG and found that carriers of the homozygous wild-type genotype *TLR2*-753 ArgArg:*TLR6*-249 ProPro have a higher chance of developing the disease, which may be explained precisely by the peculiarities of joint functioning during ligand recognition and stimulation of further immune cascade. In addition, since the *TLR2* and *TLR6* genes are within the same chromosome, we performed an analysis of the linkage disequilibrium and showed a change of the LD positions of the *TLR2*-*TLR6* genes analyzed. This means that certain alleles of two genes may appear in a single haplotype more often than would be expected with a random combination. Previously, the linkage disequilibrium for these polymorphic positions has been shown in other studies (Stashkevich et al., 2022). At the same time, we have not revealed any differences in the frequencies of haplotypes.

The relationship of *TLR3* gene polymorphism in the analyzed position is also shown for a number of ophthalmopathologies (Titi-Party et al., 2022). But stratification analysis by

ethnicity indicates that rs3775291 is associated, in particular, with all forms of macular degeneration only in Caucasians, but not in East Asians (Ma et al., 2016). The polymorphism of Leu412Phe affects the normal dimerization of TLR3, which leads to a change in protein activity necessary for proper signaling (Ranjith-Kumar et al., 2007). In glaucoma, TLR3 and TLR4 have been shown to initiate nephroptosis – the regulated proinflammatory lytic form of necrotic cell death characterized by cell swelling followed by rupture of the plasma membrane with the release of cellular contents (Basavarajappa et al., 2023). The activity of *TLR3* involved in the recognition of nucleic acids released from damaged cells is mainly associated with the early stage of glaucoma (Soto, Howell, 2014). However, the association of polymorphism of this gene with glaucoma is controversial in the literature. Several studies of the WDR36 locus of the *TLR3* gene, including SNPs rs3775291, have demonstrated its role as a modifier gene in POAG due to the clinical severity of the process (Hauser et al., 2006; Meer et al., 2021). However, a meta-analysis of 122 publications did not confirm the significant role of this polymorphic position in the genetic predisposition to POAG or its subtypes. At the same time, the authors are inclined to believe that further research is needed in specific populations (Liu et al., 2017).

rs4986790 and rs4986791 polymorphisms in exon 3 of the *TLR4* gene are among the most well-known and frequently studied SNPs. Polymorphism in these positions leads to changes in the polypeptide chains of the extracellular domain of the receptor and affects binding to the coreceptor, which leads to hyperactivity of the receptor. This can cause dysfunction of the TLR4 molecule and disrupt the host's immune system (Arbour et al., 2000; Jahantigh et al., 2013; Lin et al., 2019). Currently, the results of meta-analyses of the association of these SNPs with POAG by different research groups indicate that the data differ in different ethnic groups and further research is needed (Chaiwang, Poyomtip, 2019; Lin et al., 2019). At the same time, almost all studies have shown linkage disequilibrium of *TLR4* rs4986790 and *TLR4* rs4986791 (Guimarães et al., 2018; Kania et al., 2022), which is confirmed in our study. This indicates that the recombination of the chromosome regions on which these polymorphic markers are located is inherited as a single block. It is believed that it is the analysis of haplotypes in the presence of a high degree of multilocus LD that can significantly increase the statistical significance of the study (Jiang, et al., 2014); however, we have not revealed significant differences in the analyzed frequencies of the *TLR4* gene haplotypes.

Since the receptors of TLR1, 2, 4, 5, 6 and 10 belong to surface membrane receptors that recognize mainly lipid components of bacterial structures, and TLR3, 7, 8 and 9 are expressed on the membranes of intracellular organelles, ligands for which are components of nucleic acids of viruses (Akira et al., 2001; Sameer, Nissar, 2021), an increase in the frequency of a number of *TLR* gene genotypes in a group of patients may indirectly indicate the involvement of infectious factors in the initiation of POAG.

Conclusion

Thus, the relationship of *TLR* gene polymorphism, despite the proven importance of the participation of their protein products in the pathogenesis of glaucoma, requires additional research

taking into account the ethnic characteristics of patients. In addition, it is necessary to take into account gene-gene interactions to better understand the complex mechanisms of disease development, which will help to carry out early diagnosis and develop the necessary therapeutic strategy.

References

- Akira S., Takeda K., Kaisho T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol.* 2001;2:675-680. doi 10.1038/90609
- Arbour N.C., Lorenz E., Schutte B.C., Zabner J., Kline J.N., Jones M., Frees K., Watt J.L., Schwartz D.A. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet.* 2000; 25(2):187-191. doi 10.1038/76048
- Basavarajappa D., Galindo-Romero C., Gupta V., Agudo-Barriuso M., Gupta V.B., Graham S.L., Chitranshi N. Signalling pathways and cell death mechanisms in glaucoma: insights into the molecular pathophysiology. *Mol Aspects Med.* 2023;94:101216. doi 10.1016/j.mam.2023.101216
- Baudouin C., Kolko M., Melik-Parsadaniantz S., Messmer E.M. Inflammation in glaucoma: from the back to the front of the eye, and beyond. *Prog Retin Eye Res.* 2021;83:100916. doi 10.1016/j.preteyeres.2020.100916
- Chaiwang N., Poyomtip T. The association of toll-like receptor 4 gene polymorphisms with primary open angle glaucoma susceptibility: a meta-analysis. *Biosci Rep.* 2019;39(4):BSR20190029. doi 10.1042/BSR20190029
- Guimarães L.O., Bajayb M.M., Monteiro E.F., Wunderlich G., Santosd S.E., Kirchgattera K. Genetic ancestry effects on the distribution of toll-like receptors (TLRs) gene polymorphisms in a population of the Atlantic Forest, São Paulo, Brazil. *Hum Immunol.* 2018; 79(2):101-108. doi 10.1016/j.humimm.2017.11.007
- Hamann L., Koch A., Sur S., Hoefer N., Glaeser C., Schulz S., Gross M., Franke A., Nöthlings U., Zacharowski K., Schumann R.R. Association of a common TLR-6 polymorphism with coronary artery disease – implications for healthy ageing? *Immun Ageing.* 2013;10(1):43. doi 10.1186/1742-4933-10-43
- Hauser M.A., Allingham R.R., Linkroum K., Wang J., LaRocque-Abramson K., Figueiredo D., Santiago-Turla C., del Bono E.A., Haines J.L., Pericak-Vance M.A., Wiggs J.L. Distribution of WDR36 DNA sequence variants in patients with primary open-angle glaucoma. *Invest Ophthalmol Vis Sci.* 2006;47(6):2542-2546. doi 10.1167/iovs.05-1476
- Jahantigh D., Salimi S., Alavi-Naini R., Emamdadi A., Osquee H.O., Mashhadi F.F. Association between *TLR4* and *TLR9* gene polymorphisms with development of pulmonary tuberculosis in Zahedan, Southeastern Iran. *Sci World J.* 2013;2013:534053. doi 10.1155/2013/534053
- Jiang D., Ma G., Yang R., Li K., Fang M. Bayesian model selection for multiple QTLs mapping combining linkage disequilibrium and linkage. *Genet Res.* 2014;96:e10. doi 10.1017/S0016672314 000135
- Kania K.D., Haręza D., Wilczyński J.R., Wilczyński M., Jarych D., Malinowski A., Paradowska E. The Toll-like receptor 4 polymorphism Asp299Gly is associated with an increased risk of ovarian cancer. *Cells.* 2022;11(19):3137. doi 10.3390/cells11193137
- Lin Z., Huang S., Sun J., Xie B., Zhong Y. Associations between TLR4 polymorphisms and open angle glaucoma: a meta-analysis. *Biomed Res Int.* 2019;2019:6707650. doi 10.1155/2019/6707650
- Liu K., Wenling H., Jun Z., Yingxia Z., Hongbo C. Association of WDR36 polymorphisms with primary open angle glaucoma. A systematic review and meta-analysis. *Medicine (Baltimore).* 2017;96: e7291. doi 10.1097/MD.00000000000007291
- Luo C., Yang X., Kain A.D., Powell D.W., Kuehn M.H., Tezel G. Glaucomatous tissue stress and the regulation of immune response through glial Toll-like receptor signaling. *Invest Ophthalmol Vis Sci.* 2010;51(11):5697-5707. doi 10.1167/iovs.10-5407



- Ma L., Tang F., Chu W., Young A., Brelen M., Pang C., Chen L. Association of toll-like receptor 3 polymorphism rs3775291 with age-related macular degeneration: a systematic review and meta-analysis. *Sci Rep.* 2016;6:19718. doi 10.1038/srep19718
- Macedo A.B., Novis C.L., Bosque A. Targeting cellular and tissue HIV reservoirs with Toll-like receptor agonists. *Front Immunol.* 2019;10:2450. doi 10.3389/fimmu.2019.02450
- Meer E., Aleman T.S., Ross A.G. WDR36-associated neurodegeneration: a case report highlights possible mechanisms of normal tension glaucoma. *Genes.* 2021;12:1624. doi 10.3390/genes12101624
- Mulfaul K., Ozaki E., Fernando N., Brennan K., Chirco K., Connolly E., Greene C., Maminishkis A., Salomon R., Linetsky M., Natoli R., Mullins R., Campbell M., Doyle S. Toll-like receptor 2 facilitates oxidative damage-induced retinal degeneration. *Cell Rep.* 2020;30(7):2209-2224.e5. doi 10.1016/j.celrep.2020.01.064
- Nakamura J., Meguro A., Ota M., Nomura E., Nishide T., Kashiwagi K., Mabuchi F., Iijima H., Kawase K., Yamamoto T., Nakamura M., Negi A., Sagara T., Nishida T., Inatani M., Tanihara H., Aihara M., Araie M., Fukuchi T., Abe H., Higashide T., Sugiyama K., Kanamoto T., Kiuchi Y., Iwase A., Ohno S., Inoko H., Mizuki N. Association of toll-like receptor 2 gene polymorphisms with normal tension glaucoma. *Mol Vis.* 2009;15:2905-2910
- Narkevich A.N., Vinogradov K.A., Grjibovski A.M. Multiple comparisons in biomedical research: the problem and its solutions. *Ekologiya Cheloveka = Human Ecology.* 2020;27(10):55-64. doi 10.33396/1728-0869-2020-10-55-64 (in Russian)
- Ranjith-Kumar C.T., Miller W., Sun J., Xiong J., Santos J., Yarbrough I., Lamb R.J., Mills J., Duffy K., Hoose S., Cunningham M., Holzenburg A., Mbow M., Sarisky R., Kao C. Effects of single nucleotide polymorphisms on Toll-like receptor 3 activity and expression in cultured cells. *J Biol Chem.* 2007;282(24):17696-17705. doi 10.1074/jbc.M700209200
- Sameer A.S., Nissar S. Toll-like receptors (TLRs): structure, functions, signaling, and role of their polymorphisms in colorectal cancer susceptibility. *Biomed Res Int.* 2021;2021:1157023. doi 10.1155/2021/1157023
- Semlali A., Almutairi M., Rouabhia M., Reddy Parine N., Al Amri A., Al-Numair N.S., Hawsawi Y.M., Saud Alanazi M. Novel sequence variants in the *TLR6* gene associated with advanced breast cancer risk in the Saudi Arabian population. *PLoS One.* 2018;13(11):e0203376. doi 10.1371/journal.pone.0203376
- Soto I., Howell G.R. The complex role of neuroinflammation in glaucoma. *Cold Spring Harb Perspect Med.* 2014;4(8):a017269. doi 10.1101/cshperspect.a017269
- Stashkevich D.S., Belyaeva S.V., Evdokimov A.V. Comparative assessment of genetic polymorphism of Toll-like 2 and 6 receptors predisposing for non-specific ulcerative colitis and irritable bowel syndrome in Russians from Chelyabinsk Region. *Rossiiskii Immunologicheskii Zhurnal = Russ J Immunol.* 2022;25(3):327-332. doi 10.46235/1028-7221-1139-CAO (in Russian)
- Stewart E.A., Wei R., Branch M.J., Sidney L.E., Amoaku W.M. Expression of Toll-like receptors in human retinal and choroidal vascular endothelial cells. *Exp Eye Res.* 2015;138:114-123. doi 10.1016/j.exer.2015.06.012
- Tezel G. Molecular regulation of neuroinflammation in glaucoma: current knowledge and the ongoing search for new treatment targets. *Prog Retin Eye Res.* 2022;87:100998. doi 10.1016/j.preteyeres.2021.100998
- Titi-Lartey O., Mohammed I., Amoaku W.M. Toll-like receptor signaling pathways and the pathogenesis of retinal diseases. *Front Ophthalmol.* 2022;2:850394. doi 10.3389/fopht.2022.850394
- Törmänen S., Korppi M., Teräsjarvi J., Vononvirta J., Koponen P., Helminen M., He Q., Nuolivilta K. Polymorphism in the gene encoding toll-like receptor 10 may be associated with asthma after bronchiolitis. *Sci Rep.* 2017;7:2956. doi 10.1038/s41598-017-03429-x
- Xiong Y., Song C., Snyder G.A., Sundberg E.J., Medvedev A.E. R753Q polymorphism inhibits Toll-like receptor (TLR) 2 tyrosine phosphorylation, dimerization with TLR6, and recruitment of myeloid differentiation primary response protein 88. *J Biol Chem.* 2012;287(45):38327-38337. doi 10.1074/jbc.M112.375493
- Zhang Y., Liu J., Wang C., Liu J., Lu W. Toll-Like receptors gene polymorphisms in autoimmune disease. *Front Immunol.* 2021;12:672346. doi 10.3389/fimmu.2021.672346

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

Received April 5, 2024. Revised August 12, 2024. Accepted October 24, 2024.

doi 10.18699/vjgb-25-16

The role of *SELE* gene polymorphism in ST-elevation myocardial infarction


N.P. Babushkina ¹, A.M. Nikolaeva ², A.D. Dolbnya³, V.E. Shavrak⁴, V.V. Ryabov^{2, 3, 4}

¹ Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia

² Cardiology Research Institute, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia

³ Siberian State Medical University of the Ministry of Healthcare of the Russian Federation, Tomsk, Russia

⁴ Tomsk State University, Tomsk, Russia

 nad.babushkina@medgenetics.ru

Abstract. Ischemic heart disease (IHD) is an important medical and social problem. ST-elevation myocardial infarction (STEMI) is the most severe form of IHD, affecting all layers of the heart muscle. One of the diagnostic criteria for endothelial dysfunction in myocardial infarction is the level of sE-selectin, a cell adhesion molecule that recruits neutrophils and induces neutrophil inflammation. The aim of this study is to investigate intronic polymorphisms rs5353, rs3917412 and rs1534904 of the E-selectin coding gene *SELE* in patients with STEMI. We have analyzed a group of patients with STEMI ($n = 74$) and a population sample of Tomsk ($n = 136$) as the control group. The frequencies of the rs5353 genotypes in the *SELE* gene have shown statistically significant differences between patients and the control sample ($p = 0.004$). The CC genotype is a predisposing factor to STEMI (OR = 6.93, CI:95 % (1.84–26.04), $\chi^2 = 8.69$, $p = 0.002$). The analyzed markers were not studied previously in cardiovascular diseases (CVDs) and were rarely involved in association studies at all; there is no information on these SNPs in the leading databases. At the same time, all three variants, according to the RegulomeDB classification, belong to the functional class 1f, and are highly likely to have regulatory potential relative not only to the *SELE* gene, but also to other genes in the nearby region. The analysis of the functional significance of the studied markers has shown the presence of a region more extensive than one gene, which is co-regulated by the studied nucleotide substitutions. The association of rs5353 with STEMI identified in this study once again confirms the involvement of the *SELE* gene in the pathogenesis of CVDs. It is possible that this entire region of the genome may be involved indirectly in the pathogenesis of CVD through the systems of inflammation, immune response and DNA repair.

Key words: ST-elevation myocardial infarction; STEMI; *SELE* gene; SNP.

For citation: Babushkina N.P., Nikolaeva A.M., Dolbnya A.D., Shavrak V.E., Ryabov V.V. The role of *SELE* gene polymorphism in ST-elevation myocardial infarction. *Vavilovskii Zhurnal Genetiki i Selektzii* = *Vavilov J Genet Breed.* 2025;29(1): 135-143. doi 10.18699/vjgb-25-16

Funding. The work was carried out with partial financing of the State Assignment of the Ministry of Science and Higher Education (No. 122020300041-7 and No. 122020300043-1).

Acknowledgements. The molecular genetic study was carried out at the Center for Collective Use of Research Equipment and Experimental Biological Material "Medical Genomics" of the Research Institute of Medical Genetics of the Tomsk National Research Medical Center of the Russian Academy of Sciences.

Роль полиморфизма гена *SELE* при инфаркте миокарда с подъемом сегмента ST


Н.П. Бабушкина ¹, А.М. Николаева ², А.Д. Долбня³, В.Е. Шаврак⁴, В.В. Рябов^{2, 3, 4}

¹ Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия

² Научно-исследовательский институт кардиологии, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия

³ Сибирский государственный медицинский университет Министерства здравоохранения Российской Федерации, Томск, Россия

⁴ Национальный исследовательский Томский государственный университет, Томск, Россия

 nad.babushkina@medgenetics.ru

Аннотация. Ишемическая болезнь сердца представляет собой важную медико-социальную проблему. Наиболее тяжелой формой заболевания, с поражением всех слоев сердечной мышцы, считается инфаркт миокарда с подъемом сегмента ST (ИМпST). Одним из диагностических критериев дисфункции эндотелия при инфаркте миокарда является уровень sE-селектина – молекулы клеточной адгезии, осуществляющей рекрутинг нейтро-

филов и индукцию нейтрофильного воспаления. В настоящем исследовании изучен интронный полиморфизм (rs5353, rs3917412, rs1534904) гена *SELE*, кодирующего Е-селектин, у пациентов с ИМнСТ. Проанализированы две выборки: пациенты с ИМнСТ ($n = 74$) и популяционная выборка г. Томска ($n = 136$). По частотам генотипов rs5353 в гене *SELE* зарегистрированы статистически значимые различия между пациентами и контрольной выборкой ($p = 0.004$). Генотип СС является рискованным по отношению к ИМнСТ ($OR = 6.93$, $CI: 95\% (1.84-26.04)$, $\chi^2 = 8.69$, $p = 0.002$). Проанализированные маркеры не изучались ранее при сердечно-сосудистых заболеваниях и вообще редко привлекались к ассоциативным исследованиям; в ведущих базах данных отсутствует информация об ассоциациях этих маркеров с заболеваниями. Вместе с тем все три варианта по классификации RegulomeDB относятся к функциональному классу 1f и, соответственно, с высокой вероятностью обладают регуляторным потенциалом относительно не только гена *SELE*, но и других генов близлежащего региона. Анализ функциональной значимости изученных маркеров показал наличие более обширного, чем один ген, региона, корегулируемого данными нуклеотидными заменами. Выявленная в настоящем исследовании ассоциация rs5353 с ИМнСТ еще раз подтверждает вовлеченность гена *SELE* в развитие сердечно-сосудистых заболеваний. Не исключено, что опосредованно (через системы воспаления, иммунного ответа и репарации ДНК) весь этот регион генома может быть вовлечен в патогенез сердечно-сосудистых заболеваний.

Ключевые слова: инфаркт миокарда с подъемом сегмента ST; ИМнСТ; ген *SELE*; SNP.

Introduction

Ischemic heart disease (IHD) is an important medical and social problem, holding the leading place in the structure of mortality from cardiovascular diseases. The most life-threatening condition is the acute form of ischemia, myocardial infarction. ST-elevation myocardial infarction (STEMI) is the most severe form with damage to all layers of the heart muscle (Clinical practice guidelines..., 2020). Inflammation is one of the leading elements in the pathogenesis, course and prognosis of myocardial infarction (Kachkovsky, Ragozina, 2013; Kalinin et al., 2022; Zhang N. et al., 2022). The inflammatory response is initiated by endothelial dysfunction associated with an imbalance in the production of endothelial mediators and leading to overexpression of adhesion molecules (Kachkovsky, Ragozina, 2013; Habas, Shang, 2018; Mathur et al., 2023).

E-selectin is a surface glycoprotein, that belongs to the class of cell adhesion molecules. E-selectin is expressed only by endothelial cells and exists in two forms: a transmembrane glycoprotein and a serum sE-selectin. In endothelium that is functioning normally, the amount of the protein is so low as to be negligible. E-selectin plays a role in the adhesion of neutrophils from circulating blood to the damaged vascular wall, and also promotes the migration of monocytes into the subendothelial space (Lorenzon et al., 1998; Vestweber, Blanks, 1999; Cid et al., 2000; Blankenberg et al., 2003; Calder et al., 2013; McEver, 2015)). In addition, the mechanism of neutrophil inflammation activation induced by E-selectin (through NLRP3 inflammasome activation) has been demonstrated (Pruenster et al., 2023). Given that neutrophils are the initial cells to infiltrate the site of damage during myocardial infarction (Kalinin et al., 2022), the pathogenetic role of E-selectin, which plays a dual role in the response to damage (neutrophil recruitment and induction of neutrophil inflammation), appears to be even more significant.

De novo synthesis of E-selectin is initiated in the endothelium following stimulation with proinflammatory cytokines (TNF- α , IL-1), endotoxin, or under conditions of shear stress. Following initial exposure to the stimulus, the protein level rises within four to six hours, subsequently declining after one to two days. Therefore, E-selectin expression may reflect the

acute phase of inflammation (Kalinin et al., 2022; Uy et al., 2024). Selectins in general, and E-selectin in particular, are well-recognized markers of endothelial dysfunction (Silva et al., 2018; Mangoni, Zinellu, 2024; Wang K. et al., 2024). The measurement of sE-selectin levels is a diagnostic tool used to diagnose endothelial dysfunction in patients with heart failure, atherosclerosis, glaucoma, DM2, arterial hypertension, ACS, an indicator of myocardial damage in children with respiratory mycoplasmosis (*Mycoplasma pneumoniae*), COVID-19, and other conditions (Wang N. et al., 2001; Ueno, 2012; Sandoval-Pinto et al., 2014; Srivastava et al., 2018; Lampsas et al., 2022; Mathur et al., 2023). There is evidence to support the hypothesis that E-selectin levels are associated with the presence of atherosclerotic vascular lesions, both coronary and peripheral (Zhito et al., 2019; Kalinin et al., 2022; Mathur et al., 2023). This is likely to reflect systemic inflammation as a characteristic feature of atherosclerosis.

The role of selectins in the pathogenesis of IHD is controversial. As is the case with numerous association studies, the accumulated evidence is contradictory: in some cases, authors have reported a statistically significant increase in E-selectin levels in patients with sTable IHD, whereas in other cases, no significant differences have been observed (see review (Zhito et al., 2019)). These results are explained by small sample sizes, heterogeneity in sex, age, presence of comorbidities, and the treatment received by patients (Zhito et al., 2019).

It is noteworthy that, despite a considerable amount of information dedicated to E-selectin, the focus is significantly shifted towards biochemistry: the protein level is analyzed in various pathological conditions and its role as a diagnostic criterion is discussed in detail. Nevertheless, a number of studies have shown associations of three polymorphic variants (single nucleotide polymorphisms, SNPs) in the *SELE* gene (G98T (rs1805193) in the 5'UTR, A561C (rs5361), C1880T (rs5355) in exons 4 and 10, respectively) with severe and subclinical atherosclerosis, coronary heart disease, ischemic heart disease, myocardial infarction, ischemic stroke, Kawasaki disease, and arterial hypertension (Wenzel et al., 1994; Zheng et al., 2001; Yoshida et al., 2003; Zak et al., 2008; Mallik, Majumder, 2011; Shirakawa et al., 2012; Wang Z. et al., 2012; Zhao et al., 2012; Wang X. et al., 2013; Qin et al.,

2015; Liao B. et al., 2016; Deng et al., 2017; Vargas-Alarcon et al., 2019; Ding et al., 2021). Thus, associations of exon and promoter polymorphisms of the *SELE* gene with cardiovascular pathology are shown.

The aim of our study was to investigate the associations of intronic functionally significant polymorphic variants of the *SELE* gene with the development of ST-elevation myocardial infarction.

Material and methods

The study included 74 patients hospitalized in the Department of Emergency Cardiology of the Cardiology Research Institute of the Tomsk NRMС from 2019 to 2021. The diagnosis of primary STEMI was established in accordance with fourth Universal Definition of Myocardial Infarction (Thygesen et al., 2018). The inclusion criteria in the study were: a verified diagnosis of primary STEMI, age over 18 years and a permanent residence in the Tomsk region. The exclusion criteria were: cardiogenic shock, autoimmune, oncologic diseases, terminal chronic kidney disease, atrial fibrillation/atrial flutter, hemodynamically significant valve heart defects, marked cognitive dysfunction. The study protocol adhered to the standards established by the Declaration of Helsinki and received approval from the local ethical committee of the Cardiology Research Institute. A population sample of Russians from Tomsk (136 individuals), formed from DNA samples from the “Biobank of the Population of Northern Eurasia” of the Research Institute of Medical Genetics of the Tomsk NRMС, was used as a control. The groups of patients and the control sample were comparable in sex and age. All examined individuals were ethnically homogeneous and were represented predominantly by Russians (>95 %) from Tomsk: all of them gave informed consent.

Both study groups were predominantly comprised of men, with a male-to-female ratio of 2.1 in the STEMI group and 1.5

in the control group; there were no statistically significant differences between the groups. The mean age in the STEMI group was 61 ± 10 years (median 62.5; interquartile range [55.0–69.0]), and in the control group, 62.1 ± 7 years (median 63.0; interquartile range [57.0–68.0]); there were no statistically significant differences between the groups.

DNA from venous peripheral blood was isolated using the standard phenol-chloroform method (Sambrook, Russell, 2006). Genotyping was performed using real-time polymerase chain reaction (real-time PCR) with the BioMaster HS-qPCR (2×) PCR kit (BioLabMix, Novosibirsk), region-specific primers and TaqMan probes (manufactured by DNA-Synthesis, Moscow) (Table 1).

We selected for analysis polymorphic variants in the *SELE* gene that are eQTL variants for their own or nearby genes (according to GTExPortal (<https://www.gtexportal.org/home/>)), potentially having functional significance (according to RegulomeDB (<https://regulomedb.org/regulome-search/>)), are located in non-coding regions of the gene and have a minor allele frequency of at least 25 % in Caucasians (data from the 1000 Genomes Project, Ensemble (<https://www.ensembl.org/index.html>)). As a result, three intronic markers were analyzed (Table 1).

Association analysis was performed using standard methods of statistical analysis (χ^2 , OR with 95 % CI; differences between the compared groups were considered statistically significant at $p < 0.05$). The method of logistic regression was used to study the inheritance model. Linkage analysis (including calculation of the linkage disequilibrium coefficient (D')) was performed in the Haploview 4.2 program (Barrett et al., 2005).

