

Научный рецензируемый журнал

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.

Периодичность 8 выпусков в год

DOI 10.18699/VJGB-23-24

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 – канд. 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 (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. Розозин – 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)

Scientific Peer Reviewed Journal

VAVILOV JOURNAL OF GENETICS AND BREEDING

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

*Founded in 1997**Published 8 times annually*

DOI 10.18699/VJGB-23-24

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, Cand. 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. Bespalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia

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

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

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

A. Börner, Dr. Sci., Germany

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

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

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

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),
RussiaD.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. Shavrukov, Dr. Sci. (Biology), Australia

R.I. Sheiko, Corr. Member of the NAS of Belarus,
Dr. Sci. (Agricul.), BelarusS.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology),
RussiaV.A. Stepanov, Full Member of the RAS, Dr. Sci. (Biology),
RussiaI.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),
RussiaN.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

Генетика и селекция растений

- 189 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
SWEET транспортеры *Medicago lupulina* в арбускулярно-микоризной системе в условиях среднего уровня доступного фосфора. А.А. Крюков, А.О. Горбунова, Т.Р. Кудряшова, О.Б. Иванченко, М.Ф. Шишова, А.П. Юрков
- 197 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Влияние генов *NAM-1* на содержание белка в зерне и показатели продуктивности у линий мягкой пшеницы с интрогрессиями чужеродного генетического материала в условиях Беларуси. О.А. Орловская, С.И. Вакула, К.К. Яцевич, Л.В. Хотылева, А.В. Кильчевский
- 207 **ОБЗОР**
Морфогенетические особенности репродуктивной биологии в селекции сахарной свеклы (*Beta vulgaris* L.). Т.П. Жужжалова, А.А. Налбандян, Е.Н. Васильченко, Н.Н. Черкасова

Генетика человека

- 218 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Оценка роли отбора в эволюции митохондриальных геномов коренного населения Сибири. Б.А. Малярчук, М.В. Деренко
- 224 **ОБЗОР**
Генетический контроль N-гликозилирования белков плазмы крови человека. С.Ж. Шарапов, А.Н. Тимошук, Ю.С. Аульченко

Генетика насекомых

- 240 **ОБЗОР**
Видовая идентификация паутиных клещей (Tetranychidae: Tetranychinae): обзор методов. А.В. Разуваева, Е.Г. Ульянова, Е.С. Сколотнева, И.В. Андреева
- 250 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Влияние LIM-киназы 1 на память и брачную песню самцов дрозофилы зависит от типа нейронов. А.В. Журавлев, Е.С. Заломая, Е.С. Егорова, А.Д. Емелин, В.В. Сокурова, Е.А. Никитина, Е.В. Савватеева-Попова (на англ. языке)

Генетика микроорганизмов

- 264 **ОБЗОР**
О природе пикобирнавирусов. А.Ю. Кашиников, Н.В. Епифанова, Н.А. Новикова
- 276 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Обнаружение генных кластеров биодеструкции алканов и ароматических соединений в геноме *Rhodococcus qingshengii* VKM Ac-2784D. Ю.А. Маркова, И.С. Петрушин, Л.А. Беловежец

Биоинформатика

- 283 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Компиляция и функциональная классификация генов, ассоциированных с длиной теломер, у человека и других видов животных. Е.В. Игнатьева, Н.С. Юдин, Д.М. Ларкин

Plant genetics and breeding

- 189 ORIGINAL ARTICLE
SWEET transporters of *Medicago lupulina* in the arbuscular-mycorrhizal system in the presence of medium level of available phosphorus. A.A. Kryukov, A.O. Gorbunova, T.R. Kudriashova, O.B. Ivanchenko, M.F. Shishova, A.P. Yurkov
- 197 ORIGINAL ARTICLE
Effect of *NAM-1* genes on the protein content in grain and productivity indices in common wheat lines with foreign genetic material introgressions in the conditions of Belarus. O.A. Orlovskaya, S.I. Vakula, K.K. Yatsевич, L.V. Khotyleva, A.V. Kilchevsky
- 207 REVIEW
Morphogenetic peculiarities of reproductive biology in sugar beet (*Beta vulgaris* L.) breeding. T.P. Zhuzhzhlova, A.A. Nalbandyan, E.N. Vasilchenko, N.N. Cherkasova

Human genetics

- 218 ORIGINAL ARTICLE
Evaluating the role of selection in the evolution of mitochondrial genomes of aboriginal peoples of Siberia. B.A. Malyarchuk, M.V. Derenko
- 224 REVIEW
Genetic control of N-glycosylation of human blood plasma proteins. S.Zh. Sharapov, A.N. Timoshchuk, Y.S. Aulchenko

Insect genetics

- 240 REVIEW
Species identification of spider mites (Tetranychidae: Tetranychinae): a review of methods. A.V. Razuvaeva, E.G. Ulyanova, E.S. Skolotneva, I.V. Andreeva
- 250 ORIGINAL ARTICLE
LIM-kinase 1 effects on memory abilities and male courtship song in *Drosophila* depend on the neuronal type. A.V. Zhuravlev, E.S. Zalomaeva, E.S. Egozova, A.D. Emelin, V.V. Sokurova, E.A. Nikitina, E.V. Savvateeva-Popova

Microbial genetics

- 264 REVIEW
On the nature of picobirnaviruses. A.Yu. Kashnikov, N.V. Epifanova, N.A. Novikova
- 276 ORIGINAL ARTICLE
Detection of gene clusters for biodegradation of alkanes and aromatic compounds in the *Rhodococcus qingshengii* VKM Ac-2784D genome. Yu.A. Markova, I.S. Petrushin, L.A. Belovezhets

Bioinformatics

- 283 ORIGINAL ARTICLE
Compilation and functional classification of telomere length-associated genes in humans and other animal species. E.V. Ignatieva, N.S. Yudin, D.M. Larkin

Original Russian text <https://vavilovj-icg.ru/>

SWEET transporters of *Medicago lupulina* in the arbuscular-mycorrhizal system in the presence of medium level of available phosphorus

A.A. Kryukov¹✉, A.O. Gorbunova¹, T.R. Kudriashova^{1,2}, O.B. Ivanchenko², M.F. Shishova³, A.P. Yurkov¹

¹ All-Russia Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia

² Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

³ Saint Petersburg State University, Biological Faculty, St. Petersburg, Russia

✉ rainniar@rambler.ru

Abstract. Arbuscular mycorrhiza (AM) fungi receive photosynthetic products and sugars from plants in exchange for contributing to the uptake of minerals, especially phosphorus, from the soil. The identification of genes controlling AM symbiotic efficiency may have practical application in the creation of highly productive plant-microbe systems. The aim of our work was to evaluate the expression levels of SWEET sugar transporter genes, the only family in which sugar transporters specific to AM symbiosis can be detected. We have selected a unique “host plant–AM fungus” model system with high response to mycorrhization under medium phosphorus level. This includes a plant line which is highly responsive to inoculation by AM fungi, an ecologically obligate mycotrophic line MIS-1 from black medick (*Medicago lupulina*) and the AM fungus *Rhizophagus irregularis* strain RCAM00320, which has a high efficiency in a number of plant species. Using the selected model system, differences in the expression levels of 11 genes encoding SWEET transporters in the roots of the host plant were evaluated during the development of or in the absence of symbiosis of *M. lupulina* with *R. irregularis* at various stages of the host plant development in the presence of medium level of phosphorus available for plant nutrition in the substrate. At most stages of host plant development, mycorrhizal plants had higher expression levels of *MISWEET1b*, *MISWEET3c*, *MISWEET12* and *MISWEET13* compared to AM-less controls. Also, increased expression relative to control during mycorrhization was observed for *MISWEET11* at 2nd and 3rd leaf development stages, for *MISWEET15c* at stemming (stooling) stage, for *MISWEET1a* at 2nd leaf development, stemming and lateral branching stages. The *MISWEET1b* gene can be confidently considered a good marker with specific expression for effective development of AM symbiosis between *M. lupulina* and *R. irregularis* in the presence of medium level of phosphorus available to plants in the substrate.

Key words: arbuscular mycorrhiza; *Medicago lupulina*; *Rhizophagus irregularis*; SWEET; gene expression assessment; sugar transporter genes.

For citation: Kryukov A.A., Gorbunova A.O., Kudriashova T.R., Ivanchenko O.B., Shishova M.F., Yurkov A.P. SWEET transporters of *Medicago lupulina* in the arbuscular-mycorrhizal system in the presence of medium level of available phosphorus. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2023;27(3):189-196. DOI 10.18699/VJGB-23-25

SWEET транспортеры *Medicago lupulina* в арбускулярно-микоризной системе в условиях среднего уровня доступного фосфора

А.А. Крюков¹✉, А.О. Горбунова¹, Т.Р. Кудряшова^{1,2}, О.Б. Иванченко², М.Ф. Шишова³, А.П. Юрков¹

¹ Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии, Пушкин, Санкт-Петербург, Россия

² Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

³ Санкт-Петербургский государственный университет, биологический факультет, Санкт-Петербург, Россия

✉ rainniar@rambler.ru

Аннотация. Грибы арбускулярной микоризы (AM) помогают растениям усваивать из почвы минеральные вещества, особенно фосфор, в то время как получают от растений продукты фотосинтеза – сахара. Выявление генов, контролирующих симбиотическую эффективность AM, может иметь практическое применение при создании высокопродуктивных растительно-микробных систем. Поэтому целью работы была оценка уровней экспрессии генов семейства транспортеров сахаров SWEET – единственного семейства, где могут быть обнаружены специфические для AM-симбиоза транспортеры сахаров. Нами разработана уникальная высокочувствительная к уровню фосфора и микоризации модельная система «растение-хозяин–AM-гриб», включающая высокоотзывчивую на инокуляцию AM-грибами экологически облигатно микотрофную линию MIS-1 люцерны хмелевидной

(*Medicago lupulina*) и штамм AM-гриба RCAM00320 *Rhizophagus irregularis*, обладающий высокой эффективностью на множестве видов растений. С использованием этой модельной системы в корнях растения оценены изменения в уровне экспрессии 11 генов, кодирующих транспортеры семейства SWEET, при развитии либо при отсутствии симбиоза *M. lupulina* с *R. irregularis* в различные фазы развития растения-хозяина в условиях среднего уровня доступного для питания растений фосфора в субстрате. У микоризованных растений обнаружены более высокие уровни экспрессии в большинстве фаз развития растения-хозяина для *MISWEET1b*, *MISWEET3c*, *MISWEET12*, *MISWEET13* по сравнению с контролем без AM. Повышенная относительно контроля экспрессия при микоризации наблюдалась также и для *MISWEET11* в фазу развития 2-го и 3-го листьев, для *MISWEET15c* – в фазу стеблевания, для *MISWEET1a* – в фазу 2-го листа, стеблевания и бокового ветвления. Ген *MISWEET1b* с уверенностью можно считать хорошим маркером со специфической экспрессией для эффективного развития AM-симбиоза *M. lupulina* с *R. irregularis* в условиях среднего уровня доступного для растений фосфора в субстрате.
Ключевые слова: арбускулярная микориза; *Medicago lupulina*; *Rhizophagus irregularis*; SWEET; оценка экспрессии генов; гены транспортеров сахаров.

Introduction

Plant sugar transporters belong to three key families: Sucrose Transporters (SUT = SUC), Monosaccharide Transporters (MST, including subfamilies STP, TMT, PMT, VGT, pGlcT/SGB1, ESL, INT) and Sugars Will Eventually be Exported Transporters (SWEET). SUT transporters are involved in loading the phloem with sucrose and its long-distance transport from the leaves to other parts of the plant. These sugars are broken down into monosaccharides and transported to the cells by MST proteins.

The least studied of these groups is the SWEET family of transporters. These are non-volatile and bidirectional transmembrane transporters of various sugars in all plant organs and tissues. L.Q. Chen's was first to account for this family of transporters in 2010. It is in the SWEET family that the proteins particular to AM-symbiosis could be identified, while ones in the other two families had not yet been found (Chen et al., 2010; Doidy et al., 2019). Currently, it is believed that SWEET transporters are found in all living organisms (Feng et al., 2015). At the same time, it is noted that the number of isoforms of these transporters differs even in closely related species. The numbering of new SWEET proteins and their isoforms in other organisms is executed according to gene orthology principals as with the proteins in *Arabidopsis thaliana*.

As a result of early studies of transporters, it turned out that the *SWEET* plant genes, despite their low homology, are grouped into four clades (Chen et al., 2015). Representatives of each of the clades are observed in almost all terrestrial plants. It is believed that the representatives of the four clades are phylogenetically and functionally distinct. Thus, it is noted that representatives of clades I and II transport hexoses, clade III, mainly sucrose, and clade IV, mainly fructose (Chen et al., 2012; Feng et al., 2015).

SWEET proteins are involved in a variety of processes. In addition to the transport of sugars, they apparently participate in the transport of other substances, for example, gibberellins, as was shown for *Arabidopsis* (Kanno et al., 2016). A great deal of data in the literature deals with the functions of SWEET proteins in different plant species. For example, the MtSWEET1b transporter may supply glucose to AM fungi (An et al., 2019), LjSWEET3 mediates sucrose transport (Sugiyama et al., 2017) to nodules of *Lotus japonicus*. SWEET clade I transporters are probably involved in the supply of sugars to symbiotic systems (Doidy et al., 2019). Rhizosphere pathogens can cause increased synthesis of clade III proteins. This, in turn, leads

to additional sucrose transport to roots and contributes to the nutrition of microorganisms in the rhizosphere (Doidy et al., 2019). In 2010, L.Q. Chen et al. demonstrated that pathogenic bacteria, for example, of the genus *Xanthomonas*, can enter tissues of the host plant and induce expression of the *SWEET* genes. These encode transporters from clade III (primarily *SWEET11* and *SWEET14*) in order to produce sugars. Like symbiotic AM fungi, pathogenic fungi induce gene expression to produce sugars (Chen et al., 2010).

J. Manck-Gotzenberger and N. Requena (2016) note that the genes of many transporters have a significant level of expression in AM symbiosis, but are not necessarily particular to it. The work of A. Kafle shows that the orthologs of *SWEET1* – *MtSWEET1.2* and *PsSWEET1.2* – can be expressed both in mycorrhizal roots and in root nodules (Kafle et al., 2019). Therefore, the orthologs of the transporters of clades I (MtSWEET1-MtSWEET3) and III (MtSWEET9-MtSWEET15) are primarily considered as active participants in the symbiotic relationship “plant-AM fungus” (Kryukov et al., 2021).

The study of the role of SWEET transporters in the formation of symbiotic relationships has never been specifically directed toward PMS. In this regard, the aim of this work was to evaluate the expression of the *SWEET* genes in PMS models during mycorrhization and its respective absence at different stages of plant development.

Materials and methods

Plant and fungal material. Black medick (*Medicago lupulina* L. subsp. *vulgaris* Koch) is a widespread species of the genus *Medicago*, a self-pollinating diploid. In the present study, the authors selected the MIS-1 line as being highly responsive to mycorrhization from the black medick cultivar-population VIK32 (Yurkov et al., 2015). Barring inoculation by AM-fungus, and with low level of available inorganic phosphorus (P_i) in the soil, this line exhibits signs of dwarfism (Yurkov et al., 2015, 2020). An effective strain RCAM00320 of the AM fungus *Rhizophagus irregularis* (formerly known as *Glomus intraradices* Shenck & Smith) was used for inoculation (CIAM8 from the All-Russia Research Institute for Agricultural Microbiology (ARRIAM) collection). An accurate identification of the strain was carried out by the authors (Kryukov, Yurkov, 2018). Since AM-fungi are obligate symbionts, the strain is maintained in a cumulative culture of *Plectranthus* sp. (exact species identification is currently being

undertaken by the authors) under standard conditions in the ARRIAM Laboratory of ecology of symbiotic and associative microorganisms (Yurkov et al., 2010).

Vegetative method. The procedure for the method is described in the work of A.P. Yurkov et al. (Yurkov et al., 2015). Optimal conditions were provided for the development of AM while preventing spontaneous infection with nodule bacteria and other microorganisms. A mixture of soil and sand in a ratio of 2:1 was autoclaved twice at 134 °C, 2 atm for 1 hour with repeated autoclaving two days later; no signs of toxicity appeared after such treatment. Specimens were planted with two seedlings per one vessel filled with a soil-sand mixture (210 g). Agrochemical characteristics of the soil are given by (Yurkov et al., 2015). The content of P₂O₅ in the soil was 23 mg/kg of soil (as according to Kirsanov). Before the experiment, 0.5 doses of phosphorus were added in the form of CaH₂PO₄*2H₂O (86 mg/kg of soil) according to the prescription of D.N. Pryanishnikov (Klechkovsky, Petersburgsky, 1967). The final phosphorus content in the soil-sand mixture was 109 mg/kg and corresponded to the average P_i level; i. e. the availability of mobile phosphates in the soil in terms of the content in the Kirsanov extract according to (Sokolov, 1975); pH_{KCl} – 6.44. The first measurement of plants was carried out 21 days after sowing and inoculation, followed by measurement at key stages of black medick ontogenesis. There was a total number of 7 measurements (Supplementary Material 1)¹.

The specimens involved *Plectranthus* roots inoculated and uninoculated with *R. irregularis* strain RCAM00320. During collection, the material was frozen in liquid nitrogen and stored for up to 6 months at –80 °C.

RNA isolation and evaluation of gene expression. The selection of genes of interest was carried out based on the results of *M. truncatula* transcriptome analysis (*MtSWEET1a* = Medtr1g029380, *MtSWEET1b* = Medtr3g089125, *MtSWEET2a* = Medtr8g042490, *MtSWEET2b* = Medtr2g073190, *MtSWEET2c* = Medtr6g034600, *MtSWEET3a* = Medtr3g090940, *MtSWEET3b* = Medtr3g090950, *MtSWEET3c* = Medtr1g028460, *MtSWEET4* = Medtr4g106990, *MtSWEET5a* = Medtr6g007610, *MtSWEET5b* = Medtr6g007637, *MtSWEET5c* = Medtr6g007623, *MtSWEET5d* = Medtr6g007633, *MtSWEET6* = Medtr3g080990, *MtSWEET7* = Medtr8g099730, *MtSWEET9a* = Medtr5g092600, *MtSWEET9b* = Medtr7g007490, *MtSWEET11* = Medtr3g098930, *MtSWEET12* = Medtr8g096320, *MtSWEET13* = Medtr3g098910, *MtSWEET14* = Medtr8g096310, *MtSWEET15a* = Medtr2g007890, *MtSWEET15b* = Medtr5g067530, *MtSWEET15c* = Medtr7g405730, *MtSWEET15d* = Medtr7g405710, *MtSWEET16* = Medtr2g436310; sequence numbers from the database: <https://phytozome.jgi.doe.gov/pz/portal.html>) with subsequent selection of primer sequences for the genes of interest.

Three pairs of primers were tested for each gene. The absence of the second product was estimated based on electrophoresis and melting curves. The effectiveness of primers was calculated on the basis of real-time PCR (quantitative poly-

merase chain reaction in real time) serial dilutions of the cDNA matrix. Only primers with efficiency equal to or close to 100 % were used. The primer test was carried out for several measurement periods (Supplementary Material 2).

In 2022, the conformity and quality of primers were verified using *M. lupulina* MIS-1 transcriptomic data (MACE sequencing). Total RNA from plant material was isolated using the trizole method with modifications (MacRae, 2007). The quality of DNAase treatment was tested by PCR for RNA with the reference gene, actin tested immediately before cDNA synthesis. cDNA synthesis was carried out using the Maxima First Strand cDNA Synthesis Kit with dsDNase in accordance with the manufacturer's instructions (Thermo Scientific, USA). ~1 mcg of total RNA was selected for cDNA synthesis. cDNA quality was tested with a ubiquitin test.

Changes in gene expression were evaluated using the RT-PCR method employing the BioRad CFX-96 real-time thermal cycler (Bio-Rad, USA) and using a set of reagents for RT-PCR in the presence of the SYBR Green I dye. The parameters of the amplification cycles were as follows: 95 °C, 5 min, 1 cycle; 95 °C, 15 s, 60 °C, 30 s, 72 °C, 30 s, 40 cycles. The specificity of amplification was evaluated using melting curve analysis. Changes in the expression level of the gene under examination were compared with the expression level of the same gene in the control. Analysis was carried out using the 2^{-ΔΔCT} method. The levels of gene expression were normalized with the selected reference gene, actin according to (Yurkov et al., 2020). The PCR mix (10 ml) contained: 1 ml of 10x B + SYBR Green buffer, 1 ml of 2.5 mM dNTP, 1 ml of MgCl₂ (25 mM), 0.3 ml of each of a pair of primers (10 mM for each primer), 0.125 ml (0.625 units) SynTaq DNA polymerase (manufacturer of mix components – Synthol, Russia), 4.275 μl ddH₂O, 2 μl cDNA sample. The relative values of the cDNA gene expression level for each sample were evaluated (experiment with AM, control without AM). The biological repeatability is 3, the technical repeatability is 4 measurements.

Evaluation of parameters of symbiotic efficiency and activity. J.M. Philips and D.S. Hayman's trypan blue staining method was used for root samples (Phillips, Hayman, 1970). The parameters of AM fungus activity in the root, the mycorrhization indices, are calculated according to (Vorob'ev et al., 2016) as: *a* and *b* (abundance of arbuscules and vesicles in mycorrhized parts of the roots, respectively), and *M* (intensity of AM development in the root). They have the following calculation formulas:

$$M = \frac{(95n_5 + 70n_4 + 30n_3 + 5n_2 + 1n_1)}{N} \%, \quad (1)$$

where *n*₅ is the number of visual fields with a mycorrhiza density class – *M* = 5; *n*₄ – with *M* = 4; *n*₃ – with *M* = 3, etc.; *M* is estimated from 1 to 5 points: 1 evaluation score: 0–1 % mycorrhiza at the root in the field of view of the microscope; 2 evaluation scores: 2–10 %; 3 scores: 11–50 %; 4 scores: 51–90 %; 5 scores: 91–100 % mycorrhiza in the root.

$$a = \frac{(100mA_3 + 50MA_2 + 10MA_1)}{100} \%, \quad (2)$$

where *m*_{*A*} =

$$= \frac{(95n_5mA_i + 70n_4mA_i + 30n_3mA_i + 5n_2mA_i + 1n_1mA_i)F}{M(N - n_0)} \%, \quad (3)$$

¹ Supplementary Materials 1 and 2 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Kryukov_Engl_27_3.pdf.

where $n_i m A_j$ is the number of visual fields with $M = i$, $A = j$, F is the incidence of mycorrhizal infection (the proportion of visual fields with AM relative to the total number of visual fields in one root sample), N is the total number of viewed visual fields, n_0 is the number of visual fields without AM, n_i is the number of visual fields with a mycorrhiza density class from 1 to 5, and A_i is the arbuscule density class from 1 to 3.

$$b = \frac{(100mB_3 + 50mB_2 + 10mB_1)}{100} \% \quad (4)$$

where $n_i m B_j$ is the number of visual fields with $M = i$ and B (vesicle density class) = j is calculated similarly to the calculation for arbuscules (3); B_j is the density class of arbuscules from 1 to 3.

The symbiotic efficiency of AM was estimated as the difference in productivity index (crude weight of aboveground parts) between the variant with AM inoculation (“+AM”) and the control without AM (“without AM”), divided by the value in the variant “without AM” as a standard calculation MGR (mycorrhizal growth response) (Kaur et al., 2022). Biological repeatability in assessing the parameters of the effectiveness and activity of AM in each variant was equal to 8 plants.

Statistical analysis. ANOVA and Tukey’s HSD test ($p < 0.05$) were used as a post-hoc test to compare the differences in all indicators; Student’s t -test ($p < 0.05$) was also used to assess the significance of differences in the average values of gene expression levels between the “+AM” and “without AM” variants.

Results

The results of the evaluation of the symbiotic efficacy and parameters of mycorrhization showed that the studied symbiotic test system *Medicago lupulina* + *Rhizophagus irregularis* should be considered highly effective (with high MGR) and symbiotically active. Symbiotically active refers to the presence of active mycorrhization of roots by mycelium, arbuscules and vesicles under conditions of an average level of phosphorus available to plants in the substrate.

Analysis of *M. lupulina* mycorrhization by AM fungus *R. irregularis* showed that the intensity of mycorrhization (M; Fig. 1, b) and the abundance of vesicles (b, see Fig. 1, d) significantly decreased at the initiation of lateral branching stage (48 days). However the abundance of arbuscules (a, see Fig. 1, c), the principal symbiotic structures of AM, were maintained at a high level along with the symbiotic efficiency (MGR) calculated for the fresh weight of the aboveground parts (see Fig. 1, a).

The obtained data on microscopy and MGR evaluation indicate that highly effective and active PMS with an early and prolonged response can be used as a genetic model for the search and analysis of marker genes for the development of effective AM symbiosis. It can be accessed from the early stage (2nd leaf stage) to the late stage of the fruiting initiation in conditions of an average P_i level in the substrate. To this end, the expression of 11 genes of the SWEET family was evaluated. The relative level of transcripts (normalized values of $2^{-\Delta\Delta Ct}$) in the roots of *M. lupulina* with normalization to control without AM was evaluated (Fig. 2).

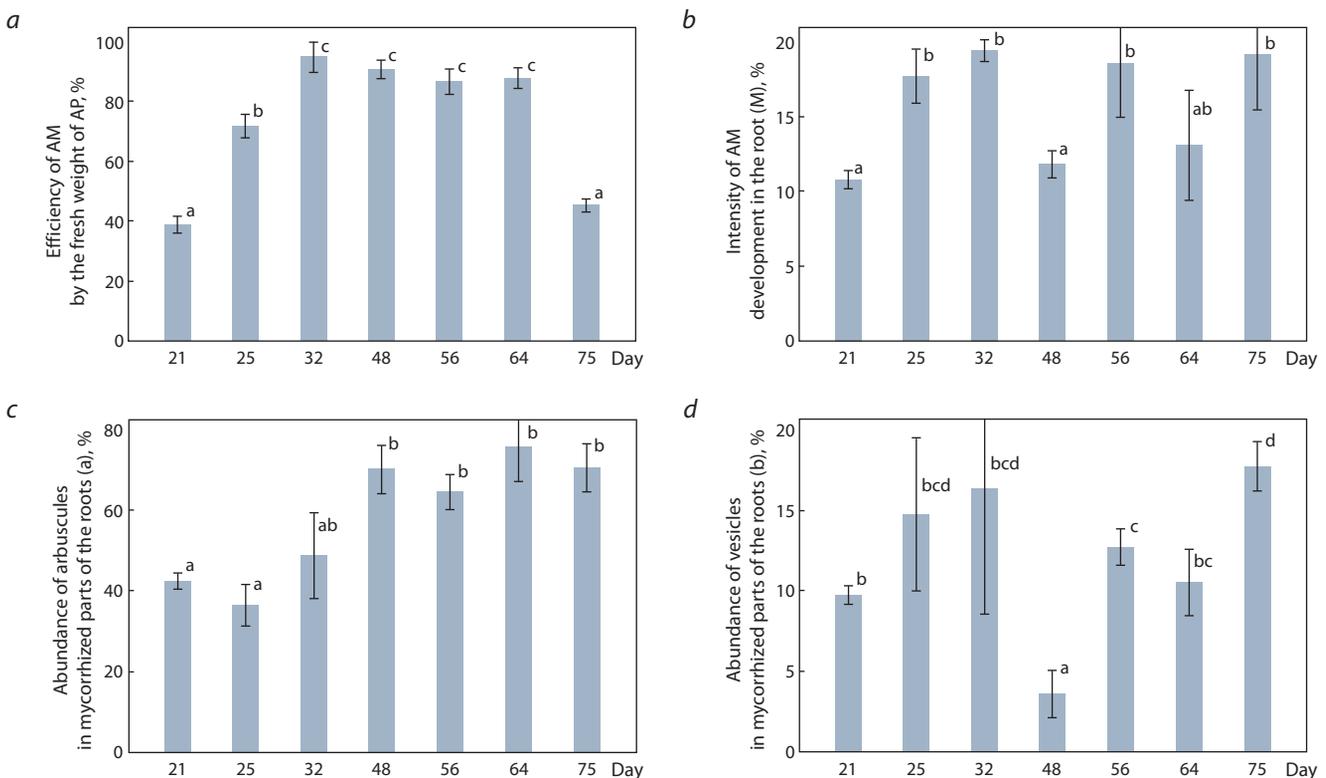


Fig. 1. The symbiotic efficiency of AM calculated from the fresh weight of the aboveground parts (a), the intensity of mycorrhization M (b), the abundance of arbuscules (c) and the abundance of vesicles (d) formed by *R. irregularis* in the roots of *M. lupulina*.

The average values with the error of the average are presented; the LF is the aboveground parts; “day” is the day after sowing and inoculation. Values with significant ($p < 0.05$) differences are marked with different letters (a, b, c, d).

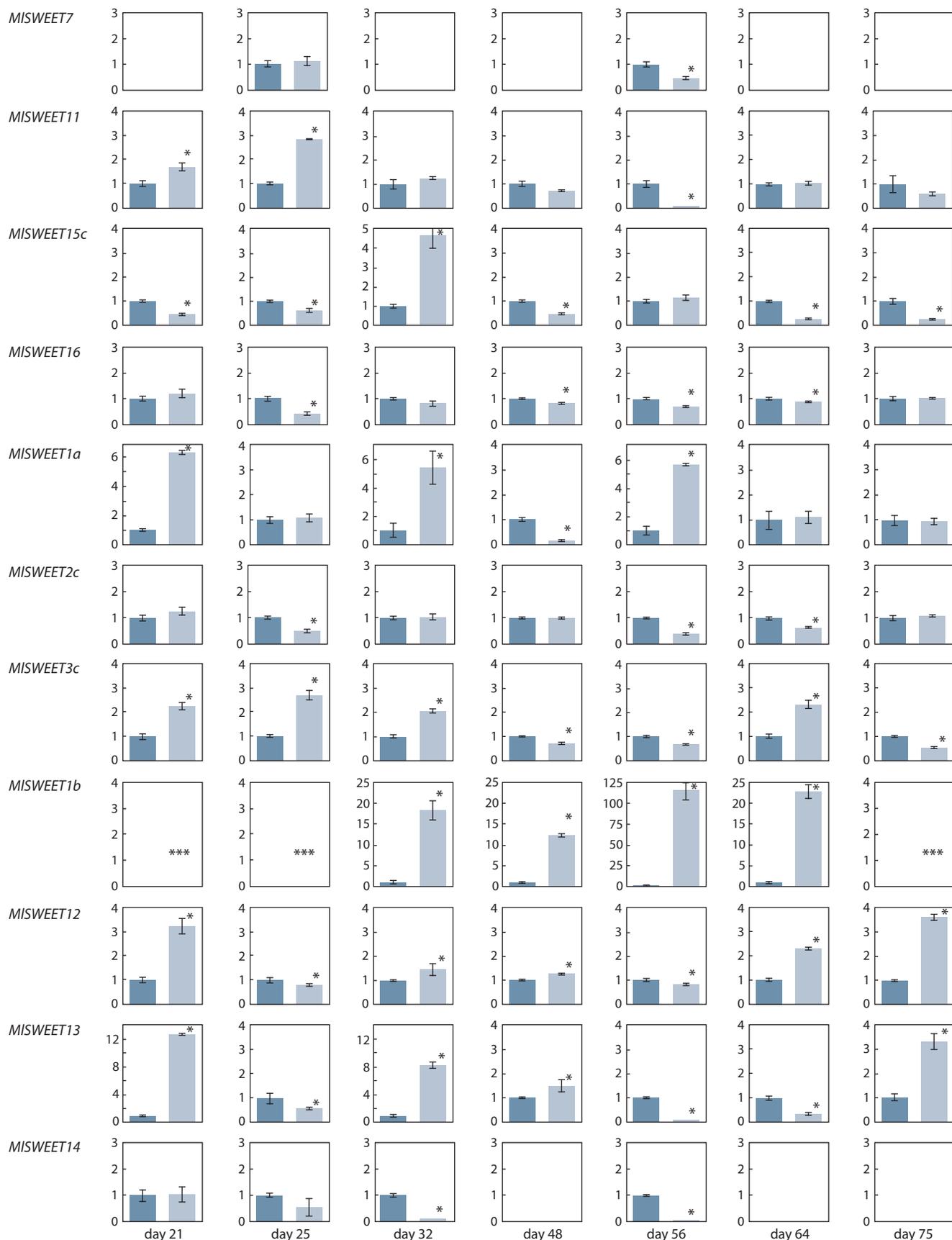


Fig. 2. Relative transcript level (normalized value of $2^{-\Delta\Delta Ct}$) of the SWEET family genes in the roots of *M. lupulina*.

The average values with the error of the average are presented. *The presence of significant ($p < 0.05$) differences between the variant “without” (dark bars) and the variant “+AM” with *R. irregularis* inoculation (light column); empty diagrams – the absence of gene expression in the variants; “day” – day after sowing and inoculation. *** Specific gene expression in the variant “with AM”.

It should be noted that some of the genes did not have a pronounced difference in variants with and without mycorrhiza. Nevertheless, results showed the presence of genes that had significantly higher expression at AM for most stages of host plant development: *MISWEET3c*, *MISWEET1b*, *MISWEET12*, *MISWEET13*. Increased expression relative to the control during mycorrhization was also observed for *MISWEET11* at the 1st and 2nd stage of analysis (in the development stage of the 2nd and 3rd leaf), for *MISWEET15c* at the 3rd stage of stemming. Significantly higher expression during mycorrhization was shown by the genes *MISWEET1a* at the 1st, 3rd and 5th stage (in the stage of the 2nd leaf, stemming and lateral branching, respectively) and *MISWEET1b* at all 7 stages of measurement. Furthermore, its expression at the 1st, 2nd and 7th stages (in the stage of the 2nd and 3rd leaves and the fruiting initiation stage, respectively) was also dependent upon mycorrhization.

Discussion

According to the results obtained, in the roots of mycorrhized *M. lupulina*, a higher expression of SWEET family genes was observed (see Fig. 2) in comparison to plants without AM: in six genes at the stage of the second leaf (21 days); in three genes at the stage of the stemming initiation (25 days); in six genes at the stage of stemming (32 days); in three genes at the stage of the lateral branching initiation (48 days); in two genes at the stage of lateral branching (56 days); in three genes at the stage of the flowering initiation (64 days); in three genes at the stage of the fruiting initiation (75 days). Reduced expression in the roots of plants with AM was characterized (see Fig. 2) by: one gene at day 21; five genes at day 25; one gene at day 32; four genes at day 48; eight genes at day 56; four genes at day 64; two genes at day 75.

Based on this, it can be assumed that during the stemming stage (32 days), the formation and maintenance of AM required enhanced transport of sugars. This led to increased expression of the largest number (6) of the SWEET family of genes in AM, and to reduced expression of only one gene during mycorrhization. On the other hand, the stage at the lateral branching initiation (48 days) was characterized by a sharp decrease in expression. Increased expression was observed in only 3 genes with AM and the number of genes with reduced expression increased to 4. The lateral branching stage of stem (56 days) was accompanied by even weaker gene expression. It should be noted that the observed relationship with the development stage of the host plant also took place in our previous studies (Yurkov et al., 2021), in which it was shown that the role of the development stage in the formation of the plant metabolome was higher than the effect of the actual inoculation by an effective AM fungus in conditions of low P_i level in the substrate. This explains why the analysis of gene expression was carried out taking into account the timing of plants entering new stages of development.

This study shows that the symbiotic pair “MIS-1 line of *M. lupulina* + RCAM00320 strain of *R. irregularis*” retains high efficiency in conditions of medium P_i (see Fig. 1, a), although this efficiency is still lower than in conditions of low P_i (Yurkov et al., 2020, 2021). Furthermore, the model PMS maintains high mycorrhization parameters throughout

the entire development from the 2nd leaf stage to the late stage of the fruiting initiation, apart from a decrease in the parameters of the intensity of mycorrhization M and the abundance of vesicles b at the lateral branching initiation (see Fig. 1, b, d) (48 days). It is noteworthy that, in comparison with the previous period, only at this stage was a lower relative expression observed in 8 out of the 11 studied genes in AM infected roots: *MISWEET1a*, *MISWEET1b*, *MISWEET3c*, *MISWEET11*, *MISWEET12*, *MISWEET13*, *MISWEET15c*, *MISWEET16* (see Fig. 2).

Studies with low P_i level in the substrate also showed significant changes at the initiation of lateral branching, but of a different nature. At this point the development of AM-symbiosis was observed with the greatest efficiency and with significant metabolic rearrangements (Yurkov et al., 2021).

Among the genes of the SWEET family examined, as a rule, only the genes *MISWEET1b*, *MISWEET3c*, *MISWEET12*, *MISWEET13* were characterized by higher expression during mycorrhization under conditions of medium P_i in the substrate (see Fig. 2).

The literature still offers insufficient and contradictory data about the localization and function of proteins of the SWEET family, but it can be assumed that some of these aspects will be identical for homologues between different plant species. Thus, it is known that VfSWEET1 in *Vernicia fordii* is found in leaves (Cao et al., 2019), SISWEET1a in *Solanum lycopersicum* in young leaves and flowers (Ho et al., 2019), TaSWEET1b1-1B in *Triticum aestivum* in the stem, StSWEET1h and StSWEET1i in leaves, stems, and root, BrSWEET1b in *Brassica rapa* in the root (Li et al., 2018); MtSWEET1b in *M. truncatula* in seeds and nodules (Hu et al., 2019). It should be noted that the *SWEET1b* gene is fixed in studies mainly in aboveground parts, when, as in our study, its specificity in AM is demonstrated in roots under growth conditions with a medium level of P_i in the substrate.

The products of the orthologue of the *MISWEET3c* gene are localized in flowers, leaf petioles, beans, stems, seeds, roots and nodules in *M. truncatula* and transfer glucose (Hu et al., 2019); but, for example, in *Lotus japonicus* they transfer sucrose, and not glucose. Other sources indicate 2-deoxyglucose (Sugiyama et al., 2017). Thus, not only localization but also the functions of the products of the SWEET family proteins are still the subject of discussion.

The products of the orthologs of the *MISWEET12* and *MISWEET13* genes carry sucrose in *Solanum lycopersicum* and *Manihot esculenta* (Cohn et al., 2014; Zhao et al., 2018). In *M. truncatula*, MtSWEET13 is found in flowers, leaves, leaf petioles, stems and seeds, MtSWEET12 is found in all organs (Hu et al., 2019). Thus, we observe that *MISWEET12* and *MISWEET13* are polyfunctional. The functions of *MISWEET12* ortholog products are sucrose outflow from leaves for phloem loading (Chen et al., 2012, 2015) and resistance to pathogens (Cohn et al., 2014; Zhao et al., 2018; Gautam et al., 2019). And the functions of *MISWEET13* ortholog products are pollen development (Sun et al., 2013), resistance to pathogens (Cohn et al., 2014; Zhao et al., 2018), possible deposition of primexin, phloem loading (Chen et al., 2012; Gautam et al., 2019), and possible transport of gibberellin (Kanno et al., 2016; Jeena et al., 2019).

In our work, *MISWEET1b* showed high expression during mycorrhization at all measurement periods (see Fig. 2). This gene can be confidently considered a good marker of the effective development of AM-symbiosis of *M. lupulina* with *R. irregularis* under conditions of an average P_i level in the substrate. It should be noted that *MISWEET1b* had a specific expression in the roots at AM on the 1st, 2nd and 7th measurements (at the stages of development of the 2nd, 3rd leaf, and fruiting) (see Fig. 2). Perhaps in the initial stages (1 and 2), this gene was specifically expressed for the formation of an effective AM symbiosis (for the redistribution of sugars in favor of the mycosymbiont). Then, at a later stage, it was expressed because of a drop in symbiotic efficiency during the transition to fruiting when the host plant again needed to maintain an effective AM symbiosis to ensure seed maturation.

Thus these observations revealed a relationship between the stage of development of the host plant *M. lupulina* and the stage of development of AM symbiosis with *R. irregularis* with the expression of a particular *MISWEET1b* gene in AM infected plants. Our data are consistent with the results of studies on other plant species. Thus, a recent work from 2016 has also mentioned orthologs of the *MISWEET1b* gene as being symbiotic with AM in *S. tuberosum* and *M. truncatula* (Manck-Götzenberger, Requena, 2016; An et al., 2019; Doidy et al., 2019; Kafle et al., 2019; Cope et al., 2022). For genes of *MISWEET3c* orthologs were found in *Lotus japonicus* and *M. truncatula* (Sugiyama et al., 2017; Cope et al., 2022), orthologs of *MISWEET12*, in *S. tuberosum* and *M. truncatula* (Manck-Götzenberger, Requena, 2016; Hennion et al., 2019; Cope et al., 2022).

Based on the results of this work and analysis of the literature, it can be assumed that under different biotic and abiotic conditions (for example, with different chemical composition of the substrate, moisture, light, etc.), the expression and function of the *SWEET* genes may differ significantly.

Conclusion

The expression of the genes of the *SWEET* family during the transition of plants from the initiation of one stage of development to another has been practically ignored in the literature. This study has managed to eliminate this drawback. For the first time the analysis of their expression in the roots of an *M. lupulina* line highly responsive to mycorrhization was performed under conditions of average P_i level in the substrate. Results showed that the expression of the *MISWEET1b* gene specifically increased with a decrease in symbiotic efficiency calculated by the weight of fresh aboveground parts. It is likely that the high expression in AM plants at early stages of development is associated with the active redistribution of sugars during the formation of effective AM. At the fruiting stage, on the other hand, it is a response to the needs of sugars for seed maturation. More than half of the studied genes also showed increased expression, among which genes such as *MISWEET3c* and *MISWEET12* should be singled out. The results obtained are consistent with the literature in that AM-specific genes of the *SWEET* family can be found among the genes of clades I and III.

Given the diversity of orthologs in other plant species, there is reason to believe that not all the genes of the *SWEET*

family have yet been identified, both in the plant we have examined, *M. lupulina*, and in other species of the *Medicago* genus. Research over the coming years will surely expand our conception of what functions *SWEET* transporters perform in *Medicago* plants.

References

- An J., Zeng T., Ji C., Graaf S., Zheng Z., Xiao T., Deng X., Xiao S., Bisseling T., Limpens E., Pan Z. A *Medicago truncatula* SWEET transporter implicated in arbuscule maintenance during arbuscular mycorrhizal symbiosis. *New Phytol.* 2019;224(1):396-408. DOI 10.1111/nph.15975.
- Cao Y., Liu W., Zhao Q., Long H., Li Z., Liu M., Zhou X., Zhang L. Integrative analysis reveals evolutionary patterns and potential functions of SWEET transporters in Euphorbiaceae. *Int. J. Biol. Macromol.* 2019;139:1-11. DOI 10.1016/j.ijbiomac.2019.07.102.
- Chen L.Q., Cheung L.S., Feng L., Tanner W., Frommer W.B. Transport of sugars. *Annu. Rev. Biochem.* 2015;84(1):865-894. DOI 10.1146/annurev-biochem-060614-033904.
- Chen L.Q., Hou B.H., Lalonde S., Takana H., Hartung M.L., Qu X.Q., Guo W.J., Kim J.G., Underwood W., Chaudhuri B., Chermak D., Antony G., White F.F., Somerville S.C., Mudgett M.B., Frommer W.B. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature.* 2010;468(7323):527-532. DOI 10.1038/nature09606.
- Chen L.Q., Qu X.Q., Hou B.H., Sosso D., Osorio S., Fernie A.R., Frommer W.B. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science.* 2012;335(6065):207-211. DOI 10.1126/science.1213351.
- Cohn M., Bart R.S., Shybut M., Dahlbeck D., Gomez M., Morbitzer R., Hou B.H., Frommer W.B., Lahaye T., Staskawicz B.J. *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Mol. Plant Microbe Interact.* 2014;27(11):1186-1198. DOI 10.1094/MPMI-06-14-0161-R.
- Cope K., Kafle A., Yakha J., Pfeffer P., Strahan G., Garcia K., Subramanian S., Bücking H. Physiological and transcriptomic response of *Medicago truncatula* to colonization by high- or low-benefit arbuscular mycorrhizal fungi. *Mycorrhiza.* 2022;32(3-4):281-303. DOI 10.1007/s00572-022-01077-2.
- Doidy J., Vidal U., Lemoine R. Sugar transporters in Fabaceae, featuring SUT MST and SWEET families of the model plant *Medicago truncatula* and the agricultural crop *Pisum sativum*. *PLoS One.* 2019;14(9):e0223173. DOI 10.1371/journal.pone.0223173.
- Feng C.Y., Han J.X., Han X.X., Jiang J. Genome-wide identification, phylogeny, and expression analysis of the SWEET gene family in tomato. *Gene.* 2015;573(2):261-272. DOI 10.1016/j.gene.2015.07.055.
- Gautam T., Saripalli G., Gahlaut V., Kumar A., Sharma P.K., Balyan H.S., Gupta P.K. Further studies on sugar transporter (SWEET) genes in wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.* 2019;46:2327-2353. DOI 10.1007/s11033-019-04691-0.
- Hennion N., Durand M., Vriet C., Doidy J., Maurousset L., Lemoine R., Portau N. Sugars en route to the roots. Transport, metabolism and storage within plant roots and towards microorganisms of the rhizosphere. *Physiol. Plant.* 2019;165(1):44-57. DOI 10.1111/pp1.12751.
- Ho L.H., Klemens P.A.W., Neuhaus H.E., Ko H.Y., Hsieh S.Y., Guo W.J. SISWEET1a is involved in glucose import to young leaves in tomato plants. *J. Exp. Bot.* 2019;70(12):3241-3254. DOI 10.1093/jxb/erz154.
- Hu B., Wu H., Huang W., Song J., Zhou Y., Lin Y. SWEET gene family in *Medicago truncatula*: genome-wide identification, expression and substrate specificity analysis. *Plants.* 2019;8(9):338. DOI 10.3390/plants8090338.
- Jeena G.S., Kumar S., Shukla R.K. Structure, evolution and diverse physiological roles of SWEET sugar transporters in plants. *Plant Mol. Biol.* 2019;100(4-5):351-365. DOI 10.1007/s11103-019-00872-4.

- Kaffe A., Garcia K., Wang X., Pfeiffer P.E., Strahan G.D., Bücking H. Nutrient demand and fungal access to resources control the carbon allocation to the symbiotic partners in tripartite interactions in *Medicago truncatula*. *Plant Cell Environ.* 2019;42(1):270-284. DOI 10.1111/pce.13359.
- Kanno Y., Oikawa T., Chiba Y., Ishimaru Y., Shimizu T., Sano N., Koshiba T., Kamiya Y., Ueda M., Seo M. AtSWEET13 and AtSWEET14 regulate gibberellin-mediated physiological processes. *Nat. Commun.* 2016;7:13245. DOI 10.1038/ncomms13245.
- Kaur S., Campbell B.J., Suseela V. Root metabolome of plant-arbuscular mycorrhizal symbiosis mirrors the mutualistic or parasitic mycorrhizal phenotype. *New Phytol.* 2022;234(2):672-687. DOI 10.1111/nph.17994.
- Klechkovskiy V.M., Petersburgskiy A.V. Agrochemistry. Moscow: Kolos Publ., 1967. (in Russian)
- Kryukov A.A., Gorbunova A.O., Kudriashova T.R., Yakhin O.I., Lubyaynov A.A., Malikov U.M., Shishova M.F., Kozhemyakov A.P., Yurkov A.P. Sugar transporters of the SWEET family and their role in arbuscular mycorrhiza. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2021;25(7):754-760. DOI 10.18699/VJ21.086. (in Russian)
- Kryukov A.A., Yurkov A.P. Optimization procedures for molecular-genetic identification of arbuscular mycorrhizal fungi in symbiotic phase on the example of two closely kindred strains. *Mikologiya i Fitopatologiya = Mycology and Phytopathology.* 2018;52(1):38-48. (in Russian)
- Li X., Si W., Qin Q., Wu H., Jiang H. Deciphering evolutionary dynamics of SWEET genes in diverse plant lineages. *Sci. Rep.* 2018; 8(1):13440. DOI 10.1038/s41598-018-31589-x.
- MacRae E. Extraction of plant RNA. *Methods Mol. Biol.* 2007;353: 15-24. DOI 10.1385/1-59745-229-7:15.
- Manck-Gotzenberger J., Requena N. Arbuscular mycorrhiza symbiosis induces a major transcriptional reprogramming of the potato SWEET sugar transporter family. *Front. Plant Sci.* 2016;7:487. DOI 10.3389/fpls.2016.00487.
- Phillips J.M., Hayman D.S. Improved procedures for clearing roots and staining parasitic and vesicular-arbuscular mycorrhizal fungi for rapid assessment of infection. *Trans. Br. Mycol. Soc.* 1970;55(1): 158-161. DOI 10.1016/S0007-1536(70)80110-3.
- Sokolov A.V. Agrochemical Methods of Soil Research. Moscow: Nauka Publ., 1975. (in Russian)
- Sugiyama A., Saida Y., Yoshimizu M., Takanashi K., Sosso D., Frommer W.B., Yazaki K. Molecular characterization of LjSWEET3, a sugar transporter in nodules of *Lotus japonicus*. *Plant Cell Physiol.* 2017;58(2):298-306. DOI 10.1093/pcp/pcw190.
- Sun M.X., Huang X.Y., Yang J., Guan Y.F., Yang Z.N. *Arabidopsis RPG1* is important for primexine deposition and functions redundantly with *RPG2* for plant fertility at the late reproductive stage. *Plant Reprod.* 2013;26(2):83-91. DOI 10.1007/s00497-012-0208-1.
- Vorob'ev N.I., Yurkov A.P., Provorov N.A. Program for calculating the mycorrhization indices of plant roots. Certificate No. 2010612112 of state registration of a computer program. December 2, 2016. Moscow, 2016. (in Russian)
- Yurkov A.P., Jacobi L.M., Gapeeva N.E., Stepanova G.V., Shishova M.F. Development of arbuscular mycorrhiza in highly responsive and mycotrophic host plant – black medick (*Medicago lupulina* L.). *Russ. J. Dev. Biol.* 2015;46:263-275. DOI 10.1134/S1062360415050082.
- Yurkov A., Kryukov A., Sherbakov A., Dobryakova K., Mikhaylova Y., Afonin A., Shishova M. AM-induced alteration in the expression of genes, encoding phosphorus transporters and enzymes of carbohydrate metabolism in *Medicago lupulina*. *Plants.* 2020;9(4):486. DOI 10.3390/plants9040486.
- Yurkov A.P., Puzanskiy R.K., Avdeeva G.S., Jacobi L.M., Gorbunova A.O., Kryukov A.A., Kozhemyakov A.P., Laktionov Y.V., Kosulnikov Y.V., Romanyuk D.A., Yemelyanov V.V., Shavarda A.L., Kirpichnikova A.A., Smolikova G.N., Shishova M.F. Mycorrhiza-induced alterations in metabolome of *Medicago lupulina* leaves during symbiosis development. *Plants.* 2021;10(11):2506. DOI 10.3390/plants10112506.
- Yurkov A.P., Shishova M.F., Semenov D.G. Features of the Development of Hop Clover with Endomycorrhizal Fungus. Saarbrücken: Lambert Academic Publishing, 2010. (in Russian)
- Zhao D., You Y., Fan H., Zhu X., Wang Y., Duan Y., Xuan Y., Chen L. The role of sugar transporter genes during early infection by root-knot nematodes. *Int. J. Mol. Sci.* 2018;19(1):302. DOI 10.3390/ijms19010302.

ORCID ID

A.A. Kryukov orcid.org/0000-0002-8715-6723
A.O. Gorbunova orcid.org/0009-0002-5624-1244
T.R. Kudriashova orcid.org/0000-0001-5120-7229
M.F. Shishova orcid.org/0000-0003-3657-2986
A.P. Yurkov orcid.org/0000-0002-2231-6466
O.B. Ivanchenko orcid.org/0000-0002-1311-1258

Acknowledgements. The work was supported by Russian Foundation for Basic Research, 20-016-00245_A (assessment of SWEET family gene expression, assessment of productivity parameters and symbiotic efficiency), Russian Foundation for Basic Research, 19-29-05275_mk (assessment of mycorrhization parameters) and Russian Science Foundation No. 22-16-00064 (selection and verification of the applicability of using primers for genes of interest).

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

Received September 23, 2022. Revised December 1, 2022. Accepted December 1, 2022.

Original Russian text <https://vavilovj-icg.ru/>

Effect of *NAM-1* genes on the protein content in grain and productivity indices in common wheat lines with foreign genetic material introgressions in the conditions of Belarus

O.A. Orlovskaya , S.I. Vakula, K.K. Yatsevich, L.V. Khotyleva, A.V. Kilchevsky

Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus
 O.Orlovskaya@igc.by

Abstract. Modern varieties of common wheat (*Triticum aestivum* L.) bred mainly for high productivity are often of low grain quality. The identification of *NAM-1* alleles associated with high grain protein content in wheat relatives has enhanced the significance of distant hybridization for the nutritional value of *T. aestivum* L. grain. In this work we aimed to study the allelic polymorphism of the *NAM-A1* and *NAM-B1* genes in wheat introgression lines and their parental forms and evaluate the effects of various *NAM-1* variants on the grain protein content and productivity traits in the field conditions of Belarus. We studied parental varieties of spring common wheat, the accessions of tetraploid and hexaploid species of the genus *Triticum* and 22 introgression lines obtained using them (2017–2021 vegetation periods). Full-length *NAM-A1* nucleotide sequences of *T. dicoccoides* k-5199, *T. dicoccum* k-45926, *T. kiharae*, and *T. spelta* k-1731 accessions were established and registered with the international molecular database GenBank. Six combinations of *NAM-A1/B1* alleles were identified in the accessions studied and their frequency of occurrence varied from 40 to 3 %. The cumulative contribution of *NAM-A1* and *NAM-B1* genes to the variability of economically important wheat traits ranged from 8–10 % (grain weight per plant and thousand kernel weight) to up to 72 % (grain protein content). For most of the traits studied, the proportion of variability determined by weather conditions was small (1.57–18.48 %). It was shown that, regardless of weather conditions, the presence of a functional *NAM-B1* allele ensures a high level of grain protein content; at the same time, it does not significantly decrease thousand kernel weight. The genotypes combining the *NAM-A1d* haplotype and a functional *NAM-B1* allele demonstrated high levels of productivity and grain protein content. The results obtained demonstrate the effective introgression of a functional *NAM-B1* allele of related species increasing the nutritional value of common wheat.

Key words: common wheat; wheat relatives; wheat introgressive lines; *NAM-1* genes; grain protein content; productivity.

For citation: Orlovskaya O.A., Vakula S.I., Yatsevich K.K., Khotyleva L.V., Kilchevsky A.V. Effect of *NAM-1* genes on the protein content in grain and productivity indices in common wheat lines with foreign genetic material introgressions in the conditions of Belarus. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2023;27(3): 197–206. DOI 10.18699/VJGB-23-26

Влияние генов *NAM-1* на содержание белка в зерне и показатели продуктивности у линий мягкой пшеницы с интрогрессиями чужеродного генетического материала в условиях Беларуси

O.A. Орловская , С.И. Вакула, К.К. Яцевич, Л.В. Хотылева, А.В. Кильчевский

Институт генетики и цитологии Национальной академии наук Беларуси, Минск, Республика Беларусь
 O.Orlovskaya@igc.by

Аннотация. Современные сорта мягкой пшеницы (*Triticum aestivum* L.), селекция которых велась в основном на увеличение продуктивности, часто имеют невысокое качество зерна. Обнаружение у сородичей пшеницы аллелей генов *NAM-1*, ассоциированных с высоким содержанием белка, увеличило значимость отдаленной гибридизации для повышения питательной ценности зерна *T. aestivum* L. Целью настоящей работы были изучение аллельного состава генов *NAM-A1* и *NAM-B1* у интрогрессивных линий пшеницы и их родительских форм и оценка эффекта различных вариантов генов *NAM-1* на содержание белка в зерне и признаки продуктивности пшеницы в полевых условиях Беларуси. Исследовали родительские сорта яровой мягкой пшеницы, образцы тетраплоидных и гексаплоидных видов рода *Triticum*, а также 22 интрогрессивные линии, полученные с их участием (вегетационный период 2017–2021 гг.). Впервые установлены и зарегистрированы в международной молекулярной базе данных GenBank полноразмерные нуклеотидные последовательности гена *NAM-A1* образцов *T. dicoccoides* к-5199, *T. dicoccum* к-45926, *T. kiharae*, *T. spelta* к-1731. В исследуемой выборке выявлено шесть комбинаций аллелей

NAM-A1/B1, частоты встречаемости которых варьировали от 40 до 3 %. Совместный вклад генов *NAM-A1* и *NAM-B1* в изменчивость хозяйственно важных признаков пшеницы составил от 8–10 % (масса зерна с растения и масса 1000 зерен) до 72 % (накопление белка в зерне). Для большинства изученных признаков доля изменчивости, обусловленная погодными условиями, была невелика (1.57–18.48 %). Показано, что наличие функционального аллеля *NAM-B1* обеспечивает высокий уровень накопления белка в зерне независимо от погодных условий и при этом не приводит к существенному снижению массы 1000 зерен. Высокие показатели продуктивности и уровня накопления белка в зерне установлены для генотипов, сочетающих гаплотип *NAM-A1d* и функциональный аллель *NAM-B1*. Полученные результаты свидетельствуют об эффективности интрогрессии функционального аллеля гена *NAM-B1* от видов-сородичей для повышения питательной ценности мягкой пшеницы. Ключевые слова: мягкая пшеница; сородичи пшеницы; интрогрессивные линии пшеницы; гены *NAM-1*; содержание белка в зерне; продуктивность.