The functional annotation of the variants was performed using the VannoPortal resource (<http://www.mulinlab.org/vportal/index.html>). In particular, the PhyloP, GerpN, and GerpS scores are provided to assess evolutionary conserva-

Table 1. The conditions for marker genotyping in the *SELE* gene

SNP	Primers and TaqMan probes*	Annealing temperature, °C
rs3917412	F: TGTAATTCTGTGTCCTGCG	55
	R: GGCTCATAGGTACACACTGGAA	
	5'-FAM-TCATTTTCATTCAAGCGACTTGCTCCAT-BHQ1-3'	
	5'-HEX-TCATTTTCATTCAAGTGACTTGCTCCAT-BHQ1-3'	
rs1534904	F: TAACTGAAGGCTCTGGGCTC	57
	R: AGACCACTCAGCATAGGCAAAG	
	5'-FAM-AACCACTGAGGATTTGAAAGAGCACCAT-BHQ1-3'	
	5'-HEX-AACCACTGAGGATTTTAAAGAGCACCAT-BHQ1-3'	
rs5353	F: AAGAAGGAAATCGTGGGTAGC	60
	R: TTCCCAAAACGTAAGTGC	
	5'-FAM-TAAGACTTTCATCATTTAGGTCAAAGAGAAA-BHQ1-3'	
	5'-HEX-TAAGACTTTCATTATTTAGGTCAAAGAGAAA-BHQ1-3'	

* Primers and samples were selected using the Vector NTI program (<http://www.informaxinc.com>).

tism. The PhyloP score is used to estimate evolutionary conservatism based on interspecies comparisons, with humans excluded from the analysis (the prefixes denote classification ranks, in this case: pri – primates) (Pollard et al., 2010; Caron et al., 2019). The GerpN and GerpS scores are based on single nucleotide analysis, namely the analysis of the homology of the locus across species (GerpN) and the analysis of the deficit or surplus in substitutions at the locus (GerpS) (Zerbino et al., 2018; Caron et al., 2019). The Phred Score is a measure of the quality of the score obtained (if Phred assigns a quality score of 20, the error probability is 1 %, score of 10 means an error probability of 10 %). To evaluate the selection effect, we use data from Neilsen’s CLR (Composite Likelihood Rate) test – Composite Likelihood method to determine the strength of positive selection (Vy, Kim, 2015).

Protein-protein interactions were evaluated using the BioGRID resource (<https://thebiogrid.org/>) (Oughtred et al., 2021). Functional enrichment analysis was performed using WebGestalt (WEB-based GENE SeT AnaLysis Toolkit) (<https://www.webgestalt.org/>) (Liao Y. et al., 2019).

Results

Genotyping of three intronic variants (rs5353, rs1534904, rs3917412) in the *SELE* gene was performed in the population sample of Tomsk and in the group of STEMI patients.

Linkage and association analysis

We revealed the complete linkage of the rs5353 marker with two other SNPs studied ($D' = 1$ in both patients and control sample). In turn, rs3917412 and rs1534904 are closely but not completely linked (D' is 0.916 in the patient group and 0.976 in the population sample). The linkage analysis of the studied markers indicates that the substitution in rs5353 occurred

within the context of the haplotype comprising the reference rs3917412 and rs1534904 alleles. As a result, we expect a multidirectional effect of the studied nucleotide substitutions (rs5353 on the one side, rs3917412 and rs1534904 on the other side) on the manifestation of pathological features.

We found statistically significant differences in the frequencies of rs5353 genotypes in the *SELE* gene between patients and the control sample ($p = 0.004$) (Table 2). According to logistic regression (Table 3), two models, codominant and recessive, were statistically significant. However, the information criteria (Akaike and Bayesian) are the lowest for the recessive model, which defines it as the best model. Consequently, the CC genotype was identified as a risk factor for myocardial infarction, occurring six times more frequently in the patient group (OR = 6.93; CI:95 % (1.84–26.04); $\chi^2 = 8.69$; $p = 0.002$) (Tables 2 and 3).

The studied samples exhibited statistically significant differences in the genotype combination ($\chi^2 = 22.76$; $df = 8$; $p = 0.004$). The observed differences can be attributed to two distinct combinations of genotypes (rs5353/rs1534904/rs3917412): CC/GG/CC predisposes to the development of myocardial infarction (OR = 6.93, CI:95 % (1.68–32.98), $\chi^2 = 8.69$, $p = 0.003$), whereas the TC/GG/CC combination is protective (OR = 0.38, CI:95 % (0.16–0.90), $\chi^2 = 5.01$, $p = 0.02$). So, the CGC haplotype is more prevalent in patients than in controls (27.7 % and 20.5 %, respectively) and the TGC haplotype is more prevalent in controls (40.4 % in patients and 49.2 % in controls), but these differences are not statistically significant.

In the sample of patients, a deviation from Hardy–Weinberg equilibrium was identified ($p = 0.012$). There is a deficiency in heterozygotes and frequent allele homozygotes, but an excess of rare homozygotes. The quality control of genotyping

Table 2. The frequencies of alleles and genotypes of markers in the *SELE* gene in the compared groups

SNP	Genotypes and rare allele	Frequencies of genotypes and rare allele, % (n)		χ^2, p
		in patients	in control sample	
rs5353	T/T	58.11 (43)	61.76 (84)	$\chi^2 = 10.85, p = 0.004$
	T/C	28.38 (21)	36.03 (49)	
	C/C	13.51 (10)	2.21 (3)	
	C allele	27.27 (41)	20.22 (55)	$\chi^2 = 2.630, p = 0.105$
rs1534904	G/G	45.95 (34)	50.74 (69)	$\chi^2 = 1.315, p = 0.518$
	G/T	47.30 (35)	39.71 (54)	
	T/T	6.76 (5)	9.56 (13)	
	T allele	30.41 (45)	29.41 (80)	$\chi^2 = 0.010, p = 0.919$
rs3917412	C/C	55.41 (41)	58.09 (79)	$\chi^2 = 0.196, p = 0.907$
	C/T	39.19 (29)	37.50 (51)	
	T/T	5.41 (4)	4.41 (6)	
	T allele	25.00 (37)	23.16 (63)	$\chi^2 = 0.092, p = 0.762$

Note. Statistically significant differences are highlighted in bold.

Table 3. Models of predisposing effect inheritance of rs5353 in the *SELE* gene

Model	Genotype	Control	STEMI	OR (95 % CI)	Statistical significance, <i>p</i> -value	AIC	BIC
Codominant	T/T	84 (61.8 %)	43 (58.1 %)	1.00	0.006	268.1	278.2
	T/C	49 (36 %)	21 (28.4 %)	0.84 (0.45–1.57)			
	C/C	3 (2.2 %)	10 (13.5 %)	6.51 (1.70–24.91)			
Dominant	T/T	84 (61.8 %)	43 (58.1 %)	1.00	0.610	276.3	283.0
	T/C-C/C	52 (38.2 %)	31 (41.9 %)	1.16 (0.65–2.07)			
Recessive	T/T-T/C	133 (97.8 %)	64 (86.5 %)	1.00	0.002	266.5	273.2
	C/C	3 (2.2 %)	10 (13.5 %)	6.93 (1.84–26.04)			
Overdominant	T/T-C/C	87 (64 %)	53 (71.6 %)	1.00	0.206	275.3	282.0
	T/C	49 (36 %)	21 (28.4 %)	0.70 (0.38–1.30)			

Note. Statistically significant differences are highlighted in bold.

(100 % re-genotyping of the patient sample) confirmed the correctness of the experiment. We can conclude that in this case there is a biological reason for the deviation from the Hardy–Weinberg equilibrium, given the *a priori* bias in the patient sample and the prevalence of the pathology-associated genotype. On the other hand, the studied sample of patients is not large, and thus the results obtained require further validation on larger samples.

Functional analysis of the studied markers

The studied markers are located in introns 1 (rs5353), 4 (rs1534904) and 5 (rs3917412) of the *SELE* gene. Notably, the degree of conservatism of the studied markers exhibits considerable variability. So, rs5353 is conservative only in primates (Phred Score = 11.93 for priPhyloP), rs1534904 is probably conservative not only in primates (Phred Score = 16.10 for priPhyloP), but also in various species in general (Phred Score = 16.58 and 11.11 for GerpN and GerpS, respectively), rs3917412 is not conservative (according to VannoPortal resource). All the studied substitutions are under positive selection (Nei's CLR test), according to information from the VannoPortal resource.

According to the RegulomeDB classification, all three of the studied markers belong to functional class 1f, i. e. they are eQTL variants in the transcription factor (TF) binding motif or in the DNAase hypersensitivity region. Indeed, the studied substitutions can theoretically (i. e. according to bioinformatic

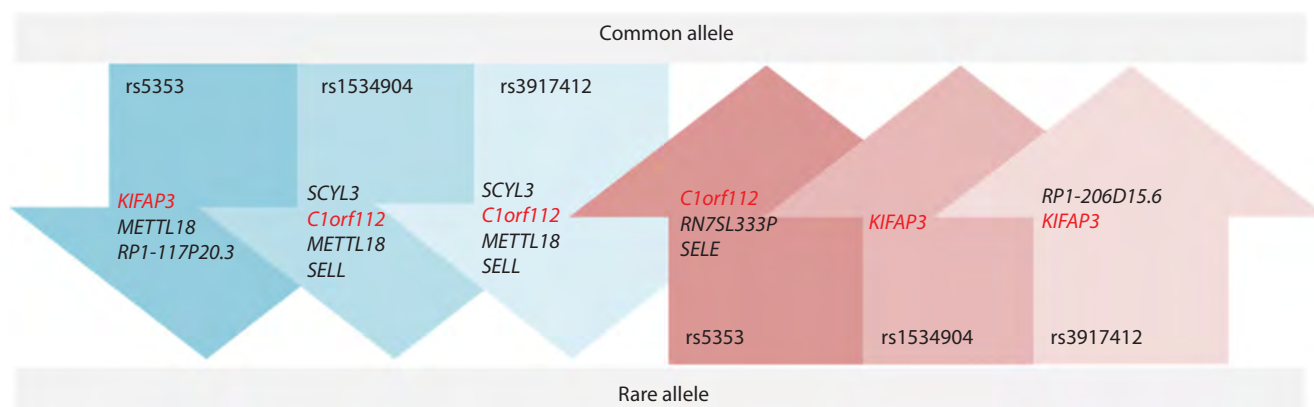
analysis) alter the affinity of a number of TFs (according to the VannoPortal resource). Thus, 24 TF binding sites have been identified for rs5353, with 17 of them being lost in the presence of an alternative allele. Nevertheless, according to experimental studies, physical interactions of these TFs with DNA in this region have not yet been detected in the studied tissues. Further seven substitutions resulted in the formation of novel transcription factor binding sites (ATF1, CEBPA, HOXA1, JUND, REST, JUNB) (Table 4). In the case of rs1534904, there are 11 theoretically variable sites, of which four are lost (not actually detected) in the presence of the alternative allele, but seven new sites emerge (PRDM1, RORC, IRF5, ZNF143, NCOR1, ZEB1, POU2F2) (Table 4). For rs3917412, theoretical calculations indicate the presence of 16 TF binding sites; the alternative allele results in the emergence of two novel sites (CHD2, PAX1) (Table 4), the disappearance of ten sites, and a decrease in affinity of four TFs. As in the previous cases, the list of theoretically variable TFs does not include physically detected interactions (according to the VannoPortal resource).

As eQTL variants, the studied markers are associated with changes in the expression levels of *SELE* and nearby genes (see the Figure); in addition, rs1534904 alters the splicing site of the *C1orf112* (*FIRRM*) gene (according to the GTEx Portal resource). The effect of the studied nucleotide substitutions can therefore be realized not only through the biochemical pathways involving E-selectin, but also through pathways

Table 4. Transcription factors for which binding sites emerge in the presence of rare alleles of the studied markers

SNP, allele	Transcription Factors
rs5353, C	ATF1, CEBPA, HOXA1, JUND, REST, JUNB
rs1534904, T	PRDM1, RORC, IRF5, ZNF143, NCOR1, ZEB1, POU2F2
rs3917412, T	CHD2, PAX1

Note. TFs expressed in vascular walls are highlighted in bold (compiled according to the VannoPortal and GTEx Portal resources).



Scheme of alterations in the expression levels of a number of genes depending on rs5353, rs1534904, rs3917412 in the *SELE* gene.
Compiled according to the GTEx Portal resource.

involving other regulated genes. To date, six regulated genes have been identified for rs5353 and rs3917412 and five for rs1534904; a total of nine genes (*FIRRM*, *KIFAP3*, *METTL18*, *RN7SL333P*, *RP1-206D15.6*, *RP1-117P20.3*, *SCYL3*, *SELE*, *SELL*) have been identified (GTEx Portal). The *FIRRM* and *KIFAP3* genes seem to be the most interesting, as the effect of the examined substitutions in the *SELE* gene on their expression corresponds with the linkage outcomes: while *KIFAP3* expression is reduced in homozygotes for the rare allele rs5353, it is elevated in homozygotes for the rs1534904 and rs3917412 derived alleles. Conversely, *FIRRM* expression is lower in carriers of the rs1534904 and rs3917412 derived alleles, but higher in carriers of the rare rs5353 allele (see the Figure).

In light of these observations, three proteins appear to be the most promising in terms of realizing the functional effect of the studied substitutions. For E-selectin, a small number of interactors are detected: there are only 19 interacting molecules including two chemical compounds and 17 proteins (according to the BioGRID resource). The WebGestalt enrichment analysis indicates that the most significant processes, with the largest number of interactor proteins involved, are hemostasis (10 proteins, $p_{adj} = 6.4E^{-7}$), platelet activation, signaling and aggregation (6 proteins, $p_{adj} = 3.6E^{-4}$) and cell surface interactions at the vascular wall (5 proteins, $p_{adj} = 3.6E^{-4}$). The results are in complete alignment with the known biochemical functions of E-selectin. For FIGNL1 Interacting Regulator of Recombination and Mitosis encoded by the *FIRRM* (*C1orf112*) gene, 60 interactor proteins are identified (according to the BioGRID resource). The most significant processes with the largest number of interactors involved include serpentine receptor ligand binding (19 proteins, $p_{adj} = 3.8E^{-13}$) and various types of signaling through these receptors (16–20 proteins each, $p_{adj} = 4.5E^{-12}$ – $4.2E^{-7}$) (according to WebGestalt). The protein product of the *FIRRM* gene has not yet been sufficiently studied, but in addition to its originally described role as a kinetochore protein (Xu et al., 2021), it is also known to be involved in DNA repair processes (Mazouzi et al., 2023; Pinedo-Carpio et al., 2023; Tischler et al., 2024). A total of 105 interactor proteins have

been identified for the *KIFAP3* gene product (kinesin-2 associated protein) according to the BioGRID resource. The results of the enrichment analysis indicate that the most significant processes with the highest number of interactors involved (by WebGestalt) are the adaptive immune response (16 proteins, $p_{adj} = 6.2E^{-3}$), different mRNA processing pathways (7–9 proteins, $p_{adj} = 6.2E^{-3}$ – $6.4E^{-3}$), and antigen presentation via major histocompatibility complex class II (7 proteins, $p_{adj} = 6.2E^{-3}$). These processes have the highest number of interactors involved, as determined by WebGestalt.

Discussion

Thus, all three proteins encoded by the *SELE*, *FIRRM*, and *KIFAP3* genes (co-regulated by the studied intronic variants) are not directly involved in the development of cardiovascular events but may be indirectly involved in the pathogenesis of CVDs through the inflammation, immune response, and DNA repair systems.

Of the nine genes co-regulated by the studied markers, four genes showed no associations with pathologies or quantitative traits identified by GWAS. Three of them do not encode proteins: two genes of long non-coding RNAs (*RP1-206D15.6*, *RP1-117P20.3*) and a pseudogene (*RN7SL333P*); and one gene (*METTL18*) encodes a methyltransferase. For the remaining five genes, GWAS demonstrated a multitude of associations (Supplementary Material, compiled from GWAS catalog)¹. Three of them (*FIRRM*, *KIFAP3*, *SELE*) are discussed in detail above; the *SELL* gene encodes L-selectin, which promotes leukocyte rolling as well as E-selectin (GeneCard, <https://www.genecards.org/>); the *SCYL3* gene encodes a pseudokinase that also plays a role in cell adhesion and migration (GeneCard, <https://www.genecards.org/>). Markers in the region of location of these five genes showed associations with various blood biochemical parameters, cell composition of blood (*FIRRM*, *KIFAP3*, *SCYL3*, *SELE*, *SELL*), as well as with amyotrophic lateral sclerosis and venous thromboembolism (*FIRRM*, *KIFAP3*), type 2 diabetes mellitus (*FIRRM*, *SELL*) (Supplementary Material). We would like to emphasize that among

¹ Supplementary Material is available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Babush_Engl_29_1.pdf

the 64 SNPs associated with various pathologies or quantitative traits according to GWAS results, there are neither the markers studied in the present study (rs5353, rs1534904 and rs3917412), nor the SNPs associated with cardiovascular diseases according to earlier studies (rs1805193, rs5361, rs5355).

It should be noted that the polymorphic variants analyzed in this study had not been previously studied in the context of CVDs and had seldom been included in associative studies (there is no information on these markers in the PubMed, DisGeNet, GWAS Catalog databases). Only one study has been found in the available literature that has demonstrated the risk effect of the GG rs3917412 genotype on the development of colon cancer (Custodio et al., 2014). At the same time, genomic estimates of pathogenicity (which reflect the probability of marker involvement in the development of multifactorial pathology, estimated by regBase (Zhang S. et al., 2019)) indicate that the analyzed variants may be involved in pathological processes. Additionally, the oncogenicity estimates indicate the probability of a “driver” effect (*likely cancer driver*) of these nucleotide substitutions for the development of oncopathology (according to the VannoPortal resource).

Conclusion

Thus, the mechanism of E-selectin involvement in the pathogenesis of STEMI is not fully understood. On the one hand, a sufficient amount of biochemical data indicates its involvement in the development of CVDs (Liao B. et al., 2016; Deng et al., 2017; Vargas-Alarcon et al., 2019; Ding et al., 2021), and primarily in the development of such IHD risk factors as atherosclerosis and DM2 (Roldán et al., 2003; McEver, 2015; Qiu et al., 2019; Mathur et al., 2023). On the other hand, the involvement of this protein in the inflammatory response suggests its involvement primarily in the recovery processes after a cardiovascular event, and only indirectly in the development of susceptibility to myocardial infarction (Ueno, 2012; Sandoval-Pinto et al., 2014; Srivastava et al., 2018). The association of rs5353 with STEMI revealed in the present study provides further confirmation of the involvement of the *SELE* gene in the development of CVDs. Additionally, the analysis demonstrates the presence of a region more extensive than one gene, which is co-regulated by the nucleotide substitutions studied. It is possible that this entire genome region may be involved in the pathogenesis of CVD indirectly, through the inflammation, immune response, and DNA repair systems.

References

- 2020 Clinical practice guidelines for Acute ST-segment elevation myocardial infarction. *Rossiskiy Kardiologicheskii Zhurnal = Russ J Cardiol.* 2020;25(11):251-310. doi 10.15829/1560-4071-2020-4103 (in Russian)
- Barrett J.C., Fry B., Maller J., Daly M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2): 263-265. doi 10.1093/bioinformatics/bth457
- Blankenberg S., Barbaux S., Tiret L. Adhesion molecules and atherosclerosis. *Atherosclerosis.* 2003;170(2):191-203. doi 10.1016/s0021-9150(03)00097-2
- Calder P.C., Ahluwalia N., Albers R., Bosco N., Bourdet-Sicard R., Haller D., Holgate S.T., Jönsson L.S., Latulippe M.E., Marcos A., Moreines J., M'Rini C., Müller M., Pawelec G., van Neerven R.J., Watzl B., Zhao J. A consideration of biomarkers to be used for evaluation of inflammation in human nutritional studies. *Br J Nutr.* 2013;109(Suppl.1):S1-S34. doi 10.1017/S0007114512005119
- Caron B., Luo Y., Rausell A. NCBoost classifies pathogenic non-coding variants in Mendelian diseases through supervised learning on purifying selection signals in humans. *Genome Biol.* 2019;20(1):32. doi 10.1186/s13059-019-1634-2
- Cid M.C., Cebrián M., Font C., Coll-Vinent B., Hernández-Rodríguez J., Esparza J., Urbano-Márquez A., Grau J.M. Cell adhesion molecules in the development of inflammatory infiltrates in giant cell arteritis: inflammation-induced angiogenesis as the preferential site of leukocyte-endothelial cell interactions. *Arthritis Rheum.* 2000;43(1): 184-194. doi 10.1002/1529-0131(200001)43:1<184::AID-ANR23>3.0.CO;2-N
- Custodio A., Moreno-Rubio J., Aparicio J., Gallego-Plazas J., Yaya R., Maurel J., Rodríguez-Salas N., Burgos E., Ramos D., Calatrava A., Andradá E., Díaz-López E., Sánchez A., Madero R., Cejas P., Feliu J. Pharmacogenetic predictors of outcome in patients with stage II and III colon cancer treated with oxaliplatin and fluoropyrimidine-based adjuvant chemotherapy. *Mol Cancer Ther.* 2014;13(9):2226-2237. doi 10.1158/1535-7163.MCT-13-1109
- Deng M.H., Lin C.W., Sun Y.N., Zeng X.L., Wen F. Role of E-selectin for diagnosing myocardial injury in paediatric patients with mycoplasma pneumoniae pneumonia. *Ann Clin Biochem.* 2017;54(1): 49-54. doi 10.1177/0004563216631570
- Ding G., Wang J., Liu K., Huang B., Deng W., He T. Association of E-selectin gene rs5361 polymorphism with ischemic stroke susceptibility: a systematic review and meta-analysis. *Int J Neurosci.* 2021; 131(5):511-517. doi 10.1080/00207454.2020.1750385
- Habas K., Shang L. Alterations in intercellular adhesion molecule 1 (ICAM-1) and vascular cell adhesion molecule 1 (VCAM-1) in human endothelial cells. *Tissue Cell.* 2018;54:139-143. doi 10.1016/j.tice.2018.09.002
- Kachkovsky M.A., Ragozina E.Yu. Assessment of systemic inflammatory reaction in acute myocardial infarction: status update on the problem. *Ratsionalnaya Farmakoterapiya v Kardiologii = Ration Pharmacother Cardiol.* 2013;9(6):690-697 (in Russian)
- Kalinin R.E., Korotkova N.V., Suchkov I.A., Mzhavanadze N.D., Ryabkov A.N. Selectins and their involvement in the pathogenesis of cardiovascular diseases. *Kazan Med J.* 2022;103(4):617-627. doi 10.17816/KMJ2022-617
- Lampas S., Tsaplaris P., Pantelidis P., Oikonomou E., Marinos G., Charalambous G., Souvliotis N., Mystakidi V.C., Goliopoulou A., Katsianos E., Siasos G., Vavuranakis M.A., Tsioufis C., Vavuranakis M., Tousoulis D. The role of endothelial related circulating biomarkers in COVID-19. A systematic review and meta-analysis. *Curr Med Chem.* 2022;29(21):3790-3805. doi 10.2174/0929867328666211026124033
- Liao B., Chen K., Xiong W., Chen R., Mai A., Xu Z., Dong S. Relationship of *SELE* A561C and G98T variants with the susceptibility to CAD. *Medicine (Baltimore).* 2016;95(8):e1255. doi 10.1097/MD.0000000000001255
- Liao Y., Wang J., Jaehnig E.J., Shi Z., Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 2019;47(W1):W199-W205. doi 10.1093/nar/gkz401
- Lorenzon P., Vecile E., Nardon E., Ferrero E., Harlan J.M., Tedesco F., Dobrina A. Endothelial cell E- and P-selectin and vascular cell adhesion molecule-1 function as signaling receptors. *J Cell Biol.* 1998; 142(5):1381-1391. doi 10.1083/jcb.142.5.1381
- Mallik S., Majumder P.P. A two-step genetic study on quantitative precursors of coronary artery disease in a homogeneous Indian population: case-control association discovery and validation by transmission-disequilibrium test. *J Biosci.* 2011;36(5):857-868. doi 10.1007/s12038-011-9148-4
- Mangoni A.A., Zinellu A. A systematic review and meta-analysis of circulating adhesion molecules in rheumatoid arthritis. *Inflamm Res.* 2024;73(3):305-327. doi 10.1007/s00011-023-01837-6