Introduction

Common wheat (*Triticum aestivum* L.) is an important agricultural crop that plays a key role in providing food to people across the globe. One of the priority directions of wheat breeding is to improve grain quality, which is primarily determined by the total protein content (Brevis et al., 2010). Complex polygenic nature of the trait “grain protein content”, its variability when exposed to external factors, as well as the negative correlation between the protein content and productivity, complicate the breeding process (Iqbal et al., 2016). In addition, low genetic diversity of modern varieties by trait limits their use in breeding programs aimed at improving the nutritional value of wheat. Many related species of *T. aestivum* are characterized by a higher grain protein content compared to cultivated varieties (Peleg et al., 2008; Kumar et al., 2019).

New opportunities that make it possible to increase the total grain protein appeared in breeding with the identification of the *Gpc-B1* locus associated with protein content in wild emmer *T. dicoccoides* (AABB genome). The locus was mapped in the short arm of chromosome 6B and upon detailed clarification of its localization site boundaries, a sequence was found identified as the *NAM-B1* gene belonging to NAC family transcription factors (Uauy et al., 2006a). The genes of this family are involved in the regulation of various plant development programs, control of defense responses to biotic and abiotic stressors, and they play an important role in plant senescence (Puranik et al., 2012; Zhao et al., 2015). In addition to the *NAM-B1* gene, common wheat also has homologous *NAM-A1* and *NAM-D1* genes on chromosomes 6A and 6D (Avni et al., 2014).

A functional *NAM-B1* allele (wild-type allele) was found in wild emmer, providing high protein content in grain. The allele includes three exons and two introns and encodes a protein of 407 amino acid residues that has the conserved N-terminal region, or the NAC domain with five subdomains and the highly variable C-terminal transcriptional activation region (Waters et al., 2009). A functional *NAM-B1* allele is not found in most modern wheat varieties. Varieties have, as a rule, a 1 bp insertion in the first exon leading to a frameshift (mutant allele) or a gene deletion (partial or complete) and, as a result, to an inactive protein or its absence (Uauy et al., 2006b). Thus, a study of 218 wheat varieties from five main regions of China did not reveal any single variety with a functional allele of the *NAM-B1* gene (Chen et al., 2017). Molecular characterization of the *NAM-1* genes of Australian common wheat varieties showed the presence of the wild allele of the *NAM-B1* gene in only 2 out of 51 varieties (Yang et al., 2018).

It was established that the *NAM-A1* gene similar to *NAM-B1* consists of three exons; it possesses typical characteristics of NAC-family genes and is involved in the regulation of the same processes as *NAM-B1*. As a result of the analysis of the association of single nucleotide variants (SNPs) of the *NAM-A1* gene with nitrogen remobilization from leaves and grain protein accumulation, two functional single nucleotide substitutions were identified: at position 722 (T/C) and at position 1509 (A/del). Based on the data obtained, a classification of *NAM-A1* haplotypes was proposed: *NAM-A1a* (722C and 1509A), *NAM-A1b* (722C and 1509del), *NAM-A1c* (722T and 1509A), and *NAM-A1d* (722T and 1509del) (Cormier et al., 2015).

The absence, in modern wheat varieties, of the functional *NAM-B1* allele, which provides for high grain protein content in various environmental conditions, has strengthened the position of distant hybridization from a perspective of increased nutritional value of wheat grain. In order to enrich and to improve the common wheat gene pool, in the crossing with *T. aestivum* L. varieties we used accessions of the species of the genus *Triticum* (*T. dicoccoides*, *T. dicoccum*, *T. durum*, *T. spelta*, and *T. kiharae*). An earlier study of genetic diversity of the collection of introgressive wheat lines using C-banding and SSR analysis showed that in the genome of hybrid lines the foreign genetic material is presented both in the form of short fragments and whole chromosomes (Orlovskaya et al., 2016, 2020).

In this work, we aimed to study the allelic composition of *NAM-A1* and *NAM-B1* genes in introgressive wheat lines and their parental forms and to evaluate the effect of different variants of *NAM-1* genes on grain protein content and wheat productivity traits in the field conditions of Belarus.

Materials and methods

The study included five varieties of spring common wheat (Rassvet, Saratovskaya 29, Festivalnaya, Belorusskaya 80, and Pitic S62); tetraploid accessions *T. dicoccoides*, *T. dicoccoides* k-5199, *T. dicoccum* k-45926, and *T. durum*, and hexaploid accessions *T. spelta* k-1731 and *T. kiharae* of the species of the genus *Triticum*, as well as 22 introgressive lines we had developed (Supplementary Material 1)¹. The accessions of foreign donors were obtained from the collection of N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR). Information about the pedigree of individual accessions is not available as unpreserved (VIR catalogue numbers are not indicated).

¹ Supplementary Materials 1–8 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Orlovskaya_Engl_27_3.pdf.

The plants were grown in the experimental fields of the Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, in 2017–2021 (Minsk), on sod-podzolic loamy sand soil. The characteristic of weather conditions in the region of our experiment in 2017–2021 is presented in Supplementary Material 2. Data on average daily temperatures and precipitation (<http://rp5.by>) were used to calculate the sum of active temperatures (SAT) and Selyaninov's hydrothermal coefficient (HTC) (Mamontova, Khromov, 1974). When harvesting, the following traits were taken into account: plant height, the number of productive shoots per plant, the length of the main spike, the number of spikelets and grains of the main spike, grain weight per spike and plant, as well as thousand kernel weight. To assess the traits, 15 plants of each genotype were randomly selected.

To sequence full-length *NAM-A1* gene sequences and the first exon of *NAM-B1*, specific primers developed by R. Yang et al. (2018), were used. The sequencing reaction was performed using the BigDye Terminator v. 3.1 Cycle Sequencing kit (Applied Biosystems, USA); the separation of sequencing reaction products was carried out using the ABI PRISM 3500 genetic analyzer (Applied Biosystems). Alignment of nucleotide sequences and homology analysis were performed using the BLAST analyzer of the National Center for Biotechnology Information, USA (<http://www.ncbi.nlm.nih.gov/BLAST>). The Chinese Spring variety was used as a reference sequence, which, according to the literature data, is the carrier of the *NAM-A1a* haplotype (722C and 1509A) and the *NAM-B1* mutant allele (T insertion at position +11) (Yang et al., 2018).

The total protein content in wheat grain was determined in accordance with GOST 10846-91 (2009) at the Central Republican Laboratory of the State Institution "State Inspectorate for Testing and Protection of Plant Varieties" (Minsk, Belarus). The essence of the method lies in the mineralization of organic matter with sulfuric acid in the presence of a catalyst with the formation of ammonium sulfate, the destruction of ammonium sulfate with alkali with the release of ammonia, the stripping of ammonia with water vapor into a solution of sulfuric or boric acids followed by titration.

The results of the experiment were summarized using descriptive statistics methods; two-way analysis of variance, regression, and correlation analyses (the Spearman's rank correlation coefficient was used). Statistical procedures were implemented in the software packages Statistica 10.0, MS Excel, and web application SNPstats. A quantitative contribution of individual factors of the dispersion analysis was calculated on the basis of the relation of the absolute factor variance to the sum of variances of that factor and other factors, according to the formulae given in (Rokitsky, 1973).

Results

Allelic polymorphism of *NAM-A1* and *NAM-B1* genes

Full-length *NAM-A1* gene sequencing was carried out in the parental varieties and accessions of species of the genus *Triticum*, in 22 introgressive lines developed on their basis. Among the parental forms, we detected haplotype *NAM-A1a* in Festivalnaya and Rassvet varieties; accessions *T. durum*,

T. dicoccum k-45926, *T. dicoccoides* k-5199, *T. dicoccoides*, and *T. kiharae*; haplotype *NAM-A1c* in *T. spelta* k-1731; and haplotype *NAM-A1d* in varieties Saratovskaya 29, Belorusskaya 80, and Pitic S62 (Supplementary Material 3).

A comparative analysis showed that *NAM-A1a* nucleotide sequences of *T. dicoccoides* k-5199, *T. dicoccoides*, and *T. kiharae* accessions did not have 100 % similarity with the *NAM-A1a* sequence of *T. aestivum* (MH160778) from the GenBank database. The sequences of wild emmer accessions (*T. dicoccoides* k-5199 and *T. dicoccoides*) that we studied differed from *NAM-A1a* of *T. aestivum* (MH160778) in two SNPs: positions 538 bp (C/A in exon 2) and 1139 bp (G/T in exon 3) (99.9 % identity level). The SNP 1139 bp of G/T results in the replacement of asparagine with tyrosine in the amino acid sequence of protein. The level of similarity of *NAM-A1a* of the *T. kiharae* accession with *NAM-A1a* from the GenBank database (MH160778) was 99.7 % and differed from it in six SNPs: in the positions of 189 bp (C/A in exon 1), 306 bp (A/C in intron 1), 1133 bp (G/A in exon 3), 1271 bp (G/T in exon 3), 1414 bp (C/G in exon 3), and 1491 bp (G/C in exon 3). Three of these SNPs lead to changes in the amino acid sequence of the protein: a G/A substitution in the position of 1133 bp leads to the replacement of alanine with threonine; G/T in the position of 1271 bp results in the replacement of alanine with serine; and G/C in the position of 1491 bp replaces glycine with alanine.

The *NAM-A1* gene sequence of *T. dicoccum* k-45926 and *T. durum* accessions was completely homologous to the *NAM-A1a* allele sequence (GenBank: MH160778); the *NAM-A1* sequence of the *T. spelta* k-1731 accession corresponded to the *NAM-A1c* allele (GenBank: MH160777). Nucleotide sequences of the *NAM-A1* gene of *T. kiharae*, *T. spelta* k-1731, *T. dicoccoides* k-5199, and *T. dicoccum* k-45926 accessions, which we described for the first time, were registered with the International GenBank Database (access codes MT572492, MT920417, MW384855, and MW384856, respectively).

Analysis of *NAM-A1* gene sequencing data in introgressive wheat lines showed that 54.6 % of the lines had the *NAM-A1d* haplotype; 36.4 % – the *NAM-A1a* haplotype; and 9.1 % – the *NAM-A1c* haplotype. It should be noted that wheat lines with foreign genetic material inherited, as a rule, the *NAM-A1* gene of the original wheat variety but there were a number of exceptions (see Supplementary Material 3). A haplotype corresponding to a related species was identified in the lines developed using *T. spelta* k-1731 (lines 1-8 and 7, *NAM-A1c*); in line 226-7 *T. durum* × Belorusskaya 80 (*NAM-A1a*); and in line 20-1 *T. kiharae* × Saratovskaya 29 (*NAM-A1a*). Among the lines developed using *T. dicoccoides*, only lines 11-1 and 13-3 obtained as a result of crossing with the Festivalnaya variety inherited the *NAM-A1* gene from wild emmer with the SNPs only characteristic of it at positions 538 bp and 1139 bp of the nucleotide gene sequence.

Analysis of the sequenograms of the first exon of the *NAM-B1* gene in the studied genotypes revealed a functional allele (*F*) in all the accessions of related species, except for *T. durum*, for which the *NAM-B1* allele was not identified, and in 5 out of 22 introgressive wheat lines (13-3 and 15-7-1 of the *T. dicoccoides* × Festivalnaya combination; 19 and 25-2

of *T. kiharae* × Saratovskaya 29; and line 7 of the *T. spelta* k-1731 × Saratovskaya 29 combination). All parental varieties and most of the wheat lines with the foreign genetic material (77.3 %) had a mutant allele (*NF*).

Effects of *NAM-A1* and *NAM-B1* genes on the productivity traits and the grain protein content

Genotyping results of common wheat introgression lines and parental forms by *NAM-A1* and *NAM-B1* genes were compared with the results of field trials and the data on grain protein content for 2017–2021. The effect of genotype, environmental factors, and their interaction were assessed using the general linear model (GLM) of two-way analysis of variance (Supplementary Material 4).

The combination of *NAM-A1/B1* genes produces a statistically significant effect on the manifestation of all nine traits studied, while exceeding a contribution of individual *NAM-I* genes to the variability of protein content, plant height, the number of spikelets per spike, and thousand kernel weight (see Supplementary Materials 4 and 5). *NAM-B1* does not significantly affect the spike length of wheat, while *NAM-A1* or the combination of *NAM-A1/B1* determine more than half of the observed trait variation.

The greatest length of the main spike relative to other haplotypes is typical for *NAM-A1c* plants (9.86 cm in average over the 5-year observation period), and in the case of the *c/F* combination, this indicator increases up to 10.87 cm. The effect of *NAM-B1* allelic variants on the variability of the number and weight of grains in the spike is almost three times higher than the effect of *NAM-A1* haplotypes, and it is one-and-a-half times higher in the case of the *NAM-A1* and *NAM-B1* combination.

The spike productivity in the group of samples with the wild *NAM-B1* allele was significantly lower than in the genotypes with a mutant allele (Supplementary Material 6). Thus, in the spike of the vast majority of genotypes with a functional allele (7 out of 10), the number of grains did not reach 30 pieces, while in the spike of genotypes with a mutant allele, as a rule, 30–40 grains were formed. A significant variation in spike productivity traits can be noted in both groups. Individual genotypes with a functional allele demonstrated high indices by these traits (lines 19 of *T. kiharae* × Saratovskaya 29 (*d/F*) and 15-7-1 of *T. dicoccoides* × Festivalnaya (*a/F*)). In all three variants of variance analysis, thousand kernel weight and the weight of grains per plant are the traits with a low (up to 10 %) contribution of genetic variances; at that, the weight of grains per plant is statistically independent of the *NAM-A1* haplotype and thousand kernel weight is independent of the *NAM-B1* allele. Variability of the protein content in grain is 70 % associated with *NAM-B1* polymorphism, and a contribution of *NAM-A1* is significantly lower and it slightly increases when considering a combination of gene alleles (see Supplementary Material 5). Thus, on average, during the 4-year period, the highest grain protein content in the groups with different *NAM-A1* haplotypes was observed in the case of *NAM-A1a* genotypes (21.53 %); in the groups with different *NAM-B1* variants – in the case of the functional *NAM-B1* allele (22.53 %); and the maximum amount of protein was noted (23.72 %) in the case of the *a/F* combination (Supplementary Material 7).

The strength and direction of the relationship between the different alleles of *NAM-I* genes and economically valuable traits was assessed using the Spearman's correlation coefficient (Table 1).

A medium-strength relationship was found between the grain protein content and the allelic variants of *NAM-I* genes, but only the correlation with *NAM-B1* is statistically significant (see Table 1). Apart from that, significant dependence was established between the *NAM-B1* allelic variants and the tilling capacity, the number and weight of grains per spike. The *NAM-A1* haplotype significantly correlated only with the spike length; association with other productivity traits was weak. Variance analysis results also showed that *NAM-B1* allelic variants produce a significantly greater impact on the variability of grain protein content, the number and weight of grains per spike, and productive tilling capacity than *NAM-A1* (see Supplementary Material 5). It should be noted that both variance and correlation analyses established a low correlation of *NAM-I* genes with the traits “plant height”, “the number of spikelets per spike”, “grain weight per plant” and “thousand kernel weight” (see Supplementary Material 5, Table 1).

In total, six combinations of *NAM-A1/B1* alleles were identified in the accessions under study and their frequencies varied from 40 (*d/NF*) to 3 % (*c/NF*). Some combinations of alleles are represented by a small number of samples (*d/F*, *c/F*, and *c/NF*), and therefore, it does not seem possible to speak about a significant difference of traits in plants with such combinations of alleles.

Mean values of productivity traits of the genotypes carrying various combinations of *NAM-A1/B1* genes are demonstrated in Table 2.

The genotypes combining different *NAM-A1* haplotypes with the functional *NAM-B1* allele were taller and of a higher tilling capacity than the plants with a corresponding haplotype combined with the mutant *NAM-B1* allele (see Table 2). The maximum values for these traits are typical for *c/F* accessions: 93.36 cm and 3.86 pcs. An opposite trend was revealed for spike productivity traits: all three *NAM-A1* haplotypes in combination with the functional *NAM-B1* allele had a low number and weight of grains per spike (see Table 2). The samples with *NAM-A1a* and *NAM-A1c* haplotypes in combination with the functional allele demonstrated a slight decrease in thousand kernel weight compared with the genotypes combining *NAM-A1a* and *NAM-A1c* and the mutant allele (see Table 2).

For the plants with the *NAM-A1d* haplotype, an increase in thousand kernel weight in the group with the functional allele was revealed (40.20 g for *d/NF* genotypes and 44.34 g for *d/F* genotypes). It can be noted that *NAM-A1a* accessions were the least productive by this trait both in the group with functional (37.97 g) and non-functional (38.93 g) *NAM-B1* alleles. The presence of the mutant allele of the *NAM-B1* gene leads to a decrease in the protein content relative to the combination with the functional allele: for *NAM-A1a* haplotypes – by 3.6 % on average; for *NAM-A1d* by 2.3 %; and for *NAM-A1c*, by 0.6 % (in the latter case, the decrease is not statistically significant). The maximum amount of protein in grain was accumulated by the lines with the *a/F* combination and the minimal amount – by the lines with the *d/NF* and *c/NF* combinations (see Table 2).

Table 1. Spearman’s correlation coefficient among the allelic *NAM-A1* and *NAM-B1* gene variants and economically important traits of investigated wheat genotypes

Gene	PH	Tillering	SL	SS	GS	GWS	GWP	TKW	Grain protein content
<i>NAM-A1</i>	-0.008	0.01	0.67*	0.06	0.08	0.22	0.20	0.18	0.43*
<i>NAM-B1</i>	0.33	0.53*	-0.16	-0.17	-0.44*	-0.47*	-0.18	-0.06	0.57*

Note. PH – plant height; SL – spike length; SS – spikelets per spike; GS – grain per spike; GWS – grain weight per spike; GWP – grain weight per plant; TKW – thousand kernel weight. * Significant at $p < 0.05$.

Table 2. Mean values of productivity traits and grain protein content in the groups of wheat genotypes with different combinations of *NAM-1* allelic variants

<i>NAM-1</i> alleles	<i>d/NF</i>	<i>a/NF</i>	<i>a/F</i>	<i>d/F</i>	<i>c/F</i>	<i>c/NF</i>
Frequency of allele combinations, %	40	30	13	7	7	3
Trait	X ± SEM					
Plant height, cm	76.19 ± 0.42	78.79 ± 0.57	79.93 ± 0.77	81.73 ± 1.09	93.36 ± 1.54	82.95 ± 1.79
Tilling capacity, <i>n</i>	2.89 ± 0.02	2.97 ± 0.03	3.29 ± 0.05	3.11 ± 0.07	3.86 ± 0.09	3.19 ± 0.09
Spike length, cm	9.46 ± 0.04	8.82 ± 0.05	7.87 ± 0.10	10.17 ± 0.09	10.75 ± 0.19	8.78 ± 0.11
Spikelets per spike, <i>n</i>	16.41 ± 0.08	16.27 ± 0.09	15.86 ± 0.20	14.58 ± 0.22	15.13 ± 0.20	15.32 ± 0.25
Grains per spike, <i>n</i>	35.55 ± 0.30	35.36 ± 0.37	28.60 ± 0.60	28.78 ± 0.78	27.84 ± 0.55	28.83 ± 0.64
Grain weight per spike, g	1.43 ± 0.01	1.38 ± 0.02	1.07 ± 0.03	1.27 ± 0.04	1.08 ± 0.03	1.18 ± 0.03
Grain weight per plant, g	3.12 ± 0.04	3.14 ± 0.05	2.64 ± 0.07	2.93 ± 0.10	3.11 ± 0.10	2.90 ± 0.10
Thousand kernel weight, g	40.20 ± 0.23	38.93 ± 0.33	37.97 ± 0.50	44.34 ± 0.54	39.12 ± 0.68	41.25 ± 0.72
Grain protein content, %	19.14 ± 0.2	20.07 ± 0.2	23.72 ± 0.3	21.37 ± 0.4	19.77 ± 0.2	19.14 ± 0.4

Note. *a* – *NAM-A1a* haplotype; *d* – *NAM-A1d* haplotype; *c* – *NAM-A1c* haplotype; *F* – functional *NAM-B1* allele; *NF* – mutant *NAM-B1* allele; *X* – mean values of traits; SEM – the standard error of the mean.

The role of weather conditions and their relationship with *NAM-1* genes regarding variability of productivity traits and grain protein content

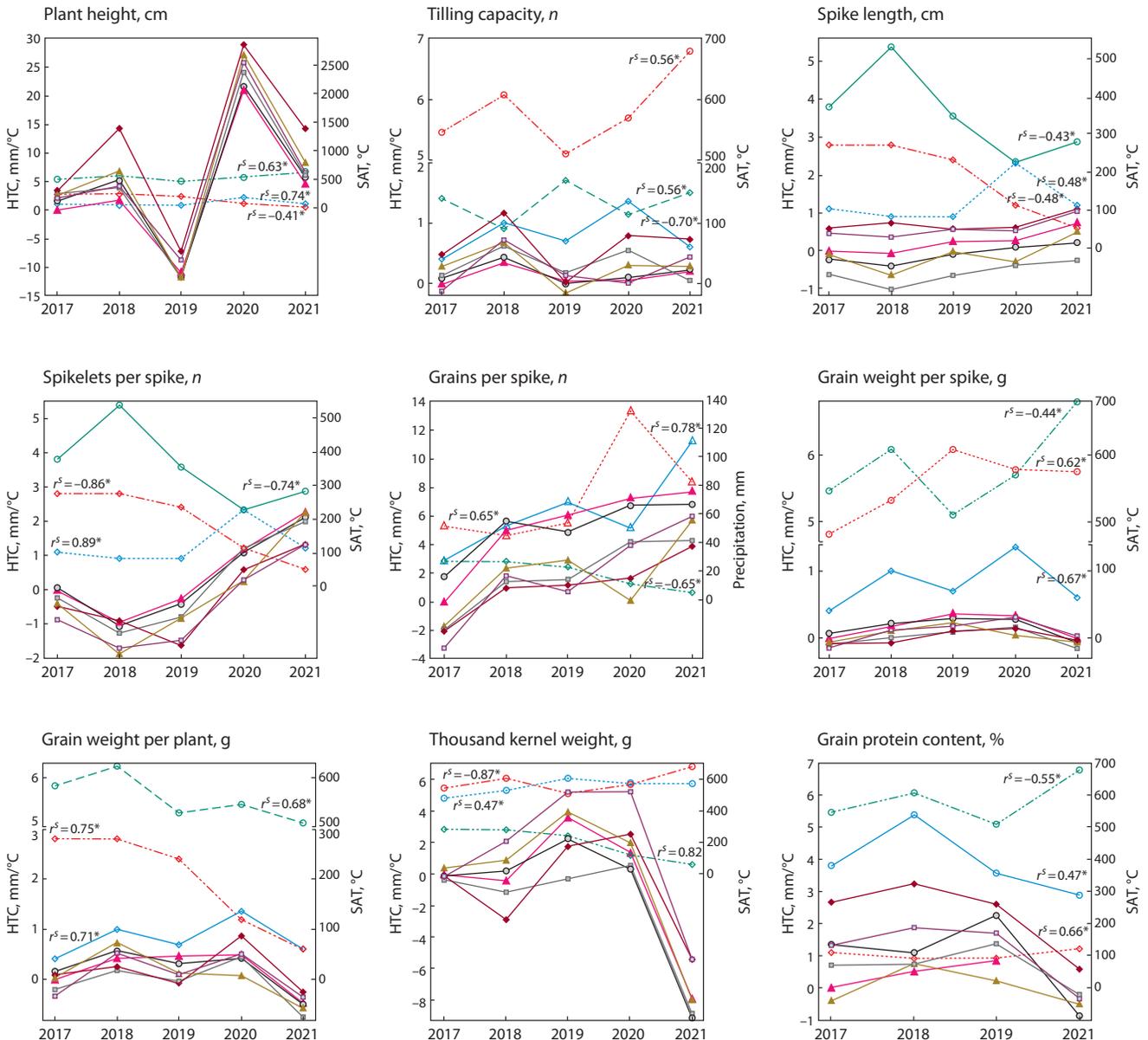
Regardless of the model used (*NAM-A1* × weather conditions, *NAM-B1* × weather conditions; and *NAM-A1/B1* × weather conditions), the high statistical significance of the contribution of weather conditions of the year of growth was shown for all the traits. However, the predominant role of this factor (> 50.0 %) was found only in the variability of traits “plant height” and “the number of spikelets per spike”. The interaction of the *NAM-A1/B1* gene combination with weather conditions significantly affects all wheat productivity indices. The contribution of this interaction is especially high (more than 60 %) in the variance of the traits “grain weight per spike and plant”, “thousand kernel weight” and “tilling capacity”. The impact of the interaction “*NAM-B1* × weather conditions” on productivity traits is lower than the contribution of other factors (see Supplementary Material 5). The only exception is the trait “number of spikelets per spike”, for which a contribution of genotype-environmental interactions with *NAM-B1* is 3 % and with *NAM-A1* it is statistically insignificant ($p = 0.23$). It should be noted that the interaction of weather conditions and all three genetic factors did not affect grain protein content.

Different genotypes may respond to changes in environmental conditions differently. Additivity of *NAM-A1/B1* effects and weather conditions on the manifestation of the traits studied

was tested using the log-additive linear regression model of the SNPstats web application. For the traits “plant height”, “tilling capacity”, “the number of spikelets per spike”, “weight of grains per plant”, and “thousand kernel weight”, the mutual enhancement of the effects of two factors was found: *NAM-A1/B1* alleles and weather conditions. The additive interaction of the *NAM-A1/B1* genotype and the environment is statistically insignificant for the variability of traits “spike length”, “the number and weight of grains in the main spike”, and “protein content”.

Variability of the studied traits of genotypes carrying various *NAM-A1/B1* combinations was assessed under the conditions of different growing seasons relative to the productivity of plants with *d/NF* alleles in 2017 (see the Figure). With a view to searching for meteorological factors that determine genotype-environmental interaction, the following was carried out: (1) a comparative analysis of the ranking of allelic combinations in 2017–2021 conditions; (2) a correlation analysis of productivity traits with meteorological parameters (see the Figure).

An analysis of correlation coefficients showed that most of the traits studied were significantly influenced by weather conditions during the grain filling stage – HTC and SAT in July. The closest association was found between the HTC in July and the plant height, the number of spikelets per spike, the weight of grains per plant and thousand kernel weight (see



The difference in productivity traits between the wheat plants with *d/NF* in 2017 and different combinations of *NAM-A1/B1* alleles under the weather conditions of Belarus in 2017–2021 (SNPStats results).