- Mathur R., Ahmid Z., Ashor A.W., Shannon O., Stephan B.C.M., Siervo M. Effects of dietary-based weight loss interventions on biomarkers of endothelial function: a systematic review and meta-analysis. *Eur J Clin Nutr.* 2023;77(10):927-940. doi 10.1038/s41430-023-01307-6
- Mazouzi A., Moser S.C., Abascal F., van den Broek B., Del Castillo Velasco-Herrera M., van der Heijden I., Hekkelman M., Drenth A.P., van der Burg E., Kroese L.J., Jalink K., Adams D.J., Jonkers J., Brummelkamp T.R. FIRRMC1orf112 mediates resolution of homologous recombination intermediates in response to DNA inter-strand crosslinks. *Sci Adv.* 2023;9(22):eadf4409. doi 10.1126/sciadv.adf4409
- McEver R.P. Selectins: initiators of leucocyte adhesion and signalling at the vascular wall. *Cardiovasc Res.* 2015;107(3):331-339. doi 10.1093/cvr/cvv154
- Oughtred R., Rust J., Chang C., Breitkreutz B.J., Stark C., Willems A., Boucher L., Leung G., Kolas N., Zhang F., Dolma S., Coulombe-Huntington J., Chatri-Aryamontri A., Dolinski K., Tyers M. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 2021;30(1):187-200. doi 10.1002/pro.3978
- Pinedo-Carpio E., Dessapt J., Beneyton A., Sacre L., Bérubé M.A., Vil-lot R., Lavoie E.G., Coulombe Y., Blondeau A., Boulais J., Malina A., Luo V.M., Lazaratos A.M., Côté J.F., Mallette F.A., Guarné A., Masson J.Y., Fradet-Turcotte A., Orthwein A. FIRRMC cooperates with FIGNL1 to promote RAD51 disassembly during DNA repair. *Sci Adv.* 2023;9(32):eadf4082. doi 10.1126/sciadv.adf4082
- Pollard K.S., Hubisz M.J., Rosenbloom K.R., Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 2010;20(1):110-121. doi 10.1101/gr.097857.109
- Pruenster M., Immler R., Roth J., Kuchler T., Bromberger T., Napoli M., Nussbaumer K., Rohwedder I., Wackerbarth L.M., Piantoni C., Hennis K., Fink D., Kallabis S., Schroll T., Masgrau-Alsina S., Budke A., Liu W., Vestweber D., Wahl-Schott C., Roth J., Meissner F., Moser M., Vogl T., Hornung V., Broz P., Sperandio M. E-selectin-mediated rapid NLRP3 inflammasome activation regulates S100A8/S100A9 release from neutrophils via transient gasdermin D pore formation. *Nat Immunol.* 2023;24(12):2021-2031. doi 10.1038/s41590-023-01656-1
- Qin L., Zhao P., Liu Z., Chang P. Associations SELE gene haplotype variant and hypertension in Mongolian and Han populations. *Intern Med.* 2015;54(3):287-293. doi 10.2169/internalmedicine.54.2797
- Qiu S., Cai X., Liu J., Yang B., Zügel M., Steinacker J.M., Sun Z., Schumann U. Association between circulating cell adhesion molecules and risk of type 2 diabetes: a meta-analysis. *Atherosclerosis.* 2019;287:147-154. doi 10.1016/j.atherosclerosis.2019.06.908
- Roldán V., Marín F., Lip G.Y., Blann A.D. Soluble E-selectin in cardiovascular disease and its risk factors. A review of the literature. *Thromb Haemost.* 2003;90(6):1007-1020. doi 10.1160/TH02-09-0083
- Sambrook J., Russell D.W. Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harbor Protocols.* 2006;2006(1):pdb.prot4455. doi 10.1101/pdb.prot4455
- Sandoval-Pinto E., Padilla-Gutiérrez J.R., Valdes-Alvarado E., García-González I.J., Valdez-Haro A., Muñoz-Valle J.F., Flores-Salinas H.E., Rivas F., Valle Y. Assessment of the E-selectin rs5361 (561A>C) polymorphism and soluble protein concentration in acute coronary syndrome: association with circulating levels. *Mediators Inflamm.* 2014;2014:158367. doi 10.1155/2014/158367
- Shirakawa T., Ikeda K., Nishimura S., Kuniba H., Nakashima K., Motomura H., Mizuno Y., Zaitsum M., Nakazato M., Maeda T., Hama-saki Y., Hara T., Moriuchi H. Lack of an association between E-selectin gene polymorphisms and risk of Kawasaki disease. *Pediatr Int.* 2012;54(4):455-460. doi 10.1111/j.1442-200X.2012.03608.x
- Silva M., Videira P.A., Sackstein R. E-selectin ligands in the human mononuclear phagocyte system: implications for infection, inflammation, and immunotherapy. *Front Immunol.* 2018;8:1878. doi 10.3389/fimmu.2017.01878
- Srivastava K., Chandra S., Narang R., Bhatia J., Saluja D. E-selectin gene in essential hypertension: a case-control study. *Eur J Clin Invest.* 2018;48(1):e12868. doi 10.1111/eci.12868
- Thygesen K., Alpert J.S., Jaffe A.S., Chaitman B.R., Bax J.J., Morrow D.A., White H.D.; Executive Group on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction. Fourth universal definition of myocardial infarction (2018). *J Am Coll Cardiol.* 2018;72(18):2231-2264. doi 10.1016/j.jacc.2018.08.1038
- Tischler J.D., Tsuchida H., Bosire R., Oda T.T., Park A., Adeyemi R.O. FLIP(C1orf112)-FIGNL1 complex regulates RAD51 chromatin association to promote viability after replication stress. *Nat Commun.* 2024;15(1):866. doi 10.1038/s41467-024-45139-9
- Ueno T. E-selectin gene and essential hypertension. *Hypertens Res.* 2012;35(4):380. doi 10.1038/hr.2011.223
- Uy G.L., DeAngelo D.J., Lozier J.N., Fisher D.M., Jonas B.A., Mag-nani J.L., Becker P.S., Lazarus H.M., Winkler I.G. Targeting hematologic malignancies by inhibiting E-selectin: a sweet spot for AML therapy? *Blood Rev.* 2024;65:101184. doi 10.1016/j.blr.2024.101184
- Vargas-Alarcon G., Perez-Mendez O., Herrera-Maya G., Posadas-Romero C., Posadas-Sanchez R., Ramirez-Bello J., Escobedo G., Fragoso J.M. The rs1805193, rs5361, and rs5355 single nucleotide polymorphisms in the *E-selectin* gene (*SEL-E*) are associated with subclinical atherosclerosis: The Genetics of Atherosclerotic Disease (GEA) Mexican study. *Immunobiology.* 2019;224(1):10-14. doi 10.1016/j.imbio.2018.11.003
- Vestweber D., Blanks J.E. Mechanisms that regulate the function of the selectins and their ligands. *Physiol Rev.* 1999;79(1):181-213. doi 10.1152/physrev.1999.79.1.181
- Vy H.M., Kim Y. A composite-likelihood method for detecting incomplete selective sweep from population genomic data. *Genetics.* 2015;200(2):633-649. doi 10.1534/genetics.115.175380
- Wang K., Lei L., Li G., Lan Y., Wang W., Zhu J., Liu Q., Ren L., Wu S. Association between ambient particulate air pollution and soluble biomarkers of endothelial function: a meta-analysis. *Toxics.* 2024;12(1):76. doi 10.3390/toxics12010076
- Wang N., Chintala S.K., Fini M.E., Schuman J.S. Activation of a tissue-specific stress response in the aqueous outflow pathway of the eye defines the glaucoma disease phenotype. *Nat Med.* 2001;7(3):304-309. doi 10.1038/85446
- Wang X., Zhang J., Du X., Song M., Jia C., Liu H. Association of A561C and G98T polymorphisms in E-selectin gene with coronary artery disease: a meta-analysis. *PLoS One.* 2013;8(11):e79301. doi 10.1371/journal.pone.0079301
- Wang Z., Xu Y., Chen S., Wang L., Ding H., Lu G., Wang D., Zhai Z., Duan J., Zhang W. A common missense single nucleotide polymorphism in the E-selectin gene is significantly associated with essential hypertension in the Han population but only weakly associated in the Uyghur population. *Hypertens Res.* 2012;35(4):413-417. doi 10.1038/hr.2011.204
- Wenzel K., Felix S., Kleber F.X., Brachold R., Menke T., Schattke S., Schulte K.L., Gläser C., Rohde K., Baumann G., Speer A. E-selectin polymorphism and atherosclerosis: an association study. *Hum Mol Genet.* 1994;3(11):1935-1937. doi 10.1093/hmg/3.11.1935
- Xu L., Ali M., Duan W., Yuan X., Garba F., Mullen M., Sun B., Poser I., Duan H., Lu J., Tian R., Ge Y., Chu L., Pan W., Wang D., Hyman A., Green H., Li L., Dou Z., Liu D., Liu X., Yao X. Feedback control of PLK1 by Apolol ensures accurate chromosome segregation. *Cell Rep.* 2021;36(2):109343. doi 10.1016/j.celrep.2021.109343
- Yoshida M., Takano Y., Sasaoka T., Izumi T., Kimura A. E-selectin polymorphism associated with myocardial infarction causes en-

- hanced leukocyte-endothelial interactions under flow conditions. *Arterioscler Thromb Vasc Biol.* 2003;23(5):783-788. doi 10.1161/01.ATV.0000067427.40133.59
- Zak I., Sarecka B., Krauze J. Synergistic effects between 561A > C and 98G > T polymorphisms of E-selectin gene and hypercholesterolemia in determining the susceptibility to coronary artery disease. *Heart Vessels.* 2008;23(4):257-263. doi 10.1007/s00380-008-1040-2
- Zerbino D.R., Achuthan P., Akanni W., Amode M.R., Barrell D., Bhai J., Billis K., ... Trevanion S.J., Aken B.L., Cunningham F., Yates A., Flicek P. Ensembl 2018. *Nucleic Acids Res.* 2018;46(D1):D754-D761. doi 10.1093/nar/gkx1098
- Zhang N., Aiyasiding X., Li W.J., Liao H.H., Tang Q.Z. Neutrophil degranulation and myocardial infarction. *Cell Commun Signal.* 2022; 20(1):50. doi 10.1186/s12964-022-00824-4
- Zhang S., He Y., Liu H., Zhai H., Huang D., Yi X., Dong X., Wang Z., Zhao K., Zhou Y., Wang J., Yao H., Xu H., Yang Z., Sham P.C., Chen K., Li M.J. regBase: whole genome base-wise aggregation and functional prediction for human non-coding regulatory variants. *Nucleic Acids Res.* 2019;47(21):e134. doi 10.1093/nar/gkz774
- Zhao D.X., Feng J., Cong S.Y., Zhang W. Association of E-selectin gene polymorphisms with ischemic stroke in a Chinese Han population. *J Neurosci Res.* 2012;90(9):1782-1787. doi 10.1002/jnr.23075
- Zheng F., Chevalier J.A., Zhang L.Q., Virgil D., Ye S.Q., Kwitovich P.O. An HphI polymorphism in the E-selectin gene is associated with premature coronary artery disease. *Clin Genet.* 2001; 59(1):58-64. doi 10.1034/j.1399-0004.2001.590110.x
- Zhito A.V., Iusupova A.O., Privalova E.V., Khabarova N.V., Belenkov Y.N. Markers of endothelial dysfunction: E-selectin, endothelin-1 and von Willebrand factor in patients with coronary heart disease, including in combination with type 2 diabetes mellitus. *Ratsionalnaya Farmakoterapia v Kardiologii = Ration Pharmacother Cardiol.* 2019;15(6):892-899. doi 10.20996/1819-6446-2019-15-6-892-899 (in Russian)

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

Received August 27, 2024. Revised October 18, 2024. Accepted October 21, 2024.


doi 10.18699/vjgb-25-17

Comparative analysis of haplotypes carrying pathogenic variants c.1545T>G, c.2027T>A and c.919-2A>G of the *SLC26A4* gene in patients with hearing loss from the Tyva Republic (Southern Siberia)

V.Yu. Danilchenko ^{1, 2}, M.V. Zysar ^{1, 2}, E.A. Panina ^{1, 2}, K.E. Orishchenko ^{1, 2}, O.L. Posukh ^{1, 2} 

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

² Novosibirsk State University, Novosibirsk, Russia

 posukh@bionet.nsc.ru

Abstract. Pathogenic variants in the *SLC26A4* gene (OMIM #605646), leading to non-syndromic recessive hearing loss type 4 (DFNB4) and Pendred syndrome, significantly contribute to the etiology of hearing loss in many populations of the world. The spectrum and prevalence of different pathogenic *SLC26A4* variants are characterized by wide ethno-geographical variability. A high frequency of some of them in certain regions of the world may indicate either their independent origin or be a consequence of the founder effect. The proportion of *SLC26A4*-associated hearing loss in Tuvian patients (the Tyva Republic, Southern Siberia) is one of the highest in the world (28.2 %) and the vast majority of mutant *SLC26A4* alleles are represented by three pathogenic variants c.919-2A>G, c.2027T>A and c.1545T>G (69.3, 17.5 and 8.0 %, respectively). Their overall carrier frequency in the Tuvian population reaches 7.1 %. The accumulation of these variants in Tuvian patients suggests a role of the founder effect in their prevalence in Tuva, which can be confirmed by the common genetic background (haplotypes) for each of them. For reconstruction of haplotypes in the carriers of variants c.1545T>G and c.2027T>A, the genotyping data of a panel of polymorphic genetic markers were used: five STRs (four of them flank the *SLC26A4* gene at different distances and one is intragenic) and nine intragenic SNPs. Comparative analysis of the reconstructed haplotypes for c.1545T>G and c.2027T>A with previously obtained data on haplotypes for the c.919-2A>G variant showed that each of the analyzed variants has a specific (similar for all carriers of a particular variant) genetic background, apparently inherited from different "founder ancestors". These data confirm the cumulative founder effect in the prevalence of pathogenic variants c.1545T>G, c.2027T>A, and c.919-2A>G of the *SLC26A4* gene in the indigenous population of the Tyva Republic. The obtained data are relevant both for predicting the prevalence of *SLC26A4*-caused hearing loss and for development of region-specific DNA diagnostics of inherited hearing loss in the Tyva Republic.

Key words: hearing loss; *SLC26A4*; pathogenic variants; STRs; SNPs; haplotypes; founder effect; Siberian populations.

For citation: Danilchenko V.Yu., Zysar M.V., Panina E.A., Orishchenko K.E., Posukh O.L. Comparative analysis of haplotypes carrying pathogenic variants c.1545T>G, c.2027T>A and c.919-2A>G of the *SLC26A4* gene in patients with hearing loss from the Tyva Republic (Southern Siberia). *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025; 29(1):144-152. doi 10.18699/vjgb-25-17


Funding. The study was carried out with the financial support of the budget project FWNR-2022-0021 of the Institute of Cytology and Genetics SB RAS and the grant FSUS-2024-0018 of the Ministry of Science and Higher Education of the Russian Federation.

Сравнительный анализ гаплотипов, несущих патогенные варианты с.1545T>G, с.2027T>A и с.919-2A>G гена *SLC26A4*, у пациентов с потерей слуха из Республики Тыва (Южная Сибирь)

В.Ю. Данильченко ^{1, 2}, М.В. Зыцарь ^{1, 2}, Е.А. Панина ^{1, 2}, К.Е. Орищенко ^{1, 2}, О.Л. Посух ^{1, 2} 

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 posukh@bionet.nsc.ru

Аннотация. Патогенные варианты в гене *SLC26A4* (OMIM #605646), приводящие к несиндромальной рецессивно наследуемой потере слуха 4-го типа (DFNB4) и синдрому Пендреда, вносят весомый вклад в этиологию потери слуха во многих популяциях мира. Спектр и распространенность различных патогенных вариантов гена *SLC26A4* характеризуются широкой этногеографической вариабельностью. Высокая частота некоторых из них в отдельных регионах мира может свидетельствовать об их независимом возникновении или же быть следствием эффекта основателя. Доля *SLC26A4*-ассоциированной потери слуха у тувинских пациентов (Республика Тыва, Южная Сибирь) является одной из самых высоких в мире (28.2 %). Подавляющее большинство мутантных *SLC26A4*-аллелей представлено тремя патогенными вариантами – с.919-2A>G, с.2027Т>A и с.1545Т>G (69.3, 17.5 и 8.0 % соответственно). Суммарная частота их гетерозиготного носительства в тувинской популяции достигает 7.1 %. Накопление этих вариантов у тувинских пациентов позволяет предположить роль эффекта основателя в их распространенности на территории Тувы, что может быть подтверждено общностью генетического окружения (гаплотипов) для каждого из них. Для реконструкции гаплотипов у носителей вариантов с.1545Т>G и с.2027Т>A были использованы данные генотипирования панели полиморфных генетических маркеров: пяти STR-маркеров (четыре из них фланкируют на разном расстоянии ген *SLC26A4* и один является внутригенным) и девяти внутригенных SNP-маркеров. Сравнительный анализ реконструированных гаплотипов для с.1545Т>G и с.2027Т>A с ранее полученными данными о гаплотипах для с.919-2A>G показал, что каждый из анализируемых вариантов имеет особое и сходное для всех носителей того или иного варианта генетическое окружение, по-видимому, унаследованное от различных «предков-основателей». Эти данные подтверждают роль кумулятивного эффекта основателя в распространенности патогенных вариантов с.1545Т>G, с.2027Т>A и с.919-2A>G гена *SLC26A4* у коренного населения Республики Тыва. Полученные данные актуальны как для прогнозирования распространенности *SLC26A4*-обусловленной потери слуха, так и для создания регион-специфичной ДНК-диагностики наследуемой потери слуха в Республике Тыва.

Ключевые слова: потеря слуха; *SLC26A4*; патогенные варианты; STR; SNP; гаплотипы; эффект основателя; популяции Сибири.

Introduction

Currently, more than 5 % of the world's population has severe or profound hearing loss, caused by both environmental and genetic factors (World Health Organization, <https://www.who.int/ru>). Genetic factors underlie more than half of all cases of congenital (or early manifestation) pathology of auditory function. Hereditary hearing loss can be one of the clinical features of many (about 400) syndromes, or an isolated (non-syndromic) pathology, which is characterized by unique genetic heterogeneity: about 200 loci have already been mapped and at least 150 genes associated with hearing loss have been identified (Hereditary Hearing Loss Homepage: <https://hereditaryhearingloss.org>, April 2024).

Wide ethnogeographical variability is known in the prevalence of various forms of inherited hearing loss caused by pathogenic variants in different “deafness genes”. The “accumulation” of some forms of hereditary hearing loss in a certain population, like a number of other monogenic diseases, can be determined by the ethnic composition of the population, isolation, marital structure, founder and bottleneck effects, as well as a possible selective advantage of heterozygotes (Scott et al., 1995; Ben Arab et al., 2004; Common et al., 2004; Zlotogora, 2007; Chong et al., 2012; Razdan et al., 2012). An important role in the prevalence of hereditary forms of deafness was probably also played by such a social factor as the long-term tradition of assortative marriages between deaf people, based on their linguistic homogamy (sign language), which led to an increase in social adaptation and genetic fitness of deaf people (Nance et al., 2000; Nance, Kearsy, 2004).

Identification of the most frequent (major) mutations in genes involved in hearing loss is an urgent task for both genetic risk assessment and medical and genetic counseling of affected families, and for developing the most effective methods of molecular diagnostics of this pathology. The high

frequency of some mutations in certain regions of the world may either indicate their independent occurrence (mutational “hot spot”) or be a consequence of the founder effect (founder mutation). The founder effect in the prevalence of mutations can be confirmed by the similarity of their genetic background (haplotypes). Haplotype reconstruction is usually carried out based on the analysis of highly polymorphic genetic markers: STRs (Short Tandem Repeats) and SNPs (Single Nucleotide Polymorphisms). Analysis of haplotypes carrying a particular mutation can allow to estimate its “age” (time of occurrence) using a “molecular clock” approach, and, in some cases, identify potential regions of its origin using information about population history.

The most significant contribution to the etiology of hearing loss in many populations of the world is made by pathogenic variants in the *GJB2* gene (OMIM #121011). The *SLC26A4* gene (solute carrier family 26, member 4, 7q22.3, OMIM #605646) is the second most significant gene, at least for Asian populations. *SLC26A4* encodes the transmembrane transport protein pendrin, which is involved in the transport of various ions and is mainly expressed in the inner ear, thyroid gland and kidneys. Pathogenic variants in the *SLC26A4* gene lead to non-syndromic recessive hearing loss (type DFNB4) and Pendred Syndrome (OMIM #274600), a recessive disease characterized by hearing loss and goiter. Patients with *SLC26A4*-associated hearing loss often have abnormalities of the bony labyrinth of the inner ear (enlarged vestibular aqueduct, Mondini dysplasia). Numerous studies have found that the prevalence of *SLC26A4*-associated hearing loss and the spectrum of pathogenic variants of this gene vary significantly in different regions of the world. It has now become apparent that the spectrum of pathogenic variants of the *SLC26A4* gene found in Asian populations is significantly different from that in populations of European origin (Park et al., 2003; Albert

et al., 2006; Du et al., 2013; Lu Y.J. et al., 2015; Tsukada et al., 2015).

Analysis of the *SLC26A4* gene, conducted during long-term studies of inherited hearing loss in Tuvinians, the indigenous population of the Tyva Republic (Southern Siberia), showed that the proportion of *SLC26A4*-associated hearing loss in Tuvinian patients is one of the highest in the world (28.2 %) (Danilchenko et al., 2021). A specific spectrum of variations in the *SLC26A4* gene was identified in Tuvinian patients, including both known pathogenic variants and a number of new variants with still uncertain clinical significance. The vast majority of mutant *SLC26A4* alleles identified in patients were represented by three pathogenic variants c.919-2A>G, c.2027T>A and c.1545T>G (69.3 %, 17.5 %, and 8.0 %, respectively), and their overall carrier frequency reached 7.1 % in the Tuvinian control sample (Danilchenko et al., 2021). The predominance of variant c.919-2A>G suggested the role of the founder effect in its accumulation in Tuvinians, and in our recent study (Danilchenko et al., 2023), we identified a similarity of STR and SNP haplotypes in all c.919-2A>G carriers, which convincingly indicates its origin from a common ancestor, thereby confirming the decisive role of the founder effect in the prevalence of this pathogenic *SLC26A4* variant in the indigenous population of the Tyva Republic.

The aim of this study is a comparative analysis of the genetic background of pathogenic variants c.1545T>G, c.2027T>A, and c.919-2A>G of the *SLC26A4* gene, identified with high frequency in the indigenous population of the Tyva Republic.

Materials and methods

Analyzed samples. Genotyping of genetic markers (STRs and SNPs) for the analysis of haplotypes of the chromosome 7 region including the *SLC26A4* gene was performed on the sample of Tuvinian patients having variant c.2027T>A in a homozygous ($n = 4$) or compound heterozygous ($n = 15$) state, or variant c.1545T>G in a compound heterozygous state ($n = 15$). For comparative analysis, previously obtained data on the structure of haplotypes in Tuvinian patients homozygous for the c.919-2A>G variant of the *SLC26A4* gene ($n = 23$) and in individuals from the control sample represented by unrelated Tuvinians ($n = 63$) were used (Danilchenko et al., 2023). Written informed consent was obtained from all individuals or their legal guardians before venous blood sampling for DNA extraction. The study was approved by the Bioethics Commission at the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia).

Historical information about the Tuvinian population. Tuvinians (Tuvans), numbering about 300,000 people in total (according to the 2021 All-Russian Census), currently live mainly in the Tyva Republic, which borders Mongolia to the south and east. In addition to the Tyva Republic and several other regions of Russia, relatively small groups of Tuvinians also live in northern Mongolia and in Xinjiang Uyghur Autonomous Region of China (Mongush, 1996; Chen et al., 2011).

Tuvinians are one of the oldest Turkic-speaking peoples inhabiting Central Asia and the Sayan-Altai region. At different times, Tuva was at the periphery of a powerful state

of Huns (2nd century BC – 1st century AD), was part of the Ancient Turkic Khaganate (6–8th centuries), of the Uyghur Khaganate (8–9th centuries), of the Yenisei Kyrgyz Khaganate (9–12th centuries), and was also incorporated in the Mongol Empire (13–14th centuries). These historical events, as well as long-term contacts with the population of neighboring regions (Turkic-, Mongolic-, Ket-, and Samoyedic-speaking tribes), had a certain influence on the formation of the Tuvinian ethnic group (Vainshtein, Mannay-Ool, 2001; Mannai-ool, 2004) and the genetic structure of the Tuvinian population.

Experimental methods. To analyze the haplotypes carrying variants c.1545T>G and c.2027T>A of the *SLC26A4* gene, we performed the genotyping of genetic markers (STRs and SNPs), which we had previously used in our study to investigate the structure of haplotypes for variant c.919-2A>G (Danilchenko et al., 2023): five STRs (D7S2420, D7S496, D7S2456, D7S525, flanking the *SLC26A4* gene at different distances, and intragenic D7S2459), as well as nine intragenic SNPs (rs2248464, rs2248465, rs3801943, rs2712212, rs2395911, rs2712211, rs3801940, rs2072064, rs2072065) (Fig. 1).

Genotyping of STRs (fragment analysis) and SNPs (Sanger sequencing) was performed on an ABI 3130XL genetic analyzer (Applied Biosystems, USA) in the SB RAS Genomics Core Facility (Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia). Details of the experimental genotyping methods are presented in the work (Danilchenko et al., 2023).

Statistical methods. A one-sided Fisher's exact test with a significance level of $p < 0.05$ was applied to compare the allele and haplotype frequencies between the examined samples.

Linkage disequilibrium between the alleles of STR markers of chromosome 7 and alleles with variants c.1545T>G or c.2027T>A of the *SLC26A4* gene was calculated using the formula $\delta = (Pd - Pn) / (1 - Pn)$, where δ is a measure of linkage disequilibrium; Pd is the frequency of the associated allele among chromosomes with variants c.1545T>G or c.2027T>A in the samples of patients; Pn is the frequency of the same allele among chromosomes without these variants in the control sample (Bengtsson, Thomson, 1981).

Reconstruction of haplotypes based on the detected alleles of STR and SNP markers in samples of the carriers of variants c.1545T>G or c.2027T>A was carried out manually, using data from the analysis of genetic markers in their relatives (when it was possible). Reconstruction of haplotypes and analysis of their frequency distribution in the control sample of Tuvinians were performed by us using software package Arlequin v.3.5.1.2 (<https://cmpg.unibe.ch/software/arlequin3512/>, Expectation-Maximization algorithm) (Danilchenko et al., 2023).

Estimation of the “age” of variants of the *SLC26A4* gene. Estimation of the “age” of a mutation is based on the expected loss of linkage between the mutation and alleles of surrounding genetic markers over time due to recombination (the “molecular clock” concept). To estimate the “age” of the analyzed variants, two methods were used: the “single marker method” based on the allelic variation of one marker (Risch et al., 1995; Slatkin, Rannala, 2000), and the second method based on haplotype data implemented by the DMLE+ v.2.3

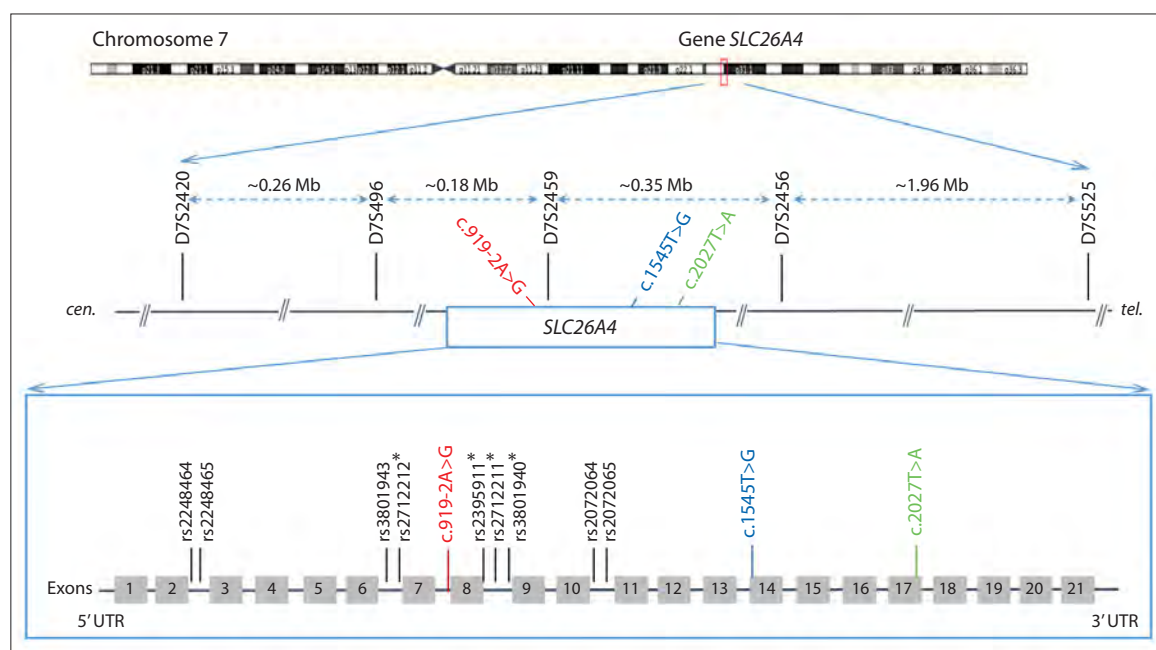


Fig. 1. Schematic structure of the *SLC26A4* gene and the location of genetic markers (five STRs and nine SNPs) which were used to reconstruct the haplotypes.