Combinations of *NAM-A1/B1* alleles: \blacktriangle *d/NF*; \circ *a/NF*; \square *a/F*; \blacklozenge *c/F*; \blacktriangledown *d/F*; \square *c/NF*. Weather conditions: \circ EHS – effective heat sum; $\circ\cdots\circ$ HTC – hydrothermal coefficient; Δ – rainfall. Vegetation periods: — May; \cdots June; $\cdots\cdots$ July; $-\cdots-$ August; r^s – the Spearman's rank correlation coefficient, which shows the strength of the relationship between productivity traits and meteorological parameters.

the Figure). During the grain filling stage, there is an increased supply of mineral and organic substances to the wheat grain; unfavorable conditions during this period significantly worsen its quality and reduce yields. The optimal HTC in July (1.2) is typical for the 2020 season. In July 2021, there was an increase in air temperature and a lack of precipitation compared to the norm, and the HTC was only 0.4, characterizing this period as dry. In other years, the HTC in July significantly exceeded the norm and amounted to 2.8, 2.8 and 2.4 in 2017, 2018 and 2019 respectively. It can be noted that the air temperature and precipitation throughout the growing season of 2020 were close to the climatic norm, which contributed to the maximum realization of wheat productivity (Supplementary Material 8).

The conditions of 2021 (soil waterlogging in May; drought in July; and rains during grain ripening and harvesting) led to a significant decrease in the yields of the genotypes under study (see Supplementary Material 8).

“Plant height” is a trait with an additive contribution of *NAM-1* genes and weather conditions. The presence of the functional *NAM-B1* allele and the *NAM-A1c* haplotype is associated with a statistically significant increase in culm. Maintained soil moisture (HTC > 1) during the “booting–flowering” period stimulates plant growth, while excessive precipitation in July–August, on the contrary, negatively affects plant height. For example, the average height of wheat plants in 2019 (cold and wet summer) decreased by 10.6–14.5 cm,

and in 2020 (the year close to the climatic norm) it increased by more than 20 cm relative to 2017 values. Plants with the *NAM-A1c* haplotype are the least sensitive to environmental factors that reduce the stem height, while the growth of the culm reaches 10–25 cm under favorable conditions.

It was shown that the tillering capacity of plants is negatively affected by soil waterlogging and a large amount of precipitation in August, which is associated with the inhibition of secondary growth processes, lodging and disruption of gas exchange in the root system. The trait is positively correlated with the HTC of May and the SAT of July; that is, the absence of frosts and drought at the tillering stage and the high intensity of photosynthesis at the maturation stage. Plants with the functional *NAM-B1* allele form more stems under favorable conditions and significantly reduce tillering capacity under spring drought conditions (see the Figure).

SNPstats models failed to explain spike length variability either by the additive interaction of *NAM-A1* and *NAM-B1* alleles or the interaction of genetic factors and growing conditions. Both the spike length and the number of spikelets per spike are influenced by the conditions observed at the stages of booting and heading, namely, the HTC in June. Warm and dry conditions in May–June 2018 were accompanied by a decrease in the average length and the number of spikelets per spike. However, for late-ripening genotypes with the *c/F* combination, a decrease in the number of spikelets turned out to be statistically insignificant. The number and weight of grains in the main spike depend on the optimal moisture conditions at the stage of germination and tillering, the sum of active temperatures at the stage of heading and flowering. Low yields of the main spike were noted in the conditions of 2017 and 2021, which were characterized by a low hydrothermal coefficient in May, but differed in temperature and hydrological regimes at the subsequent stages of the growing season. In 2021, the maximum number of grains formed in the main spike was noted and in 2017 – the minimum one (see the Figure).

The weight of grains per plant and thousand kernel weight are the most important indicators of wheat yield. In cases of a significant difference in the mean values of traits in the plants carrying *NAM-A1/B1* combination variants, genetic factors do not produce an additive effect. Regardless of the year of cultivation, *NAM-A1a* and *NAM-A1c* haplotype combinations with a functional *NAM-B1* variant are associated with a slight decrease in thousand kernel weight relative to combinations with a non-functional gene variant. There were no significant differences found between *d/F* and *d/NF* genotypes by seed productivity traits (see the Figure). Late-ripening genotypes carrying the *NAM-A1c* haplotype significantly increase thousand kernel weight in the case of a combination of high HTC in July and August, which was observed in 2019–2020 (see the Figure).

For the formation of grain, wheat plants remobilize nitrogen and carbohydrates from the flag leaf, and therefore, the protein content in grain depends on the intensity of photosynthesis and the photosynthetic surface area (Lawlor et al., 1989). High HTC in June may produce a negative effect on the protein content through a decrease in these parameters – low solar insolation as a result of high cloudiness and an increase in the lamina damage area caused by phytopathogens under high

humidity conditions. Under unfavorable conditions for overall productivity in 2021, plants with the functional *NAM-B1* allele accumulated more protein than the genotypes with a mutant gene variant (see the Figure).

Discussion

The common wheat varieties we studied had the *NAM-A1d* or *NAM-A1a* haplotype and the mutant allele *NAM-B1*, which is consistent with the literature data. For example, in the collection of Australian wheat varieties, accessions with haplotypes *NAM-A1a* (50.1 %) and *NAM-A1d* (28 %) were the most frequently occurring, while *NAM-A1b* (1.9 %) was the least common (Yang et al., 2018). F. Cormier et al. (2015) revealed in their studies that the *NAM-A1d* haplotype is typical for most modern European wheat varieties, while *NAM-A1a* is more common among the varieties with high baking properties. Wheat samples with the *NAM-A1b* haplotype were not found in the collection we studied. The works of foreign scientists provide data on the low frequency of occurrence of this haplotype. For example, in the collection of 795 wheat accessions, only one accession with *NAM-A1b* was found. There is an assumption that this haplotype has appeared recently as a result of recombination between *NAM-A1a* and *NAM-A1d* (Cormier et al., 2015).

The presence of functional *NAM-B1* alleles only in wheat relatives among the parental forms we studied is confirmed by the data of other researchers. Thus, in the work of C. Uauy et al. (2006b), a functional *NAM-B1* allele was found in all 42 analyzed accessions of *T. dicoccoides* and in 17 out of 19 accessions of *T. dicoccum* (Schrank) Schuebl., while all 57 studied varieties of durum wheat and 34 of common wheat either contained a 1 bp insertion or had a gene deletion. Among 367 common wheat accessions of the INRA core collection (France) selected from 3942 genotypes of different geographic origin, only 5 contained functional *NAM-B1* alleles (Hagenblad et al., 2012). Due to the fact that currently cultivated varieties are, as a rule, missing a functional *NAM-B1* allele, five introgressive lines of wheat with the allele that we have developed are of great interest from the point of view of their ability to enhance the quality of wheat grain.

Analysis of differences in the mean values of quantitative traits in the groups carrying different *NAM-A1* haplotypes showed that during all years of the experiment, *NAM-A1a* plants had a short spike, low grain weight per plant and thousand kernel weight, and high protein content in grain; *NAM-A1c* – the maximum plant height, tillering capacity and spike length but low grain content and seed weight per spike; *NAM-A1d* – the minimum plant height, number of productive shoots, and the maximum values of spike and plant productivity traits. A number of studies have shown that the presence of the *NAM-A1a* haplotype is associated with a shorter period of grain filling and a more intense process of nitrogen remobilization, which leads to increased protein content; but, at the same time, to a decrease in the number of grains per spike and thousand kernel weight, while the presence of *NAM-A1c* or *NAM-A1d* leads to an increase in the grain filling period, which results in an increase in the amount of nitrogen uptake and wheat yield (Cormier et al., 2015; Alhabbar et al., 2018). There is an assumption that *NAM-A1a* is a functional variant

of the *NAM-A1* gene, which is rarely found in elite modern wheat varieties the breeding of which was carried out mainly for productivity purposes.

The data we obtained on a high level of protein accumulation in grain in wheat genotypes with the wild *NAM-B1* allele coincide with the results of many foreign scientists. The study of a series of almost isogenic lines based on common and durum wheat in different countries across the globe (the USA, Argentina, India, China, Australia, etc.) allowed us to conclude that the introgression of a functional *NAM-B1* allele into the genome of cultivated wheats of both ploidy levels leads to an increase in the content of protein and key minerals in grain, an improved harvesting nitrogen index, and increased protein harvest (Tabbita et al., 2013; Maphosa et al., 2014; Mishra et al., 2015; Kuhn et al., 2016).

Throughout the entire observation period, the genotypes with a functional *NAM-B1* allele were characterized by higher plant height and tillering capacity, but lower indicators by spike and plant productivity traits compared to the genotypes carrying a nonfunctional allele. However, we did not find any significant impact of the *NAM-B1* allele state on thousand kernel weight. In the literature, there are data on both positive and negative effects of the wild-type *NAM-B1* allele on the main components of wheat productivity (Carter et al., 2012; Maphosa et al., 2014; Kuhn et al., 2016). A significant increase in the productive stems in common wheat lines with a functional *NAM-B1* allele that we established was also described in the works of other scientists (Tabbita et al., 2013; Vishwakarma et al., 2016).

Tillering capacity is known to be determined by many environmental factors, including nitrogen availability (Wang, Below, 1996). It is possible that a functional *NAM-B1* allele contributes to the formation of productive stems due to the fact that it improves nitrogen metabolism (Tabbita et al., 2013). According to the review article that summarizes the data of 25 studies on the influence of allelic variants of the *NAM-B1* gene on 50 wheat traits, 36 % of the studies did not show any significant differences in thousand kernel weight between the genotypes with different allele variants of this gene; correspondingly, 23 and 41 % of the studies revealed both a significant decrease and a significant increase in this indicator in the lines with a functional allele. It should be noted that the majority of studies (79 %) did not establish a statistically significant effect of *NAM-B1* polymorphism on wheat yield, and only 4 % showed a decrease in the yield of lines with a functional allele (Tabbita et al., 2017). This fact is explained by a positive effect of a functional allele on the formation of productive stems, since it is precisely due to high tillering capacity that no significant decrease takes place in grain yield, even despite low spike productivity (Tabbita et al., 2013).

Evaluation of the effect of six combinations of *NAM-A1/B1* alleles on the level of manifestation of a number of economically valuable wheat traits showed that the maximum height and tillering capacity are characteristic of *c/F* genotypes; *d/NF* genotypes are responsible for high productivity of spikes and plants; and the highest grain protein concentration is determined by *a/F* (see Table 2). Similar results are presented in (Alhabbar et al., 2018), regarding the influence of the allelic composition of *NAM-1* genes on the efficiency of nitrogen

use, productivity and protein content in wheat grain. In this study, the Mace variety, which has a non-functional *NAM-B1* allele and the *NAM-A1d* haplotype, significantly outperformed other genotypes in terms of yield but had the minimum grain protein content. Varieties combining a functional allele with different *NAM-A1* haplotypes had high grain protein content, while a negative correlation was found between the mutant allele and grain protein content, regardless of the *NAM-A1* haplotype (Alhabbar et al., 2018).

The combination of *NAM-A1/B1* genes had a significant effect on the formation of all the traits of wheat under study, while the grain weight per plant was statistically independent of the *NAM-A1* haplotype, and the spike length and thousand kernel weight were statistically independent of the *NAM-B1* gene allele. It is also important to note a small joint contribution of these two genetic factors to the variability of most productivity traits studied (7.59–18.75 %), but the predominance of *NAM-A1/B1* in the variability of protein content in grain (72 %) should be mentioned at that.

It is known that, in addition to the genetic control, crop and its components are significantly influenced by environmental factors (Ahmed et al., 2020; Kronenberg et al., 2021). Our study shows a high statistical significance of the contribution of weather conditions to the variability of wheat quantitative traits. A particularly high role of this factor (more than 50 %) was revealed for the dispersion of traits with a wide reaction rate and a high variation coefficient strongly dependent on the ambient temperature and the amount of precipitation – “plant height” and “the number of spikelets per spike”. For the rest of the traits studied, the share of variability determined by weather conditions was significantly lower and was in the range of 1.57–18.48 %; and the impact of the “*NAM-A1/B1* × weather conditions” interaction was of great importance (> 60 %). According to the data we obtained, the effect of the “*NAM-B1* × weather conditions” interaction on productivity traits is lower than the contribution of other factors; and for the traits “the length of spike”, “the number of grains per spike”, and “the grain protein content”, it is statistically insignificant (see Supplementary Material 5).

It should be emphasized that *NAM-B1* has a high degree of impact on the level of protein accumulation (70 %), which, along with a low contribution of weather conditions and the absence of genotype-environment interaction, indicates the effectiveness of functional allele introgression for the improvement of the quality of wheat grain in Belarus. A number of works by foreign scientists have shown a significant impact of the “*NAM-B1* × environment” interaction on wheat productivity and grain quality traits. For example, the study by (Carter et al., 2012), established that differences in environmental conditions affect the expression of the *NAM-B1* gene, which limits the use of functional allele introgression for increasing the level of protein accumulation in spring wheat grains in the regions with a short vegetation period. However, when studying the effects of *NAM-B1* on the total protein content in grain and the main yielding components of common wheat in Argentina, for most of the traits studied (including thousand kernel weight and protein content), no significant interaction “*NAM-B1* × environment” and “*NAM-B1* × genotype” was shown, while the impact of “genotype” and “environment” factors was statistically significant (Tabbita et al., 2013).

Conclusion

For the first time, full-length nucleotide *NAM-A1* sequences of accessions *T. kiharae*, *T. spelta* k-1731, *T. dicoccoides* k-5199, and *T. dicoccum* k-45926 were identified and registered with the international molecular database GenBank. The studied accessions of related common wheat species had the *NAM-A1a* haplotype, except for *T. spelta* k-1731 (*NAM-A1c*). Both *NAM-A1d* and *NAM-A1a* haplotypes are characteristic of *T. aestivum* varieties. Among the parental forms, a functional allele of the *NAM-B1* gene was found only in the accessions of related species. Introgressive lines inherited, as a rule, the variants of *NAM-1* genes of the original wheat variety. Out of 22 introgressive lines, the *NAM-A1* haplotype of related species was identified in 6 lines, while a functional *NAM-B1* allele was detected in 5. Line 13-3 *T. dicoccoides* × Festivalnaya (*a/F*) and line 7 *T. spelta* k-1731 × Saratovskaya 29 (*c/F*) had the *NAM-A1* haplotype and the *NAM-B1* allele of *T. dicoccoides* and *T. spelta*, correspondingly.

The genotyping results of introgressive lines of common wheat and parental forms by *NAM-A1* and *NAM-B1* genes were compared with the results of field trials in the conditions of Belarus and with the analysis data on protein content in grain of 2017–2021. The combination of *NAM-A1/B1* genes had a significant effect on the formation of all the traits of wheat studied, while the grain weight per plant was statistically independent of the *NAM-A1* haplotype, and the spike length and thousand kernel weight were statistically independent of the *NAM-B1* allele. A joint contribution of these two genetic factors to the variability of economically valuable traits ranges from 8–10 % (grain weight per plant and thousand kernel weight) to 72 % (grain protein content).

For most of the traits studied, the proportion of variability determined by weather conditions was small (1.57–18.48 %). The closest correlation was established between the majority of traits studied and the HTC during the grain filling phase. The interaction “*NAM-A1/B1* × weather conditions” determines 65–71 % of the variability of wheat grain productivity traits, while it is not significant for the grain protein content. The contribution of the “*NAM-B1* × weather conditions” interaction to quantitative traits is lower than the contribution of other factors; and for the traits “spike length”, “the number of grains per spike” and “the grain protein content” it is statistically insignificant. It was found that the presence of a functional *NAM-B1* allele provides for a high level of protein accumulation in grain, regardless of weather conditions, and at the same time, it does not lead to a significant decrease in thousand kernel weight.

Evaluation of the effect of six combinations of *NAM-A1/B1* alleles on the level of manifestation of a number of economically valuable traits of wheat showed that high productivity of spike and plant but a low level of protein content in grain are characteristic of *d/NF* genotypes, while the highest protein concentration and low indicators by main productivity traits are characteristic of *a/F*. The optimal combination of the wheat traits studied was established for *d/F* genotypes. The results obtained prove the effectiveness of the introgression of a functional *NAM-B1* allele of related species for increased nutritional value of common wheat.

References

- Ahmed H.G.M., Sajjad M., Zeng Y., Iqbal M., Khan S.H., Ullah A., Akhtar M.N. Genome-wide association mapping through 90K SNP array for quality and yield attributes in bread wheat against water-deficit conditions. *Agriculture*. 2020;10(9):392. DOI 10.3390/agriculture10090392.
- Alhabbar Z., Yang R., Juhasz A., She M., Anwar M., Sultana N., Diepeveen D., Ma W., Islam S. *NAM* gene allelic composition and its relation to grain-filling duration and nitrogen utilization efficiency of Australian wheat. *PLoS One*. 2018;13(10):e0205448. DOI 10.1371/journal.pone.0205448.
- Avni R., Zhao R., Pearce S., Jun Y., Uauy C., Tabbita F., Fahima T., Slade A., Dubcovsky J., Distelfeld A. Functional characterization of *GPC-1* genes in hexaploid wheat. *Planta*. 2014;239(2):313-324. DOI 10.1007/s00425-013-1977-y.
- Brevis J.C., Morris C.F., Manthey F., Dubcovsky J. Effect of the grain protein content locus *Gpc-B1* on bread and pasta quality. *J. Cereal Sci.* 2010;51:357-365. DOI 10.1016/j.jcs.2010.02.004.
- Carter A.H., Santra D.K., Kidwell K.K. Assessment of the effects of the *Gpc-B1* allele on senescence rate, grain protein concentration and mineral content in hard red spring wheat (*Triticum aestivum* L.) from the Pacific Northwest region of the USA. *Plant Breed.* 2012; 131(1):62-68. DOI 10.1111/j.1439-0523.2011.01900.x.
- Chen X., Song G., Zhang S., Li Yu., Gao J., Islam S., Ma W., Li G., Ji W. The allelic distribution and variation analysis of the *NAM-B1* gene in Chinese wheat cultivars. *J. Integr. Agric.* 2017;16(6):1294-1303. DOI 10.1016/S2095-3119(16)61459-4.
- Cormier F., Throut M., Ravel C., Le Gouis J., Leveugle M., Lafarge S., Exbrayat F., Duranton N., Praud S. Detection of *NAM-A1* natural variants in bread wheat reveals differences in haplotype distribution between a worldwide core collection and European elite germplasm. *Agronomy*. 2015;5(2):143-151. DOI 10.3390/agronomy5020143.
- GOST 10846-91. Interstate standard. Grain and products of its processing. Method for determination of protein. Moscow: Standartinform Publ., 2009. (in Russian)
- Hagenblad J., Aspland L., Balfourier F., Ravel C., Leino M.W. Strong presence of the high grain protein content allele *NAM-B1* in Fennoscandian wheat. *Theor. Appl. Genet.* 2012;125(8):1677-1686. DOI 10.1007/s00122-012-1943-2.
- Iqbal M., Moakhar N.P., Strenzke K., Haile T., Pozniak C., Hucl P., Spaner D. Genetic improvement in grain yield and other traits of wheat grown in Western Canada. *Crop Sci.* 2016;56(2):613-624. DOI 10.2135/cropsci2015.06.0348.
- Kronenberg L., Yates S., Boer M.P., Kirchgessner N., Walter A., Hund A. Temperature response of wheat affects final height and the timing of stem elongation under field conditions. *J. Exp. Bot.* 2021; 72(2):700-717. DOI 10.1093/jxb/era471.
- Kuhn J.C., Stubbs T.L., Carter A.H. Effect of the *Gpc-B1* allele in hard red winter wheat in the US Pacific Northwest. *Crop Sci.* 2016;56(3): 1009-1017. DOI 10.2135/cropsci2015.08.0470.
- Kumar A., Kapoor P., Chunduri V., Sharma S., Garg M. Potential of *Aegilops* sp. for improvement of grain processing and nutritional quality in wheat (*Triticum aestivum*). *Front. Plant Sci.* 2019;10:308. DOI 10.3389/fpls.2019.00308.
- Lawlor D.W., Kontturi M., Young A.T. Photosynthesis by flag leaves of wheat in relation to protein, ribulose biphosphate carboxylase activity and nitrogen supply. *J. Exp. Bot.* 1989;40(1):43-52. DOI 10.1093/jxb/40.1.43.
- Mamontova L.I., Khromov S.P. Meteorological Dictionary. Leningrad: Gidrometeoizdat Publ., 1974. (in Russian)
- Maphosa L., Collins N.C., Taylor J., Mather D.E. Post-anthesis heat and a *Gpc-B1* introgression have similar but non-additive effects in bread wheat. *Funct. Plant Biol.* 2014;41(9):1002-1008. DOI 10.1071/FP14060.
- Mishra V.K., Pushpendra K.G., Balasubramaniam A., Ramesh C., Neeraj K.V., Manish K.V., Punam S., Arun K.J. Introgression of a

- gene for high grain protein content (*Gpc-B1*) into two leading cultivars of wheat in Eastern Gangetic Plains of India through marker assisted backcross breeding. *J. Plant Breed. Crop Sci.* 2015;7(8): 292-300. DOI 10.5897/JPBCS2015.0514.
- Orlovskaya O., Dubovets N., Solovey L., Leonova I. Molecular cytological analysis of alien introgressions in common wheat lines derived from the cross of *Triticum aestivum* with *T. kiharae*. *BMC Plant Biol.* 2020;20(Suppl. 1):201. DOI 10.1186/s12870-020-02407-2.
- Orlovskaya O.A., Leonova I.N., Adonina I.G., Salina E.A., Khotyleva L.V., Shumny V.K. Molecular cytogenetic analysis of triticale and wheat lines with introgressions of the genetic material of *Triticaceae* tribe species. *Russ. J. Genet. Appl. Res.* 2016;6(5):527-536. DOI 10.1134/S2079059716050087.
- Peleg Z., Saranga Y., Yazici A., Fahima T., Ozturk L., Cakmak I. Grain zinc, iron and protein concentrations and zinc-efficiency in wild emmer wheat under contrasting irrigation regimes. *Plant Soil.* 2008; 306:57-67. DOI 10.1007/s11104-007-9417-z.
- Puranik S., Sahu P.P., Srivastava P.S., Prasad M. NAC proteins: regulation and role in stress tolerance. *Trends Plant Sci.* 2012;17(6):369-381. DOI 10.1016/j.tplants.2012.02.004.
- Rokitsky P.F. Biological Statistics. Moscow: Vysshaya Shkola Publ., 1973. (in Russian)
- Tabbitta F., Lewis S., Vouilloz J.P., Ortega M.A., Kade M., Abbate P.E., Barneix A.J. Effects of the *Gpc-B1* locus on high grain protein content introgressed into Argentinean wheat germplasm. *Plant Breed.* 2013;132(1):48-52. DOI 10.1111/pbr.12011.
- Tabbitta F., Pearce S., Barneix A.J. Breeding for increased grain protein and micronutrient content in wheat: ten years of the *GPC-B1* gene. *J. Cereal Sci.* 2017;73:183-191. DOI 10.1016/j.jcs.2017.01.003.
- Uauy C., Brevis J.C., Dubcovsky J. The high grain protein content gene *Gpc-B1* accelerates senescence and has pleiotropic effects on protein content in wheat. *J. Exp. Bot.* 2006a;57(11):2785-2794. DOI 10.1093/jxb/erl047.
- Uauy C., Distelfeld A., Fahima T., Blechl A., Dubkovsky J. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science.* 2006b;314(5803):1298-1301. DOI 10.1126/science.1133649.
- Vishwakarma M.K., Arun B., Mishra V.K., Yadav P.S., Kumar H., Joshi A.K. Marker-assisted improvement of grain protein content and grain weight in Indian bread wheat. *Euphytica.* 2016;208(2): 313-321. DOI 10.1007/s10681-015-1598-6.
- Wang X., Below F.E. Cytokinins in enhanced growth and tillering of wheat induced by mixed nitrogen source. *Crop Sci.* 1996;36(1): 121-126. DOI 10.2135/CROPSCI1996.0011183X003600010022x.
- Waters B.M., Uauy C., Dubcovsky J., Grusak M.A. Wheat (*Triticum aestivum*) NAM proteins regulate the translocation of iron, zinc, and nitrogen compounds from vegetative tissues to grain. *J. Exp. Bot.* 2009;60(15):4263-4274. DOI 10.1093/jxb/erp257.
- Yang R., Juhasz A., Zhang Y., Chen X., Zhang Y., She M., Zhang J., Maddern R., Edwards I., Diepeveen D., Islam S., Ma W. Molecular characterisation of the *NAM-1* genes in bread wheat in Australia. *Crop Pasture Sci.* 2018;69(12):1173-1181. DOI 10.1071/CP18273.
- Zhao D., Derkx A.P., Liu D.C., Buchner P., Hawkesford M.J. Overexpression of a NAC transcription factor delays leaf senescence and increases grain nitrogen concentration in wheat. *Plant Biol.* 2015; 17(4):904-913. DOI 10.1111/plb.12296.

ORCID ID

O.A. Orlovskaya orcid.org/0000-0002-1187-1317
S.I. Vakula orcid.org/0000-0002-2242-7107
L.V. Khotyleva orcid.org/0000-0003-0295-5022
A.V. Kilchevsky orcid.org/0000-0002-0175-9786

Acknowledgements. The work was carried out under the support of project 2.1.2. of the State Program for Fundamental Research "Biotechnology-2."

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

Received March 24, 2022. Revised January 26, 2023. Accepted January 30, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Morphogenetic peculiarities of reproductive biology in sugar beet (*Beta vulgaris* L.) breeding

T.P. Zhuzhzhhalova , A.A. Nalbandyan, E.N. Vasilchenko , N.N. Cherkasova

The A.L. Mazlumov All-Russian Research Institute of Sugar Beet and Sugar, vil. VNISS, Ramonsky district, Voronezh region, Russia
 biotechnologiya@mail.ru; vasilchenko@inbox.ru

Abstract. This review considers the processes of morphogenesis used in the development of propagation methods and the creation of a new starting material for sugar beet. It has been demonstrated that methods of particulation, *in vitro* microcloning and cell breeding that reflect non-sexual forms of plant reproduction increase the effectiveness of breeding experiments. The review describes the *in vitro* culture methods maintaining a tendency in plants for vegetative propagation and stimulating increase in genetic variability of properties when mutagens such as ethyl methanesulfonate, alien genetic structures with *mf2* and *mf3* bacterial genes in *Agrobacterium tumefaciens* strains, and selective agents (Cd^{++} ions and abscisic acid) are incorporated into plant cells. It presents the results of using fluorescent microscopy, cytophotometry, biochemical analysis and determining the level of phytohormones and content of nucleic acids in nuclei for forecasting the seed setting ability. It has demonstrated that long self-pollination of plants causes decrease in fertility of pollen grains, resulting in the sterilization of male gametes and the appearance of pistillody flowers. Self-fertile plants isolated from these lines serve as sterility fixers, while the apomixis elements increased the ovule number, additional embryo sacs and embryos. A role of apomixis in contributing to variability in the onto- and phylogenetic development of plants has been substantiated. The review reflects the morphological features of the *in vitro* development of sexual and somatic cells in embryos during the formation of seedlings based on floral and vegetative embryoidogeny. Use of the SNP and SSR (Unigenes) molecular-genetic markers having a high polymorphism level has appeared effective to characterize the developed breeding material and hybrid components when carrying out crossings. The study of sugar beet starting materials for the presence of TRs mini-satellite loci making it possible to reveal O-type plants-pollinators (sterility fixing agent) and MS-form plants are of interest for breeding as well. The selected material can be widely used in breeding to produce hybrids, allowing for a 2–3-fold reduction of the development period. The review also discusses the prospects for the development and implementation of new methods and original schemes in sugar beet genetics, biotechnology and breeding.

Key words: sugar beet; reproduction; morphogenesis; embryoidogeny; molecular markers; self-incompatibility; cytoplasmic male sterility (CMS); apomixis.

For citation: Zhuzhzhhalova T.P., Nalbandyan A.A., Vasilchenko E.N., Cherkasova N.N. Morphogenetic peculiarities of reproductive biology in sugar beet (*Beta vulgaris* L.) breeding. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(3):207-217. DOI 10.18699/VJGB-23-27

Морфогенетические особенности репродуктивной биологии в селекции сахарной свеклы (*Beta vulgaris* L.)