Pathogenic variants c.919-2A>G, c.1545T>G and c.2027T>A of the *SLC26A4* gene are highlighted in color. The distances between STRs (bp) are given in megabases (Mb). * – four SNPs from (Wu et al., 2005) used for comparative analysis. cen. – centromere, tel. – telomere. The figure was modified from (Danilchenko et al., 2023).

program (Disequilibrium Mapping and Likelihood Estimation, DMLE+ v.2.3: <http://dmle.org/>) (details of the methods used are given in Supplementary Material 1)¹. The “age” of a variant was determined by estimating the number of generations (g) and years (assuming that g = 25 years) that have passed since its occurrence.

Results

In our recent study (Danilchenko et al., 2021), we analyzed the *SLC26A4* gene using Sanger sequencing in patients with hearing loss belonging to the Tuvinians, an indigenous Siberian Turkic-speaking people (the Tyva Republic, Southern Siberia). Biallelic pathogenic *SLC26A4* variants were detected in 28.2 % (62 out of 220) of the patients included in the study. This rate of *SLC26A4*-associated hearing loss was one of the highest among all populations in the world. The vast majority of the detected mutant *SLC26A4* alleles were represented by three pathogenic variants c.919-2A>G, c.2027T>A, and c.1545T>G.

Variant c.919-2A>G

Most Tuvinian patients were homozygous or compound heterozygous for pathogenic variant c.919-2A>G. The proportion of c.919-2A>G was 69.3 % among all mutant *SLC26A4* alleles identified in Tuvinian patients, and its carrier frequency in the Tuvinian population was 5.1 % (Danilchenko et al., 2021). Variant c.919-2A>G is located in the canonical (-2) 3' splice acceptor site in the intronic region between exons 7 and 8 and

leads to splicing abnormalities (Yang J.J. et al., 2005; Lu Y.C. et al., 2011; Wasano et al., 2020).

Numerous studies have shown that the c.919-2A>G variant is highly prevalent in patients from Asian regions (mainland China, Taiwan, Mongolia, Korea, and Japan) and is found with the highest frequency in China and Mongolia, while in other regions of the world, this variant is extremely rare or absent (Park et al., 2003; Wu et al., 2005; Albert et al., 2006; Dai et al., 2008; Du et al., 2013; Yang X.L. et al., 2013; Lu Y.J. et al., 2015; Tsukada et al., 2015; Erdenechuluun et al., 2018).

Variant c.2027T>A

Variant c.2027T>A (p.Leu676Gln) of the *SLC26A4* gene was found in homozygous or compound heterozygous state in 19 Tuvinian patients and was the second most frequent pathogenic variant (17.5 %) after c.919-2A>G (69.3 %) among all mutant variants of the *SLC26A4* gene identified in Tuvinian patients (Danilchenko et al., 2021).

Variant c.2027T>A leads to the replacement of leucine with glutamine at amino acid position 676 (p.Leu676Gln) in the highly conserved region of the STAS domain in the COOH-terminal part of the pendrin protein molecule. Experimental studies have shown that this variant leads to retention of the mutant protein in the intracellular space and disruption of its function (Gillam et al., 2004; Yoon et al., 2008). The c.2027T>A variant was detected at low frequency (only in isolated patients in compound heterozygous or heterozygous state) in China, Korea, and Mongolia (Park et al., 2003; Choi et al., 2009; Huang et al., 2011; Chai et al., 2013; Erdenechuluun et al., 2018).

¹ Supplementary Materials 1 and 2 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Danil_Engl_29_1.pdf

Variant c.1545T>G

Variant c.1545T>G is a new, previously undescribed missense variant in exon 14 of the *SLC26A4* gene, apparently resulting in the substitution of phenylalanine with leucine at amino acid position 515 (p.Phe515Leu), was found in a compound heterozygous state in 15 Tuvanian patients originating from ten unrelated families. The carrier frequency of c.1545T>G in the Tuvanian control sample was 2.0 %. Segregation of c.1545T>G with hearing loss in the pedigrees of patients, a significant excess of its frequency in the sample of patients compared to the control sample ($p = 0.03391$), the results of predictive computer programs and the absence of this variant in the world's human genomic databases support its pathogenic significance (Danilchenko et al., 2021).

Reconstruction of STR haplotypes for variants c.1545T>G and c.2027T>A of the *SLC26A4* gene

To reconstruct the haplotypes of the chromosome 7 region carrying pathogenic variants c.1545T>G and c.2027T>A of the *SLC26A4* gene, five STRs (D7S2420, D7S496, D7S2459, D7S2456, D7S525) were genotyped in unrelated carriers of these variants (Fig. 1). These STRs were previously used in the analysis of haplotypes carrying the pathogenic variant c.919-2A>G (Danilchenko et al., 2023). The results of genotyping of STR markers, in comparison with the data obtained on the Tuvanian control sample, are presented in Supplementary Material 2 (Tables S1 and S2). All STRs in the control sample were previously found to be highly polymorphic: D7S2420 – 10 alleles, D7S496 – 10 alleles, D7S2459 – 7 alleles, D7S2456 – 5 alleles, D7S525 – 8 alleles (Danilchenko et al., 2023). All STRs in the carriers of variant c.1545T>G were monomorphic (only one allele for each STR marker) (Table S1). In the carriers of variant c.2027T>A, four STRs were monomorphic (D7S2420, D7S496, D7S2459, and D7S2456), but for the distal marker D7S525, three different alleles (221, 227, 231) were detected, with frequencies of 0.4000, 0.1333 and 0.4667, respectively (Table S2). Comparative analysis of the frequencies of D7S525 alleles in the carriers of c.2027T>A and in the control sample revealed statistically significant differences ($p < 0.05$) in the frequencies of alleles 227 and 231 (Table S2).

Genotyping of STRs in the carriers of variant c.1545T>G revealed complete linkage of this variant with alleles 286 (D7S2420), 118 (D7S496), 147 (D7S2459), 244 (D7S2456), 229 (D7S525); thus, variant c.1545T>G is characterized by a single haplotype 286-118-147-c.1545T>G-244-229, the size of which, determined by distal markers D7S2420 and D7S525, is ~ 2.75 Mb. STRs analysis in the carriers of c.2027T>A revealed complete linkage of this variant with alleles 280 (D7S2420), 118 (D7S496), 141 (D7S2459), 244 (D7S2456), but the presence of three different alleles (221, 227, 231) at the distal marker D7S525 suggests the presence of three different haplotypes for variant c.2027T>A.

STR haplotypes reconstructed for variants c.1545T>G and c.2027T>A, in comparison with STR haplotypes for variant c.919-2A>G, are presented in Figure 2.

We compared the structure and frequency of the STR haplotypes found for variants c.1545T>G and c.2027T>A with the STR haplotypes that were previously identified for variant c.919-2A>G (Danilchenko et al., 2023) (Table 1). It

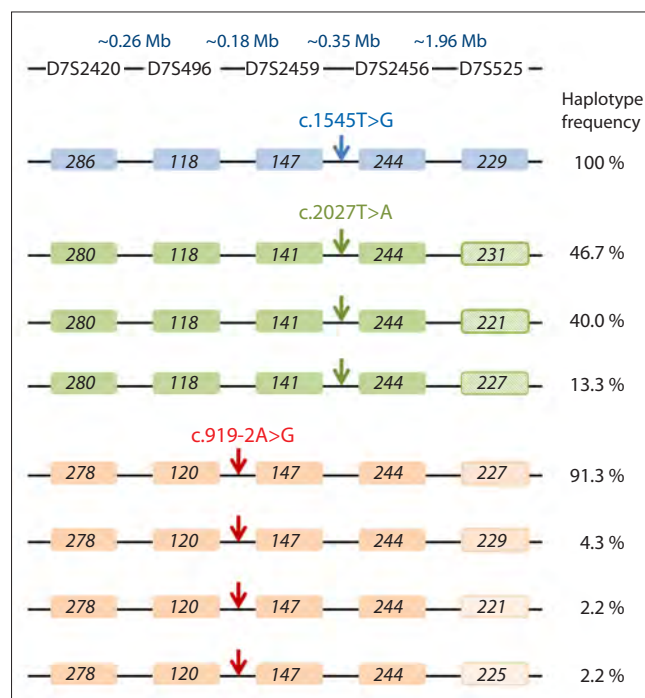


Fig. 2. Schematic representation of the STR haplotypes in the carriers of variants c.1545T>G and c.2027T>A in comparison with the STR haplotypes for variant c.919-2A>G (Danilchenko et al., 2023).

The localization of each of the analyzed variants (c.1545T>G, c.2027T>A or c.919-2A>G) is shown by an arrow.

should be noted that the STR haplotypes found for all three variants c.1545T>G, c.2027T>A and c.919-2A>G differ in allelic composition, which indicates a pronounced specificity of the genetic background for each of them. In addition, a comparison of the frequency of the main STR haplotypes in the samples of carriers of c.1545T>G, c.2027T>A and c.919-2A>G (groups of Tuvanian patients with hearing loss) and in the Tuvanian control sample revealed statistically significant differences (Table 1).

Reconstruction of SNP haplotypes for variants c.1545T>G and c.2027T>A of the *SLC26A4* gene

To study the fine structure of haplotypes including variants c.1545T>G or c.2027T>A of the *SLC26A4* gene, nine intra-genic SNPs (rs2248464, rs2248465, rs3801943, rs2712212, rs2395911, rs2712211, rs3801940, rs2072064, and rs2072065) were genotyped in the carriers of these variants. These SNPs were previously analyzed to reconstruct the genetic background (haplotypes) of variant c.919-2A>G in its homozygous carriers (Danilchenko et al., 2023). Four of them (rs2712212, rs2395911, rs2712211, and rs3801940) were included for comparative analysis with the data from the study by C.C. Wu et al. (2005), where they were used to establish the structure of haplotypes carrying variant c.919-2A>G in Taiwanese patients with hearing impairment (Fig. 1). The structure of SNP haplotypes for variants c.1545T>G and c.2027T>A is presented in Figure 3 in comparison with the SNP haplotype for variant c.919-2A>G (Danilchenko et al., 2023).

All carriers of variant c.1545T>G had a single SNP haplotype A-C-T-G-T-C-G-T-T (100 %), while the frequency

Table 1. The frequencies of STR haplotypes found on the mutant chromosomes carrying pathogenic variants c.919-2A>G, c.1545T>G or c.2027T>A of the *SLC26A4* gene, compared with the normal chromosomes

STR-haplotypes	Haplotype frequency, %		χ^2	p	References
D7S2420-D7S496-D7S2459-D7S2456-D7S525 (~2.75 Mb)	Mutant chromosomes	Normal chromosomes			
D7S2420-D7S496-D7S2459-/c.1545T>G/-D7S2456-D7S525					
286-118-147-244-229	1.0	0.0079	110	<10 ⁻¹³	This study
Other haplotypes	0.0	0.9921	–	–	
D7S2420-D7S496-D7S2459-/c.2027T>A/-D7S2456-D7S525					
280-118-141-244-231	0.4667	0.0	52	<10 ⁻⁷	This study
280-118-141-244-221	0.4000	0.0079	36	<10 ⁻⁵	
280-118-141-244-227	0.1333	0.0159	3.1	0.0561	
Other haplotypes	0.0	0.9762	–	–	
D7S2420-D7S496-/c.919-2A>G/-D7S2459-D7S2456-D7S525					
278-120-147-244-227	0.9130	0.0	150	<10 ⁻³⁵	Danilchenko et al., 2023
278-120-147-244-229	0.0435	0.0	2.4	0.0704	
278-120-147-244-221	0.0217	0.0	0.28	0.2674	
278-120-147-244-225	0.0217	0.0	0.28	0.2674	
Other haplotypes	0.0	1.0	–	–	

Note. Designations of the STR alleles included in haplotypes correspond to the size of the PCR products (in nucleotides). The most common haplotypes and statistically significant ($p < 0.05$) differences in haplotype frequencies are shown in bold.

of this haplotype in the Tuvanian control sample was 3.8 % (data not shown). All carriers of variant c.2027T>A also had a single SNP haplotype T-C-T-A-T-C-C-T-C (100 %), the frequency of which in the control sample was 1.7 % (data not shown). Previously, a single haplotype A-C-T-A-G-G-C-A-C (100 %) was also identified in all carriers of variant c.919-2A>G, and its frequency in the Tuvanian control sample was 2.8 % (Danilchenko et al., 2023). In addition, we previously established the identity of a small (~4.5 kb) “internal” SNP haplotype A-G-G-C formed by four SNPs (rs2712212, rs2395911, rs2712211, and rs3801940) in the carriers of variant c.919-2A>G – Tuvinians (Danilchenko et al., 2023) and Han Chinese from Taiwan (Wu et al., 2005), which suggests their common origin. However, this SNP haplotype was not detected in the carriers of variants c.1545T>G and c.2027T>A (Fig. 3). Thus, we can conclude that the haplotypes formed by the alleles of SNP markers for each of the three analyzed pathogenic variants of the *SLC26A4* gene are highly specific.

Estimation of the “age” of variants

c.1545T>G and c.2027T>A of the *SLC26A4* gene

In our recent study (Danilchenko et al., 2023), two methods were used to estimate the “age” of pathogenic variant c.919-2A>G of the *SLC26A4* gene: by the “single marker method”, which is based on the analysis of alleles of the most distal markers exhibiting significant linkage disequilibrium, and by using the DMLE+ v.2.3 program, where the “age” of a variant is estimated based on the reconstructed haplotypes. In this study, we applied both of these methods to estimate the

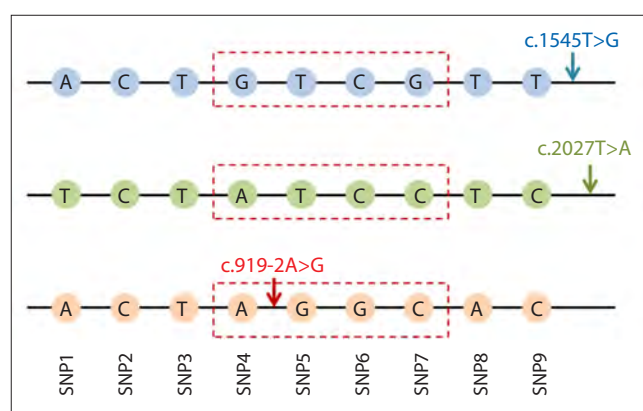


Fig. 3. Schematic representation of SNP haplotypes in the carriers of variants c.1545T>G and c.2027T>A in comparison with SNP haplotypes for variant c.919-2A>G (Danilchenko et al., 2023).

SNP marker designations: SNP1 – rs2248464, SNP2 – rs2248465, SNP3 – rs3801943, SNP4 – rs2712212, SNP5 – rs2395911, SNP6 – rs2712211, SNP7 – rs3801940, SNP8 – rs2072064, SNP9 – rs2072065. The red dotted lines highlight the four SNP markers analyzed in the carriers of variant c.919-2A>G in Taiwan (Wu et al., 2005). The localization of each of the analyzed variants (c.1545T>G, c.2027T>A, or c.919-2A>G) is shown by an arrow.

“age” of variants c.1545T>G and c.2027T>A of the *SLC26A4* gene (Table 2).

All carriers of variant c.1545T>G had an identical STR haplotype 286-118-147-c.1545T>G-244-229. Two haplotypes, 280-118-141-c.2027T>A-244-231 and 280-118-141-

Table 2. Comparative assessment of the “age” of variants c.1545T>G, c.2027T>A, and c.919-2A>G of the *SLC26A4* gene based on the STR markers

Pathogenic <i>SLC26A4</i> variant	<i>d</i>	The single-marker method		The DMLE+ calculation		References
		<i>g</i>	“Age” (years)	<i>g</i> (95 % CI)	“Age” (95 % CI) (years)	
c.1545T>G	0.05	–	–	76–163	1,900–4,075	This study
	0.10			48–95	1,200–2,375	
	0.20			27–52	675–1,300	
c.2027T>A	0.05	51	1,275	65–140	1,625–3,500	This study
	0.10	50	1,250	38–83	950–2,075	
	0.20	47	1,175	24–48	600–1,200	
c.919-2A>G	0.05	22	550	103–198	2,575–4,950	Danilchenko et al., 2023
	0.10	21	525	63–107	1,575–2,675	
	0.20	17	425	35–59	875–1,475	

Note. To assess the “age” of variants c.2027T>A and c.919-2A>G by the “single marker method”, the alleles of the distal STR marker D7S525 were used, and for the assessment by the DMLE+ v.2.3 program, STR haplotypes were used. *d* – different (0.05, 0.10 or 0.20) population growth rates; *g* – number of generations; “age” – *g* × 25 years; CI – confidence interval.

c.2027T>A-244-221, the structural differences of which are determined by the presence of different alleles (231 and 221) of the distal STR marker D7S525, were the most frequent for variant c.2027T>A (0.4667 and 0.4000, respectively). The obtained data allow us to tentatively estimate the time of occurrence of variants c.1545T>G and c.2027T>A in Tuvinians, the indigenous population of the Tyva Republic. We were unable to estimate the “age” of variant c.1545T>G using the “single marker method” due to the lack of recombination in all analyzed STR markers, but such estimates were obtained by the DMLE+ v.2.3 program (Table 2). To estimate the “age” of variant c.2027T>A by the “single marker method”, allele 231 of the distal STR marker D7S525 (~2.32 Mb from c.2027T>A), found in significant linkage disequilibrium with c.2027T>A, was used (Table S2).

The methods used to estimate the “age” of mutations are sensitive to the demographic parameters of the population, in particular, to the population growth rates at different historical stages of its development. Since there are no accurate data on changes in the size of the indigenous population of Tuva (Tuvinians) at the early stages of its formation, we used three different population growth rates for our calculations (*d* = 0.05, 0.10, and 0.20) (Table 2). It should be noted that the data on the “age” of variants c.2027T>A and c.919-2A>G obtained by the “single marker method” differ from the time ranges obtained by the DMLE+ v.2.3 program, apparently “underestimating” it at all three population growth rates (*d* = 0.05, 0.10, 0.20). In addition, the observed overlapping of the time intervals obtained by the DMLE+ v.2.3 program for each of the analyzed variants at all population growth rates (*d* = 0.05, 0.10, 0.20) (Table 2) does not allow us to conclude which of the analyzed variants is “older”.

Discussion

This work provides data on the haplotype structure for pathogenic variants c.1545T>G and c.2027T>A of the *SLC26A4* gene, identified in a study of hereditary deafness in Tuvinians, the indigenous population of the Tyva Republic (Southern Siberia) (Danilchenko et al., 2021). Variant c.1545T>G was

discovered for the first time; this variant has not been recorded in other regions of the world. All carriers of c.1545T>G were found to have highly specific STR and SNP haplotypes: STR haplotype 286-118-147-**c.1545T>G**-244-229 (100 %) and SNP haplotype A-C-T-G-T-C-G-T-T-**c.1545T>G** (100 %); the frequency of them in the Tuvinian control sample is less than 1 and 3.8 %, respectively. Thus, these data provide convincing evidence of a single origin of variant c.1545T>G and the role of the founder effect in its prevalence among the indigenous population of Tuva. Variant c.2027T>A is second in frequency among all pathogenic variants of the *SLC26A4* gene identified in Tuvinian patients; at the same time, this variant is found only in isolated patients from China, Korea and Mongolia (Park et al., 2003; Choi et al., 2009; Chai et al., 2013; Erdenechuluun et al., 2018; Kun et al., 2024). In addition, we also found this variant in several patients from the Altai Republic, which borders the Tyva Republic (Danilchenko et al., 2021). In contrast to variant c.1545T>G, three STR haplotypes were identified in the c.2027T>A carriers: 280-118-141-**c.2027T>A**-244-231 (46.7 %), 280-118-141-**c.2027T>A**-244-221 (40.0 %), and 280-118-141-**c.2027T>A**-244-227 (13.3 %), that differ only by alleles of the distal STR marker D7S525.

The use of a set of polymorphic genetic markers identical to that previously used by us in the study of haplotypes of pathogenic variant c.919-2A>G of the *SLC26A4* gene, the most common in Tuvinian patients (Danilchenko et al., 2023), allowed us to conduct a correct comparison of the structure of STR and SNP haplotypes for all three pathogenic variants (c.1545T>G, c.2027T>A, and c.919-2A>G). Comparative analysis showed that the composition of alleles of the genetic markers included in the haplotypes is different and highly specific for each of them. Thus, we can conclude that each of the analyzed variants has a special (similar for all carriers of a particular variant) genetic background, apparently inherited from different “founder ancestors”.

We have roughly estimated the “age” of variants c.1545T>G, c.2027T>A and c.919-2A>G, but due to the limited information on the demographic changes in the Tuvinian population throughout its history, the obtained time intervals of the

appearance of these variants in the indigenous population of Tuva should be considered only as approximate ones. Nevertheless, it can be cautiously assumed that variants c.1545T>G, c.2027T>A and c.919-2A>G are not “young” (recently emerged) mutations, and the wide time intervals of their occurrence overlap at almost all population growth rates ($d = 0.05, 0.10$ and 0.20) (Table 2).

Data on the haplotype structure for variant c.1545T>G and its prevalence, limited only to the territory of Tuva, as well as historical information on ethnogenesis of the indigenous population of Tuva, suggest that this variant could have arisen as a result of a unique mutational event that occurred after the main formation of the Tuvanian ethnic group at the end of the 13th–14th centuries. It is more difficult to draw conclusions about the origin of variant c.2027T>A in Tuvinians. This variant is found with low frequency in patients from neighboring Mongolia and China, but, unfortunately, there are no data on the structure of the genetic background of c.2027T>A in its carriers from these regions, which excludes comparative analysis. As for variant c.919-2A>G, which is the most frequent in Tuvinians, we previously (Danilchenko et al., 2023) found the identity of the “internal” SNP haplotype A-G-G-C (Fig. 3), found in Tuvanian patients homozygous for c.919-2A>G, and the haplotype formed by the same SNPs in the c.919-2A>G carriers from Taiwan (Han Chinese) (Wu et al., 2005). These data support the presence of a common ancestor for the “Tuvanian” and “Chinese” founder chromosomes with c.919-2A>G. Considering the obtained results, as well as the territorial distribution of variant c.919-2A>G, with a maximum frequency in Tuvinians (Southern Siberia) and in Chinese and Mongols (East and Central Asia), we suggested that variant c.919-2A>G could have arisen in geographically close territories of these regions and subsequently spread to other regions of Asia (Danilchenko et al., 2023).

Conclusion

We analyzed the haplotype structure for pathogenic variants c.1545T>G and c.2027T>A of the *SLC26A4* gene, found with high frequency in Tuvanian patients with hearing loss (the Tyva Republic, Southern Siberia). Comparative analysis of the reconstructed haplotypes for c.1545T>G and c.2027T>A with previously obtained data on the haplotypes for variant c.919-2A>G showed that each of the analyzed variants has a specific genetic background which is similar for all carriers of a particular variant, apparently inherited from different “founder ancestors”. Thus, evidence was obtained for the role of the cumulative founder effect in the prevalence of these pathogenic variants of the *SLC26A4* gene in the indigenous population of the Tyva Republic. The obtained data are relevant both for predicting the prevalence of *SLC26A4*-associated hearing loss and for developing region-specific DNA diagnostics of inherited hearing loss in the Tyva Republic.

References

Albert S., Blons H., Jonard L., Feldmann D., Chauvin P., Loundon N., Sergent-Allaoui A., Houang M., Joannard A., Schmerber S., Delobel B., Leman J., Journel H., Catros H., Dollfus H., Eliot M.M., David A., Calais C., Drouin-Garraud V., Obstoy M.F., Tran Ba Huy P., Lacombe D., Duriez F., Francannet C., Bitoun P., Petit C., Garabédian E.N., Couderc R., Marlin S., Denoyelle F. *SLC26A4* gene is frequently involved in nonsyndromic hearing impairment with en-

larged vestibular aqueduct in Caucasian populations. *Eur J Hum Genet.* 2006;14(6):773-779. doi 10.1038/sj.ejhg.5201611

Ben Arab S., Masmoudi S., Beltaief N., Hachicha S., Ayadi H. Consanguinity and endogamy in Northern Tunisia and its impact on non-syndromic deafness. *Genet Epidemiol.* 2004;27(1):74-79. doi 10.1002/gepi.10321

Bengtsson B.O., Thomson G. Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens.* 1981;18(5):356-363. doi 10.1111/j.1399-0039.1981.tb01404.x

Chai Y., Huang Z., Tao Z., Li X., Li L., Li Y., Wu H., Yang T. Molecular etiology of hearing impairment associated with nonsyndromic enlarged vestibular aqueduct in East China. *Am J Med Genet A.* 2013;161A(9):2226-2233. doi 10.1002/ajmg.a.36068

Chen Z., Zhang Y., Fan A., Zhang Y., Wu Y., Zhao Q., Zhou Y., Zhou C., Bawudong M., Mao X., Ma Y., Yang L., Ding Y., Wang X., Rao S. Brief communication: Y-chromosome haplogroup analysis indicates that Chinese Tuvans share distinctive affinity with Siberian Tuvans. *Am J Phys Anthropol.* 2011;144(3):492-497. doi 10.1002/ajpa.21453

Choi B.Y., Stewart A.K., Nishimura K.K., Cha W.J., Seong M.W., Park S.S., Kim S.W., Chun Y.S., Chung J.W., Park S.N., Chang S.O., Kim C.S., Alper S.L., Griffith A.J., Oh S.H. Efficient molecular genetic diagnosis of enlarged vestibular aqueducts in East Asians. *Genet Test Mol Biomarkers.* 2009;13(5):679-687. doi 10.1089/gtmb.2009.0054

Chong J.X., Ouwenga R., Anderson R.L., Waggoner D.J., Ober C. A population-based study of autosomal-recessive disease-causing mutations in a founder population. *Am J Hum Genet.* 2012;91(4):608-620. doi 10.1016/j.ajhg.2012.08.007

Common J.E., Di W.L., Davies D., Kelsell D.P. Further evidence for heterozygote advantage of *GJB2* deafness mutations: a link with cell survival. *J Med Genet.* 2004;41(7):573-575. doi 10.1136/jmg.2003.017632

Dai P., Li Q., Huang D., Yuan Y., Kang D., Miller D.T., Shao H., Zhu Q., He J., Yu F., Liu X., Han B., Yuan H., Platt O.S., Han D., Wu B.L. *SLC26A4* c.919-2A>G varies among Chinese ethnic groups as a cause of hearing loss. *Genet Med.* 2008;10(8):586-592. doi 10.1097/gim.0b013e31817d2ef1

Danilchenko V.Y., Zytars M.V., Maslova E.A., Bady-Khoo M.S., Barashkov N.A., Morozov I.V., Bondar A.A., Posukh O.L. Different rates of the *SLC26A4*-related hearing loss in two indigenous peoples of southern Siberia (Russia). *Diagnostics (Basel).* 2021;11(12):2378. doi 10.3390/diagnostics11122378

Danilchenko V.Y., Zytars M.V., Maslova E.A., Orishchenko K.E., Posukh O.L. Insight into the natural history of pathogenic variant c.919-2A>G in the *SLC26A4* gene involved in hearing loss: the evidence for its common origin in southern Siberia (Russia). *Genes (Basel).* 2023;14(4):928. doi 10.3390/genes14040928

Du W., Guo Y., Wang C., Wang Y., Liu X. A systematic review and meta-analysis of common mutations of *SLC26A4* gene in Asian populations. *Int J Pediatr Otorhinolaryngol.* 2013;77(10):1670-1676. doi 10.1016/j.ijporl.2013.07.023

Erdenechuluun J., Lin Y.H., Ganbat K., Bataakhuu D., Makhbal Z., Tsai C.Y., Lin Y.H., Chan Y.H., Hsu C.J., Hsu W.C., Chen P.L., Wu C.C. Unique spectra of deafness-associated mutations in Mongolians provide insights into the genetic relationships among Eurasian populations. *PLoS One.* 2018;13(12):e0209797. doi 10.1371/journal.pone.0209797

Gillam M.P., Sidhaye A.R., Lee E.J., Rutishauser J., Stephan C.W., Kopp P. Functional characterization of pendrin in a polarized cell system. Evidence for pendrin-mediated apical iodide efflux. *J Biol Chem.* 2004;279(13):13004-13010. doi 10.1074/jbc.M313648200

Huang S., Han D., Yuan Y., Wang G., Kang D., Zhang X., Yan X., Meng X., Dong M., Dai P. Extremely discrepant mutation spectrum of *SLC26A4* between Chinese patients with isolated Mondini deformity and enlarged vestibular aqueduct. *J Transl Med.* 2011;9:167. doi 10.1186/1479-5876-9-167

- Kun L., Jiexiang H., Hua L., Junlin H., Yijun R., Lixian Z., Ming-qiao C. Genetic screening of 15 hearing loss variants in 77,647 neonates with clinical follow-up. *Mol Genet Genomic Med.* 2024;12(1): e2324. doi 10.1002/mgg3.2324
- Lu Y.C., Wu C.C., Shen W.S., Yang T.H., Yeh T.H., Chen P.J., Yu I.S., Lin S.W., Wong J.M., Chang Q., Lin X., Hsu C.J. Establishment of a knock-in mouse model with the *SLC26A4* c.919-2A>G mutation and characterization of its pathology. *PLoS One.* 2011;6(7):e22150. doi 10.1371/journal.pone.0022150
- Lu Y.J., Yao J., Wei Q.J., Xing G.Q., Cao X. Diagnostic value of *SLC26A4* mutation status in hereditary hearing loss with EVA: a PRISMA-compliant meta-analysis. *Medicine (Baltimore).* 2015; 94(50):e2248. doi 10.1097/MD.0000000000002248
- Mannai-ool M.Kh. Tuvan People. The Origin and Formation of the Ethnos. Novosibirsk: Nauka Publ., 2004 (in Russian)
- Mongush M.V. Tuvans of Mongolia and China. *Int J Cent Asian Stud.* 1996;1:225-243
- Nance W.E., Kearsey M.J. Relevance of connexin deafness (DFNB1) to human evolution. *Am J Hum Genet.* 2004;74(6):1081-1087. doi 10.1086/420979
- Nance W.E., Liu X.Z., Pandya A. Relation between choice of partner and high frequency of connexin-26 deafness. *Lancet.* 2000; 356(9228):500-501. doi 10.1016/S0140-6736(00)02565-4
- Park H.J., Shaikat S., Liu X.Z., Hahn S.H., Naz S., Ghosh M., Kim H.N., Moon S.K., Abe S., Tukamoto K., Riazuddin S., Kabra M., Erdenetungalag R., Radnaabazar J., Khan S., Pandya A., Usami S.I., Nance W.E., Wilcox E.R., Riazuddin S., Griffith A.J. Origins and frequencies of *SLC26A4* (*PDS*) mutations in east and south Asians: global implications for the epidemiology of deafness. *J Med Genet.* 2003;40(4):242-248. doi 10.1136/jmg.40.4.242
- Razdan S., Raina S.K., Pandita K.K., Razdan S., Nanda R., Kaul R., Dogra S. Inbreeding as a cause for deafness: Dadhkai study. *Indian J Hum Genet.* 2012;18(1):71-74. doi 10.4103/0971-6866.96655
- Risch N., de Leon D., Ozelius L., Kramer P., Almasy L., Singer B., Fahn S., Breakefield X., Bressman S. Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet.* 1995;9(2):152-159. doi 10.1038/ng0295-152
- Scott D.A., Carmi R., Elbedour K., Duyk G.M., Stone E.M., Sheffield V.C. Nonsyndromic autosomal recessive deafness is linked to the DFNB1 locus in a large inbred Bedouin family from Israel. *Am J Hum Genet.* 1995;57(4):965-698.
- Slatkin M., Rannala B. Estimating allele age. *Annu Rev Genomics Hum Genet.* 2000;1:225-249. doi 10.1146/annurev.genom.1.1.225
- Tsukada K., Nishio S.Y., Hattori M., Usami S. Ethnic-specific spectrum of *GJB2* and *SLC26A4* mutations: their origin and a literature review. *Ann Otol Rhinol Laryngol.* 2015;124(1):61S-76S. doi 10.1177/0003489415575060
- Vainshtein S.I., Mannay-Ool M.H. (Eds) History of Tyva. Vol. 1. Novosibirsk: Nauka Publ., 2001 (in Russian)
- Wasano K., Takahashi S., Rosenberg S.K., Kojima T., Mutai H., Matsunaga T., Ogawa K., Homma K. Systematic quantification of the anion transport function of pendrin (*SLC26A4*) and its disease-associated variants. *Hum Mutat.* 2020;41(1):316-331. doi 10.1002/humu.23930
- Wu C.C., Yeh T.H., Chen P.J., Hsu C.J. Prevalent *SLC26A4* mutations in patients with enlarged vestibular aqueduct and/or Mondini dysplasia: a unique spectrum of mutations in Taiwan, including a frequent founder mutation. *Laryngoscope.* 2005;115(6):1060-1064. doi 10.1097/01.MLG.0000163339.61909.D0
- Yang J.J., Tsai C.C., Hsu H.M., Shiao J.Y., Su C.C., Li S.Y. Hearing loss associated with enlarged vestibular aqueduct and Mondini dysplasia is caused by splice-site mutation in the *PDS* gene. *Hear Res.* 2005;199(1-2):22-30. doi 10.1016/j.heares.2004.08.007
- Yang X.L., Bai-Cheng X., Chen X.J., Pan-Pan B., Jian-Li M., Xiao-Wen L., Zhang Z.W., Wan D., Zhu Y.M., Guo Y.F. Common molecular etiology of patients with nonsyndromic hearing loss in Tibetan, Tu nationality, and Mongolian patients in the northwest of China. *Acta Otolaryngol.* 2013;133(9):930-934. doi 10.3109/00016489.2013.795288
- Yoon J.S., Park H.J., Yoo S.Y., Namkung W., Jo M.J., Koo S.K., Park H.Y., Lee W.S., Kim K.H., Lee M.G. Heterogeneity in the processing defect of *SLC26A4* mutants. *J Med Genet.* 2008;45(7):411-419. doi 10.1136/jmg.2007.054635
- Zlotogora J. Multiple mutations responsible for frequent genetic diseases in isolated populations. *Eur J Hum Genet.* 2007;15(3):272-278. doi 10.1038/sj.ejhg.5201760

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

Received August 15, 2024. Revised November 7, 2024. Accepted November 8, 2024.

doi 10.18699/vjgb-25-18

MiceDEGdb: a knowledge base on differentially expressed mouse genes as a model object in biomedical research

O.A. Podkolodnaya ¹, I.V. Chadaeva ^{1, 2}, S.V. Filonov¹, N.L. Podkolodnyy ^{1, 3}, D.A. Rasskazov ¹, N.N. Tverdokhle¹,
K.A. Zolotareva ¹, A.G. Bogomolov ^{1, 4}, E.Yu. Kondratyuk ^{1, 5}, D.Yu. Oshchepkov ^{1, 2, 4}, M.P. Ponomarenko ^{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 Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

⁴ Novosibirsk State University, Novosibirsk, Russia

⁵ Siberian Federal Scientific Centre of Agro-BioTechnologies of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk region, Russia

 pon@bionet.nsc.ru

Abstract. The fundamental understanding of many biological processes that unfold in a human body has become possible due to experimental studies on animal models. The backbone of modern biomedical research is the use of mouse models for studying important pathophysiological mechanisms, assessing new therapeutic approaches and making decisions on acceptance or rejection of new candidate medicines in preclinical trials. The use of mice is advantageous because they have small size, are easy to keep and to genetically modify. Mice make up more than 90 % of the rodents used for pharmaceutical research. We present the pilot version of MiceDEGdb, a knowledge base on the genes that are differentially expressed in the mouse used as a model object in biomedical research. MiceDEGdb is a collection of published data on gene expression in mouse strains used for studying age-related diseases, such as hypertension, periodontal disease, bone fragility, renal fibrosis, smooth muscle remodeling, heart failure and circadian rhythm disorder. The pilot release of MiceDEGdb contains 21,754 DEGs representing 9,769 unique *Mus musculus* genes the transcription levels whereof were found as being changed in 25 RNA-seq experiments involving eight tissues – gum, bone, kidney, right ventricle, aortic arch, hippocampus, skeletal muscle and uterus – in six genetic mouse strains (C57BL/6J, Ren1cCre|ZsGreen, B6.129S7(Cg)-Polgtm1Prol/J, BPN/3J, BPH/2J and Kunming) used as models of eight human diseases – all these data were based on information in 10 original articles. MiceDEGdb is novel in that it features a curated annotation of changes in the expression levels of mouse DEGs using independent biomedical publications about same-direction changes in the expression levels of human homologs in patients with one disease or the other. In its pilot release, MiceDEGdb documented 85,092 such annotations for 318 human genes in 895 diseases, as suggest to 912 scientific articles referenced by their PubMed ID. The information contained in MiceDEGdb may be of interest to geneticists, molecular biologists, bioinformatics scientists, clinicians, pharmacologists and genetic advisors in personalized medicine. MiceDEGdb is freely available at <https://www.sysbio.ru/MiceDEGdb>.











Key words: knowledge base; DEG; mouse *Mus musculus*; mouse models of disease; age frustration; infectious diseases; circadian rhythm; RNA-seq.

For citation: Podkolodnaya O.A., Chadaeva I.V., Filonov S.V., Podkolodnyy N.L., Rasskazov D.A., Tverdokhle N.N., Zolotareva K.A., Bogomolov A.G., Kondratyuk E.Yu., Oshchepkov D.Yu., Ponomarenko M.P. MiceDEGdb: a knowledge base on differentially expressed mouse genes as a model object in biomedical research. *Vavilovskii Zhurnal Genetiki i Selektcii* = *Vavilov J Genet Breed*. 2025;29(1):153-161. doi 10.18699/vjgb-25-18

Funding. The work was supported by State Budget Projects FWNR-2022-0020 and FWNM-2025-0005.

Acknowledgements. Authors express their gratitude to the multi-access center Bioinformatics for access to computational resources.

База знаний MiceDEGdb по дифференциально экспрессирующимся генам мыши как модельного объекта биомедицинских исследований

О.А. Подкологная ¹, И.В. Чадаева ^{1, 2}, С.В. Филонов¹, Н.Л. Подкологный ^{1, 3}, Д.А. Рассказов ¹, Н.Н. Твердохлеб¹,
К.А. Золотарева ¹, А.Г. Богомолов ^{1, 4}, Е.Ю. Кондратюк ^{1, 5}, Д.Ю. Ощепков ^{1, 2, 4}, М.П. Пономаренко ^{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 Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия

⁴ Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

⁵ Сибирский федеральный научный центр агробиотехнологий Российской академии наук, р. п. Краснообск, Новосибирская область, Россия

 pon@bionet.nsc.ru

© Podkolodnaya O.A., Chadaeva I.V., Filonov S.V., Podkolodnyy N.L., Rasskazov D.A., Tverdokhle N.N., Zolotareva K.A., Bogomolov A.G., Kondratyuk E.Yu., Oshchepkov D.Yu., Ponomarenko M.P., 2025

This work is licensed under a Creative Commons Attribution 4.0 License

Аннотация. Фундаментальное понимание многих биологических процессов, происходящих в организме человека, стало возможным благодаря экспериментальным исследованиям на животных моделях. Основным стержнем современных биомедицинских исследований является использование мышиных моделей для изучения важных патофизиологических механизмов, оценки новых терапевтических подходов и принятия решений о переходе или отказе от новых кандидатов на препараты в доклинических испытаниях. Преимущество задействия мышей заключается в их небольших размерах, простых условиях содержания и относительно легкой генетической модификации. В настоящее время более 90 % грызунов, используемых для фармацевтических исследований, – мыши. В представленной работе создана пилотная версия базы знаний MiceDEGdb по дифференциально экспрессирующимся генам (ДЭГ) мыши как модельного объекта в биомедицинских исследованиях. Она представляет собой коллекцию опубликованных данных по экспрессии генов у мышей разных линий, предназначенных для изучения возрастных заболеваний: гипертонии, пародонтита, хрупкости кости, фиброза почки, ремоделирования гладких мышц, сердечной недостаточности, нарушения циркадного ритма. Пилотный выпуск базы знаний MiceDEGdb содержит 21 754 ДЭГ, представляющих 9769 уникальных генов *Mus musculus*, у которых изменяется уровень транскрипции в 25 экспериментах по технологии RNA-seq с использованием восьми тканей (десна, кость, почка, правый желудочек сердца, дуга аорты, гиппокамп, скелетная мышца и матка) в шести генетических линиях мышей, C57BL/6J, Ren1cCre|ZsGreen, B6.129S7(Cg)-Polgtm1Prol/J, BPN/3J, BPH/2J и Kunming, в качестве моделей восьми заболеваний человека согласно 10 оригинальным статьям. Новшеством MiceDEGdb в сравнении с другими базами данных о ДЭГ мышей является курируемая аннотация отклонений ДЭГ мыши от соответствующей нормы с помощью независимых биомедицинских публикаций о сопоставленных изменениях экспрессии гомологичных генов человека у пациентов с теми или иными заболеваниями относительно условно здоровых добровольцев. В пилотном выпуске MiceDEGdb документировано 85 092 таких аннотаций для 318 генов человека при 895 заболеваниях согласно 912 научным статьям, цитируемым с помощью их идентификаторов PubMed ID. Информационное содержание MiceDEGdb может быть интересным для генетиков, молекулярных биологов, биоинформатиков, клиницистов, фармацевтов и генетических консультантов по персонализированной медицине. База знаний MiceDEGdb находится в свободном доступе по гиперссылке: <https://www.sysbio.ru/MiceDEGdb>.
Ключевые слова: база знаний; ДЭГ; мышь, *Mus musculus*; мышиные модели заболеваний; возрастные расстройства; инфекционные заболевания; циркадный ритм; RNA-seq.

Introduction

The use of animals is absolute to biomedical research aimed at studying biological processes (Lukacs et al., 1996), pathogenesis of diseases (Conti et al., 2002) and therapeutic interventions (Chuang et al., 2002) as well as assessing the safety, toxicity and carcinogenicity of candidate medicines (Segalat, 2007). At the same time, the relevance of animal disease models is established according to strict criteria of consistency between the animals' conditions being studied and the symptoms the patients of interest are experiencing (Gryksa et al., 2023). To be able to interpret the results of observations made using animal models of human diseases, one should have not only knowledge of the processes being studied and pathophysiology, but also the ability to recognize spontaneous, background and associated conditions that may bias the results (White et al., 2016).

At present, more than 90 % of the pharmaceutical studies involves laboratory mouse strains (Vandamme, 2014). They are cheaper to keep than, for example, primates and can give birth every two months – these two qualities make them so popular among the researchers (Girard et al., 2009).

Although animal models still play an important role in assessing the efficiency and safety of new interventions in anti-cancer therapy, their use is often limited by genetic, molecular and physiological factors. Despite successful preclinical testing, 85 % of novel medicines fail during phase 1 of clinical trials: only half of those that advance to phase 3 become licensed. The use of mice as model organisms in biomedical research is deemed to be the option of choice because of their close genetic and physiological similarity with humans

(Swindell et al., 2012) and because their genome is rather easy to manipulate (Monteiro et al., 2023). The latter advantage becomes more and more relevant with the advancement of genome editing methods (Bruter et al., 2024).

This inclines the researchers to move more actively to the “humanized mouse” platform, which is a good setting to use for studying the mechanisms of physiological processes (Yong et al., 2018), for exploring the pathogenesis of infectious diseases (Yajima et al., 2008; Frias-Staheli et al., 2014; Amaladoss et al., 2015; Keng et al., 2016), autoimmune diseases (Zayoud et al., 2013; Viehmann Milam et al., 2014; Gunawan et al., 2017) and cancers (Chuprin et al., 2023; Liu L. et al., 2024), and for developing anti-cancer therapies (Petrova et al., 2022). In some cases, for example, in microbiome research, wild mice should be preferred to their laboratory conspecifics, who have long been under artificial selection for the ability to breed in cages and on an *ad libitum* diet, which may bias the results (Hild et al., 2021).

Finally, the life sciences of the post-genome era increasingly thrive on the so-called Big Data about the differential gene expression in certain mouse tissues in the norm and pathology. The multidimensionality of Big Data requires for them to be systematized, analyzed and searched for patterns using bioinformatics methods (Liu B et al., 2024). Two sources of information become critically important for post-genome medicine and pharmacology: 1) clinical data on patients vs. unaffected volunteers and 2) experimental data on animals used as models of human diseases (Krause et al., 2023), which calls for the need of data processing resources to integrate data coming from these sources.

In one of our previous works (Chadaeva et al., 2023), we reported RatDEGdb, a freely available knowledge base on the genes that are differentially expressed in the rat used as a model object in biomedical research. As a logical step in expanding the capabilities of this series of information resources, we created MiceDEGdb, a knowledge base on DEGs in mouse strains developed in a range of scientific organizations and used as biomedical models. MiceDEGdb is freely available at <https://www.sysbio.ru/MiceDEGdb>.

Materials and methods

Searching PubMed for information on differentially expressed mouse genes. The experimentally identified genes that are differentially expressed in several laboratory mouse strains used as biomedical disease models were taken as published in the original articles that we found by querying ["mice" "RNA-Seq" "disease"] in PubMed (Lu, 2011).

MiceDEGdb. MiceDEGdb includes three tables (see the schema in Figure 1). The mouse DEG information found as described in the previous subsection was put into a relational table, MiceDEGs. Next, we copied the relational table HumanDisorder from the one of our previous developments,

Human_SNP_TATAdb database (Filonov et al., 2023); the table is explained at the bottom right of Figure 1.

Further, we linked MiceDEGs and HumanDisorder using the MiceDEGdb's unique relational table HumanMiceHomolog based on the entries in the "Paralogs" section of GeneCards, a freely available database (Stelzer et al., 2016). Finally, we conversed MiceDEGs, HumanMiceHomolog, HumanDisorder and the links between them (the links are pointed to by arrows in Figure 1) into MiceDEGdb, which is freely available at <https://www.sysbio.ru/MiceDEGdb>, using MariaDB 10.2.12 (MariaDB Corp AB, Finland), a freely available open-source database management system (DBMS).

A model assessing the effect of circadian rhythm disorder on human health. Estimates of age-related changes in the expression levels of mammalian core circadian genes were obtained using a computational model explained and validated elsewhere (Podkolodnyy et al., 2016). The outstanding feature of this model is that the interactions between the core circadian oscillator and the NAD⁺/SIRT1 pathway are taken into account through the use of the following modules: 1) a pathway associated with SIRT1-promoted acetylation and

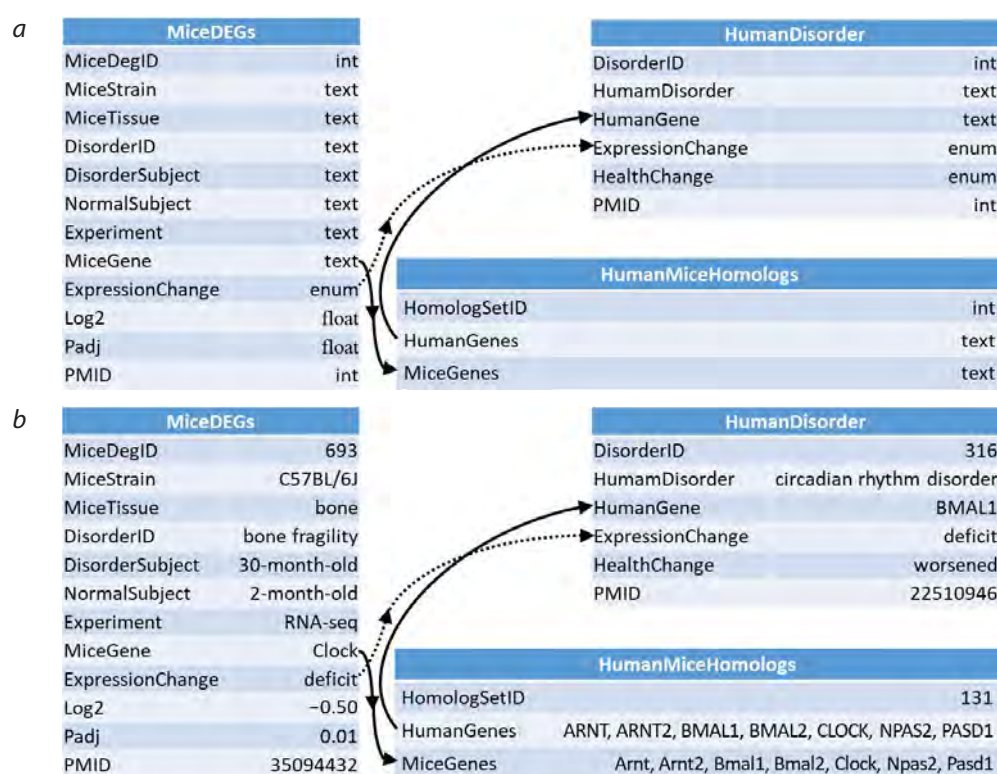


Fig. 1. MiceDEGdb, a knowledge base on the genes that are differentially expressed in the mouse used as a model object in biomedical research into human diseases.

a – flow chart; **b** – sample entries. Designations: MiceDEGs, HumanDisorder and HumanMiceHomologs are three unique relational tables in MiceDEGdb. In each of these tables, the left column contains the name of fields, such as “mouse gene”, “mouse strain”, “tissue”, “disease” and “the identifier of the article, to which the experimental data belong, in PubMed” (Lu, 2011); the right column is for the types of data, such as integer (int), real number (float), binary (enum) or string (text). Arrows point to relational references (solid arrows) between experimental data on the DEGs in the mouse used as a biomedical model of a particular human disease on the one hand and, on the other hand, the same-direction changes (dotted arrows) in the expression levels of human genes homologous to the DEGs in people suffering from one disease or the other – all these data were based on information in original articles found in PubMed (Lu, 2011) and referenced accordingly.

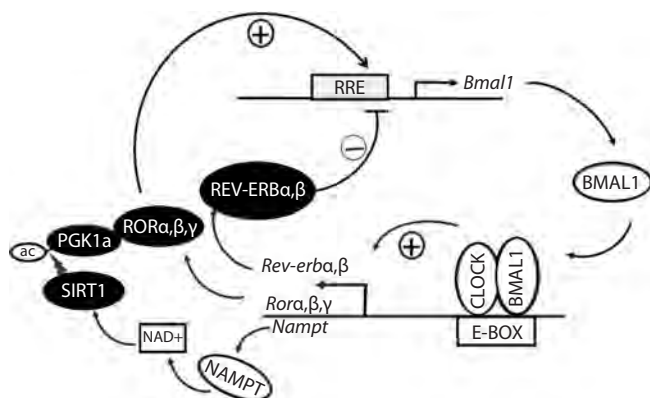


Fig. 2. Gene regulatory network associated with the effect of the deacetylase SIRT1 on activation of *Bmal1* transcription and inhibition of the CLOCK/BMAL1 function.

The oscillating feedback loop that increases *Bmal1* expression and the expression of genes targeted by the transcription factor CLOCK/BMAL1 is factored in by the computational model of age-related changes in the function of the core circadian oscillator (Podkolodnyy et al., 2016): a point worth making.

degradation of the PER2 protein; 2) a gene regulatory network associated with the effect of the deacetylase Sirt1 on the transcription of the mouse gene *Bmal1* and inhibition of the CLOCK/BMAL1 function associated with the E-BOX through histone deacetylation (Fig. 2); 3) a pathway associated with the effect of Sirt1 on the rate at which CLOCK/BMAL1 unbinds from the E-BOX; and 4) the Nampt/NAD⁺/Sirt1 pathway.

In our model, the mechanism of transcriptional regulation of the *Nampt* gene depends on the presence of three copies of the E-BOX in its promoter, similarly to the mechanism of regulation of the *Per1*, *Per2* and *Cry1* genes in a model of the core circadian oscillator by J.K. Kim and D.B. Forger (2012).

Results

Mouse DEGs as biomedical models of age-related diseases

Figure 3 shows the results for the *in silico* modeling of changes in *Bmal1* mRNA concentrations in mice using a computational model by N.L. Podkolodnyy and the co-workers (2016): the concentration levels decrease as the mice grow older.