Т.П. Жужжалова , А.А. Налбандян, Е.Н. Васильченко , Н.Н. Черкасова

Всероссийский научно-исследовательский институт сахарной свеклы и сахара им. А.Л. Мазлумова, пос. ВНИИСС, Рамонский район, Воронежская область, Россия

 biotechnologiya@mail.ru; vasilchenko@inbox.ru

Аннотация. В обзоре рассмотрены процессы морфогенеза, задействованные при разработке методов размножения и создания нового исходного материала сахарной свеклы. Показано, что методы партикуляции, микроклонирования *in vitro* и клеточной селекции, отражая неполовые формы репродукции растений, повышают результативность селекционных экспериментов. Приведены данные применения методов культуры *in vitro*, сохраняющих склонность растений к вегетативному размножению и стимулирующих повышение генетической изменчивости при использовании мутагенов этилметансульфоната, методов генетической трансформации (бактериальные гены *mf2* и *mf3*) и селективных агентов (ионы Cd^{++} и абсцизовая кислота). Отражены результаты применения методов оценки степени самонесовместимости растений: флуоресцентной микроскопии, цитофотометрии, и определения уровня фитогормонов и содержания нуклеиновых кислот в ядрах клеток для прогнозирования завязываемости семян. Выявлено, что длительное самоопыление растений вызывает снижение фертильности пыльцевых зерен, стерилизацию мужских гамет и образование

пистиллоидных цветков. Самофертильные растения, выделенные из этих линий, служат закрепителями стерильности. В обзоре продемонстрирована роль апомиксиса в реализации изменчивости при онто- и филогенетическом развитии растений. Отражены морфологические особенности развития половых и соматических клеток зародышей *in vitro* при формировании проростков на основе флоральной и вегетативной эмбриодогении. Применение молекулярно-генетических маркеров SNP и SSR, обладающих высоким уровнем полиморфизма, оказалось эффективным для характеристики создаваемого селекционного материала гибридов при проведении скрещиваний. Интерес для селекции представляют исследования исходных материалов сахарной свеклы на наличие минисателлитных локусов TRs, позволившие выявлять растения-опылители O-типа (закрепитель стерильности) и растения MC (мужскостерильные) формы. Созданный материал может найти широкое использование в селекционной работе при создании гибридных форм, сокращая сроки селекции в 2–3 раза. Обсуждаются перспективы разработки и возможности применения новых методов и оригинальных схем в генетике, биотехнологии и селекции сахарной свеклы.

Ключевые слова: сахарная свекла; репродукция; морфогенез; эмбриодогения; молекулярные маркеры; самонесовместимость; цитоплазматическая мужская стерильность (ЦМС); апомиксис.

Introduction

The first studies into the reproductive properties of sugar beet date back to the early XX century, when the systematic characterization of the species was carried out and its evolutionary paths described. It was also when the investigation of embryological patterns and morphological properties of plant cells, organs, and tissues started. The researchers began to focus on the developmental nuances of sugar beet flowers, anthers, and pollen grains. Studies into anatomical and embryological development patterns of ovules, embryos, and seeds were initiated, handbooks of anatomical charts for *Beta vulgaris* representing the key morphological traits of the species were prepared. Later, more in-depth studies focused on formation of flowers and their elements, reduction division stages, pollen development stages, embryo sac development, as well as embryo and seed formation. These efforts made possible the large-scale research work on genetics and breeding of sugar beet, laid foundation for genetic breeding programs, and eventually stimulated the creation of the first multigerm sugar beet cultivars (Mazlumov, 1970).

The late XX century marked a transition from population breeding to hybrid breeding for cytoplasmic male sterility (CMS), monogermity, and increased ploidy level. This major shift was facilitated by the results of studies into karyotype structure of the genus *Beta*, mitotic peculiarities, development of generative organs, as well as fertilization and embryogenesis in development of polyploids and self-pollinated lines. These findings showed feasibility of heterosis breeding for cultivation of hybrids with high crop yields and sugar contents (Balkov et al., 2017). Cultivation of high-yielding genotypes with valuable traits was also driven by the studies of reproductive processes ensuring genetic variability based on morphological, cytological, and biochemical traits (Simko et al., 2012; Stevanato et al., 2015).

Reproductive biology studies the development of plant organisms at molecular and subcellular, cellular, organismic, and population levels to determine how their traits are propagated to the next generation (Maletskiy, 2005). This is why sugar beet breeding material is analyzed using a wide range of methods intended for various reproduction types, including genetic markers (Banzal et al., 2014), cytology, embryology, physiology, and morphology (Paesold et al., 2012; Tomaszewska-Sowa et al., 2017). The studies are designed taking into account the nuances of plant reproduction having to do with multiple variants of procreation (sexual, asexual, and apomictic), various morphogenetic pathways (embryogenesis, embryoidogenesis, and hemmorhizogenesis), and reproduction types (seed or vegetative) (Batygina, Vinogradova, 2007). This approach makes it possible to predict the progeny genotype and shows new opportunities for controlling specific ontogenetic stages and designing new breeding programs (Maletskaya, Maletskiy, 2010; Bartenev et al., 2018).

Genetic marker methods appear to be the most common techniques in plant breeding today. The high occurrence rate of molecular markers in the genome, studied using the universal and continuously evolving analytical methods makes the use of DNA markers viable for breeding purposes (Khlestkina, 2013). They make it possible to not only accurately and quickly identify genetic variability in sugar beet populations, lines, and hybrids, but also determine the breeding value of the experimentally obtained plant material (Nalbandyan et al., 2020). Molecular markers ISSR (Inter-Simple Sequence Repeats), SSR (Simple Sequence Repeats), and SNP (Single Nucleotide Polymorphism) are considered the most efficient.

The goal of the present paper is to review the data and methods (both proprietary and available in the literature) used for characterization of the morphogenetic potential of sugar beet at various development stages.

Reproduction of vegetative plant tissues

Particulation. A common way to fix the most valuable traits and attributes of sugar beet breeding material is to use the plant's capability of vegetative reproduction, which is due to the morphological and anatomical peculiarities of the roots, stems, and shoot buds, as well as to their ontogenetic transformation.

Growth and development processes in plants after seed germination include root and leaf system formation. At this stage, the crown, i. e. the upper part of the root, represents a reduced dome-shaped stem with a rosette leaf arrangement. The apical bud is located at the center of the rosette, and the axillary buds formed by organogenesis at the base of leaves.

A successful vegetative reproduction in winter is ensured by preliminary vernalization promoting epigenetic changes in the gene expression driving the post-vernalization development of plants (Shevelukha, 1992; Lutova et al., 2010). The observed activation of meristematic tissues (phellogen and cambium) stimulating the development of apical and axillary buds and then central and lateral shoots ensures rapid reproduction of the breeding material through traumatic particulation (Barykina et al., 2004). Here, the selected super-elite roots (pedigree) were divided into 4–8 groups, with one group cloned, and the rest grown for seeds.

The main shortcomings of the method were the significant time it took to achieve material uniformity (up to 5 years), the limited number of reproduced clones, and high labor intensity. At the same time, this approach made it possible to preserve the genotypes with high crop yields and sugar contents relatively pure for several generations and thus found its use in breeding schemes based on individual selection combined with cloning and hybridization (Balkov et al., 2017).

The reproduction process was later intensified by cloning the plants using rooted lateral shoots (with 3–8 leaves) developing from axillary buds. The method ensured a significant increase in number of rooted shoots and grown plants, but gained no traction in sugar beet breeding.

The particulation ability of sugar beet is a backup reproduction mechanism, which combined with seed reproduction ensures the organism's flexibility, survival, and overall preservation of the *B. vulgaris* species.

Microcloning. Biotechnology methods for mass vegetative reproduction *in vitro* appear to show the most promise in breeding (Murashige, 1974). These methods rest upon the ability of somatic cells to recover the organism's continuity with all its traits while cultivation (Atanasov, 1993).

It was found that a plant can develop from vegetative cells in two ways: either formed from callus tissue through regeneration of buds and embryoids or from apical meristems, axillary and dormant buds as a result of activation processes (Katayeva, Butenko, 1983; Rusea et al., 2020). However, long-term subculturing of the callus led to changes in ploidy, accumulation of genetic mutations, and loss of morphological potential. Shoot meristems, laminae, petioles, anthers, ovules, embryos, and seeds are the most common explants for micropropagation (Leyke et al., 1980). Hormonal activation of this material leads to the development of shoot buds and seedlings (Ilyenko, 1983; Seman, Farago, 1988).

The most convenient way to refine sugar beet microcloning process is through the use of apical meristem due to its cells' high genetic stability and high morphogenetic ability (Znamenskaya, 2010). Apical meristem is responsible for creation of the above-ground organs of plants, developing in a modular fashion (a shoot element includes a leaf, an internode, and an axillary bud), and meristematic stem cells can differentiate and stimulate the development of new organs (Lutova et al., 2010). Another linchpin of the method is totipotency, a unique property of plant cells allowing them to recreate a plant organism with all of its traits, when isolated from a parent organism and affected by exogenous factors of a culture medium (Shevelukha, 1992).

Cultivation of the starting material for explantation under cover, which made the contamination of explants 1.5 times as low, was a necessary condition for microcloning refinement (Znamenskaya, Zhuzhzhhalova, 1989). The next mandatory stage of the cultivation process was using the optimized sterilization protocol for apical explant meristems that were soaked in 0.05 % corrosive sublimate solution for 1–3 hours.

Maintaining the optimal hormonal balance of 0.2–0.3 mg/l of gibberellin, 6-benzylaminopurine (BAP), and kinetin (1:1:1) in the culture medium stimulated morphogenesis. Root formation was induced by adding 1 mg/l of indole butyric and indole acetic acids into the medium. With these contents, the number of lateral shoots per explant increased to 9–10, microclone height – to 73.4 mm, and root formation – to 74.3 %. This method made it possible to obtain an unlimited number of clones from one donor by halving the breeding cycle duration and thus circumventing the biennial life cycle of sugar beet (Znamenskaya, Zhuzhzhhalova, 1989). The totipotency allowed the progeny of the derived lines to preserve genotype, phenotype, ploidy, monogermity, and CMS fixing ability in the field.

To ensure the long-term use of the obtained strain material in breeding practice, the three-year cycle of breeding, biotechnological, and agronomic techniques is used (Kolesnikova, Zhuzhzhhalova, 2018). At the first stage, the selected hybrid genotypes are transferred to an *in vitro* culture and then transplanted to the gene bank of elite clones to ensure the long-term preservation of valuable traits of the material. The second stage includes micropropagation of elite regenerants, rooting, and transplanting into greenhouse soil substrate to form stecklings (small roots) of approximately 100 g in size. At the third stage, the plants are grown for elite seeds. The seeds of a simple hybrid and a heterosis pollinator are harvested separately. This approach has contributed significantly to the development of new-generation high-yielding hybrids as a result of breeding and seed farming efforts at the A.L. Mazlumov All-Russian Research Institute of Sugar Beet and Sugar (Ramon, Voronezh Region) and is applied commercially.

Cell breeding. Breeding at the cellular level includes development of the methods for cultivation of isolated plant cells, organs, and tissues to increase hereditary variability. Through that, the methods involving the introduction of various chemical and physical factors, such as mutagenic substances, genetic constructs with alien genes, selective agents, etc., into culture media stimulate variability in plant organisms and increase their stability against environmental stresses (Dubrovna, 2017; Kourelis, van der Hoorn, 2018).

The experiments showed that exposure to 2 to 6 mM ethyl methanesulphonate mutagen (EMS) for 30 minutes has a stimulating effect on sugar beet petioles (Hohmann et al., 2005; Vasilchenko et al., 2020b). The diploid material demonstrated high regeneration ability of 38.1–53.1 %, and in haploids it reached 71.3–87.5 %. This method also ensured genetic variability of regenerants by increasing ploidy level of haploids to that of diploids and polyploids. Molecular genetic analysis using the specific SSR marker mSSCIR 47 identified 9 lines, where exposure to mutagen caused genotypic changes. Cultivation of the identified lines under cover made it possible to preserve up to 92 % of microclones and obtain small roots 30 to 89 g in size. (Vasilchenko et al., 2020a). The plant material will be further studied at the A.L. Mazlumov All-Russian Research Institute of Sugar Beet and Sugar using the TILLING method (Targeting Induced Local Lesions in Genomes) developed by C.M. Mc Callum et al. (2000).

The use of genetic constructs *Agrobacterium tumefaciens* with bacterial genes *mf2* (*Bacillus thuringiensis*) and *mf3* (*Pseudomonas fluorescens*) inducing non-specific resistance against fungal and viral pathogen also proved to be effective (Dzhavakhia et al., 2005;

Vasilchenko et al., 2009). Cocultivation of sugar beet tissues and *A. tumefaciens* with alien genes *mf2* and *mf3* produced Fusarium rot resistant GM crops, whose transgenic nature was confirmed by molecular and biochemical analysis, as well as by phytopathological assessment under cover (Khussein et al., 2013; Vasilchenko et al., 2014). The total of seven new sugar beet lines with *mf2* gene and eight lines with *mf3* gene have been derived so far.

The effect of heavy metal ions, particularly Cd^{2+} , on the cultivated seedlings was noteworthy (Cherkasova et al., 2020). Several samplings with cadmium acetate ($Cd(CH_3CO_2)_2$) content of 4 mM showed high adaptability and survivability of regenerants up to 74.8 %, which indicated their high resistance. Molecular genetic analysis of these microclones showed the presence of significant SNPs in *MTP4*, which presumably played a major part in development of resistance against this particular abiotic stress (Nalbandyan et al., 2019; Khussein et al., 2021). The obtained results have allowed researchers at the A.L. Mazlumov All-Russian Research Institute of Sugar Beet and Sugar create and test isogenic lines of sugar beet showing promise in terms of heavy metal resistance.

It is known that abscisic acid (ABA) acts as a stress hormone in plants, causing stomatal conductance and transpiration reduction as a drought resistance mechanism (Lutova et al., 2010). ABA controls abiotic stresses in plants (salinity, drought, and temperature) by putting plant tissues (cells, buds, and leaves) into resting state via stomatal closure. It is also possible that the role of this hormone is to trigger the resistance mechanism, which then develops on its own regardless of the hormone's content (Titov et al., 2007). This is why the investigation of ABA properties *in vitro* may facilitate the development of drought-resistant breeding material. The optimal ABA content (2.0–3.0 mg/l) ensuring the ratio of drought-resistant sugar beet regenerants at 70.0–87.5 % has been found in (Cherkasova et al., 2018, 2021). Thus, abscisic acid can be used to simulate droughts, when developing the sugar beet strains tolerant to osmotic stresses.

Improving the adaptive properties of regenerants in the reproductive cycle by increasing their genetic variability makes it possible to obtain new starting material and quickly reproduce it.

Reproductive ability of generative organs

Self- and cross-incompatibility

Sugar beet is a monoecious hermaphroditic species characterized by cross-pollination and the self-incompatibility preventing self-fertilization. Most plants in

sugar beet populations of mixed origin (65.3–78.1 %) are self-incompatible. These plants do not set seeds in isolation after 1–2 inbreedings. The occurrence rate of self-compatible (self-fertile) plants in varietal populations is 4.3–12.9 %, given the pollen grain fertility of up to 83.6–98.2 %. Self-fertile strains are typically characterized by higher crop yields and sugar contents, but some evidence shows that inbreeding depression is not completely suppressed.

Experiments showed that in some strains the self-incompatibility of individual plants was barely affected by environmental changes in cultivation conditions (Kornienko, Znamenskaya, 2001). For instance, plants cultivated in different environmental areas only showed minor differences in seed setting, with the average value of 7.4 % in Ramon (Voronezh Region) and 6.3 % in Przhhevsk (Kyrgyzstan). At the same time, a number of authors demonstrated that seed setting in self-pollinating plants at lower temperatures was significantly higher than that at higher temperatures. This discovery formed the basis for the seed reproduction method at low temperatures used to create a vast collection of inbred sugar beet strains (Maletskii, Maletskaya, 1996). The discrepancies in the results of various studies indicate the complexity of seed reproduction processes in sugar beet. Genetic and physiological mechanisms underlying these processes require additional research and remain relevant in the context of sexual reproduction.

According to the relevant genetic data, the incompatibility in sugar beet is a rather complex gametophytic system based on the complementary interaction of alleles 2–4 of S-loci (Larsen, 1977). These data were taken into consideration in production of strain material by self-pollination with circumvention of physiological self- and cross-incompatibility barriers for successful transition from population breeding to interstrain heterotic hybrid breeding.

The primary internal factor affecting the seed setting under inbreeding is genetic and hormonal regulation of the fertilization process. The seed setting in self-compatible strains remained at the self-pollination level or decreased slightly. Intrastrain crosses in partially compatible strains increased the seed setting by a multiple of 1.5–13. Selection of individual plants from strains of interest differing in pollen recognition by the stigma and pollination ability turned out to be a promising technique as well. This selection has made it possible to produce strain material with prevalent traits of the female or male parent to increase the effectiveness of close (intrastrain) crosses (Kornienko et al., 2014).

To determine the degree of incompatibility and investigate the pollen tube growth, fluorescence microscopy modified for sugar beet studies was used (Zaykovskaya, Zhuzhzhlova, 1976; Vaisman et al., 1984). The method was applied to identify potential seed productivity of plants as early as the flowering stage, and thus made it possible to determine the seed setting and assign pairs while of breeding material hybridization based on pollen tube growth activity. The incompatibility response localization manifested itself clearly in the morphological changes in pollen tubes, thickenings, bulgings, inhibited growth activity at the stigma or in the calcium oxalate deposition area in the ovary. Presumably, Ca^{++} ions play a major part in pollen tube growth, since it is one of the first pollen germination regulators (Bednarska, 1989). No other element facilitating germination in absence of Ca^{++} was found.

Based on period of self-incompatibility gene activity and localization of gene products, sporophytic and gametophytic incompatibility types were defined. From the morphological perspective, sporophytic responses are localized at the stigma, and gametophytic ones – in the style. The current concept is that the self-incompatibility response localization in sugar beet matches that in the species with gametophytically controlled self-incompatibility (Vishnyakova, 1998).

Nucleic acid content in ovary and ovule cell nuclei observed by cytophotometry was another indication of incompatibility gene activity in sugar beet (Zhuzhzhlova et al., 2007). Compatible pollination was usually reflected in these cells by increase in nuclear RNA content and total nuclear nucleic acid content. The RNA/DNA ratio under incompatible pollination remained the same as in non-pollinated pistils. A significant increase in RNA content of up to 36 % was recorded after the third day of delayed pollination. It is possible that genetically determined incompatibility mechanism manifesting itself at this stage in DNA molecules indicates functional changes in ovule nuclei associated with fertilization process, as in other plants (Kovaleva, 1991). Thus, we may assume that fertilization stimulates transcription and reduplication processes facilitating increased RNA contents and total nuclear nucleic acid contents. Weakening of the incompatibility barrier correlates with RNA accumulation in cell nuclei. The data obtained also show that genetic and physiological incompatibility properties are currently understudied and still remain a relevant issue.

The further studies showed that pollen tube growth processes were also accompanied by complex hormonal rearrangement in the stigma and style. This is confirmed

by the experiments determining contents of endogenous phytohormones, i. e. ethylene, IAA, ABA, gibberellins, and cytokines, in sporophytic tissues of the pistil (Kovaleva et al., 2016). This is why the inner workings of the plant reproduction system, and especially the role of self- and cross-incompatibility, will still be relevant research issues in the near future.

The use of this trait was a reasonable way to overcome the incompatibility barrier in order to advance the research of heterosis breeding, including controlled crosses and production of interstrain hybrids using the fertile base.

Cytoplasmic male sterility

Normal (N) and sterile (S) are two types of cytoplasm found in sugar beet, and their effects are regulated by two recessive genes x and z (Owen, 1952). Complete male sterility is designated as $Sxxzz$. The presence of dominant genes in the heterozygous state produces semi-sterile forms $SXxzz$ and $SXxZz$. However, the causes of CMS in sugar beet have not been studied completely and still require more in-depth research (Dymshits et al., 2010).

In the context of breeding, an CMS trait is fixed and maintained in the selected sugar beet strains by inbreeding with an CMS maintainer of high pollen fertility and seed setting (Balkov et al., 2017). Peculiarities of genetic variability in breeding material are identified using DNA markers. Micro-satellite analysis (SSR) made it possible to identify polymorphisms of MS strains, multigerm pollinators (MP) of sugar beet, and their hybrids (Nalbandyan et al., 2021). This analysis was facilitated by determination of polymorphism content (PIC) for each locus. The use of Unigenes SSR markers (9 primer pairs) made it possible to clearly differentiate the material based on calculation of genetic distances and identify parent components best suited for crosses. Primers with the highest SSR polymorphism such as Unigene 22373, Unigene 27833, Unigene 16898, Unigene 7492, Unigene 24552 were recommended for identification of sugar beet breeding materials (Nalbandyan et al., 2021).

The studies into presence of CMS-associated mini-satellite locus TRs in the starting material were of interest in terms of sugar beet breeding (Nishizava et al., 2000; Bragin et al., 2011). Experimental screening of the sugar beet breeding material based on mini-satellite loci TR1 and TR3 showed the presence of N- and S-type mitochondrial genomes (Fedulova et al., 2022). MS-plants included amplicons with lengths of 400 bp. O-type pollinator plants (sterility maintainers) were

characterized by DNA amplicon lengths of 700 bp. for TR1 primer and 500 bp for TR3. The selected material makes the breeding cycle 2–3 times as short and may therefore become widely used in hybrid breeding.

It was observed in breeding research that male sterility occurring in self-fertile strains after multiple inbreedings manifests itself in formation of pistillate flowers (Oshevnev et al., 1986a, b). Stamens and anthers in these flowers are replaced with pistillate structures, i. e. stigmas or ovules without embryo sacs. Phenotypic ratio determined for the progeny of strain material in process of breeding showed that pistillate trait was controlled monogenically (Oshevnev, Gribanova, 2010). Strains with sterile anthers were used as female parents for interstrain sugar beet hybrids. Fertile plants with high self-compatibility (self-fertility) selected from this material were used as pollen sterility maintainers in hybrid breeding. The scheme for producing O-type strains based on the pistillate trait was the finishing stage of a consistent transition to breeding for heterosis (Oshevnev, Gribanova, 2010).

The results obtained have shown that sterilization of male gametophyte may follow different scenarios, but eventually results in gynodioecy, i. e. the occurrence of plants with hermaphrodite and female flowers in the population. The further investigation of this phenomenon is of great significance for genetic research.

Apomixis (agamospermy)

The use of apomixis (agamospermy) is among the most significant trends in plant breeding. This reproduction mode with no gamete fusion in some cases ensures maternal inheritance, uniformity of progeny, and high yields in an unlimited line of generations. Thus, the analysis of morphogenetic processes and investigation of embryological development nuances in agamosperous sugar beet strains have a potential to reveal new breeding applications of apomictic reproduction.

Genetic polymorphism of plant reproduction modes with no gamete fusion is typically represented by two main types of apomixis known as **gametophytic** and **sporophytic**, and by **embryoidogeny**, i. e. somatic embryo formation in the ovule or on vegetative organs.

Gametophytic apomixis. Cytological studies showed that sugar beet plants tend to form embryos with haploid set of chromosomes by apogamy from embryo sac nuclei, i. e. synergids or antipodals. However gametophytic apomixis in sugar beet manifested itself in the form of diplospory (Fomenko et al., 2003), which could possibly be explained by the use of breeding material obtained by artificial pollination with pollen exposed

to high doses of gamma radiation ranging from 1500 to 3000 Gy (Agafonov et al., 1992). Cytological study of the plants obtained from this material demonstrated that meiosis in megasporogenesis did not produce megaspore tetrads. As a result, embryo sac cells with a non-reduced number of chromosomes formed directly from the megasporocyte, which commonly causes unstable genetic changes. At the same time, the method made it possible to create several sugar beet gamma-strains with signs of facultative apomixis and CMS fixation ability. The hybrids obtained using the apomictic strain γ -MC-2113 showed crop yields of up to 122.2 % and sugar contents of up to 104.5 % compared to the control (Bogomolov, 2010). This plant material is currently used in breeding.

Sporophytic apomixis. This apogamy type is based on embryo development from somatic cells of a female parent with a diploid set of chromosomes taken from nucellus, integuments, or endosperm of the female parent or daughter embryo cells created as a result of fertilization (Batygina, 2010). The experiments in sugar beet found multiple instances of budding or ovule splitting into two or more new ones. As a result, several embryos with different inheritance, i. e. a sexual embryo and embryos formed from nucellar or integumentary somatic cells, developed in seeds. The embryos developing from endosperm nuclei were typically observed in the chalazal end of the embryo sac in process of creating polyploid or aneuploid sugar beet strains (Yarmolyuk et al., 1990).

The discovered embryo development peculiarities have shown that coexistence of sexual and asexual modes of seed reproduction leads to significant differences in quality of seed material (Zaykovskaya, Pereyat'ko, 1977). The genetic diversity of strain material in these cases causes some issues for hybrid breeding.

Embryoidogeny. This plant reproduction type represents a special category of vegetative reproduction with *embryoid* (somatic embryo) as its structural unit. Similarly to sexual embryos, it is an initial stage of a new organism, rather than its part (bud, leaf, or root) as observed in cutting propagation by gemmorrhizogenesis (Batygina, 2010). Floral and vegetative embryoidogeny types are defined based on origin and location of somatic embryos on the parent plant.

Floral embryoidogeny in sugar beet represents embryo development from haploid nuclei of the embryo sac (egg cell, synergids, and antipodals) and diploid tissue cells of the ovule (nucellus, integuments, and endosperm). Embryological studies into *in vitro* cultivation of non-fertilized ovules made it possible to identify

the formation stages of haploid embryos (Podvigina, 2010; Tomaszewska-Sowa et al., 2017). It was found that division of haploid cells of the egg apparatus or antipodals produced a multicellular proembryo similar in its development to a sexual embryo. This technique made it possible to obtain genetically and morphologically diverse material with high homozygosity from donor plants (5–6 times as fast). The present studies made it possible to accelerate the creation of homozygous lines, which were then widely used in sugar beet breeding (Batygina, 2010; Kikindonov et al., 2016; Lamaoni et al., 2018; Pazuki et al., 2018).

Vegetative embryoidogeny is based on embryo development from somatic cells of vegetative organs. *In vitro* cultivation of sugar beet petioles on the Murashige–Skoog medium supplemented with 6-BAP (0.2–0.3 mg/l) and 2.4-D (0.1 mg/l) stimulated morphological development of seedlings from epidermal cells of the petiole similar to formation stages of gametes in embryo sac. At first, formation of one or several initial cells similar to embryo sac gametes was observed. These cells had denser cytoplasm and a large nucleus with a nucleolus. Then, transverse fission of the initial cell occurred forming the future root cells. Subsequent divisions produced embryo-like structures, i. e. embryoids, transforming into seedlings (Bogomolova, Zhuzhzhlova, 1998). The data obtained agree with the findings of foreign authors, who also observed the development of structural elements similar to hypocotyl with leaves from epidermal cells at the adaxial surface of the cultivated sugar beet petioles (Mahmoud et al., 2017).

Parthenogenesis. This apomictic reproduction type implies embryo development from an egg cell not involving a male gamete or from sperm not involving an egg cell (Maletskiy, 2005). According to the relevant concepts, parthenogenetic development of a sugar beet embryo is attributed to epigenetic inheritance and variability. It is associated with the internal or external signals received by ovule cells of flowering plants and making them change their development program. Investigation of sugar beet strains with defected pollen (CMS) in the field showed that their seed reproduction rate was on par with the biparental mode or outperformed it. Parthenogenetic seed progeny demonstrated high occurrence rate of seeds with haploid sets of chromosomes. Haploids amounted to 3–10 % of the total germinated seeds, which is four orders higher than normal and matches the occurrence rate of haploids in the culture medium *in vitro* (Maletskiy et al., 2015). At the same time, it should be noted that

obtaining haploids in parthenogenetic progeny is 3–4 orders cheaper than *in vitro*.

The further research of parthenogenesis may show new opportunities in commercial breeding due to the significant simplification and cost-effectiveness of MS-hybrid breeding processes. Thus, we may assume that this way of obtaining haploids may become one of the most efficient sugar beet breeding techniques. Initiation of various reproduction modes in sugar beet plants is ensured by potential capabilities of the reproductive system of the species, which makes it possible to maintain species and population homeostasis in general.