To verify the results of the *in silico* modeling of age-related changes to the core circadian oscillator, we searched PubMed and found 10 original articles with relevance to the matter (see the Table). They presented the results of 25 RNA-Seq experiments with eight tissues (gum, bone, kidney, right ventricle, aortic arch, hippocampus, skeletal muscle and uterus) in six genetic mouse strains (C57BL/6J, Ren1cCre/ZsGreen, B6.129S7(Cg)-Polgtm1Prol/J, BPN/3J, BPH/2J and Kunming) used as models of eight human age-related diseases, including arterial hypertension, periodontal disease, bone fragility, renal fibrosis, smooth muscle remodeling, heart failure and circadian rhythm disorder. The total number of mouse DEGs was 21,754 representing 9,769 unique genes from among 22,283 annotated protein-coding genes in the reference genome GRCm38.p6 of the *Mus musculus* labora-

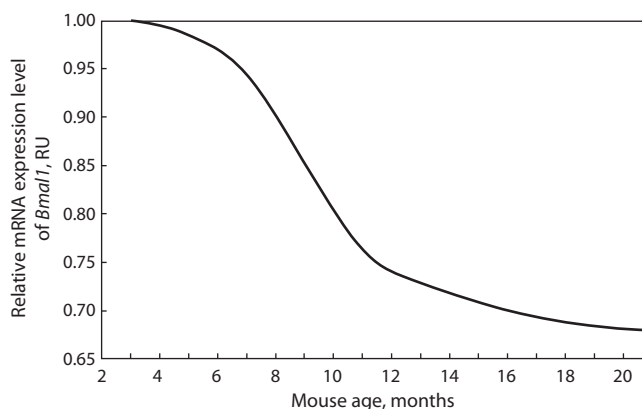


Fig. 3. Results for the *in silico* modeling of changes in *Bmal1* mRNA concentrations in mice using a computational model by Podkolodnyy and the co-workers (2016): the concentration levels decrease as the mice grow older.

tory strain C57BL/6J (Sarsani et al., 2019) (see the bottom row of the Table).

MiceDEGdb

Figure 4 shows how MiceDEGdb can be worked with. As a sample mouse gene, we took *Clock*. This gene was reported as being expressed at lower levels in 30-month-old male mice noted for bone fragility than in healthy males aged two months (Kaya et al., 2022).

As can be seen from Figure 4, a decrease in expression levels of a human *CLOCK* gene homologous to the mouse *Clock* gene was observed in patients with intestinal inflammation (Giebfried, Lorentz, 2023), circadian rhythm disorder (Oishi et al., 2005; Roybal et al., 2007), obstructive pulmonary disease and cellular senescence (Li L. et al., 2022), which are age-related disorders (Jacenic et al., 2019; Li Z. et al., 2021; Neba Ambe et al., 2022; Siniscalchi et al., 2024).

Additionally, the right half of Figure 4 contains the annotation that resulted from our work with experimental data on an age-related growth of bone fragility in mice concurrently with a decrease in the expression levels of the mouse gene *Clock* (Kaya et al., 2022) in terms of a decrease in the expression levels of *BMAL1*, a human paralog to *Clock*, according to the GeneCards database (Stelzer et al., 2016), in the following age-related human diseases: cancer (Elshazley et al., 2012), circadian rhythm disorder and Parkinson's disease (Ding et al., 2011). This information serves to verify our *in silico* predictions (Fig. 3).

The pilot release of MiceDEGdb contains 85,092 such annotations for 318 human genes, changes in the expression levels of which have clinical manifestations in 895 diseases, as suggest 912 original articles referenced by their PubMed ID. The information contained in MiceDEGdb may be of interest to geneticists, molecular biologists, bioinformatics scientists, clinicians, pharmacologists and genetic advisors in personalized medicine.

MiceDEGdb is freely available at URL=<https://www.sysbio.ru/MiceDEGdb>.

DEGs in mice used as model animals in biomedicine: revealed by RNA-Seq and documented in MiceDEGdb

No.	Strain	Tissue	Syndrome	Model	Norm	N _{DEG}	Reference
1	C57BL/6J	Gum	Periodontitis	Affected mice	Healthy mice	43	Chen Z. et al., 2023
2	C57BL/6J	Bone	Fragility	30 months	2 months	3,725	Kaya et al., 2022
3	C57BL/6J	Bone	Fragility	23 months	2 months	1,151	
4	C57BL/6J	Bone	Fragility	11 months	2 months	1,011	
5	RC ZG	Kidney	Fibrosis	27 months, ♀♂	2 months	43	Wang et al., 2018
6	RC ZG	Kidney	Fibrosis	27 months, ♀	2 months	100	
7	RC ZG	Kidney	Fibrosis	27 months, ♂	2 months	349	
8	C57BL/6J	Kidney	Aging	24 months, ♂	3 months	599	Li J. et al., 2022
9	PolGMut	RV	HF	PolG:D257A, ♀	C57BL/6J, ♀	402	Gorr et al., 2022
10	C57BL/6J	AASM	Aging	26 months, ♂♂	6 months, ♂♂	23	Kiss et al., 2022
11	C57BL/6J	Hippocampus	CRD	Biorhythm 8:8, ♀	Norm 12:12, ♀	158	Fang et al., 2021
12	C57BL/6J	Muscle	Aging	20 weeks, WT, 60 %	20 weeks, WT, al	1,178	Myers et al., 2021
13	C57BL/6J	Muscle	Aging	80 weeks, WT, 60 %	80 weeks, WT, al	747	
14	C57BL/6J	Muscle	Aging	20 weeks, KO, 60 %	20 weeks, KO, al	2,323	
15	C57BL/6J	Muscle	Aging	20 weeks, KI, 60 %	20 weeks, KI, al	1,919	
16	C57BL/6J	Muscle	Aging	80 weeks, KO, 60 %	20 weeks, KO, al	721	
17	C57BL/6J	Muscle	Aging	80 weeks, KI, 60 %	20 weeks, KI, al	2,641	
18	C57BL/6J	Muscle	Aging	80 weeks, KO, al	80 weeks, WT, al	1,976	
19	C57BL/6J	Muscle	Aging	80 weeks, WT, al	80 weeks, KI, al	445	
20	C57BL/6J	Muscle	Aging	20 weeks, KO, al	20 weeks, WT, al	1,152	
21	C57BL/6J	Muscle	Aging	20 weeks, WT, al	20 weeks, KI, al	135	
22	BPH/2J	Kidney	Hypertension	BPH/2J, hypertension	BPN/3J, norm	883	Puig et al., 2010
23	Kunming	Uterus	Toxoplasmosis	Infection, before pregnancy	W/o infection	10	Zhou et al., 2020
24	Kunming	Uterus	Toxoplasmosis	Infection, before embryo implantation	W/o infection	10	
25	Kunming	Uterus	Toxoplasmosis	Infection, after embryo implantation	W/o infection	10	
Σ	6 strains	8 tissues	8 diseases	25 models	17 models	21,754	10 articles

Note. Mouse strain: RC|ZG – Ren1cCre|ZsGreen; PolGMut – B6.129S7(Cg)-Polgtm1Prol/J. Sex: ♀ – females; ♂ – males; ♂♂ – parabionts (surgically integrated blood systems). Tissues: RV – right ventricle; AASM – aortic arch smooth muscle. Diseases: HF – heart failure; CRD – circadian rhythm disorder; WT – wild type; KO – the *Sirt1* gene knocked-out; KI – the *Sirt1* gene knocked-in; al – food *ad libitum*.

Discussion

To show how MiceDEGdb, a knowledge base on the genes that are differentially expressed in the mouse used as a model object in biomedical research, works, we considered DEGs associated with ageing-related bone fragility in C57BL/6 mice aged from 2 to 23 and 30 months (Kaya et al., 2022).

Our attention was drawn to the differential expression levels of the *Clock* and *Bmal1* genes encoding the components of

the transcription factor *Clock/Bmal1*, one of the central components of the mammalian circadian oscillator, because the circadian clock system is known to be a factor of bone health (Swanson et al., 2018). Mice with the *Clock* gene knocked out show a reduction in bone density (Yuan et al., 2017). Mice with the *Bmal1* gene knocked out are noticed to have a reduction in bone weight and density (Chen G. et al., 2020, Kikyō et al., 2024). *Bmal1* regulates osteoclast differentiation

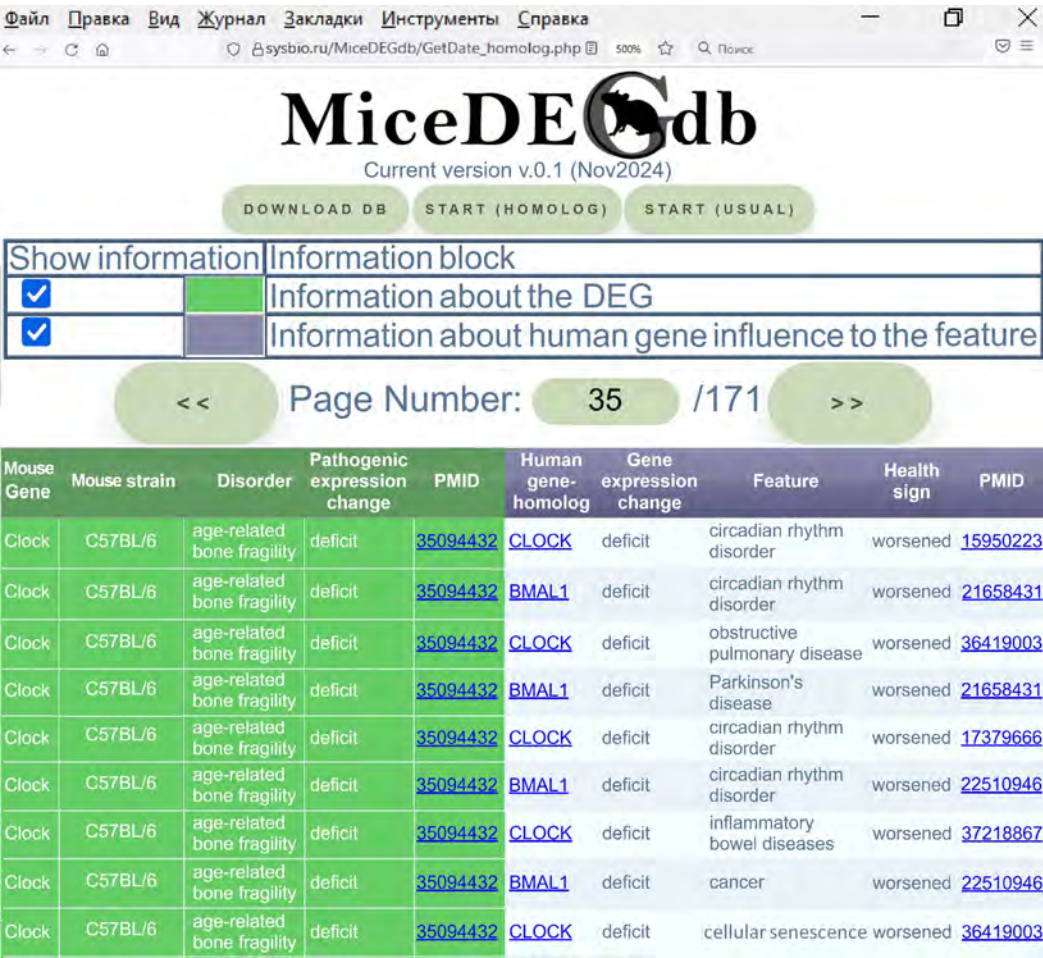


Fig. 4. A session of MiceDEGdb, a knowledge base on the genes that are differentially expressed in the mouse used as a model object in biomedical research, for verification of the results of *in silico* modeling against independent experimental data.

and bone resorption through direct and indirect mechanisms (Chen G. et al., 2020).

The measures of differential expression of the *Clock* and *Bmal1* genes in the C57BL/6 mice were significantly lower in the older than the younger group (Kaya et al., 2022) (Fig. 4, referenced by PubMed ID = 35094432).

As is known, aging is accompanied by circadian rhythm disorder, which coordinates virtually every process in living organisms, including bone tissue modeling and remodeling. This received further support from the results of our computational modeling which showed, in particular, that some parameters of the circadian rhythm and the expression levels of the circadian oscillator components substantially change with age (Podkolodnyy et al., 2016).

We searched PubMed for publications about same-direction changes in the expression levels of the mouse gene *Clock* and the human gene *Bmal1* in patients with various diseases. Note that that the as-published decrease in the expression levels of these genes is typical of age-related human pathologies, such as cancer, inflammation, neurodegenerative diseases, diabetes, circadian rhythm disorder and misregulated cellular senescence (Fig. 4). The MiceDEGdb outputs of analysis of

DEGs associated with the aging-related bone fragility showed that interpreting DEGs with the use of additional information in scientific publications and the results of mathematical modeling gives quite a harmonized view of age-related changes.

Finally, MiceDEGdb as a knowledge base on the mouse used as a model of human diseases is a logical step in expanding the family of databases on animal DEGs created and used for biomedical and pharmaceutical purposes. MiceDEGdb is, in a way, “sequel” to RatDEGdb (Chadaeva et al., 2023) on the ISIAH and OXYS rats, unique strains that have been developed at the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia) and that represent genetic models of arterial hypertension and premature aging, respectively, as well as related diseases.

Conclusion

The MiceDEGdb knowledge base is a collection of experimental data and a toolbox for interactive analysis as part of the genomic studies in the mouse used as a model object in biomedical research.

The existing medical databases focus on the human genome (Sun et al., 2022), and so MiceDEGdb, which holds data on the mouse as the most frequently used laboratory animal in biomedical and pharmaceutical research, should be a valuable add-on to them.

We are planning to keep updating MiceDEGdb with the main focus on the mouse gene expression data coming from the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia). The MiceDEGdb interface (Fig. 4) will be improved following identification, accumulation and systematization of the most trending search queries.

References

- Amaladoss A., Chen Q., Liu M., Dummmler S.K., Dao M., Suresh S., Chen J., Preiser P.R. *De novo* generated human red blood cells in humanized mice support *Plasmodium falciparum* infection. *PLoS One*. 2015;10(6):e0129825. doi 10.1371/journal.pone.0129825
- Bruter A.V., Varlamova E.A., Okulova Y.D., Tatarskiy V.V., Silaeva Y.Y., Filatov M.A. Genetically modified mice as a tool for the study of human diseases. *Mol Biol Rep*. 2024;51(1):135. doi 10.1007/s11033-023-09066-0
- Chadaeva I.V., Filonov S.V., Zolotareva K.A., Khandaev B.M., Ershov N.I., Podkolodnyy N.L., Kozhemyakina R.V., Rasskazov D.A., Bogomolov A.G., Kondratyuk E.Yu., Klimova N.V., Shikhevich S.G., Ryazanova M.A., Fedoseeva L.A., Redina O.E., Kozhevnikova O.S., Stefanova N.A., Kolosova N.G., Markel A.L., Ponomarenko M.P., Oshchepkov D.Yu. RatDEGdb: a knowledge base of differentially expressed genes in the rat as a model object in biomedical research. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2023;27(7):794-806. doi 10.18699/VJGB-23-92
- Chen G., Tang Q., Yu S., Xie Y., Sun J., Li S., Chen L. The biological function of BMAL1 in skeleton development and disorders. *Life Sci*. 2020;253:117636. doi 10.1016/j.lfs.2020.117636
- Chen Z., Huang Z., Zhao X., Zhou Y., Zhang P., Li Y. Transcriptome analysis of differentially expressed genes involved in the inflammatory status of gingiva in aged mice. *Oral Dis*. 2023;29(4):1757-1769. doi 10.1111/odi.14222
- Chuang D.M., Chen R.W., Chalecka-Franaszek E., Ren M., Hashimoto R., Senatorov V., Kanai H., Hough C., Hiroi T., Leeds P. Neuroprotective effects of lithium in cultured cells and animal models of diseases. *Bipolar Disord*. 2002;4(2):129-136. doi 10.1034/j.1399-5618.2002.01179.x
- Chuprin J., Buettner H., Seedhom M.O., Greiner D.L., Keck J.G., Ishikawa F., Shultz L.D., Brehm M.A. Humanized mouse models for immuno-oncology research. *Nat Rev Clin Oncol*. 2023;20(3):192-206. doi 10.1038/s41571-022-00721-2
- Conti L., Reitano E., Cattaneo E. Neural stem cell systems: diversities and properties after transplantation in animal models of diseases. *Brain Pathol*. 2006;16(2):143-154. doi 10.1111/j.1750-3639.2006.00009.x
- Ding H., Liu S., Yuan Y., Lin Q., Chan P., Cai Y. Decreased expression of *Bmal2* in patients with Parkinson's disease. *Neurosci Lett*. 2011;499(3):186-188. doi 10.1016/j.neulet.2011.05.058
- Elshazley M., Sato M., Hase T., Yamashita R., Yoshida K., Toyokuni S., Ishiguro F., Osada H., Sekido Y., Yokoi K., Usami N., Shames D.S., Kondo M., Gazdar A.F., Minna J.D., Hasegawa Y. The circadian clock gene *BMAL1* is a novel therapeutic target for malignant pleural mesothelioma. *Int J Cancer*. 2012;131(12):2820-2831. doi 10.1002/ijc.27598
- Fang K., Liu D., Pathak S.S., Yang B., Li J., Karthikeyan R., Chao O.Y., Yang Y.M., Jin V.X., Cao R. Disruption of circadian rhythms by ambient light during neurodevelopment leads to autistic-like molecular and behavioral alterations in adult mice. *Cells*. 2021;10(12):3314. doi 10.3390/cells10123314
- Filonov S.V., Podkolodnyy N.L., Podkolodnaya O.A., Tverdokhlebo N.N., Ponomarenko P.M., Rasskazov D.A., Bogomolov A.G., Ponomarenko M.P. Human_SNP_TATAdb: a database of SNPs that statistically significantly change the affinity of the TATA-binding protein to human gene promoters: genome-wide analysis and use cases. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2023;27(7):728-736. doi 10.18699/VJGB-23-85
- Frias-Staheli N., Dörner M., Marukian S., Billerbeck E., Labitt R.N., Rice C.M., Ploss A. Utility of humanized BLT mice for analysis of dengue virus infection and antiviral drug testing. *J Virol*. 2014;88(4):2205-2218. doi 10.1128/JVI.03085-13
- Giebried J., Lorentz A. Relationship between the biological clock and inflammatory bowel disease. *Clocks Sleep*. 2023;5(2):260-275. doi 10.3390/clockssleep5020021
- Girard C.A., Wunderlich F.T., Shimomura K., Collins S., Kaizik S., Proks P., Abdulkader F., Clark A., Ball V., Zubcevic L., Bentley L., Clark R., Church C., Hugill A., Galvanovskis J., Cox R., Rorsman P., Bruning J.C., Ashcroft F.M. Expression of an activating mutation in the gene encoding the K_{ATP} channel subunit Kir6.2 in mouse pancreatic beta cells recapitulates neonatal diabetes. *J Clin Invest*. 2009;119(1):80-90. doi 10.1172/jci35772
- Gorr M.W., Francois A., Marcho L.M., Saldana T., McGrail E., Sun N., Stratton M.S. Molecular signature of cardiac remodeling associated with Polymerase Gamma mutation. *Life Sci*. 2022;298:120469. doi 10.1016/j.lfs.2022.120469
- Gryksa K., Schmidtner A.K., Masis-Calvo M., Rodriguez-Villagra O.A., Havasi A., Wirobski G., Maloumbry R., Jagle H., Bosch O.J., Slatery D.A., Neumann I.D. Selective breeding of rats for high (HAB) and low (LAB) anxiety-related behaviour: A unique model for comorbid depression and social dysfunctions. *Neurosci Biobehav Rev*. 2023;152:105292. doi 10.1016/j.neubiorev.2023.105292
- Gunawan M., Her Z., Liu M., Tan S.Y., Chan X.Y., Tan W.W.S., Dharmaraaja S., Fan Y., Ong C.B., Loh E., Chang K.T.E., Tan T.C., Chan J.K.Y., Chen Q. A novel human systemic lupus erythematosus model in humanised mice. *Sci Rep*. 2017;7(1):16642. doi 10.1038/s41598-017-16999-7
- Hild B., Dreier M.S., Oh J.H., McCulloch J.A., Badger J.H., Guo J., Thefaine C.E., Umarova R., Hall K.D., Gavrilova O., Rosshart S.P., Trinchieri G., Rehmann B. Neonatal exposure to a wild-derived microbiome protects mice against diet-induced obesity. *Nat Metab*. 2021;3(8):1042-1057. doi 10.1038/s42255-021-00439-y
- Jacenic D., Cygankiewicz A.I., Mokrowiecka A., Malecka-Panas E., Fichna J., Krajewska W.M. Sex- and age-related estrogen signaling alteration in inflammatory bowel diseases: modulatory role of estrogen receptors. *Int J Mol Sci*. 2019;20(13):3175. doi 10.3390/ijms2013175
- Kaya S., Schurman C.A., Dole N.S., Evans D.S., Alliston T. Prioritization of genes relevant to bone fragility through the unbiased integration of aging mouse bone transcriptomics and human GWAS analyses. *J Bone Miner Res*. 2022;37(4):804-817. doi 10.1002/jbmr.4516
- Keng C.T., Sze C.W., Zheng D., Zheng Z., Yong K.S., Tan S.Q., Ong J.J., Tan S.Y., Loh E., Upadya M.H., Quick C.H., Hotta H., Lim S.G., Tan T.C., Chang K.T., Hong W., Chen J., Tan Y.J., Chen Q. Characterisation of liver pathogenesis, human immune responses and drug testing in a humanised mouse model of HCV infection. *Gut*. 2016;65(10):1744-1753. doi 10.1136/gutjnl-2014-307856
- Kikyo N. Circadian regulation of bone remodeling. *Int J Mol Sci*. 2024;25(9):4717. doi 10.3390/ijms25094717
- Kim J.K., Forger D.B. A mechanism for robust circadian timekeeping via stoichiometric balance. *Mol Syst Biol*. 2012;8:630. doi 10.1038/msb.2012.62

- Kiss T., Nyul-Toth A., Gulej R., Tarantini S., Csipo T., Mukli P., Ungvari A., Balasubramanian P., Yabluchanskiy A., Benyo Z., Conley S.M., Wren J.D., Garman L., Huffman D.M., Csiszar A., Ungvari Z. Old blood from heterochronic parabionts accelerates vascular aging in young mice: transcriptomic signature of pathologic smooth muscle remodeling. *GeroScience*. 2022;44(2):953-981. doi 10.1007/s11357-022-00519-1
- Krause C., Suwada K., Blomme E.A.G., Kowalkowski K., Liguori M.J., Mahalingaiah P.K., Mittelstadt S., Peterson R., Rendino L., Vo A., Van Vleet T.R. Preclinical species gene expression database: development and meta-analysis. *Front Genet*. 2023;13:1078050. doi 10.3389/fgene.2022.1078050
- Li J., Gao F., Wei L., Chen L., Qu N., Zeng L., Luo Y., Huang X., Jiang H. Predict the role of lncRNA in kidney aging based on RNA sequencing. *BMC Genomics*. 2022;23(1):254. doi 10.1186/s12864-022-08479-8
- Li L., Zhang M., Zhao C., Cheng Y., Liu C., Shi M. Circadian clock gene Clock-Bmal1 regulates cellular senescence in Chronic obstructive pulmonary disease. *BMC Pulm Med*. 2022;22(1):435. doi 10.1186/s12890-022-02237-y
- Li Z., Zhang Z., Ren Y., Wang Y., Fang J., Yue H., Ma S., Guan F. Aging and age-related diseases: from mechanisms to therapeutic strategies. *Biogerontology*. 2021;22(2):165-187. doi 10.1007/s10522-021-09910-5
- Liu B., Cui D., Liu J., Shi J.S. Transcriptome analysis of the aged SAMP8 mouse model of Alzheimer's disease reveals novel molecular targets of formononetin protection. *Front Pharmacol*. 2024;15:1440515. doi 10.3389/fphar.2024.1440515
- Liu L., van Schaik T.A., Chen K.S., Rossignoli F., Borges P., Vrbanc V., Wakimoto H., Shah K. Establishment and immune phenotyping of patient-derived glioblastoma models in humanized mice. *Front Immunol*. 2024;14:1324618. doi 10.3389/fimmu.2023.1324618
- Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011;2011:baq036. doi 10.1093/database/baq036
- Lukacs N.W., Strieter R.M., Standiford T.J., Kunkel S.L. Characterization of chemokine function in animal models of diseases. *Methods*. 1996;10(1):158-165. doi 10.1006/meth.1996.0090
- Monteiro C.J., Heery D.M., Whitchurch J.B. Modern approaches to mouse genome editing using the CRISPR-Cas toolbox and their applications in functional genomics and translational research. *Adv Exp Med Biol*. 2023;1429:13-40. doi 10.1007/978-3-031-33325-5_2
- Myers M.J., Shaik F., Shaik F., Alway S.E., Mohamed J.S. Skeletal muscle gene expression profile in response to caloric restriction and aging: a role for Sirt1. *Genes (Basel)*. 2021;12(5):691. doi 10.3390/genes12050691
- Neba Ambe G.N.N., Breda C., Bhambra A.S., Arroo R.R.J. Effect of the citrus flavone nobiletin on circadian rhythms and metabolic syndrome. *Molecules*. 2022;27(22):7727. doi 10.3390/molecules27227727
- Oishi K., Ohkura N., Amagai N., Ishida N. Involvement of circadian clock gene *Clock* in diabetes-induced circadian augmentation of plasminogen activator inhibitor-1 (*PAI-1*) expression in the mouse heart. *FEBS Lett*. 2005;579(17):3555-3559. doi 10.1016/j.febslet.2005.05.027
- 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., Ostannin 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
- Podkolodny N.L., Tverdokhle N.N., Podkolodnaya O.A. Computational model for mammalian circadian oscillator: interacting with NAD+/SIRT1 pathway and age-related changes in gene expression of circadian oscillator. *Vavilovskii Zhurnal Genetiki i Selektzii = Vavilov J Genet Breed*. 2016;20(6):848-856. doi 10.18699/VJ16.201
- Puig O., Wang I.M., Cheng P., Zhou P., Roy S., Cully D., Peters M., Benita Y., Thompson J., Cai T.Q. Transcriptome profiling and network analysis of genetically hypertensive mice identifies potential pharmacological targets of hypertension. *Physiol Genomics*. 2010;42A(1):24-32. doi 10.1152/physiolgenomics.00010.2010
- Roybal K., Theobald D., Graham A., DiNieri J.A., Russo S.J., Krishnan V., Chakravarty S., Peevey J., Oehrlein N., Birnbaum S., Vitaterna M.H., Orsulak P., Takahashi J.S., Nestler E.J., Carlezon W.A. Jr., McClung C.A. Mania-like behavior induced by disruption of *CLOCK*. *Proc Natl Acad Sci USA*. 2007;104(15):6406-6411. doi 10.1073/pnas.0609625104
- Sarsani V.K., Raghupathy N., Fiddes I.T., Armstrong J., Thibaud-Nissen F., Zinder O., Bolisetty M., Howe K., Hinerfeld D., Ruan X., Rowe L., Barter M., Ananda G., Paten B., Weinstock G.M., Churchill G.A., Wiles M.V., Schneider V.A., Srivastava A., Reinholdt L.G. The genome of C57BL/6J "Eve", the mother of the laboratory mouse genome reference strain. *G3 (Bethesda)*. 2019;9(6):1795-1805. doi 10.1534/g3.119.400071
- Segalat L. Invertebrate animal models of diseases as screening tools in drug discovery. *ACS Chem Biol*. 2007;2(4):231-236. doi 10.1021/cb700009m
- Siniscalchi C., Nouvenne A., Cerundolo N., Meschi T., Ticinesi A.; on behalf of the Parma Post-Graduate Specialization School in Emergency-Urgency Medicine Interest Group on Thoracic Ultrasound. Diaphragm ultrasound in different clinical scenarios: a review with a focus on older patients. *Geriatrics (Basel)*. 2024;9(3):70. doi 10.3390/geriatrics9030070
- Stelzer G., Rosen N., Plaschkes I., Zimmerman S., Twik M., Fishilevich S., Stein T.I., Nudel R., Lieder I., Mazon Y., Kaplan S., Dahary D., Warshawsky D., Guan-Golan Y., Kohn A., Rappaport N., Safran M., Lancet D. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*. 2016;54:1.30.1-1.30.33. doi 10.1002/cpbi.5
- Sun S., Wang Y., Maslov A.Y., Dong X., Vijg J. SomaMutDB: a database of somatic mutations in normal human tissues. *Nucleic Acids Res*. 2022;50(D1):D1100-D1108. doi 10.1093/nar/gkab914
- Swanson C.M., Kohrt W.M., Buxton O.M., Everson C.A., Wright K.P. Jr., Orwoll E.S., Shea S.A. The importance of the circadian system & sleep for bone health. *Metabolism*. 2018;84:28-43. doi 10.1016/j.metabol.2017.12.002
- Swindell W.R., Johnston A., Sun L., Xing X., Fisher G.J., Bulyk M.L., Elder J.T., Gudjonsson J.E. Meta-profiles of gene expression during aging: limited similarities between mouse and human and an unexpectedly decreased inflammatory signature. *PLoS One*. 2012;7(3):e33204. doi 10.1371/journal.pone.0033204
- Vandamme T.F. Use of rodents as models of human diseases. *J Pharm Bioallied Sci*. 2014;6(1):2-9. doi 10.4103/0975-7406.124301
- Viehmann Milam A.A., Maher S.E., Gibson J.A., Lebastchi J., Wen L., Ruddle N.H., Herold K.C., Bothwell A.L. A humanized mouse model of autoimmune insulinitis. *Diabetes*. 2014;63(5):1712-1724. doi 10.2337/db13-1141
- Wang Y., Eng D.G., Pippin J.W., Gharib S.A., McClelland A., Gross K.W., Shankland S.J. Sex differences in transcriptomic profiles in aged kidney cells of renin lineage. *Aging (Albany NY)*. 2018;10(4):606-621. doi 10.18632/aging.101416
- White P.L., Wiederhold N.P., Loeffler J., Najvar L.K., Melchers W., Herrera M., Bretagne S., Wickes B., Kirkpatrick W.R., Barnes R.A., Donnelly J.P., Patterson T.F. Comparison of nonculture