Conclusion

The wide variety of ways to develop new traits and attributes in plants is the key feature of reproductive biology techniques for sugar beet. The pregenerative development stage turned out to be best suited for vegetative development and production of a new breeding material capable of maintaining valuable traits using particulation and micropropagation. These methods are backup reproduction mechanisms, which, combined with seed reproduction, ensure the organism's flexibility and make it possible to preserve the genotype, phenotype, ploidy level, and other valuable traits in the field.

Cell breeding methods produce the material with altered ploidy that is resistant against environmental stresses. Morphogenetic reproduction techniques producing strain material with modified traits and facilitating the transition from population to hybrid breeding may be considered the best studied ones. A number of breeding processes and mechanisms, such as self- and cross-incompatibility, CMS, and apomixis have contributed significantly to the development of various trends in reproduction systems. Experimental studies based around apogamy, diplospory, and parthenogenesis have demonstrated new opportunities in commercial breeding and biotechnology. However, the agamospermy methods remain insufficiently studied to be applied in commercial breeding.

The studies into apomictic reproduction of sugar beet have shown the most promise in plants with CMS due to the fact that inheritance of individual traits in hybrid progeny is driven by gene transfer from female parents. Cytoplasmic male sterility is accompanied by the presence of two types of intracellular organelles in cytoplasm (mitochondria and chloroplasts with their own genetic material in the form of mtDNA and cpDNA). Cytological and molecular genetic studies of sugar beet with CMS are of interest, because genotypic and phenotypic polymorphism can be identified in aga-

mospermy progeny. This will open new opportunities for applying innovative technologies to produce homozygous sugar beet material with new traits for breeding purposes.

References

- Agafonov N.S., Bogomolov M.A., Zhuzhzhhalova T.P., Kornienko A.V., Fedulova T.P., Popova I.R. Method for obtaining sugar beet homozygous lines. Inventor's certificate 1708210 USSR. MKI AOLH 1/04. No. 4818227/13. Applied 12.03.90. Published 30.01.92. Bulletin No. 4. 1992. (in Russian)
- Atanasov A. Biotechnology in Plant Industry. Novosibirsk, 1993. (in Russian)
- Balkov I.Ya., Karakotov S.D., Logvinov A.V., Logvinov V.A., Mishchenko V.N. Evolution of Sugar Beet: from Garden Forms to Modern Profitable Hybrids. Shchelkovo, 2017. (in Russian)
- Banzal K.C., Lenka S.K., Mondal T.K. Genomic resources for breeding crops with enhanced abiotic stress tolerance. *Plant Breed.* 2014; 133(1):1-11. DOI 10.1111/pbr.12117.
- Bartenev I.I., Gavrin D.S., Nechaeva O.M., Senyutin A.A. Heterogeneity of seed plant populations and quality indicators of sugar beet seeds. *Sakhar = Sugar.* 2018;10:46-49. (in Russian)
- Barykina R.P., Veselova T.D., Devyatov A.G., Dzhaliyeva Kh.Kh., Ilyina G.M., Chubatova N.V. Handbook of Botanical Microprocedures: Grounds and Methods. Moscow: Moscow State University Publ., 2004. (in Russian)
- Batygina T.B. Nontraditional views of reproduction types and methods. Phenomenon of embryoidogeny, a new way of flowering plant propagation. In: Developmental Biology: Morphogenesis of Reproductive Structures and the Role of Somatic and Stem Cells in Development and Evolution: Proc. Intern. Conf., Dedicated to the 50th Anniversary of the Laboratory of Embryology and Reproductive Biology. St. Petersburg, December 13-16, 2010. Moscow, 2010;26-31. (in Russian)
- Batygina T.B., Vinogradova G.Yu. Phenomenon of polyembryony. Genetic heterogeneity of seeds. *Russ. J. Dev. Biol.* 2007; 38:126-151. DOI 10.1134/S1062360407030022.
- Bednarska E. The effect of intracellular calcium level regulators on the synthesis of pollen tube callose in *Oenothera biennis* L. *Acte Soc. Bot. Pol.* 1989;58(2):199-210. DOI 10.5586/asbp.1989.016.
- Bogomolov M.A. Induced apomixis and its use in sugar beet breeding. In: Encyclopedia of the Beta Genus: Biology, Genetics, and Breeding of Beet. Novosibirsk: Sova Publ., 2010; 504-513. (in Russian)
- Bogomolova N.M., Zhuzhzhhalova T.P. Study of somatic embryoidogenesis in sugar beet (*B. vulgaris* L.). In: Issues of Botanical Sciences at the Turn of the 21st Century. V. 1. St. Petersburg: Botanical Institute RAS Publ., 1998;103. (in Russian)
- Bragin A.G., Ivanov M.K., Fedoseeva L.A., Dymshits G.M. Analysis of mitochondrial DNA heteroplasmy in fertile and owen CMS sugar beet (*Beta vulgaris*) plants. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding.* 2011;15(3): 524-530. (in Russian)

- Cherkasova N.N., Zhuzhzhhalova T.P., Kolesnikova E.O. Development of a technology to select sugar beet *in vitro* regenerants for tolerance of acidity and drought. *Sakhar = Sugar*. 2018;10:43-45. (in Russian)
- Cherkasova N.N., Zhuzhzhhalova T.P., Kolesnikova E.O. Influence of ionic toxicity on morphological development of sugar beet regenerants. *Sakharnaya Svekla = Sugar Beet*. 2020;5:33-35. DOI 10.25802/SB.2020.94.56.003. (in Russian)
- Cherkasova N.N., Zhuzhzhhalova T.P., Tkachenko O.V. Increasing the resistance of sugar beet plants to drought *in vitro* conditions. *Sakharnaya Svekla = Sugar Beet*. 2021;8:12-14. DOI 10.25802/SB.2021. 52.25.002. (in Russian)
- Dubrovna O.V. *In vitro* selection of wheat for resistance to abiotic stress factors. *Fiziologiya Rastenij i Genetika = Plant Physiology and Genetics*. 2017;49(4):279-292. (in Russian)
- Dymshits E.M., Bragin A.G., Ivanov M.K. Molecular aspects of cytoplasmic male sterility. In: Encyclopedia of the *Beta* Genus. Biology, Genetics, and Breeding of Beet. Novosibirsk: Sova Publ., 2010;228-247. (in Russian)
- Dzhavakhia V., Filippov A., Skryabin K., Voinova T., Kouznetsova M., Shulga O., Shumilina D., Kromina K., Pridaniko M., Battchikova N., Korpela T. Proteins inducing multiple resistance of plants to phytopathogens and pests. US patent No. WO 2005061533. Intern. Filing Date 17.12.2004. Publ. Date 07.07.2005.
- Fedulova T.P., Nalbandyan A.A., Duvanova T.N. Screening of sugar beet parental materials for the presence of TRs minisatellite loci associated with CMS. *Sakhar = Sugar*. 2022;3:38-41. DOI 10.24412/ 2413-5518-2022-3-38-41. (in Russian)
- Fomenko N.P., Zhuzhzhhalova T.P., Fedulova T.P., Bogomolov M.A. Cytoembryological and morphological characterization of plants in apomictic gamma lines of sugar beet. In: Factors of the Experimental Evolution of Organisms. Kiev: Agrarnaya Nauka Publ., 2003; 212-217. (in Russian)
- Hohmann U., Jacobs G., Jung C. An EMC mutagenesis protocol for sugar beet and isolation non-bolting mutants. *Plant Breed*. 2005;124(4):317-321. DOI 10.1111/j.1439-0523.2005.01126.x.
- Ilyenko I.I. Microclonal propagation and maintainance of sugar beet breeding material in *in vitro* culture. *Fiziologiya i Biokhimiya Kulturnykh Rasteniy = Physiology and Biochemistry of Cultivated Plants*. 1983;15(4):351-355. (in Russian)
- Karakotov S.D., Apasov I.V., Nalbandyan A.A., Vasilchenko E.N., Fedulova T.P. Modern issues of sugar beet (*Beta vulgaris* L.) hybrid breeding. *Vavilivskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2021; 25(4):394-400. DOI 10.18699/VJ2.043. (in Russian)
- Katayeva N.V., Butenko R.G. Clonal Micropropagation of Plants. Moscow: Nauka Publ., 1983. (in Russian)
- Khlestkina E.K. Molecular markers in genetic studies and breeding. *Vavilivskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):1044-1054. (in Russian)
- Khussein A.S., Mikheeva N.R., Nalbandyan A.A., Cherkasova N.N. Screening of sugar beet regenerants for the heavy metal resistance MTP4 gene. *Biotehnologiya = Biotechnology*. 2021;37(4):14-19. DOI 10.21519/0234-2758-2021-37-4-14-19. (in Russian)
- Khussein A.S., Nalbandyan A.A., Bogacheva N.N., Vasilchenko E.N., Fedulova T.P. Microsatellite analysis of transgenic forms of sugar beet. *Doklady Rossiyskoy Akademii Sel'skokhozyaystvennykh Nauk = Proceedings of the Russian Academy of Agricultural Sciences*. 2013;2:17-20. (in Russian)
- Kikindonov G., Kikindonov Tz., Erchev S. Economical qualities of crosses between doubled haploid sugar beet lines. *Agric. Sci. Technol*. 2016;8(2):107-110. DOI 10.15547/ast 2016.02.018.
- Kolesnikova E.O., Zhuzhzhhalova T.P. Microcloning and maintenance of lines material in sugar beet breeding process. In: Proc. of IV (XII) Int. Botanical Conf. of Young Scientists in St. Petersburg, April 22–28, 2018. St. Petersburg: Komarov Botanical Institute of the RAS, 2018;267-268.
- Kornienko A.V., Podvigina O.A., Zhuzhzhhalova T.P., Fedulova T.P., Bogomolov M.A., Oshevnev V.P., Butorina A.K. High-priority research directions in genetics and the breeding of the sugar beet (*Beta vulgaris* L.) in the 21st century. *Russ. J. Genet*. 2014;50(11):1137-1148. DOI 10.1134/S1022795414110064.
- Kornienko A.V., Znamenskaya V.V. New Methods for Obtaining Sugar Beet Parent Material. Ramon', 2001. (in Russian)
- Kourelis J., van der Hoorn R.A.L. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell*. 2018;30(2):285-299. DOI 10.1105/tpc.17.00579.
- Kovaleva L.V. Intercellular interactions of the pollen-pistil system in the progamic phase of fertilization. *Uspekhi Sovremennoj Biologii = Advances in Current Biology*. 1991;111(5):782-796. (in Russian)
- Kovaleva L.V., Zakharova E.V., Voronkov A.S., Timofeyeva G.V., Minkina Yu.V. Plant hormones and polar growth of pollen tubes. In: Embryology, Genetics, and Biotechnology. Proceedings of the 5th International School for Young Scientists (St. Petersburg, October 9-14, 2016). St. Petersburg: Levsha Publ., 2016;107.
- Lamaoui M., Jemo M., Datla R., Bekkaoui F. Heat and drought stresses in crops and approaches for their mitigation. *Front. Chem*. 2018;6: 26. DOI 10.3389/fchem.2018.00026.
- Larsen K. Self-incompatibility in *Beta vulgaris* L. Four gametophytic, complementary S-loci in sugar beet. *Hereditas*. 1977; 85(2):227-248. DOI 10.1111/j.1601-5223.1977.tb00971.x.
- Leyke G., Labes R., Ertel K., Peteresdorf M. Using of Tissue and Organ Cultures in Plant Breeding and Planting Material Production. Moscow: Kolos Publ., 1980. (in Russian)
- Lutova L.A., Ezhova T.A., Dodueva I.E., Osipova M.A. Genetics of Plant Development. St. Petersburg: N-L Publ., 2010. (in Russian)
- Mahmoud K., Najar A., Jedid E., Jemai N., Jemmali A. Tissue culture techniques for clonal propagation, viral sanitation and germplasm improvement in strawberry (*Fragaria* × *ananas-sa* Duch.). *J. New Sci*. 2017;47(2):2564-2576.
- Maletskaya E.I., Maletskiy S.I. Haploidy in apozygotic progeny of sugar beet. In: Encyclopedia of the *Beta* Genus: Biology, Genetics, and Breeding of Beet. Novosibirsk: Sova Publ., 2010;466-472. (in Russian)
- Maletskiy S.I. Evolution Biology. Dictionary of Terms. Novosibirsk: ICG SB RAS Publ., 2005. (in Russian)

- Maletskii S.I., Maletskaya E.I. Self-fertility and agamospermy in sugar beet (*Beta vulgaris* L.). *Russ. J. Genet.* 1996;32(12): 1431-1437.
- Maletskiy S.I., Maletskaya E.I., Yudanov S.S. New technology of seeds reproduction in sugar beets (parthenogenetic mode). *Trudy Kubanskogo Gosudarstvennogo Agrarnogo Universiteta = Works of the Kuban State Agrarian University.* 2015;3(54):204-213. (in Russian)
- Mazlumov A.L. Sugar Beet Breeding. Moscow: Kolos Publ., 1970. (in Russian)
- McCallum C.M., Comai L., Greene E.A., Henikoff S. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* 2000;123(2):439-442. DOI 10.1104/pp.123.2.439.
- Murashige T. Plant propagation through tissue cultures. *Ann. Rev. Plant Physiol.* 1974;25:135-166. DOI 10.1146/annurev.pp.25.060174.001031.
- Nalbandyan A.A., Fedulova T.P., Cherepukhina I.V., Kryukova T.I., Oshevnev V.P., Gribova N.P. DNA markers in sugar beet breeding. *Sakharnaya Svekla = Sugar Beet.* 2021;2:10-14. DOI 10.25802/SB.2021.57.87.001. (in Russian)
- Nalbandyan A.A., Fedulova T.P., Khussein A.S. Molecular selection of sugar beet breeding material with genes of resistance to biotic stresses. *Rossiyskaya Sel'skokhozyaystvennaya Nauka = Russian Agricultural Science.* 2019;1:16-20. DOI 10.31857/S2500-2627 2019116-20. (in Russian)
- Nalbandyan A.A., Khussein A.S., Fedulova T.P., Cherepukhina I.V., Kryukova T.I., Rudenko T.S., Mikheeva N.R., Moiseenko A.V. Differentiation of sugar beet cultivars by SSR markers to create promising hybrids. *Rossiyskaya Sel'skokhozyaystvennaya Nauka = Russian Agricultural Science.* 2020; 4:18-21. DOI 10.31857/S2500262720040043. (in Russian)
- Nishizawa S., Kubo T., Mikami T. Variable number of tandem repeat loci in the mitochondrial genomes of beets. *Curr. Genet.* 2000; 37(1):34-38. DOI 10.1007/s002940050005.
- Oshevnev V.P., Cherepukhin E.I., Balkov I.Ya. Use of the self-incompatibility trait in sugar beet breeding on the CMS base. *Sel'skokhozyaystvennaya Biologiya = Agricultural Biology.* 1986a;21(8):90-92. (in Russian)
- Oshevnev V.P., Gribova N.P. Breeding of O-type self-compatible pollinators in sugar beet. In: Encyclopedia of the *Beta* Genus: Biology, Genetics, and Breeding of Beet. Novosibirsk: Sova Publ., 2010;542-554. (in Russian)
- Oshevnev V.P., Gribova N.P., Zhuzhzhhalova T.P., Cherepukhin E.I. Formation of pistillodes in sugar beet. *Genetika = Genetics.* 1986b;22(4):889-891. (in Russian)
- Owen F.V. Mendelian male sterility in sugar beets. *Am. Soc. Sugar Beet Technol.* 1952;7:371-376.
- Paesold S., Borchardt D., Schmidt T., Dechtyeva D. A sugar beet (*Beta vulgaris* L.) reference FISH karyotype for chromosome and chromosome-arm identification, integration of genetic linkage groups and analysis of major repeat family distribution. *Plant J.* 2012;72(4):600-611. DOI 10.1111/j.1365-3113.2012.05102.x.
- Pazuki A., Aflaki F., Gürel E., Ergül A., Gürel S. Gynogenesis induction in sugar beet (*Beta vulgaris*) improved by 6-benzylaminopurine (BAP) and synergized with cold pretreatment. *Sugar Tech.* 2018;20:69-77. DOI 10.1007/s12355-017-0522-x.
- Podvigina O.A. *In vitro* induction of haploidy from unfertilized sugar beet ovules. In: Encyclopedia of the *Beta* Genus: Biology, Genetics, and Breeding of Beet. Novosibirsk: Sova Publ., 2010;455-465. (in Russian)
- Rusea I., Popescu A., Valentina I., Hoza D. Micropropagation of strawberry cv. Magic. *Ann. Univ. Craiova.* 2020;XXIV(LX): 218-223.
- Seman I., Farago J. Research biotechnological methods in sugar beet breeding in the research and breeding institute at Bucany. *Potsdam. Forsh. B.* 1988;57:51-56.
- Shevelukha V.S. Plant Growth and its Regulation in Ontogenesis. Moscow: Nauka Publ., 1992. (in Russian)
- Simko I., Eujayl I., van Hintum T.J. Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. *Plant Sci.* 2012;184: 54-62. DOI 10.1016/j.plantsci.2011.12.009.
- Stevanato P., Trebbi D., Panella L., Richardson K., Droccanello C., Pakish L., Fenwick A., Saccomani M. Identification and validation of a SNP marker linked to the gene *HsBvm-1* for nematode resistance in sugar beet. *Plant Mol. Biol. Rep.* 2015;33:474-479. DOI 10.1007/s11105-014-0763-8.
- Titov A.F., Talanova V.V., Kaznina N.M., Laydinen G.F. Heavy metal tolerance in plants. Petrozavodsk: Karelian Scientific Centre Publ., 2007. (in Russian)
- Tomaszevska-Sowa M., Figas A.S., Gatz A. Histological analysis of organogenesis and somatic embryogenesis during shoot formation in sugar beet (*B. vulgaris* L.) via gynogenesis. *Polish J. Nat. Sci.* 2017;32(4):705-717.
- Vaisman N.Ya., Zhuzhzhhalova T.P., Agafonov N.S. Cytology of incompatibility in sugar beet. In: Sugar Beet Genetics. Novosibirsk: Nauka Publ., 1984;121-129. (in Russian)
- Vasilchenko E.N., Zhuzhzhhalova T.P., Fedorin D.N., Dzhavakhiya V.G. Use of the *MF2* gene for transformation of sugar beet plants. In: Modern Immunology Studies and Their Role in Development of New Varieties and Intensification of Plant Growing. Bolshie Vyazemy, 2009;162-168. (in Russian)
- Vasilchenko E.N., Zhuzhzhhalova T.P., Fedorin D.N., Zemlyanukhina O.A. Molecular and biochemical properties of transgenic sugar beet plants. In: Proceedings of the 5th All-Russian Symposium "Transgenic Plants: Development Technologies, Biochemical Properties, Use, and Biosafety". Moscow. December 1-4, 2014. Moscow, 2014;63-66. (in Russian)
- Vasilchenko E.N., Zhuzhzhhalova T.P., Kolesnikova E.O. Rapid raise of new sugar beet (*B. vulgaris* L.) homozygous lines. *Sakhar = Shugar.* 2020a;2:30-32. DOI 10.24411/2413-5518-2020-10203. (in Russian)
- Vasilchenko E.N., Zhuzhzhhalova T.P., Tkachenko O.V. Changes in morphological characteristics of sugar beet under the action of ethylmethanesulphonate. *Sakharnaya Svekla = Sugar Beet.* 2020b;8:2-6. DOI 10.25802/SB.2020.52.38.001. (in Russian)
- Vishnyakova M.A. Possible ways of structural mechanisms underlying the evolution of self-incompatibility reaction in angiosperm plants. In: Issues of Botanical Sciences at the Turn of the 21st Century. St. Petersburg: Botanical Institute RAS Publ., 1998;107-108. (in Russian)
- Yarmolyuk G.I., Shiryayeva E.I., Kulik A.G. Endospermal embryony, a new form of apomixis in sugar beet. *Doklady*

- VASKHNIL = Reports of the Academy of Agricultural Sciences*. 1990;3-4:8-11. (in Russian)
- Zaykovskaya N.E., Peretyat'ko N.A. Apomictic development of sugar beet seeds. In: *Fundamentals and Practical Methods of the Growing of Sugar Beet and Other Crops*. Kiev: All-Ukraine Institute of Breeding, 1977;19-21. (in Russian)
- Zaykovskaya N.E., Zhuzhhalova T.P. Development of pollen tubes in isolated self-fertile and self-sterile sugar beet lines. *Tsitologiya i Genetika = Cytology and Genetics*. 1976; 10(1):57-61. (in Russian)
- Zhuzhhalova T.P., Znamenskaya V.V., Podvigina O.A., Yarmolyuk G.I. *Reproduction Biology of Sugar Beet*. Voronezh: Sotrudnichestvo Publ., 2007. (in Russian)
- Znamenskaya V.V. *In vitro* microcloning as a method of maintenance and propagation of sugar beet lines. In: *Encyclopedia of the Beta Genus: Biology, Genetics, and Breeding of Beet*. Novosibirsk: Sova Publ., 2010;420-437. (in Russian)
- Znamenskaya V.V., Zhuzhhalova T.P. Use of microclonal propagation and maintenance of valuable sugar beet genotypes to accelerate breeding. In: *Use of Biotechnological Methods in Sugar Beet Breeding*. Kiev, 1989;52-58. (in Russian)

ORCID ID

A.A. Nalbandyan orcid.org/0000-0001-5959-047X

Acknowledgements. The work was carried out within the framework of state assignment of the Ministry of Science and Higher Education of the Russian Federation.

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

Received January 2, 2022. Revised February 2, 2023. Accepted February 2, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Evaluating the role of selection in the evolution of mitochondrial genomes of aboriginal peoples of Siberia

B.A. Malyarchuk , M.V. Derenko

Institute of Biological Problems of the North of the Far Eastern Branch of the Russian Academy of Sciences, Magadan, Russia

 malyarchuk@ibpn.ru

Abstract. Studies of the nature of mitochondrial DNA (mtDNA) variability in human populations have shown that protein-coding genes are under negative (purifying) selection, since their mutation spectra are characterized by a pronounced predominance of synonymous substitutions over non-synonymous ones ($Ka/Ks < 1$). Meanwhile, a number of studies have shown that the adaptation of populations to various environmental conditions may be accompanied by a relaxation of negative selection in some mtDNA genes. For example, it was previously found that in Arctic populations, negative selection is relaxed in the mitochondrial *ATP6* gene, which encodes one of the subunits of ATP synthase. In this work, we performed a Ka/Ks analysis of mitochondrial genes in large samples of three regional population groups in Eurasia: Siberia ($N = 803$), Western Asia/Transcaucasia ($N = 753$), and Eastern Europe ($N = 707$). The main goal of this work is to search for traces of adaptive evolution in the mtDNA genes of aboriginal peoples of Siberia represented by populations of the north (Koryaks, Evens) and the south of Siberia and the adjacent territory of Northeast China (Buryats, Barghuts, Khamnigans). Using standard Ka/Ks analysis, it was found that all mtDNA genes in all studied regional population groups are subject to negative selection. The highest Ka/Ks values in different regional samples were found in almost the same set of genes encoding subunits of ATP synthase (*ATP6*, *ATP8*), NADH dehydrogenase complex (*ND1*, *ND2*, *ND3*), and cytochrome *b_c* complex (*CYB*). The highest Ka/Ks value, indicating a relaxation of negative selection, was found in the *ATP6* gene in the Siberian group. The results of the analysis performed using the FUBAR method (HyPhy software package) and aimed at searching for mtDNA codons under the influence of selection also showed the predominance of negative selection over positive selection in all population groups. In Siberian populations, nucleotide sites that are under positive selection and associated with mtDNA haplogroups were registered not in the north (which is expected under the assumption of adaptive evolution of mtDNA), but in the south of Siberia.

Key words: mitochondrial DNA; natural selection; Ka/Ks -testing; human populations; Siberia.

For citation: Malyarchuk B.A., Derenko M.V. Evaluating the role of selection in the evolution of mitochondrial genomes of aboriginal peoples of Siberia. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(3): 218-223. DOI 10.18699/VJGB-23-28

Оценка роли отбора в эволюции митохондриальных геномов коренного населения Сибири

Б.А. Малярчук , М.В. Деренко

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия

 malyarchuk@ibpn.ru

Аннотация. Исследования характера изменчивости митохондриальной ДНК (мтДНК) в популяциях человека выявили, что белок-кодирующие гены находятся под действием отрицательного (очищающего) отбора, поскольку мутационные спектры генов мтДНК характеризуются выраженным преобладанием синонимичных замен над несинонимичными (величина параметра $Ka/Ks < 1$). Между тем в ряде исследований показано, что адаптация популяций к различным условиям природной среды может сопровождаться ослаблением отрицательного отбора в некоторых генах мтДНК. Так, ранее было установлено, что в арктических популяциях отрицательный отбор ослаблен в митохондриальном гене *ATP6*, кодирующем одну из субъединиц АТФ-синтазы. В настоящей работе проведен Ka/Ks -анализ митохондриальных генов в больших выборках трех региональных групп населения Евразии: Сибири ($N = 803$), Западной Азии/Закавказья ($N = 753$) и Восточной Европы ($N = 707$). Основная цель работы – поиск следов адаптивной эволюции в генах мтДНК коренного населения Сибири, представленного населением севера (коряки, эвены) и юга Сибири и прилегающей территории Северо-Восточного Китая (буряты, баргуты, хамнигане). С помощью стандартного Ka/Ks -анализа установлено, что все гены мтДНК во всех изученных региональных группах населения испытывают действие отрицательного отбора. Наиболее высокие значения Ka/Ks в различных региональных выборках обнаружены практически в одном и том же наборе генов, кодирующих субъединицы АТФ-синтазы (*ATP6*, *ATP8*), НАДН-дегидрогеназного комплекса (*ND1*, *ND2*, *ND3*) и ци-

тохром *bc₁*-комплекса (*CYB*). Самое высокое значение Ka/Ks , указывающее на ослабление отрицательного отбора, выявлено в гене *ATP6* в сибирской группе. Результаты анализа, выполненного с помощью метода FUBAR (пакет программ HyPhy) и направленного на поиск кодонов мтДНК, находящихся под действием отбора, также показали преобладание влияния отрицательного отбора над положительным отбором во всех группах населения. В сибирских популяциях нуклеотидные позиции, находящиеся под действием положительного отбора и ассоциированные с гаплогруппами мтДНК, зарегистрированы не на севере (что ожидается в предположении адаптивной эволюции мтДНК), а на юге Сибири.

Ключевые слова: митохондриальная ДНК; естественный отбор; Ka/Ks -тесты; популяции человека; Сибирь.

Introduction

Mitochondrial DNA (mtDNA) is a valuable tool for studying the evolutionary history of humans, which is associated with such features of the mitochondrial genome as maternal inheritance without recombinations and a high mutation rate compared to the nuclear genome (Brown et al., 1979; Giles et al., 1980). The gradual accumulation of mutations in mtDNA haplotypes leads to the formation of groups of phylogenetically related haplotypes (i. e., mtDNA haplogroups), which are characterized by a population-specific distribution (Wallace, 1995). At the beginning of studies, continental macrohaplogroups were discovered, and later, as the resolution of mtDNA analysis increased – from sequencing of certain mtDNA regions to sequencing of complete mitogenomes – ethnospecific mtDNA haplogroups were identified (Olivieri et al., 2017; Derenko et al., 2019; García et al., 2020).

The results of population genetic studies of the last 20 years point to the great importance of negative (purifying) selection in human mitochondrial genome evolution (Mishmar et al., 2003; Elson et al., 2004; Kivisild et al., 2006; Ingman, Gyllensten, 2007; Sun et al., 2007; Derenko, Malyarchuk, 2010; Eltsov et al., 2010; Malyarchuk, 2011; Litvinov et al., 2020). This is primarily due to significance of this genetic system, which ensures the effective functioning of the mitochondrial respiratory chain. Genes encoding subunits of protein complexes of the respiratory chain (NADH-ubiquinone-oxidoreductase, cytochrome *bc₁*, cytochrome *c*-oxidase, ATP-synthase) make up about 70 % of the mitochondrial genome. The high stability of these genes is due to the significant prevalence of synonymous substitutions in various mtDNA genes (Ks) over non-synonymous ones (Ka), leading to amino acid substitutions.

Meanwhile, early studies have shown that mitochondrial genes may be subject to positive selection, leading to the prevalence of non-synonymous substitutions over synonymous ones, due to the adaptation of human populations to various natural environmental conditions (Mishmar et al., 2003). Thus, a deviation from the neutral model of mtDNA variability was found in different population groups of Eurasia and America. An analysis of the distribution of Ka/Ks values has shown that negative selection is relaxed in the mitochondrial *ATP6* gene in the Arctic zone, in the *CYB* gene in the temperate zone (Europe), and in the *COI* and *ND3* genes in the tropics (Mishmar et al., 2003; Ingman, Gyllensten, 2007).

The prevalence of elevated Ka/Ks values in the *ATP6* gene in the Arctic zone (in Siberian and North American populations) in comparison with other regions was explained by the adaptation of populations to the Far North environmental conditions (Mishmar et al., 2003). The *ATP6* gene encodes

subunit 6 of mitochondrial ATP synthase, which is involved in the coupling of ATP production and heat to maintain body temperature, and therefore it is suggested that polymorphic variants that reduce coupling efficiency may be beneficial under cold stress conditions, as they increase heat production and overall metabolic rate (Mishmar et al., 2003).

Later, it was also shown that higher Ka/Ks values in the *ATP6* gene prevail in East Asians (Elson et al., 2004; Sun et al., 2007). In another study of the mitochondrial genomes of the North Asian populations, the highest Ka/Ks values were also found in the *ATP6* gene (Ingman, Gyllensten, 2007). The accumulation of non-synonymous mutations in this gene was considered by the authors as an evolutionarily slow process of gradual relaxation of negative selection over many thousands of years. This scenario is also supported by evidence that some ancient non-synonymous substitutions were defining mtDNA haplogroups that have become widespread in northern Asia (Ruiz-Pesini et al., 2004). For example, there is the G8584A substitution, which defines the M8 macrohaplogroup and its predominantly North Asian haplogroup C, the C8794T substitution, which determines the haplogroup A, and the A8701G substitution, which delineates the macrohaplogroup N as a whole. It is assumed that such mtDNA replacements are associated with changes in energy metabolism and, thus, contributed to adaptation to northern conditions, being potential candidates for adaptive selection (Ruiz-Pesini et al., 2004).

Another scenario, as noted above, is that mutational changes in the *ATP6* gene occurred as a result of relaxed negative selection, and the increase in the frequency of these polymorphic variants in northern populations was facilitated by genetic drift, the effects of which are better manifested in populations with a small effective size (Ingman, Gyllensten, 2007).

The results of Ka/Ks analysis of mitochondrial genes in cancer tissues demonstrated a significant relaxation of negative selection in many mtDNA genes under conditions of aerobic glycolysis, which actively occurs in cancer cells (Stafford, Chen-Quin, 2010; Liu et al., 2012; Skonieczna et al., 2018). When comparing the Ka/Ks spectra in healthy and cancerous tissues, it turned out that in the mitochondrial genes of affected cells in various types of cancer, a statistically significant relaxation of negative selection is observed in all genes, except for *ATP6* and *ATP8*, as well as *ND3* and *CO2* (Liu et al., 2012). With respect to the genes encoding subunits of ATP synthase, this means that the mitochondria of healthy cells are likely to be characterized by such a significant relaxation of negative selection that it practically does not differ from that under conditions of carcinogenesis. However, the reasons for such behavior of the mitochondrial *ATP6* and *ATP8* genes in the norm are not fully understood.

Thus, the results of studies of the evolution of protein-coding mtDNA genes in human populations testify to a rather high stability of mitochondrial genes; however, the differences that were revealed between populations indicate a possible influence of positive selection on some mtDNA genes due to the adaptation of populations to different climatic conditions. In a number of studies, this issue was considered based on populations of East Asia, including the aboriginal populations of Siberia, but the sample sizes of populations studied were insufficiently representative (less than 100 complete mitogenomes) (Mishmar et al., 2003; Elson et al., 2004; Kivisild et al., 2006; Ingman, Gyllensten, 2007; Sun et al., 2007).

In this paper, we present more detailed information on the effect of selection on the mitochondrial genomes of human populations based on the results of Ka/Ks analysis of mtDNA genes in aboriginal populations of Siberia ($N = 803$) in comparison with populations of Western Asia and Transcaucasia ($N = 753$) and Eastern Europe ($N = 707$).

Materials and methods

Whole mtDNA genome data from Siberian and East Asian populations published in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) were analyzed. The data are represented by the Koryaks ($N = 154$) and Evens ($N = 219$) from the northern part of Siberia, and by the Mongolic-speaking Buryats, Barghuts and Khamnigans ($N = 430$) from the southern part of Siberia and adjacent territories of Northeast China. For comparison, we used data on the whole mitogenome variability in populations of Western Asia (Persians, Qashqais, Lebanese) and Transcaucasia (Armenians and Azeri) ($N = 753$), as well as in populations of Eastern Europe (Russians, Ukrainians, Volga Tatars and Estonians) ($N = 707$).

We analyzed the distribution of Ka/Ks values (the ratio of the number of non-synonymous substitutions for a non-synonymous site (Ka) to the number of synonymous substitutions for a synonymous site (Ks)) in the mtDNA L-strand encoded genes *ND1*, *ND2*, *CO1*, *CO2*, *ATP8*, *ATP6*, *CO3*, *ND3*, *ND4L*, *ND4*, *ND5* and *CYB*. For Ka/Ks analysis, we used the programs of the package DnaSP v. 5 (Librado, Rozas, 2009). The effect of negative selection is assumed at $Ka/Ks < 1$ and positive selection at $Ka/Ks > 1$. To analyze the effect of selection on mtDNA protein-coding genes, the HyPhy software package was also used (<http://www.hyphy.org>) (Kosakovskiy Pond et al., 2005). To identify codons under the influence of negative and positive selection, the FUBAR method (Fast Unconstrained Bayesian AppRoximation) was used. This method allows you to quickly analyze large sets of molecular data using the hierarchical Bayesian method and the Monte Carlo method for Markov chains (MCMC) (Murrel et al., 2013).

Results and discussion

The results of the analysis of the distribution of Ka/Ks values in the protein-coding genes of the mitochondrial genome in aboriginal populations of Siberia demonstrate that in all but one of the cases, the values of this parameter are below 1, which indicates the effect of negative selection on mtDNA genes (Table 1). The highest Ka/Ks values were detected in the *ATP6* gene. Moreover, among the Koryaks, a Paleo-Asiatic people that originated in Northeast Asia under extreme

environmental conditions, the Ka/Ks value exceeds 1, which indicates the effect of positive selection on this mitochondrial gene.

Table 2 shows Ka/Ks values in three regional population groups. In the Siberian group of populations, the highest values were found in the *ATP6*, *ATP8*, *ND2*, and *CYB* genes; in the populations of Western Asia and Transcaucasia – in the *ATP6*, *ATP8*, *ND1*, *ND2*, and *CYB* genes; in the populations of Eastern Europe – in the *ATP6*, *ATP8*, *ND1*, *ND3*, and *CYB* genes. Therefore, the results of analysis indicate that in different regions of Eurasia the highest Ka/Ks values are found in about the same sets of mitochondrial genes. The maximum values of this parameter were revealed in the genes encoding subunits of ATP synthase, which is consistent with the results of previous studies (Mishmar et al., 2003; Ingman, Gyllensten, 2007; Sun et al., 2007) and points to a relaxation of negative selection in the *ATP6* and *ATP8* genes, especially in Siberian populations.

To assess selective pressure acting on individual mtDNA sites (with taking into account their location in the phylogenetic tree of mtDNA haplotypes), we used hierarchical Bayesian analysis implemented in the FUBAR program of the HyPhy package (<http://www.hyphy.org>). This method has a higher efficiency of detecting codons that are under the influence of positive and negative selection – for example, in comparison with the FEL (Fixed Effects Likelihood) and MEME (Mixed Effects Model of Evolution) methods of the HyPhy package, which are also widely used to study selective processes (Murrel et al., 2012, 2013).

Our study demonstrated that in Siberian populations, 11.4% (411) codons, which are roughly evenly distributed over the mtDNA genes, are under the influence of negative selection. The effect of positive selection was found only in 4 codons of the *ND5* and *CYB* genes (Table 3).

When only the populations of the northeastern part of Siberia (Koryaks and Evens) are analyzed, another codon, which is characterized by a borderline posterior probability value of 0.9, is revealed in the *ND4* gene (see Table 3). In this case, the nucleotide substitution in the *ND4* gene, leading to the N390S amino acid substitution, determines the C5a2 phylogenetic cluster, which is interesting in that it is distributed mainly among the Koryaks.

All other substitutions are found in mtDNA clusters characterized by a more southern distribution – they are associated either with major mtDNA haplogroups widespread in East and South Asia (for example, N9a and F1a'c'f) or in Western Eurasia (for example, H11, K, J1c), or with relatively small East Asian mtDNA haplogroups found also in Buryats, Barghuts, and Khamnigans (D4j1a, D4g2a1) (see Table 3).

Thus, oddly enough, despite the expected effect of positive selection on individual mtDNA sites due to the adaptation of aboriginal populations of Northeastern Siberia to a cold climate, the effect of positive selection was found only in southernmost Siberian populations. Similar conclusions follow from the results of the analysis of mtDNA protein-coding genes in Siberian populations obtained using other methods, FEL and MEME (results not shown).

For comparison, data sets for populations of Western Asia/Transcaucasia and Eastern Europe were also analyzed (see

Table 1. Ka/Ks values for mtDNA genes in aboriginal populations of Siberia

mtDNA gene	Koryaks (N = 154)	Evens (N = 219)	Mongolic-speaking populations (N = 430)
ND1	0.011	0.075	0.109
ND2	0.217	0.188	0.21
CO1	0.008	0.032	0.032
CO2	0.049	0.004	0.042
ATP8	0.018	0.72	0.617
ATP6	1.33	0.71	0.934
CO3	0.00003	0.00003	0.032
ND3	0.548	0.04	0.16
ND4L	0.00001	0.16	0.13
ND4	0.225	0.28	0.12
ND5	0.184	0.075	0.068
CYB	0.4	0.28	0.146

Table 2. Ka/Ks values for mtDNA genes in regional populations

mtDNA gene	Siberia (N = 803)	Western Asia and Transcaucasia (N = 753)	Eastern Europe (N = 707)
ND1	0.075	0.31	0.33
ND2	0.203	0.263	0.2
CO1	0.028	0.05	0.025
CO2	0.031	0.042	0.053
ATP8	0.56	0.28	0.71
ATP6	0.932	0.397	0.34
CO3	0.022	0.11	0.19
ND3	0.174	0.31	0.38
ND4L	0.115	0.072	0.045
ND4	0.18	0.025	0.028
ND5	0.092	0.022	0.037
CYB	0.209	0.31	0.47

Table 3. mtDNA nucleotide positions affected by positive selection in regional population groups (FUBAR method)

Gene, codon, substitution	Nucleotide position, substitution	α	β	PP ($\alpha < \beta$)	BF ($\alpha < \beta$)	mtDNA haplogroups associated with certain substitution
Populations of Siberia						
ND5, T8A	12358, A > G	2.234	14.049	0.915	31.42	N9a, D4j1a
ND5, I257V	13105, A > G	2.375	16.015	0.92	33.31	D4g2a1
ND5, A475T	13759, G > A	2.155	12.928	0.917	32.21	F1a'c'f, H11
CYB, F18L	14798, T > C	2.429	33.074	0.969	89.47	K, J1c
ND4, N390S	11928, A > G	3.058	22.089	0.9	15.04	C5a2
Populations of Western Asia and Transcaucasia						
ND2, A331T	5460, G > A	1.379	28.771	0.991	467.9	H1e, J1b1, K1a12, W
CO3, F251L	9957, T > C	1.037	7.623	0.907	42.49	N1b1a3
ND4, L89P	11025, T > C	0.97	16.616	0.959	102.8	K1a2, N1b1, N1a3
ND5, F478L	13768, T > C	1.746	11.402	0.907	42.21	U3b
CYB, T7I	14766, C > T	1.034	7.652	0.93	57.4	HV
Populations of Eastern Europe						
ATP6, I121V	8887, A > G	1.198	7.492	0.902	33.35	W1e, R1
ND4, L89P	11025, T > C	1.139	21.786	0.968	111.1	K1a2, N1a3
ND5, I257V	13105, A > G	2.34	13.831	0.902	33.64	R1a1, U4d2
ND5, T534M	13934, C > T	2.172	15.421	0.923	43.35	J1c3, U3a
CYB, L258P	15519, T > C	1.137	21.875	0.969	112.0	H34

Note. α – rate of synonymous substitutions; β – rate of non-synonymous substitutions; PP – posterior probability; BF – Bayes factor. PP for codons, which are under the influence of positive selection, is > 0.9.

Table 3). In the first case, 19.5 % (700) codons were found under the influence of negative selection, in the second case – 16.4 % (589) codons. Under the influence of positive selection, five codons were identified in both regional population groups (see Table 3). All substitutions are associated with mtDNA haplogroups widespread in Western Eurasia, and therefore it is difficult to accept (at least in the absence of a special analysis) that the fixation of these substitutions in haplogroup trunks occurred due to the adaptation of populations to natural environmental conditions. It should be noted that in two cases there is evidence of the influence of positive selection on the same codon in different geographic regions: a nucleotide substitution at position 11025, which determines the haplogroups K1a2 and N1a3, in the populations of Western Asia/Transcaucasia and Eastern Europe, and a nucleotide substitution at position 13105, which defines haplogroup D4g2a1 in Siberian populations and haplogroups R1a1 and U4d2 in populations of Eastern Europe (see Table 3).

Conclusion

Thus, our study aimed at the effects of selection on mtDNA genes in different regional groups of Eurasia using standard Ka/Ks analysis showed that all mtDNA genes are characterized by low values of this parameter ($Ka/Ks < 1$), indicating the influence of negative selection. The highest Ka/Ks values in different regional population groups were found in almost the same set of genes encoding subunits of ATP synthase (*ATP6*, *ATP8*), NADH dehydrogenase complex (*ND1*, *ND2*, *ND3*), and cytochrome *bc₁* complex (*CYB*). The highest value of Ka/Ks, indicating a relaxation of negative selection, was found in the *ATP6* gene in Siberian populations; moreover, in Koryaks, the effect of positive selection on this gene was formally recorded ($Ka/Ks = 1.33$).

Meanwhile, the results of the analysis aimed at searching for mtDNA codons affected by selection showed a multiple prevailing negative selection over positive one in all population groups under study. In Siberian populations, codons affected by positive selection and associated with mtDNA haplogroups have been revealed only in populations of the southern part of Siberia and the adjacent territory of Northeast China (among the Buryats, Barghuts, and Khamnigans). In the regional groups of Eurasian populations, codons of this kind were found in different mtDNA genes (*ND2*, *ND4*, *ND5*, *CO3*, *CYB*), but in the *ATP6* gene a single codon (at position 121) was detected in the East European group of populations rather than in the Siberian one. Apparently, further studies of the direction and strength of natural selection on mitochondrial genomes in different regional population groups of Eurasia are required.

References

Brown W.M., George M.Jr., Wilson A.C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*. 1979;76(4):1967-1971. DOI 10.1073/pnas.76.4.1967.

Derenko M., Denisova G., Malyarchuk B., Hovhannisyan A., Khachatryan Z., Hrechdakian P., Litvinov A., Yepiskoposyan L. Insights into matrilineal genetic structure, differentiation and ancestry of Armenians based on complete mitogenome data. *Mol. Genet. Genom.* 2019;294(6):1547-1559. DOI 10.1007/s00438-019-01596-2.

Derenko M.V., Malyarchuk B.A. Molecular Phylogeography of Populations of Northern Eurasia Based on Mitochondrial DNA Variability data. Magadan: SVNC DVO RAN, 2010. (in Russian)

Elson J.L., Turnbull D.M., Howell N. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am. J. Hum. Genet.* 2004;74(4):229-238. DOI 10.1086/381505.

Eltsov N.P., Volodko N.V., Starikovskaya E.B., Mazunin I.O., Sukernik R.I. The role of natural selection in the evolution of mitochondrial haplogroups in Northeastern Eurasia. *Rus. J. Genet.* 2010; 46(9):1105-1107. DOI 10.1134/S1022795410090243.

García Ó., Alonso S., Huber N., Bodner M., Parson W. Forensically relevant phylogeographic evaluation of mitogenome variation in the Basque Country. *Forensic Sci. Int. Genet.* 2020;46(5):102260. DOI 10.1016/j.fsigen.2020.102260.

Giles R.E., Blanc H., Cann H.M., Wallace D.C. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*. 1980;77(11):6715-6719. DOI 10.1073/pnas.77.11.6715.

Ingman M., Gyllensten U. Rate variation between mitochondrial domains and adaptive evolution in humans. *Hum. Mol. Genet.* 2007; 16(19):2281-2287. DOI 10.1093/hmg/ddm180.

Kivisild T., Shen P., Wall D.P., Do B., Sung R., Davis K., Passarino G., Underhill P.A., Scharfe C., Torroni A., Scozzari R., Modiano D., Coppa A., de Knijff P., Feldman M., Cavalli-Sforza L.L., Oefner P.J. The role of selection in the evolution of human mitochondrial genomes. *Genetics*. 2006;172(1):373-387. DOI 10.1534/genetics.105.043901.

Kosakovsky Pond S.L., Frost S.D.W., Muse S.V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005;21(5):676-679. DOI 10.1093/bioinformatics/bti079.

Librado P., Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25(11):1451-1452. DOI 10.1093/bioinformatics/btp187.

Litvinov A.N., Malyarchuk B.A., Derenko M.V. The nature of the molecular evolution of the mitochondrial genomes of the Russian population of East Europe. *Vestnik Severo-Vostochnogo Nauchnogo Centra DVO RAN = The Bulletin of the North-East Scientific Center*. 2020;2:107-113. DOI 10.34078/1814-0998-2020-2-107-113. (in Russian)

Liu J., Wang L.-D., Sun Y.-B., Li E.-M., Xu L.-Y., Zhang Y.-P., Yao Y.-G., Kong Q.-P. Deciphering the signature of selective constraints on cancerous mitochondrial genome. *Mol. Biol. Evol.* 2012; 29(4):1255-1261. DOI 10.1093/molbev/msr290.

Malyarchuk B.A. Adaptive evolution signals in mitochondrial genes of Europeans. *Biochemistry (Moscow)*. 2011;76(6):702-706. DOI 10.1134/S0006297911060113.

Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S., Brandon M., Easley K., Chen E., Brown M.D., Sukernik R.I., Olckers A., Wallace D.C. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA*. 2003;100(1):171-176. DOI 10.1073/pnas.0136972100.

Murrell B., Moola S., Mabona A., Weighill T., Sheward D., Kosakovsky Pond S.L., Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* 2013;30(5): 1196-1205. DOI 10.1093/molbev/mst030.

Murrell B., Wertheim J.O., Moola S., Weighill T., Scheffler K., Kosakovsky Pond S.L. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012;8(7):e1002764. DOI 10.1371/journal.pgen.1002764.

Olivieri A., Sidore C., Achilli A., Angius A., Posth C., Furtwängler A., Brandini S., Capodiferno M.R., Gandini F., Zoledziewska M., Pitzalis M., Maschio A., Busonero F., Lai L., Skeates R., Gradoli M.G., Beckett J., Marongiu M., Mazzarello V., Marongiu P., Rubino S., Rito T., Macaulay V., Semino O., Pala M., Abecasis G.R., Schlessinger D., Conde-Sousa E., Soares P., Richards M.B., Cucca F., Torroni A. Mitogenome diversity in Sardinians: a genetic window

- onto an Island's past. *Mol. Biol. Evol.* 2017;34(5):1230-1239. DOI 10.1093/molbev/msx082.
- Ruiz-Pesini E., Mishmar D., Brandon M., Procaccio V., Wallace D.C. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*. 2004;303(5655):223-226. DOI 10.1126/science.1088434.
- Skonieczna K., Malyarchuk B., Jawieñ A., Marszałek A., Banaszkiwicz Z., Jarmocik P., Grzybowski T. Mitogenomic differences between the normal and tumor cells of colorectal cancer patients. *Hum. Mutat.* 2018;39(5):691-701. DOI 10.1002/humu.23402.
- Stafford P., Chen-Quin E. The pattern of natural selection in somatic cancer mutations of human mtDNA. *J. Hum. Genet.* 2010;55(9):605-612. DOI 10.1038/jhg.2010.76.
- Sun C., Kong Q.-P., Zhang Y.-P. The role of climate in human mitochondrial DNA evolution: a reappraisal. *Genomics*. 2007;89(3):338-342. DOI 10.1016/j.ygeno.2006.11.005.
- Wallace D.C. 1994 William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am. J. Hum. Genet.* 1995;57(2):201-223. PMID 1801540.

ORCID ID

B.A. Malyarchuk orcid.org/0000-0002-0304-0652
M.V. Derenko orcid.org/0000-0002-1849-784X

Acknowledgements. The study was supported by a grant from the Russian Science Foundation No. 22-24-00264 (<https://rscf.ru/project/22-24-00264/>).

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

Received July 9, 2022. Revised July 14, 2022. Accepted July 14, 2022.

Original Russian text <https://vavilovj-icg.ru/>

Genetic control of N-glycosylation of human blood plasma proteins

S.Zh. Sharapov¹, A.N. Timoshchuk¹, Y.S. Aulchenko^{1, 2} 

¹ MSU Institute for Artificial Intelligence, Lomonosov Moscow State University, Moscow, Russia

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

 yurii@bionet.nsc.ru

Abstract. Glycosylation is an important protein modification, which influences the physical and chemical properties as well as biological function of these proteins. Large-scale population studies have shown that the levels of various plasma protein N-glycans are associated with many multifactorial human diseases. Observed associations between protein glycosylation levels and human diseases have led to the conclusion that N-glycans can be considered a potential source of biomarkers and therapeutic targets. Although biochemical pathways of glycosylation are well studied, the understanding of the mechanisms underlying general and tissue-specific regulation of these biochemical reactions *in vivo* is limited. This complicates both the interpretation of the observed associations between protein glycosylation levels and human diseases, and the development of glycan-based biomarkers and therapeutics. By the beginning of the 2010s, high-throughput methods of N-glycome profiling had become available, allowing research into the genetic control of N-glycosylation using quantitative genetics methods, including genome-wide association studies (GWAS). Application of these methods has made it possible to find previously unknown regulators of N-glycosylation and expanded the understanding of the role of N-glycans in the control of multifactorial diseases and human complex traits. The present review considers the current knowledge of the genetic control of variability in the levels of N-glycosylation of plasma proteins in human populations. It briefly describes the most popular physical-chemical methods of N-glycome profiling and the databases that contain genes involved in the biosynthesis of N-glycans. It also reviews the results of studies of environmental and genetic factors contributing to the variability of N-glycans as well as the mapping results of the genomic loci of N-glycans by GWAS. The results of functional *in vitro* and *in silico* studies are described. The review summarizes the current progress in human glycomics and suggests possible directions for further research.

Key words: glycome; glycans; N-glycosylation; genetics; GWAS.

For citation: Sharapov S.Zh., Timoshchuk A.N., Aulchenko Y.S. Genetic control of N-glycosylation of human blood plasma proteins. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(3):224-239. DOI 10.18699/VJGB-23-29

Генетический контроль N-гликозилирования белков плазмы крови человека

С.Ж. Шарапов¹, А.Н. Тимошук¹, Ю.С. Аульченко^{1, 2} 

¹ Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем

Московского государственного университета им. М.В. Ломоносова, Москва, Россия

² Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

 yurii@bionet.nsc.ru

Аннотация. Гликозилирование является важной модификацией белков, которая влияет как на их физико-химические свойства, так и на выполняемые ими биологические функции. Масштабные популяционные исследования показали, что уровни различных N-гликанов белков плазмы крови ассоциированы с риском развития ряда мультифакторных заболеваний человека. Найденные ассоциации стали основанием для рассмотрения N-гликанов в качестве потенциального источника биомаркеров и терапевтических мишеней. Биохимические пути N-гликозилирования хорошо изучены, однако понимание механизмов общей и тканеспецифической регуляции этих биохимических реакций *in vivo* весьма ограничено. Это затрудняет как интерпретацию наблюдаемых ассоциаций уровней N-гликанов с заболеваниями человека, так и разработку биомаркеров и молекулярных мишеней на их основе. Прогресс в области технологий анализа N-гликозилирования белков позволил к началу 2010-х годов проводить исследования регуляции N-гликозилирования с помощью методов генетического анализа, в том числе полногеномного исследования генетических ассоциаций. Применение этих методов дает возможность находить новые, ранее неизвестные регуляторы N-гликозилирования и расширяет представление о роли N-гликанов в контроле мультифакторных заболеваний и комплексных признаков человека. В данном обзоре мы рассматриваем современное состояние исследований генетического контроля популяционной изменчивости уровней N-гликозилирования белков плазмы крови человека. Описаны современные физико-химические методы измерения N-гликомного профиля, приведены базы данных, содержащие гены, вовлеченные в биосин-

тез N-гликанов. Систематизированы результаты исследований вклада средовых и генетических факторов в популяционную изменчивость N-гликанов, а также результаты картирования геномных локусов N-гликанов методом полногеномного исследования ассоциаций. Представлены результаты последующих функциональных исследований *in vitro* и *in silico*, позволивших предложить новые гены-кандидаты, регулирующие N-гликозилирование белков. В заключение кратко показан текущий прогресс в области гликогеномики человека и описаны возможные пути дальнейших исследований N-гликома.

Ключевые слова: гликом; гликаны; N-гликозилирование; генетика; ПГИА.

Glycomics as a branch of glycobiology

Glycans, also known as poly- or oligosaccharides, are polymers consisting of monosaccharides joined together by glycosidic linkages. Glycans may covalently bind to proteins and lipids by glycosidic bonds to produce glycoproteins and glycolipids, respectively. Glycosylation is one of the most common (Craveur et al., 2014) post- and co-translational protein modifications. It is found that about 20 % of all proteins in nature are glycosylated (Khoury et al., 2011). Meanwhile, over 40 % (by weight) of all proteins in human blood plasma are N-glycosylated (Clerc et al., 2016). Glycosylation affects not only physical and chemical properties of proteins, such as solubility, spatial configuration, folding, etc. (Varki, 1993; Ohtsubo, Marth, 2006; Skropeta, 2009), but their biological function as well. Glycoconjugates, i. e. glycoproteins and glycolipids with covalently bound glycans, are present in cells of all multicellular organisms (Varki, Kornfeld, 2015).

Glycoproteins and glycolipids at the surfaces of cell membranes are involved in various cellular interactions, including cell-cell, cell-extracellular matrix, and cell-macromolecule interactions, as well as interactions between organisms (host-parasite, symbiont-symbiont, etc.) (Ohtsubo, Marth, 2006; Skropeta, 2009; Lauc et al., 2016; Poole et al., 2018), which plays a major part in development and functioning of multicellular organisms (Gagneux et al., 2015).

Numerous studies into the chemical structure of glycans and their metabolism have been carried out in the early 20th century. At the time, however, glycans were primarily considered as structural elements and energy sources for living systems. An explosive development of chemical, physical, and molecular biological methods in glycan research has given birth to a new branch of molecular biology called glycobiology. This domain includes studies of chemical and physical properties of glycans, enzymology of glycan synthesis and degradation, their evolution, mechanisms of glycan recognition by proteins, and the role of glycans in functioning of biological systems, development of human diseases and biological traits, as well as development of new methods for management, prophylaxis, diagnostics, and prediction of diseases (Varki, 2017). Today, glycobiology is a rapidly developing science, and its findings are of great significance for many related fields, including biomedicine and biotechnology (Nikolac Perkovic et al., 2014; Varki, Kornfeld, 2017).

Similarly to genomics, transcriptomics, proteomics, and metabolomics, glycomics is a systematic investigation into glycome, i. e. a variety of all glycans and their contents in a

given specimen, whether it be a cell culture, tissue, organ, or the whole organism. The diversity of possible glycoconjugates is beyond imagination. Although the number of monomers incorporated into glycan structure is relatively small, monomers may form various glycosidic bonds to create an abundance of possible glycans. The diversity of glycoconjugates is further increased due to the possibility of a protein having not one, but several glycosylation sites.