- blood-based tests for diagnosing invasive aspergillosis in an animal model. *J Clin Microbiol.* 2016;54(4):960-966. doi 10.1128/jcm.03233-15
- Yajima M., Imadome K., Nakagawa A., Watanabe S., Terashima K., Nakamura H., Ito M., Shimizu N., Honda M., Yamamoto N., Fujiwara S. A new humanized mouse model of Epstein–Barr virus infection that reproduces persistent infection, lymphoproliferative disorder, and cell-mediated and humoral immune responses. *J Infect Dis.* 2008;198(5):673-682. doi 10.1086/590502
- Yong K.S.M., Her Z., Chen Q. Humanized mice as unique tools for human-specific studies. *Arch Immunol Ther Exp (Warsz).* 2018;66(4):245-266. doi 10.1007/s00005-018-0506-x
- Yuan G., Hua B., Yang Y., Xu L., Cai T., Sun N., Yan Z., Lu C., Qian R. The circadian gene *Clock* regulates bone formation via PDIA3. *J Bone Miner Res.* 2017;32(4):861-871. doi 10.1002/jbmr.3046
- Zayoud M., El Malki K., Frauenknecht K., Trinschek B., Kloos L., Karram K., Wanke F., Georgescu J., Hartwig U.F., Sommer C., Jonuleit H., Waisman A., Kurschus F.C. Subclinical CNS inflammation as response to a myelin antigen in humanized mice. *J Neuroimmune Pharmacol.* 2013; 8(4):1037-1047. doi 10.1007/s11481-013-9466-4
- Zhou X., Zhang X.X., Mahmood Y.S., Hernandez J.A., Li G.F., Huang W.Y., Wang Y.P., Zheng Y.X., Li X.M., Yuan Z.G. A transcriptome analysis: various reasons of adverse pregnancy outcomes caused by acute *Toxoplasma gondii* infection. *Front Physiol.* 2020; 11:115. doi 10.3389/fphys.2020.00115

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

Received November 13, 2024. Revised December 9, 2024. Accepted December 16, 2024.

doi 10.18699/vjgb-25-19

Computer reconstruction of gene networks controlling anxiety levels in humans and laboratory mice

E.G. Vergunov ^{1, 2, 3} , V.A. Savostyanov ⁴, A.A. Makarova ¹, E.I. Nikolaeva ⁴, A.N. Savostyanov ^{1, 2, 3}

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

² Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

³ Novosibirsk State University, Novosibirsk, Russia

⁴ Herzen University, St. Petersburg, Russia

 vergounov@gmail.com

Abstract. Anxiety is a normotypic human condition, and like any other emotion has an adaptive value. But excessively high or low anxiety has negative consequences for adaptation, which primarily determines the importance of studying these two extreme conditions. At the same time, it is known that the perception of aversive stimuli associated with anxiety leads to changes in the activity of the brain's cingulate cortex. The advantage of animals as models in studying the genetic bases of anxiety in humans is in the ability to subtly control the external conditions of formation of a certain state, the availability of brain tissues, and the ability to create and study transgenic models, including through the use of differentially expressed genes of small laboratory animals from the family Muridae with low and high anxiety. Within the framework of the translational approach, a three-domain potential gene network, which is associated with generalized anxiety in humans, was reconstructed using mouse models with different levels of anxiety by automatically analyzing the texts of scientific articles. One domain is associated with reduced anxiety in humans, the second with increased anxiety, and the third is a dispatcher who activates one of the two domains depending on the status of the organism (genetic, epigenetic, physiological). Stages of work: (I) A list of genes expressed in the cingulate cortex of the wild type CD-1 mouse line from the NCBI GEO database (experiment GSE29014). Using the tools of this database, differences in gene expression levels were revealed in groups of mice with low and high (relatively normal) anxiety. (II) Search for orthologs of DEG in humans and mice associated with anxiety in the OMA Orthology database. (III) Computer reconstruction using the ANDSystem cognitive system based on (a) human orthologous genes from stage (III), (b) human genes from the MalaCards database associated with human anxiety. The proven methods of the translational approach for the reconstruction of gene networks for behavior regulation can be used to identify molecular genetic markers of human personality traits, propensity to psychopathology.

Key words: differentially expressed genes; cingulate cortex; automatic text analysis; scientific publications; computer reconstruction; gene networks; mouse model with high-normal-low anxiety behavior.

For citation: Vergunov E.G., Savostyanov V.A., Makarova A.A., Nikolaeva E.I., Savostyanov A.N. Computer reconstruction of gene networks controlling anxiety levels in humans and laboratory mice. *Vavilovskii Zhurnal Genetiki i Selektzii* = *Vavilov J Genet Breed*. 2025;29(1):162-170. doi 10.18699/vjgb-25-19

Funding. The work was conducted under the budget project No. FWNR-2022-0020.

Компьютерная реконструкция и анализ генных сетей, контролирующих уровень тревожности у лабораторных мышей и человека

Е.Г. Вергунов ^{1, 2, 3} , В.А. Савостьянов ⁴, А.А. Макарова ¹, Е.И. Николаева ⁴, А.Н. Савостьянов ^{1, 2, 3}

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

² Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия

³ Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

⁴ Российский государственный педагогический университет им. А.И. Герцена, Санкт-Петербург, Россия

 vergounov@gmail.com

Аннотация. Тревожность – это нормотипичное состояние человека, которое, как и любая другая эмоция, имеет адаптивное значение. Но состояние чрезмерно высокой или низкой тревожности влечет за собой негативные последствия для адаптации, что в первую очередь обуславливает важность изучения этих двух крайних состояний. При этом известно, что в условиях восприятия авersive стимулов, ассоциированных с тревожностью, изменяется активность поясной коры мозга. Преимущество животных как моделей при изучении генети-

ческих оснований тревожности у человека связано с возможностью тонко контролировать внешние условия формирования определенного состояния, доступностью тканей мозга и возможностью создавать и изучать трансгенные модели, в том числе с использованием дифференциально экспрессирующихся генов мелких лабораторных животных из семейства мышиных с низкой и высокой тревожностью. В рамках трансляционного подхода была реконструирована трехдоменная потенциальная генная сеть, которая ассоциирована с генерализованной тревожностью у человека, по моделям мышей с разным уровнем тревожности путем автоматического анализа текстов научных статей. Один домен ассоциирован с пониженной тревожностью у человека, второй – с повышенной, третий служит диспетчером, который активирует один из двух доменов в зависимости от статуса организма (генетического, эпигенетического, физиологического). Этапы работы: (I) из базы данных NCBI GEO взят список генов, экспрессирующихся в поясной коре головного мозга линии мышей дикого типа CD-1 (эксперимент GSE29014). С помощью инструментов этой базы выявлены различия в уровнях экспрессии генов в группах мышей с низкой и высокой (относительно нормальной) тревожностью; (II) поиск ортологов ДЭГ у человека и мышей, ассоциированных с тревожностью в базе данных OMA Orthology; (III) компьютерная реконструкция с помощью когнитивной системы ANDSystem на основе генов-ортологов человека из этапа (II), генов человека из базы данных MalaCards, ассоциированных с тревожностью человека. Апробированные методы трансляционного подхода для реконструкции генных сетей регуляции поведения могут использоваться для выявления молекулярно-генетических маркеров черт личности человека, склонности к психопатологии.

Ключевые слова: дифференциально экспрессирующиеся гены; поясная кора головного мозга; автоматический анализ текстов; научные публикации; компьютерная реконструкция; генные сети; модель мышей с поведением высокой-нормальной-низкой тревожности.

Introduction

Anxiety is a normotypical human condition (Malezieux et al., 2023), and like any other emotion has adaptive value (Stein, Bouwer, 1997). However, a state of excessive anxiety or the complete absence of it has negative consequences for adaptation (Penninx et al., 2021). It is considered proven (Hettema et al., 2001) that a combination of genetic and environmental factors is the cause of extreme variations in the expression of anxiety.

The advantage of animals as models in studying the genetic basis of anxiety in humans is due to the ability to precisely control the external conditions for the formation of a particular state, because of the availability of brain tissue, and the ability to create and study transgenic models (Vandamme, 2014; Chadaeva et al., 2023; Krause et al., 2023), including using differentially expressed genes (DEGs) of small laboratory animals from the low anxiety behavior (LAB) and high anxiety behavior (HAB) mouse families (Gryksa et al., 2023). When comparing animal and human models, genetic studies of humans with generalized anxiety disorder are compared with rodent models obtained by exposure to stressful stimuli (Koskinen, Hovatta, 2023). The relevance of such models for understanding the molecular basis of anxiety is evident.

Currently, a large number of genes are considered in explaining anxiety in humans (Otowa et al., 2016; Koskinen, Hovatta, 2023; Mucha et al., 2023). The molecular mechanisms of anxiety in both humans and animals are related to differential gene activity of neurotransmitter systems, predominantly serotonin and dopamine, as well as the involvement of other monoamines and GABA (Morris-Rosendahl, 2002; Nuss, 2015; Gottschalk, Domschke, 2017; Galyamina et al., 2018; Moraes et al., 2024; Strom et al., 2024). At the same time, the role of genetic polymorphisms in determining the level of anxiety has been noted (Sen et al., 2004; Ivanov et al., 2019).

Genetic markers of anxiety in mice and humans are similar in many ways, which allows the results obtained in animals to

be transferred to understanding the mechanisms of anxiety in humans (Hovatta, Barlow, 2008; Hettema et al., 2011; Brasher et al., 2023). It has been pointed out that the manifestation of genetic polymorphisms is highly modified by sociocultural factors and, in general, the relationship between anxiety and genotype in humans is significantly modulated by environmental conditions (Schinka et al., 2004; Ebstein, 2006; Meng et al., 2024; Petrican et al., 2024).

Attempts to identify genetic markers of behavioral traits based on the analysis of candidate genes are usually ineffective due to the fact that there are no single genes that unambiguously determine behavior (Duncan et al., 2014; Bruzzzone et al., 2024). This is explained by the fact that the formation of organisms' phenotypic characteristics is controlled not by individual genes, but by gene networks – groups of coordinately functioning genes interacting with each other through their products – RNA, proteins, and metabolites (Kolchanov et al., 2000, 2013). It is gene networks, functioning on the basis of information encoded in genomes, that ensure the formation of all phenotypic traits of organisms (molecular, biochemical, cellular, physiological, morphological, etc.) (Kolchanov et al., 2013).

We believe that the reconstruction and analysis of gene networks are promising approaches to understanding the molecular genetic mechanisms underlying the formation of human personality characteristics, including those that, like anxiety, are induced by environmental factors. Reconstruction of gene networks and their functional modules is based on molecular genetic data presented in scientific publications and factographic databases, such as human and animal genome sequencing data, information on differentially expressed genes, allelic polymorphisms associated with target phenotypic characteristics of organisms, and others (Mostafavi et al., 2008; Krämer et al., 2014; Szklarczyk et al., 2015; Chen et al., 2016; Ivanisenko et al., 2022).

However, the reconstruction of human anxiety gene networks cannot be performed on the basis of *in vivo* experimental

studies that would require sampling of biological brain tissues to obtain molecular genetic data. Therefore, in this work, a translational approach was used based on the analysis of data obtained by L. Czibere et al. (2011) in experiments on mice, in which they studied the differential expression of genes (DEGs) in the cingulate cortex of a line of wild-type CD-1 mice with different levels of anxiety.

This experiment showed that mice with high anxiety exhibited a more passive coping strategy than mice with low anxiety, which is reminiscent of the clinical comorbidity of anxiety and depression (their co-occurrence) observed in psychiatric patients (Czibere et al., 2011). This was the rationale for using mouse DEG data to reconstruct human gene networks involved in the control of different levels of anxiety. The details of this translational approach are described below.

In the reconstructed potential human gene network, we identified three functional domains, one of which is responsible for the reaction of reduced anxiety, another domain is responsible for the reaction of increased anxiety, and the third plays the role of a dispatcher that activates one of the other two domains depending on the genetic, epigenetic, physiological status of the organism and environmental conditions.

Materials and methods

Experimental data. In this study, we used the data from the work of L. Czibere et al. (2011), in which 25 specimens of wild-type mice of the CD-1 line of one generation (*Mus musculus* Linnaeus, 1758; <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=10090>) were subjected to stress exposure (swimming in cold water for 10 min). Afterward, using the MouseWG-6 v1.1 Expression BeadChip-system expression chip (46,132 samples), L. Czibere et al. (2011) assessed gene expression levels in the cingulate cortex of these mice. Experimental animals were divided on the basis of behavioral tests in the sleeves of an elevated cross-shaped maze into three groups: with low (low anxiety behavior, LAB), normal (normal anxiety behavior, NAB) and high (high anxiety behavior, HAB) anxiety (Czibere et al., 2011). The results of the experiment are presented in the NCBI GEO database with the index GSE29014 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29014>).

Computer analysis methods. The list of genes expressed, according to (Czibere et al., 2011), in the cingulate cortex of mice (experiment GSE29014) was taken from the NCBI GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29014>). Identification of differences in gene expression levels between groups of mice with different levels of anxiety was performed using the NCBI GEO toolkit (<https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE29014>). The OMA Orthology database (<https://omabrowser.org/oma/home/>) was used to search for orthologs of DEGs in humans and mice associated with anxiety.

Reconstruction of potential human gene networks associated with high- and low-level generalized anxiety states was performed on the basis of human genes orthologous to mouse genes differentially expressed in the cingulate cortex. For this purpose, we used ANDSystem, a cognitive system developed

at the Institute of Cytology and Genetics SB RAS (Ivanisenko et al., 2019), which uses machine reading and artificial intelligence methods to automatically extract knowledge and facts from large genetic data sources – texts of tens of millions of scientific articles and patents and thousands of factographic databases. The ANDSystem knowledge base currently contains information on 2 million genes and proteins, 46 thousand diseases, tens of thousands metabolites and biological processes, and tens of millions intermolecular interactions (Ivanisenko et al., 2024).

Results

The basic framework for data analysis, starting with the generation of a list of DEGs in the cingulate cortex of mouse line CD-1 for high (HAB) and low (LAB) anxiety groups, which includes a search for orthologs of differentially expressed genes in humans and mice associated with anxiety, and culminating in the reconstruction of potential human gene networks associated with anxiety levels, is shown in Figure 1. Let us review the main results of this approach.

Obtaining a list of DEGs in the cingulate cortex of CD-1 line mice for HAB and LAB anxiety groups

First of all, we searched for differentially expressed genes in the cingulate cortex of CD-1 mice that distinguish (a) the high anxiety group (HAB) from the normal anxiety group (NAB) and (b) the low anxiety group (LAB) from the normal anxiety group (NAB). When comparing the HAB and NAB groups, 185 DEGs were identified, and when comparing the LAB and NAB groups, 193 DEGs were identified (Fig. 1). The number of total DEGs in mice identified in the HAB/NAB and LAB/NAB comparisons is 133. Assessing the significance of such a strong overlap using a Bonferroni-corrected hypergeometric distribution for multiple comparisons yields a $P_{adj} < 8.4 \cdot 10^{-5}$ (Fig. 1).

It can be hypothesized that the stress responses of the two compared pairs of groups of mice corresponding to increased or decreased anxiety are parts of some large gene network that determines the level of anxiety in the stress response.

Search for orthologs of DEGs in humans and mice associated with anxiety

Identification of human genes orthologous to mouse DEGs, identified by comparing gene networks responsible for differences in anxiety levels between the LAB/NAB and HAB/NAB groups of mice, was performed using the OMA Orthology database (<https://omabrowser.org/oma/home/>). For this purpose, a Python script was written that compared mouse ID genes with human orthologs and produced IDs for human genes. In total, such comparisons identified 8 human orthologous genes based on DEGs for LAB/NAB mice and 16 based on DEGs for HAB/NAB mice. The number of human orthologous genes common to the two lists is 5. Assessing the significance of the overlap using a Bonferroni-corrected hypergeometric distribution for multiple comparisons yields a value of $P_{adj} < 0.024 < 0.05$ (Fig. 1).

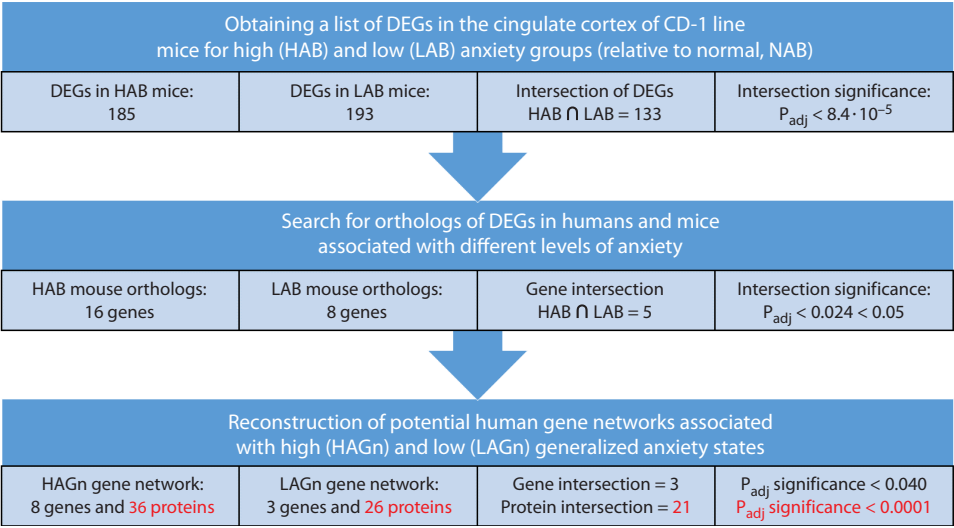


Fig. 1. Basic steps in reconstructing a potential human gene network associated with high and low levels of anxiety.

Reconstruction of potential human gene networks associated with high- and low-level generalized anxiety states

This task was solved using the cognitive system ANDSystem. Two types of data were used. First of all, a sample containing the previously identified 19 different human orthologous genes. And, in addition, 176 human genes associated with anxiety and depressive spectrum disorders, which were extracted from the MalaCards database (<https://www.malacards.org/card/anxiety#Genes>).

On this basis, two potential human gene networks associated with (a) high anxiety (HAGn, High Anxiety Gene Network, containing 8 genes, 36 proteins), and (b) low anxiety (LAGn, Low Anxiety Gene Network, containing 3 genes, 26 proteins) were reconstructed using ANDSystem.

The LAGn gene network responsible for the state of low anxiety level contains a large cluster of 10 interacting proteins and genes and 5 isolated small-sized clusters (Fig. 2).

In the HAGn gene network responsible for the state of high anxiety, first, a large cluster of 32 interacting proteins and genes is distinguished, followed by a medium-sized cluster of 7 proteins and genes, as well as 2 isolated smaller clusters (Fig. 3).

Note that the large HAGn cluster (Fig. 3, I–III) includes the entire three LAGn clusters (Fig. 2, I–III), and the medium-sized HAGn cluster (Fig. 3, IV) includes the entire LAGn cluster (Fig. 2, IV). Two HAGn clusters (Fig. 3, VII, VIII) and two LAGn clusters (Fig. 2, V, VI) have no counterparts among clusters of the other gene network. Although clusters IV, VII, VIII may have overlapping proteins with other clusters in the other gene network, even then their roles in linkages in “their” clusters are different from their roles in clusters in the other gene network.

Both networks (LAGn and HAGn) have 3 genes in common. Assessment of the significance of such intersection of LAGn and HAGn according to the hypergeometric distribution

with Bonferroni correction for multiple comparisons gives $P_{adj} < 0.040$ (< 0.050) (Fig. 1). Both networks (LAGn and HAGn) share 21 common proteins. Assessment of the significance of such intersection of LAGn and HAGn according to the hypergeometric distribution with Bonferroni correction for multiple comparisons gives $P_{adj} < 0.0001$ (Fig. 1).

Discussion

It can be assumed that the 3 identified genes common to the two networks (LAGn and HAGn) form a special gene network – GnI (Gene Network Interface), which regulates the interaction between the gene networks LAGn and HAGn, which are responsible for the formation of low and high anxiety states. Figure 4 shows a qualitative schematic of the interaction between LAGn, HAGn and GnI:

- Domain 1 (part of LAGn) is responsible for the low anxiety response;
- Domain 2 (part of HAGn) is responsible for the higher anxiety response;
- Domain 3 (GnI, the common part for both LAGn and HAGn) acts as an interface between domains 1 and 2. It plays the role of a dispatcher that activates domain 1 or domain 2 depending on the genetic, epigenetic, physiological status of the organism. A discussion of the approach based on the existence of such a dispatcher is given in (Shin et al., 2024).

As our DEG analysis based on the GSE29014 experiment shows, a similar three-domain structure is evident for the interactions between two sets of genes associated with low (LAB) and high (HAB) anxiety in mice, as well as orthologous genes (Fig. 1). As Figure 4 shows, the interactions of LAGn, HAGn, and GnI are complex and need further dedicated study.

We chose the cingulate cortex in our work to identify genes, the expression of which after stress response is associated with an increase or decrease in the level of anxiety in experi-

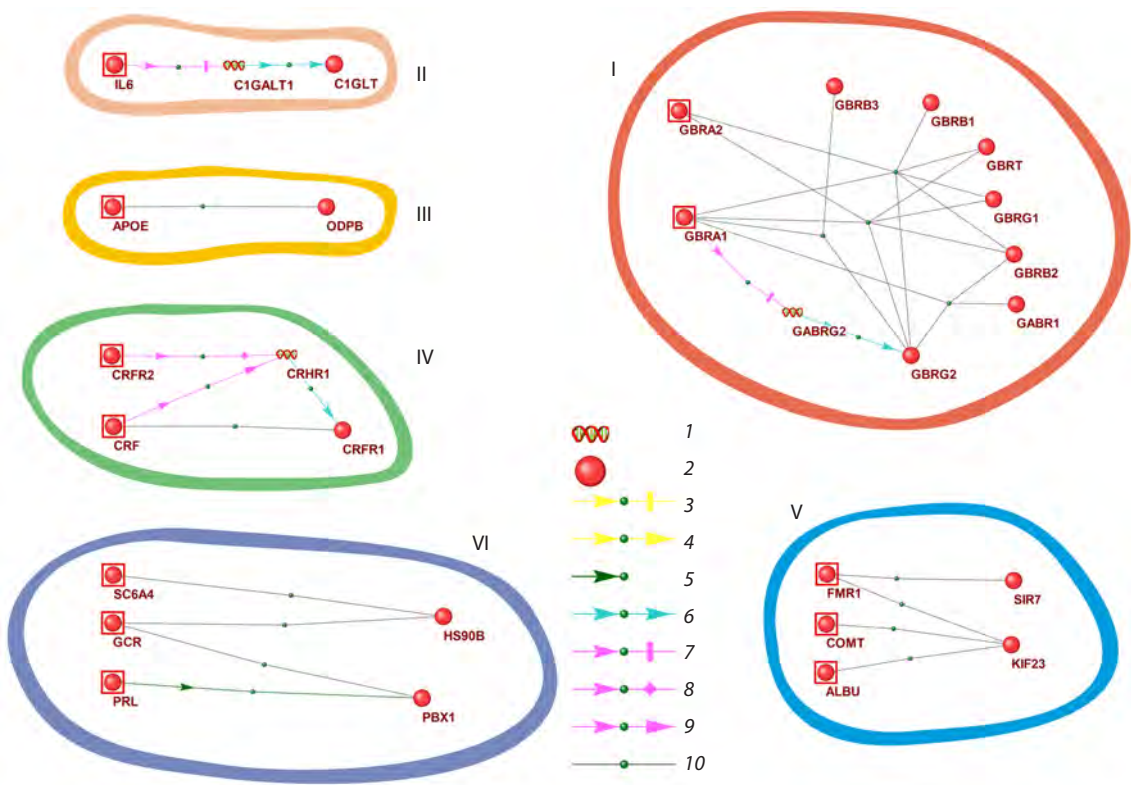


Fig. 2. Visualization of the potential human LAGn gene network responsible for low anxiety state in humans. Roman numerals indicate isolated clusters. Arabic numerals denote: 1 – gene, 2 – protein, 3 – suppression of protein activity, 4 – enhancement of protein activity, 5 – catalytic reaction, 6 – expression, 7 – suppression of gene expression, 8 – regulation of gene expression, 9 – enhancement of gene expression, 10 – protein-protein interaction.

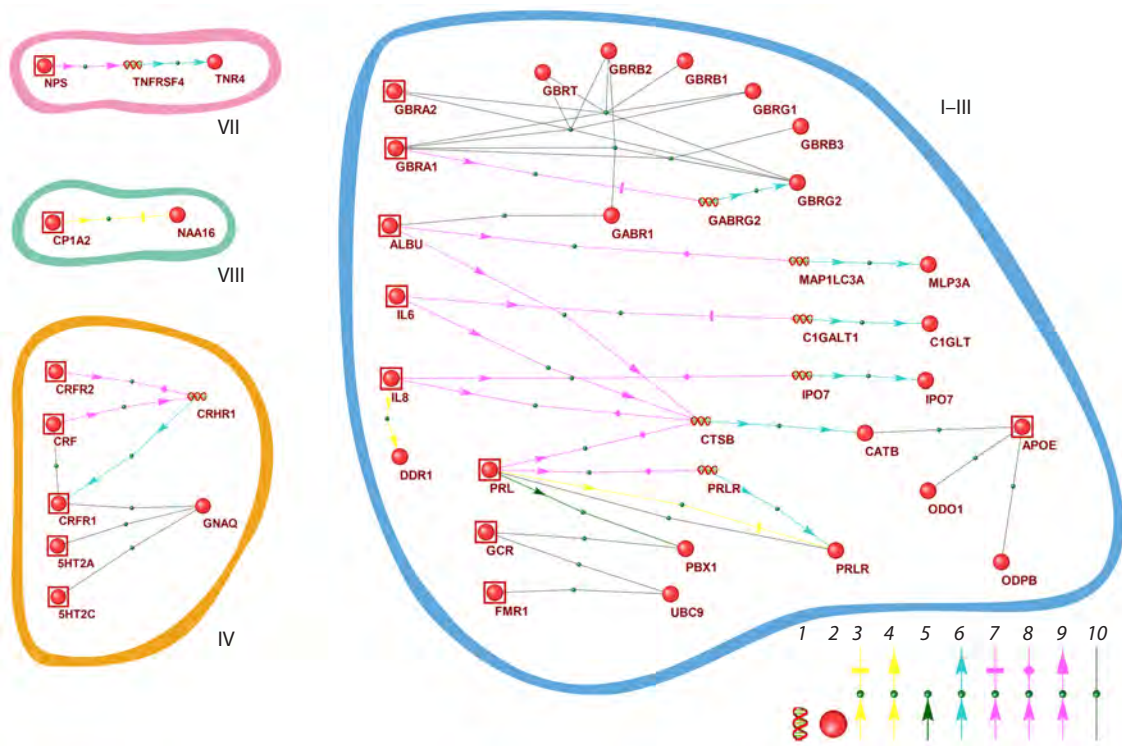


Fig. 3. Visualization of the potential human HAGn gene network responsible for the state of high anxiety in humans. The labeling is analogous to that presented in Figure 2.

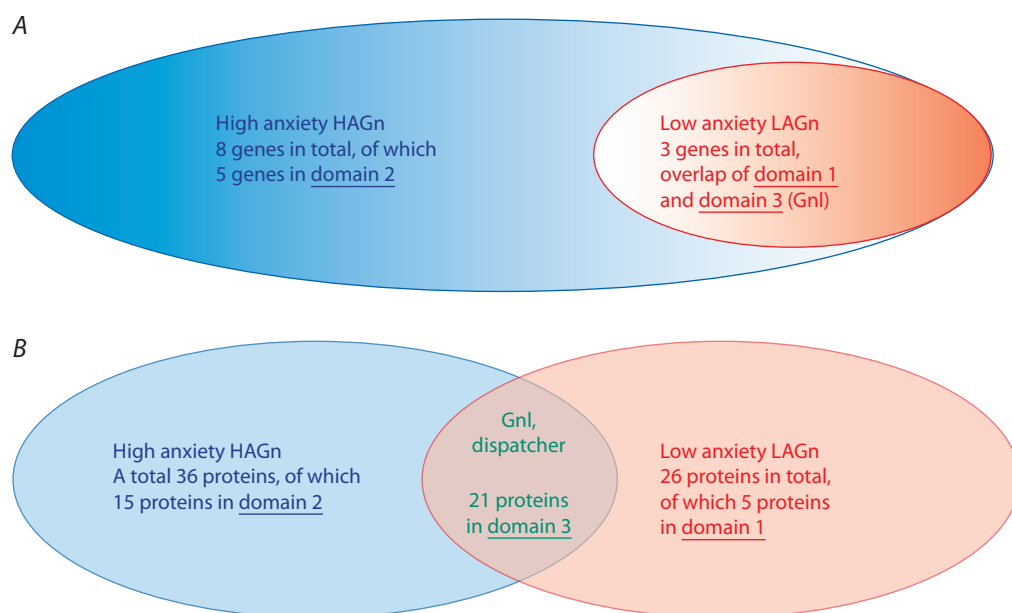


Fig. 4. Qualitative scheme of interaction between LAGn (gene network of reduced anxiety), HAGn (gene network of increased anxiety) and Gnl (dispatcher that activates domain 1 or domain 2).

A – distribution of genes, B – distribution of proteins encoded by the genes.

mental mice, because fMRI studies (de la Pena-Arteaga et al., 2024) had shown altered activity of the cingulate cortex under conditions of perception of aversive stimuli associated with anxiety.

Let us draw attention to the fact that the experiments conducted by L. Czibere et al. (2011) on a genetic line of wild-type CD-1 mice revealed two opposite reactions to the same stressor. This can be explained by the presence of latent genomic variability in the population of the examined mice (the presence of a spectrum of polymorphisms or epigenetic modifications affecting a variety of genomic loci). Perhaps this may explain the fact that the degree of anxiety is a continuum, the scores of which continuously vary from low through medium to high values (Friligkou et al., 2024).

Our analysis showed that the qualitative differences in mice between low (LAB) and normal (NAB) levels of anxiety on the one hand, and high (HAB) and normal (NAB) on the other hand, revealed in the experiment of L. Czibere et al. (2011), may lie in gene networks functioning in the cingulate cortex that provide contrasting states of anxiety relative to normal.

Earlier genetic studies have shown the association of anxiety with genes for brain monoamine systems (Lesch et al., 1996; Murphy et al., 2013). Polymorphisms in serotonin system genes, including genes encoding serotonin receptors and serotonin transporters, are associated with different levels of anxiety (Purves et al., 2020).

The set of human genes we have identified as part of the reconstructed potential gene networks includes genes of monoamine brain systems. These include, for example, serotonin receptors 5HT2A and 5HT2C (a potential gene network domain for high anxiety states). These receptors belong to the G-protein-coupled receptor (GPCR) superfamily and, through interaction with GPCRs, transmit extracellular signals to the

interior of cells. The receptors mediate the effects of a large number of compounds affecting depression, schizophrenia, anxiety, hallucinations, dysthymia, sleep patterns, eating behavior, and neuroendocrine functions (Van Oekelen et al., 2003). This is in good agreement with the monoamine hypothesis of anxiety (Morris-Rosendahl, 2002; Gottschalk, Domschke, 2017; Hirai et al., 2024).

Our reconstructed potential human gene networks also include interactions with genes encoding proteins such as COMT or APoE, which are not related to neurotransmitters but are associated with anxiety and depression through involvement in the regulation of a wide range of metabolic processes (Koskinen, Hovatta, 2023).

It has been established (Gunthert et al., 2007) that there is a functional relationship between genetic polymorphisms and anxiety levels for groups of people living in different environmental conditions. Environmental factors have been shown to interact with genetic markers of anxiety in complex ways, in some cases leading to inversion of allelic polymorphism effects when living conditions change (Schinka et al., 2004; Sen et al., 2004; Ivanov et al., 2019; Meng et al., 2024; Petrican et al., 2024).

It can be hypothesized that the level and directionality of anxiety as a stress response depends on: (a) genes directly involved in neural signal processing; (b) genes regulating other body functions (metabolic, physiological...); (c) the presence of hidden genomic variability – epigenetic modifications, polymorphisms, etc. in the above two groups of genes (a) and (b).

It is known that the results obtained on animal models in drug development cannot always be adequately extrapolated to humans (Hackam, Redelmeier, 2006). There may also be a concern that a study on 25 individuals from a single genera-

tion of a wild mouse line may lead to simplistic conclusions and limited understanding of the complex network of genes involved in anxiety, and any errors or inaccuracies in the original data may lead to incorrect conclusions about the role of genes in anxiety.

In our approbation of the translational approach, such issues were addressed as follows: human and mouse orthologous genes obtained from a list of mouse cingulate cortex DEGs were matched to a set from the MalaCards database (176 human genes that are associated with generalized anxiety and anxiety and depressive spectrum disorders for humans). The MalaCards database provides a set of references to papers describing relevant experiments, allowing validation for each case. After such a comparison, the reconstruction of potential (i. e. assuming special further study) gene networks for humans was carried out with the help of the ANDSystem cognitive system on the basis of automatic analysis (and resolution of inaccuracies and contradictions found in them) of 6 million texts of articles from leading publications on biological topics. Thus, the impact of inaccuracy or insufficiency of the original data in the LAB and NAB mouse groups was reduced to a negligible level in our approbation.

Conclusion

Based on the software resources used in our work and the generated algorithm for analyzing differential expression of genes data, we developed a software module for computer reconstruction of gene networks involved in the regulation of stress response leading to anxiety of different levels.

Within the framework of the translational approach, a three-domain potential gene network, which is associated with generalized anxiety in humans, was reconstructed using mouse models with different levels of anxiety by automatically analyzing the texts of scientific articles. One domain is associated with reduced anxiety in humans, the second with increased anxiety, and the third is a dispatcher who activates one of the two domains depending on the status of the organism (genetic, epigenetic, physiological).

We believe that this approach can be modified to reconstruct gene networks controlling anxiety and other behavioral reactions in stress responses of other types.

Limitations of the present study

The human multidomain gene network we reconstructed, which is associated with generalized anxiety, is potential, that is, it implies dedicated further study and refinement. Thus, this paper takes an initial step in investigating the domains of the gene network that is associated with human anxiety.

References

Brasher M.S., Mize T.J., Thomas A.L., Hoeffler C.A., Ehringer M.A., Evans L.M. Testing associations between human anxiety and genes previously implicated by mouse anxiety models. *Genes Brain Behav.* 2023;22(6):e12851. doi 10.1111/gbb.12851

Bruzzzone S.E.P., Ozenne B., Fisher P.M., Ortega G., Jensen P.S., Dam V.H., Svarer C., Knudsen G.M., Lesch K.P., Frokier V.G. No association between peripheral serotonin-gene-related DNA methylation and brain serotonin neurotransmission in the healthy and depressed state. *Clin Epigenet.* 2024;416(1):71. doi 10.1186/s13148-024-01678-y

Chadaeva I.V., Filonov S.V., Zolotareva K.A., Khandaev B.M., Ershov N.I., Podkolodnyy N.L., Kozhemyakina R.V., Rasskazov D.A., Bogomolov A.G., Kondratyuk E.Y., Klimova N.V., Shikhevich S.G., Ryazanova M.A., Fedoseeva L.A., Redina O.E., Kozhevnikova O.S., Stefanova N.A., Kolosova N.G., Markel A.L., Ponomarenko M.P., Oshchepkov D.Y. RatDEGdb: a knowledge base of differentially expressed genes in the rat as a model object in biomedical research. *Vavilovskii Zhurnal Genet Selektsii = Vavilov J Genet Breed.* 2023;27(7):794-806. doi 10.18699/VJGB-23-92

Chen Y., Elenee Argentinis J.D., Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther.* 2016;38(4):688-701. doi 10.1016/j.clinthera.2015.12.001

Czibere L., Baur L.A., Wittmann A., Gemmeke K., Steiner A., Weber P., Pütz B., Ahmad N., Bunck M., Graf C., Widner R., Kühne C., Panhuysen M., Hamsch B., Rieder G., Reinheckel T., Peters C., Holsboer F., Landgraf R., Deussing J.M. Profiling trait anxiety: transcriptome analysis reveals cathepsin B (*Ctsb*) as a novel candidate gene for emotionality in mice. *PLoS One.* 2011;6(8):e23604. doi 10.1371/journal.pone.0023604

de la Pena-Arteaga V., Chavarria-Elizondo P., Juaneda-Segui A., Martinez-Zalacain I., Morgado P., Menchon J.M., Pico-Perez M., Fullana M.A., Soriano-Mas C. Trait anxiety is associated with attentional brain networks. *Eur Neuropsychopharmacol.* 2024;83:19-26. doi 10.1016/j.euroneuro.2024.02.013

Duncan L.E., Pollastri A.R., Smoller J.W. Mind the gap: why many geneticists and psychological scientists have discrepant views about gene-environment interaction (G×E) research. *Am Psychol.* 2014;69(3):249-268. doi 10.1037/a0036320

Ebstein R.P. The molecular genetic architecture of human personality: beyond self-report questionnaires. *Mol Psychiatry.* 2006;11(5):427-445. doi 10.1038/sj.mp.4001814

Frilgkou E., Løkhammer S., Cabrera-Mendoza B., Shen J., He J., Deiana G., Zanoaga M.D., Asgel Z., Pilcher A., Di Lascio L., Makharashvili A., Koller D., Tylee D.S., Pathak G.A., Polimanti R. Gene discovery and biological insights into anxiety disorders from a large-scale multi-ancestry genome-wide association study. *Nat Genet.* 2024;56(10):2036-2045. doi 10.1038/s41588-024-01908-2

Galyamina A.G., Kovalenko I.L., Smagin D.A., Kudryavtseva N.N. Changes in the expression of neurotransmitter system genes in the ventral tegmental area in depressed mice: RNA-SEQ data. *Neurosci Behav Physiol.* 2018;48(5):591-602. doi 10.1007/s11055-018-0605-5

Gottschalk M.G., Domschke K. Genetics of generalized anxiety disorder and related traits. *Dialogues Clin Neurosci.* 2017;19(2):159-168. doi 10.31887/DCNS.2017.19.2/kdomschke

Gryksa K., Schmidtner A.K., Masis-Calvo M., Rodriguez-Villagra O.A., Havasi A., Wirobski G., Maloumy R., Jagle H., Bosch O.J., Slatery D.A., Neumann I.D. Selective breeding of rats for high (HAB) and low (LAB) anxiety-related behaviour: a unique model for comorbid depression and social dysfunctions. *Neurosci Biobehav Rev.* 2023;152:105292. doi 10.1016/j.neubiorev.2023.105292

Gunther K., Conner T.S., Armeli S., Tennen H., Covault J., Kranzler H.R. Serotonin transporter gene polymorphism (5-HTTLPR) and anxiety reactivity in daily life: a daily process approach to gene-environment interaction. *Psychosom Med.* 2007;69(8):762-768. doi 10.1097/PSY.0b013e318157ad42

Hackam D.G., Redelmeier D.A. Translation of research evidence from animals to humans. *JAMA.* 2006;296(14):1727-1732. doi 10.1001/jama.296.14.1731

Hettema J.M., Neale M.C., Kendler K.S. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am J Psychiatry.* 2001;158(10):1568-1578. doi 10.1176/appi.ajp.158.10.1568

Hettema J.M., Webb B.T., Guo A.-Y., Zhao Z., Maher B.S., Chen X., An S.-S., Sun C., Aggen S.H., Kendler K.S., Kuo P.-H., Otowa T.,

- Flint J., van den Oord E.J. Prioritization and association analysis of murine-derived candidate genes in anxiety-spectrum disorders. *Biol Psychiatry*. 2011;70(9):888-896. doi 10.1016/j.biopsych.2011.07.012
- Hirai A., Toda C., Yohannes Y.B., Collins N., Tambai M., Nomiya K., Eguchi A., Hoshi N., Hirano T., Nakayama S.M.M., Ishizuka M., Ikenaka Y. Role of brain monoamines in acetamiprid-induced anxiety-like behavior. *Toxicology*. 2024;505:153839. doi 10.1016/j.tox.2024.153839
- Hovatta I., Barlow C. Molecular genetics of anxiety in mice and men. *Ann Med*. 2008;40(2):92-109. doi 10.1080/07853890701747096
- Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl.1):34. doi 10.1186/s12859-018-2567-6
- Ivanisenko V.A., Gaisler E.V., Basov N.V., Rogachev A.D., Cherezis S.V., Ivanisenko T.V., Demenkov P.S., Mishchenko E.L., Khripko O.P., Khripko Y.I., Voevoda S.M., Karpenko T.N., Velichko A.J., Voevoda M.I., Kolchanov N.A., Pokrovsky A.G. Plasma metabolomics and gene regulatory networks analysis reveal the role of non-structural SARS-CoV-2 viral proteins in metabolic dysregulation in COVID-19 patients. *Sci Rep*. 2022;12(1):19977. doi 10.1038/s41598-022-24170-0
- Ivanisenko V.A., Rogachev A.D., Makarova A.-L.A., Basov N.V., Gaisler E.V., Kuzmicheva I.N., Demenkov P.S., Venzel A.S., Ivanisenko T.V., Antropova E.A., Kolchanov N.A., Plesko V.V., Moroz G.B., Lomivorotov V.V., Pokrovsky A.G. AI-assisted identification of primary and secondary metabolomic markers for postoperative delirium. *Int J Mol Sci*. 2024;25(21):11847. doi 10.3390/ijms252111847
- Ivanov R., Zamyatin V., Klimenko A., Matushkin Y., Savostyanov A., Lashin S. Reconstruction and analysis of gene networks of human neurotransmitter systems reveal genes with contentious manifestation for anxiety, depression, and intellectual disabilities. *Genes*. 2019;10(9):699. doi 10.3390/genes10090699
- Kolchanov N.A., Anan'ko E.A., Kolpakov F.A., Podkolodnaya O.A., Ignat'eva E.V., Goryachkovskaya T.N., Stepanenko I.L. Gene networks. *Mol Biol*. 2000;34(4):449-460. doi 10.1007/BF02759554
- Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov J Genet Breed*. 2013;17(4/2):833-850 (in Russian)
- Koskinen M.-K., Hovatta I. Genetic insights into the neurobiology of anxiety. *Trends Neurosci*. 2023;46(4):318-331. doi 10.1016/j.tins.2023.01.007
- Krämer A., Green J., Pollard J., Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):523-530. doi 10.1093/bioinformatics/btt703
- Krause C., Suwada K., Blomme E.A.G., Kowalkowski K., Liguori M.J., Mahalingaiah P.K., Mittelstadt S., Peterson R., Rendino L., Vo A., Van Vleet T.R. Preclinical species gene expression database: development and meta-analysis. *Front Genet*. 2023;13:1078050. doi 10.3389/fgene.2022.1078050
- Lesch K.P., Bengel D., Heils A., Sabol S.Z., Greenberg B.D., Petri S., Benjamin J., Muller C.R., Hamer D.H., 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
- Malezieux M., Klein A.S., Gogolla N. Neural circuits for emotion. *Annu Rev Neurosci*. 2023;46:211-231. doi 10.1146/annurev-neuro-111020-103314
- Meng P., Pan C., Qin X., Cai Q., Zhao Y., Wei W., Cheng S., Yang X., Cheng B., Liu L., He D., Shi S., Chu X., Zhang N., Jia Y., Wen Y., Liu H., Zhang F. A genome-wide gene-environmental interaction study identified novel loci for the relationship between ambient air pollution exposure and depression, anxiety. *Ecotoxicol Environ Saf*. 2024;285:117121. doi 10.1016/j.ecoenv.2024.117121
- Moraes A.C.N., Wijaya C., Freire R., Quagliato L.A., Nardi A.E., Kyriakoulis P. Neurochemical and genetic factors in panic disorder: a systematic review. *Transl Psychiatry*. 2024;14:294. doi 10.1038/s41398-024-02966-0
- Morris-Rosendahl D.J. Are there anxious genes? *Dialogues Clin Neurosci*. 2002;4(3):251-260. doi 10.31887/DCNS.2002.4.3/dmrosendahl
- Mostafavi S., Ray D., Warde-Farley D., Grouios C., Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*. 2008;9(Suppl.1):S4. doi 10.1186/gb-2008-9-s1-s4
- Mucha M., Skrzypiec A.E., Kolenchery J.B., Brambilla V., Patel S., Labrador-Ramos A., Kudla L., Murrall K., Skene N., Dymicka-Piekarska V., Klejman A., Przewlocki R., Mosienko V., Pawlak R. miR-483-5p offsets functional and behavioural effects of stress in male mice through synapse-targeted repression of *Pgap2* in the basolateral amygdala. *Nat Commun*. 2023;14:2134. doi 10.1038/s41467-023-37688-2
- Murphy D.L., Moya P.R., Fox M.A., Rubenstein L.M., Wendland J.R., Timpano K.R. Anxiety and affective disorder comorbidity related to serotonin and other neurotransmitter systems: obsessive-compulsive disorder as an example of overlapping clinical and genetic heterogeneity. *Phil Trans R Soc B*. 2013;368:20120435 doi 10.1098/rstb.2012.0435
- Nuss P. Anxiety disorders and GABA neurotransmission: a disturbance of modulation. *Neuropsychiatr Dis Treat*. 2015;17(11):165-175. doi 10.2147/NDT.S58841
- Otowa T., Hek K., Lee M., Byrne E.M., Mirza S.S., Nivard M.G., Bigdeli T., ... Maher B.S., van den Oord E.J., Wray N.R., Tiemeier H., Hettima J.M. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry*. 2016;21(10):1391-1399. doi 10.1038/mp.2015.197
- Penninx B.W., Pine D.S., Holmes E.A., Reif A. Anxiety disorders. *Lancet*. 2021;397(10277):914-927. doi 10.1016/S0140-6736(21)00359-7
- Petrican R., Fornito A., Boyland E. Lifestyle factors counteract the neurodevelopmental impact of genetic risk for accelerated brain aging in adolescence. *Biol Psychiatry*. 2024;95(5):453-464. doi 10.1016/j.biopsych.2023.06.023
- Purves K.L., Coleman J.R.I., Meier S.M., Rayner C., Davis K.A.S., Cheesman R., Bækvad-Hansen M., Børghlum A.D., Wan Cho S., Jürgen Deckert J., Gaspar H.A., Bybjerg-Grauholm J., Hettima J.M., Hotopf M., Hougaard D., Hübel C., Kan C., McIntosh A.M., Mors O., Bo Mortensen P., Nordentoft M., Werge T., Nicodemus K.K., Mattheisen M., Breen G., Eley T.C. A major role for common genetic variation in anxiety disorders. *Mol Psychiatry*. 2020;25(12):3292-3303. doi 10.1038/s41380-019-0559-1
- Schinka J.A., Busch R.M., Robichaux-Keene N. A meta-analysis of the association between the serotonin transporter gene polymorphism 5-HTTLPR and trait anxiety. *Mol Psychiatry*. 2004;9(2):197-202. doi 10.1038/sj.mp.4001405
- Sen S., Burmeister M., Ghosh D. Meta-analysis of the association between a serotonin transporter promoter polymorphism 5-HTTLPR and anxiety-related personality traits. *Am J Med Genet*. 2004;127B(1):85-89. doi 10.1002/ajmg.b.20158
- Shin D., Gong J.-R., Jeong S.D., Cho Y., Kim H.-P., Kim T.-Y., Cho K.-H. Attractor landscape analysis reveals a reversion switch in

- the transition of colorectal tumorigenesis. *Adv Sci.* 2024;2412503. doi 10.1002/advs.202412503
- Stein D.J., Bouwer C. A neuro-evolutionary approach to the anxiety disorders. *J Anxiety Disord.* 1997;11(4):409-429. doi 10.1016/s0887-6185(97)00019-4
- Strom N.I., Verhulst B., Bacanu S.A., Cheesman R., Purves K.L., Gedik H., Mitchell B.L., ... Million Veteran Program, FinnGen, 23andMe; Deckert J., Eley T.C., Mattheisen M., Hettema J.M. Genome-wide association study of major anxiety disorders in 122,341 European-ancestry cases identifies 58 loci and highlights GABAergic signaling. *medRxiv.* 2024. doi 10.1101/2024.07.03.24309466
- Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., Simonovic M., Roth A., Santos A., Tsafou K.P., Kuhn M., Bork P., Jensen L.J., von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(D1):D447-D452. doi 10.1093/nar/gku1003
- Vandamme T.F. Use of rodents as models of human diseases. *J Pharm Bioallied Sci.* 2014;6(1):2-9. doi 10.4103/0975-7406.124301
- Van Oekelen D., Luyten W.H., Leysen J.E. 5-HT_{2A} and 5-HT_{2C} receptors and their atypical regulation properties. *Life Sci.* 2003;72(22):2429-2449. doi 10.1016/s0024-3205(03)00141-3

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

Received November 9, 2024. Revised December 16, 2024. Accepted December 16, 2024.

Прием статей через электронную редакцию на сайте <http://vavilov.elpub.ru/index.php/jour>
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)»
до 2011 г. выходил под названием «Информационный вестник ВОГиС»/
“The Herald of Vavilov Society for Geneticists and Breeding Scientists”.

Сетевое издание «Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» – реестровая запись СМЭ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Russian Science Citation Index, Российский индекс научного цитирования, ВИНТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:
русскоязычная версия – на сайте <https://vavilovj-icg.ru/>
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440
англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, <https://www.ncbi.nlm.nih.gov/pmc/journals/3805/>

При перепечатке материалов ссылка обязательна.

✉ email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.

Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.

Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.

Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.

Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.

Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

.....
Дата публикации 00.02.2025. Формат 60 × 84 1/8. Уч.-изд. л. 00.0.
.....