In the early XXI century, high-throughput physical and chemical methods were developed, which made it possible to carry out large-scale cohort studies to discover the associations between N-glycome and human diseases and biological traits. Currently, the associations of N-glycans with many multifactorial diseases in humans (Gudelj et al., 2018a; Dotz, Wuhrer, 2019; Reily et al., 2019), including type 2 diabetes and monogenic forms of diabetes (Thanabalasingham et al., 2013; Keser et al., 2017), rheumatoid arthritis (Gudelj et al., 2018b), the Parkinson's disease (Russell et al., 2017), inflammatory bowel disease (Trbojević Akmačić et al., 2015; Clerc et al., 2018), as well as cardiovascular (Connelly et al., 2016; Wang et al., 2016) and oncological diseases (Fuster, Esko, 2005; Saldova et al., 2014; Mehta et al., 2015; Taniguchi, Kizuka, 2015), have been identified.

The results of observational studies of the associations between protein glycosylation and human diseases do not shed a light on cause-and-effect relationships between them and molecular biological mechanisms underlying these relationships. From 1983 onward, there have been a number of studies into the functional consequences of changes in protein glycosylation (Anthony et al., 2012; Cobb, 2020). Glycosylation of immunoglobulin G (IgG) appears to be the most deeply studied in this regard, due to its importance for adaptive immunity. The IgG molecule has a conserved glycosylation site Asn297, located in the conserved domain CH2 of the heavy chain. This domain plays a major part in binding to Fcγ receptors, which in turn affects the effector function of IgG. It was shown that a decrease in IgG fucosylation level intensified antibody-dependent cellular cytotoxicity (Peipp et al., 2008). The further crystallography investigation demonstrated that the lack of IgG fucosylation led to higher affinity between the Fc domain of IgG and the receptor FcγRIIIA, while the presence of fucose in IgG glycan resulted in steric hindrances during the interactions (Mizushima et al., 2011).

Given the above, a conclusion can be made about the fundamental importance of glycosylation research for the problems of diagnostics, prediction, prophylaxis, and management of human diseases. In 2012, the National Aca-

demy of Sciences of the United States presented a report on the necessity of a large-scale glycome study, the reasoning being that glycans are directly involved in pathogenesis of almost all the known diseases (http://www.nap.edu/catalog.php?record_id=13446).

Structure and diversity of glycans

Carbohydrates are among the main groups of macromolecules identified in biology, along with proteins, lipids, and nucleic acids. The polymerization ability and a large number of chiral atoms allow monosaccharides to form a large variety of stereo- and regioisomers. Four main groups of carbohydrates are determined based on the degree of polymerization as follows: monosaccharides (glucose, fructose, galactose, etc.), disaccharides (molecules consisting of two monosaccharides joined together by a glycosidic bond, e. g. sucrose, lactose, and maltose), polysaccharides with repeating units forming linear or branching compounds (O-antigens of bacteria, amylose, cellulose, and chitin), and glycans, i. e. complex oligosaccharides with non-repetitive units, which can be free or bound to proteins or lipids (glycoproteins, proteoglycans, and glycolipids).

Protein-bound glycans are in turn divided into N-, O-, and C-glycans. N-glycans form a glycosidic bond to the nitrogen atoms (N) of asparagine amino acid; O-glycans – to the hydroxyl groups of serine and threonine amino acids, and C-glycans – to carbon atoms of the tryptophan amino acid. C-glycosylation is rarely observed compared to N- and O-glycosylation (Chauhan et al., 2013).

An important difference between N- and O-glycosylation is that the N-glycosidic bond only forms with the asparagine of the Asn-X-Ser/Thr motif, where “X” may represent any amino acid, except for proline, whereas no such motif is known for O-glycans. In addition, there is a PNGase F enzyme that specifically cleaves the N-glycosidic bond between a glycan and a protein, which leads to the release of N-glycans into solution for further analysis (Vilaj et al., 2021). As opposed to N-glycosylation, O-glycosylation site does not have a consensus sequence, and the available methods for O-glycan isolation (beta elimination) show lower specificity compared to that for N-glycans (Mulloy et al., 2015). It is part of the reason why technologies and protocols of glycan structure identification and high-throughput analysis are currently more refined for N-glycans, which are the subject matter of this survey.

The most common monomers in N-glycans include monosaccharides, such as mannose, fucose, galactose, N-acetylglucosamine (GlcNAc), and N-acetylneuraminic acid (Fig. 1). N-glycan structure always includes a backbone (Man α 1-3(Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β 1-Asn-X-Ser/Thr) the other monomers bind to by glycosidic linkages. Glycans may have a branching structure with one or more branches called antennae. Monomers, such as galactose, N-acetylneuraminic acid, or fucose, can bind to any of the antennae. Fucose can also bind directly to the backbone. Negatively charged N-acetylneuraminic acid is the only monomer in N-glycans that carries a charge. The glycans not containing N-acetylneuraminic acid are neutral-

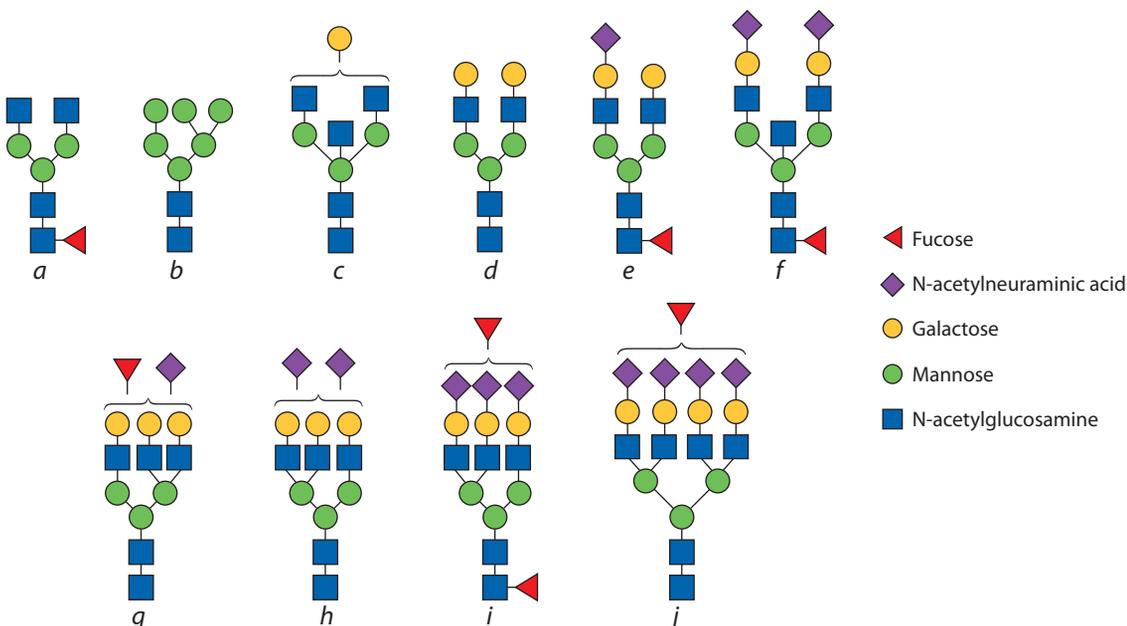


Fig. 1. Examples of N-glycan structures: structure *b* represents a glycan with a large number of mannose residues, while the other glycans have a complex structure. Structures *a*, *c*, *d-f* are biantennary glycans, *g-i* are triantennary glycans, and structure *j* is a tetraantennary glycan.

N-glycan backbone fucosylation is seen in structures *a*, *e*, *f*, and *i*; antennary fucosylation (fucose is bound to an antenna) is seen in structures *g*, *i*, and *j*. A bisecting N-acetylglucosamine (GlcNAc) residue is seen in structures *c* and *f*. Galactose and N-acetylneuraminic acid residues are shown in structures *c-j* and *e-j* respectively.

ly charged. N-acetylneuraminic acid in N-glycans is always bound to a galactose residue.

The Oxford glycan notation, one of the most commonly used glycan nomenclatures (Harvey et al., 2009), operates as follows:

1. The letter “F” at the very beginning of the name indicates the presence of fucose bound to the backbone.
2. It is followed by the “AN” sequence, where N is the number of antennae (branches) in the glycan structure.
3. Then, if the sugar backbone is bisected, the letter “B” (bisecting) is added.
4. If antennary branches are fucosylated, the letter “F” is added.
5. If the glycan structure includes galactose bound to one or several antennae, then the sequence G[n1,n2,...]N follows, where N is the number of galactose residues in a glycan, and n1 indicates the carbon atom of galactose, with which glycosidic bond is formed.
6. If the glycan structure includes N-acetylneuraminic acid bound to one or several antennae, then the sequence S[n1,n2,...]N follows, where N is the number of N-acetylneuraminic acid residues in a glycan, and n1 indicates the carbon atom of N-acetylneuraminic acid, with which glycosidic bond is formed.

For example, the designation FA2 shows the presence of a fucose residue bound to the backbone and two antennae in the glycan structure. The designation A3BG3S1 shows that glycan structure includes three antennae, sugar backbone bisection, three-antenna galactosylation, and one-antenna sialylation.

Genes involved in biological pathways of N-glycosylation

As opposed to mRNA and proteins encoded in the genomic DNA sequence and synthesized as a result of matrix processes, a glycan structure is not encoded in the genome directly, and its biosynthesis is a branching network of biochemical reactions (Lombard et al., 2014). The final structure of a glycan is determined by the interaction of a multitude of molecules and factors, including substrates and their accessibility, the enzyme activity associated with glycan biosynthesis and degradation, their localization and competition for substrate, and transport proteins (Kukuruzinska, Lennon, 1998; Nairn et al., 2008, 2012; Moremen et al., 2012). It was also shown that structure and diversity of N-glycans present in specific cells and tissues is partially regulated at the level of gene transcription encoding the proteins involved in glycan synthesis and degradation (Nairn et al., 2008, 2012; Moremen et al., 2012).

Glycan biosynthesis occurs in the endoplasmic reticulum (ER) and Golgi apparatus (GA). The KEGG database currently includes the data on over 300 enzymes involved in glycan synthesis and degradation processes (Kanehisa et al., 2017). Glycosyltransferases transporting activated monosaccharides to growing glycans are among the key enzymes directly involved in N-glycan biosynthesis. The dedicated CAZy database (Lombard et al., 2014) provides annotation and classification for over 200 glycosyltrans-

ferases, at least 40 of them associated with the protein N-glycosylation pathway. In addition, K.S. Egorova et al. developed the CSDB_GT database (Egorova et al., 2021) including the experimentally confirmed CAZy activities. N-glycan biosynthesis stages and the respective glycosyltransferase genes are described in detail in surveys (Saito, Ishii, 2002; Mohanty et al., 2020).

Physical and chemical methods for high-throughput N-glycan sequencing

Rapid development of glycobiology combined with the huge success of epidemiological population studies boosted the development of high-throughput glycome profiling methods for blood plasma proteins. In the recent decade, several high-throughput N-glycome profiling methods have been developed (Huffman et al., 2014), specifically high and ultra-high performance liquid chromatography (HPLC and UHPLC), multiplex capillary gel electrophoresis with laser-induced fluorescence detection, (xCGE-LIF), liquid chromatography electrospray mass spectrometry (LC-MS), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Representative N-glycosylation profiles of human blood plasma proteins obtained using three different methods are presented in Fig. 2.

A detailed comparison of the five most common N-glycome profiling methods for human blood plasma proteins is presented in the review (de Haan et al., 2022). Despite the differences in technologies, all these methods include several key stages, such as sample preparation (cell culture, tissue, organ, or organism), N-glycan isolation (for example by cleaving from glycoconjugates), separation of N-glycans and content measurement (absolute and relative values) (Huffman et al., 2014).

Each of these glycan analysis methods has its advantages and shortcomings. UHPLC and xCGE-LIF are used to analyze glycans cleaved from proteins, while MALDI-TOF-MS and LC-MS based on mass spectrometry make it possible to analyze glycopeptides with the protein regions containing covalently bound glycans, which provides valuable information on glycosylation of specific proteins. Compared to UHPLC and xCGE-LIF, MALDI-TOF-MS and LC-MS perform better with regard to distinguishing glycans with different molecular weights, but are unable to distinguish glycan stereoisomers. UHPLC and xCGE-LIF provide more accurate quantitative estimates due to their high resolution. In addition, they are characterized by high performance and lower initial costs compared to the methods based on mass spectrometry.

UHPLC has turned out to be the most popular high-throughput N-glycome profiling method for blood plasma proteins among the listed above (Akmačić et al., 2015) due to its relative cheapness, improved resolution (compared to HPLC), and high performance. By the time this review was composed (October 2022), human blood plasma glycome had been studied in about 200,000 samples all over the world, with about 80 % of the samples studied using UHPLC (G. Lauc, personal message).

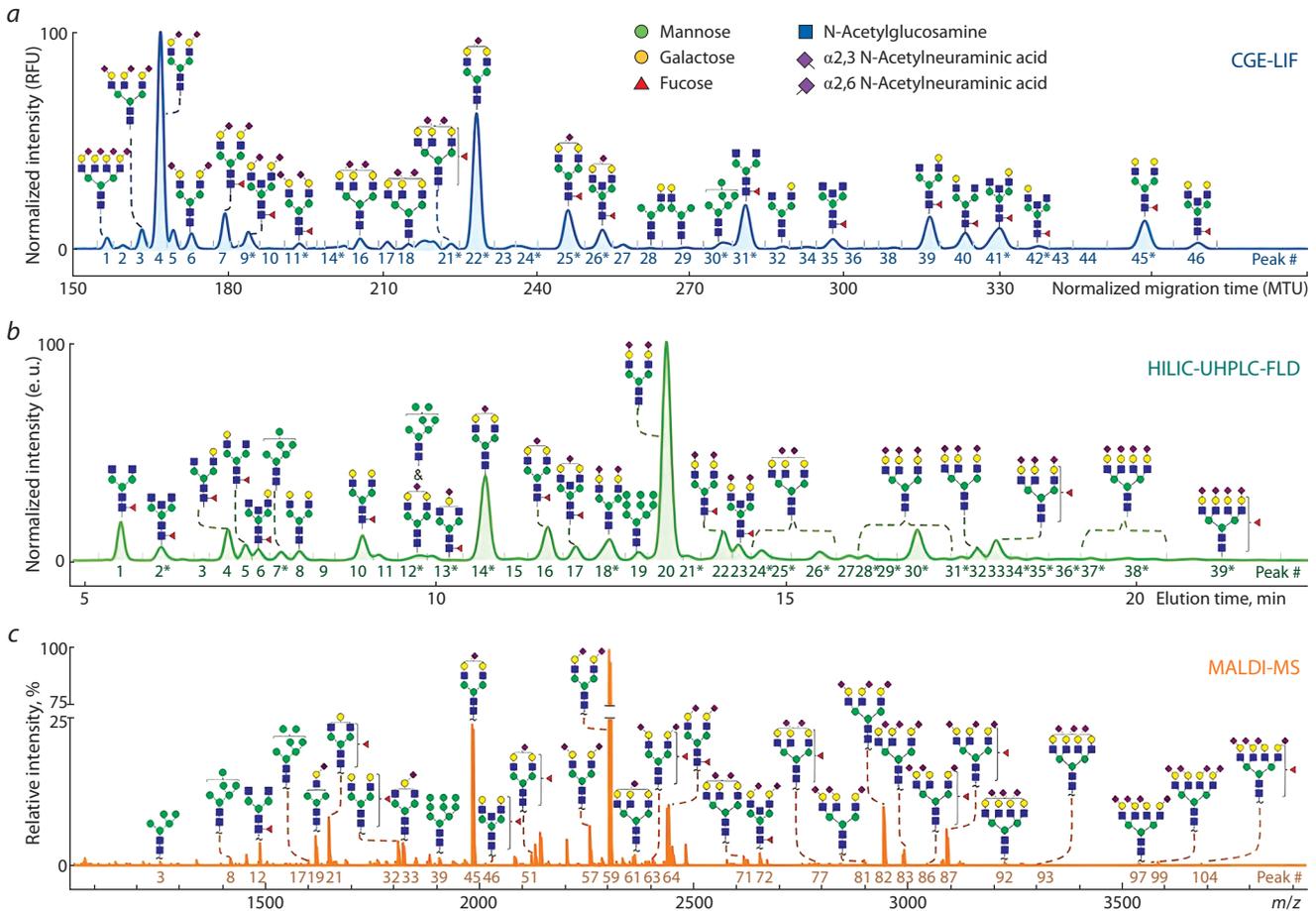


Fig. 2. N-glycosylation profiles of blood plasma proteins obtained using UHPLC, xCGE-LIF, and MALDI-MS.

N-glycome profile obtained using: *a* – xCGE-LIF (Reiding et al., 2019); *b* – UHPLC (Reiding et al., 2019; Zaytseva et al., 2020); *c* – MALDI-MS after differential esterification of N-acetylneuraminic acid (Vreeker et al., 2018; Reiding et al., 2019; Zaytseva et al., 2020). Modified after (de Haan et al., 2022).

Heritability of N-glycosylation levels of human blood plasma proteins

By the early 2010s, the development of high-throughput glycome profiling and genetic analysis methods had made it possible to carry out the first research efforts in genetic control of glycosylation based on the findings of cohort studies. There were a number of reasons why N-glycome of blood plasma became the main research focus: first, compared to other human tissues, blood plasma is a more accessible subject matter; second, as said above, the technologies for N-glycan level measurement and structure identification were more refined. The most common glycoproteins studied in human blood plasma were immunoglobulins G, A, and M, fibrinogen, transferrin, haptoglobin, etc. (Clerc et al., 2016), while liver cells and antibody-producing cells were the main glycoprotein source (Uhlén et al., 2015; Clerc et al., 2016).

Population variability of human blood plasma glycans, their heritability (trait dispersion due to genetic differences), as well as the effects of various environmental factors on glycan levels were studied in (Knezević et al., 2009). Glycan levels were measured using HPLC. The authors of the paper made several major conclusions. First, high population vari-

ability of glycosylation levels was discovered. Second, the authors discovered the significant effect biological sex and age had on glycan levels. Third, heritability of glycan levels varied (the average heritability index $h^2 = 34.7\%$ and the standard deviation of 15.5%), which implies that glycans were controlled by both genetic and environmental factors.

In (Zaytseva et al., 2020), the authors assessed the heritability of 39 N-glycan traits measured using UHPLC. It was shown that the heritability was over 50% (average heritability index $h^2 = 48.0\%$ and the standard deviation of 17.7%) for 24 out of 39 traits, which confirmed the hypothesis on the significant effect both environmental and genetic factors had on blood plasma glycome. In (Clerc et al., 2016; Uhlén et al., 2019), the authors pointed out the highest heritability (>50%) in biantennary glycans with backbone fucosylation and reduced sialylation of antennary chains included in immunoglobulins, primarily IgG being the most common glycoprotein among all human blood plasma proteins. Average and high heritability (30–62%) was observed in bi- and triantennary glycans with high sialylation of antennary chains. In (Jain et al., 2011), the authors assumed that high heritability in this case might be explained by the pre-

sence of these structures in a large number of glycoproteins (transferrin, hemopexin, alpha-1-antitrypsin, alpha-1-acid glycoprotein) causing errors in estimating genetic factors for each of them in isolation, and by the fact that these glycans were primarily contained in glycoproteins synthesized by liver cells, specifically acute-phase proteins, the glycosylation of which was significantly affected by the environment.

Despite the fact that heritability studies have made it possible to estimate the portion of trait variability controlled by genome, they have not revealed specific genomic regions affecting the manifestation of traits. The latter may be found using quantitative trait gene mapping methods, in particular genome-wide association studies.

Genome-wide association studies of N-glycan levels associated with blood plasma proteins

Genome-wide association study (GWAS) is the most common method for mapping loci of human diseases and complex traits. This method implies the analysis of associations between a large number (hundreds of thousands to tens of millions) of genetic markers distributed across the whole genome and the studied trait. Typically, large samples (several thousand to millions) of species or individuals are analyzed. The availability of these data makes it possible to essentially test the whole genome for associations with the studied trait and find new previously undiscovered associations between loci and traits. GWAS studies are usually designed around several samples. The findings from samples are combined using genome-wide meta-analysis techniques (Winkler et al., 2014), which increases the total sample size and the statistical power of the association analysis.

The presence of the association between a genomic locus and the studied trait does not by itself clarify the molecular biological mechanism underlying the discovered association. The discovered loci may contain from one to tens of genes, but they can also include none (Fig. 3) (Visscher et al., 2012, 2017). There is a multitude of reasons why an association can occur, i. e. the presence of encoding substitutions in the locus affecting the structure and functioning of the gene product (protein or RNA) or the presence of substitutions affecting the specificity of binding between transcription factors and regulatory regions. The number of functional variants may vary from one to many (Yang et al., 2012).

Identification of functional genes in the discovered loci and the mechanisms of their effect on the studied traits is the critical problem of functional studies performed using molecular and cellular biology methods. Here, the number of possible hypotheses to be tested grows geometrically (in theory) depending on the number of possible molecular association mechanisms. Taking into account the complexity, expensiveness, and labor intensity of molecular and cellular biological methods, primary bioinformatic prioritization of hypotheses on the association mechanisms becomes extremely important. Numerous methods for *in silico* functional annotation have been developed (Yang et al., 2012; Bulik-Sullivan et al., 2015; Pers et al., 2015; McLaren et al.,

2016; Staley et al., 2016; Zhu et al., 2016; Pasaniuc, Price, 2017; Hemani et al., 2018) making it possible to prioritize the hypotheses on association mechanisms, thereby increasing the efficiency of future molecular and biological research.

The subject matter in the available studies of genetic control of glycosylation using the GWAS approach was as follows: the total N-glycome of human blood plasma proteins (the subject matter of this review) (Lauc et al., 2010a, b; Huffman et al., 2011; Sharapov et al., 2019, 2020), N-glycome of immunoglobulin G, i. e. the most common N-glycoprotein in blood plasma (Lauc et al., 2013; Shen et al., 2017; Wahl et al., 2018; Klarić et al., 2020; Shadrina et al., 2021), and N-glycome of transferrin (Landini et al., 2022) secreted by liver.

At present, the results of five GWAS studies of the total N-glycome of human blood plasma proteins are available (Lauc et al., 2010a, b; Huffman et al., 2011; Sharapov et al., 2019, 2020).

GWAS studies of the total N-glycome of human blood plasma proteins

The first GWAS studies into N-glycosylation levels of human proteins were performed in 2010–2011 (Lauc et al., 2010a, b; Huffman et al., 2011). The authors used HPLC to analyze glycosylation levels, and the marker density of the genetic data was relatively low by today's standards at up to 2.5 million SNPs per genome. Six loci (*FUT8*, *HNF1A*, *FUT3/FUT5/FUT6*, *MGAT5*, *B3GAT1*, *SLC9A9*) associated with N-glycosylation of human blood plasma proteins were identified in these GWAS studies. It should be noted that none of the studies used independent samples to confirm the results.

The study published in 2019 (Sharapov et al., 2019) used the data from the TwinsUK Registry (Spector, Williams, 2006; Moayyeri et al., 2013). The genome-wide genotyping data and the primary UHPLC data on the N-glycome of blood plasma proteins were available for 2763 participants. The SNP number was increased from 2.5 to 8.5 million by means of imputation using the data of the 1000 Genomes Project and the appropriate quality control. As a result, the association was confirmed for 5 out of 6 previously identified loci (except for *SLC9A9*), and 10 new loci were discovered.

Based on four studies (Lauc et al., 2010a, b; Huffman et al., 2011; Sharapov et al., 2019), associations with 16 loci were found, with 15 of them confirmed later in (Sharapov et al., 2020) (see the Table) using the largest (at the time of the study) collection of genomic and glycomic data for 4802 specimens from four samples, namely EPIC-Potsdam, PainOmics, SOCCS, and SABRE, described in detail in the appendices (Sharapov et al., 2020). To put it briefly, the participants of the aforementioned studies were genotyped using the following DNA chips: EPIC-Potsdam (Human660W, 560,000 SNP, HumanCoreExome, 410,000 SNP, InfiniumOmniExpressExome, 850,000 SNP), PainOmics (Illumina HumanCore BeadChip, 720,000 SNP, Illumina

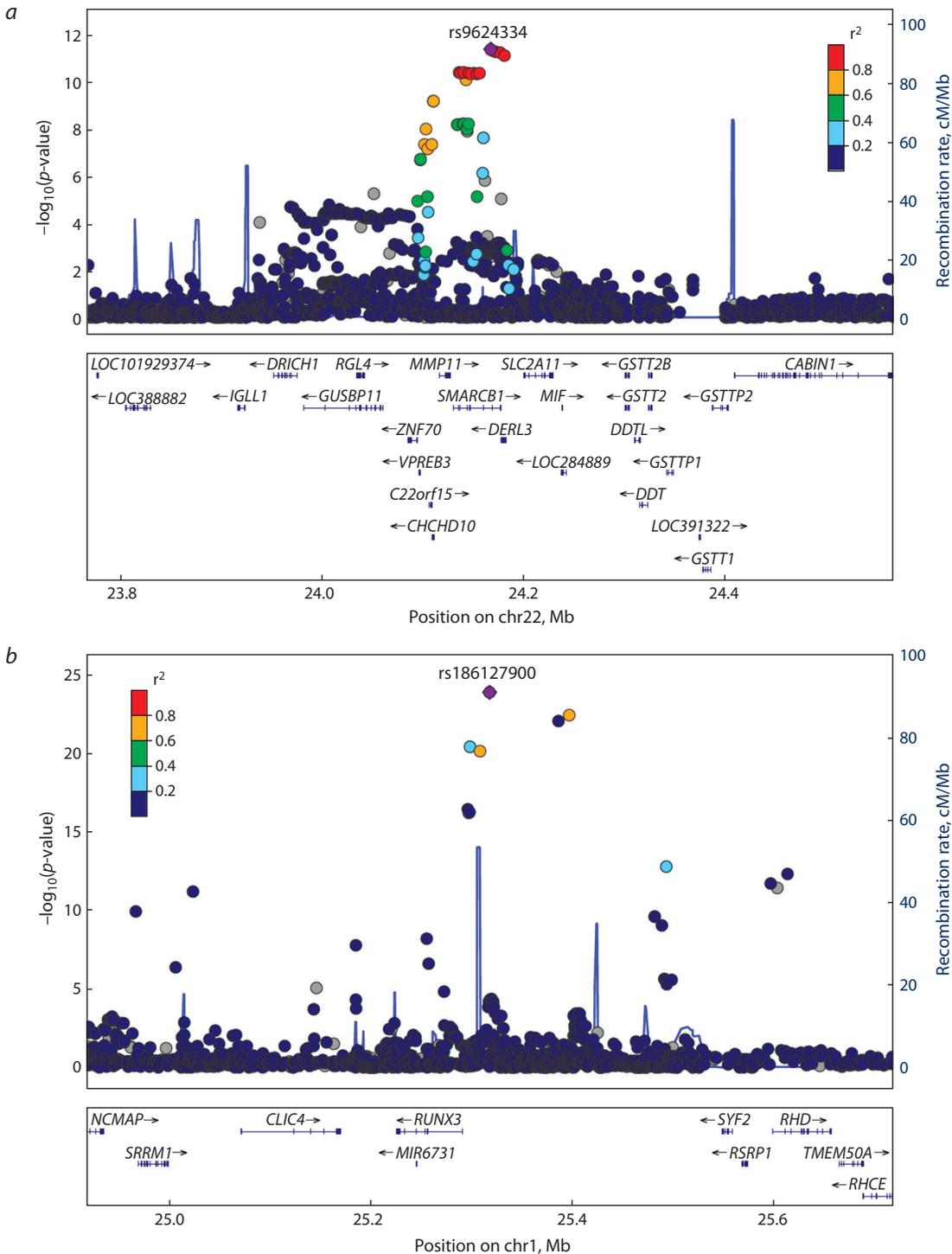


Fig. 3. Examples of regional association plots visualizing the association between the trait and genetic markers in the locus. A negative decimal logarithm of the p -value is plotted on the Y-axis. Genomic coordinates of the genetic marker (SNP) are plotted on the X-axis. The association signal may be located in the encoding region of several genes (a) or may not include any genes at all (b).

GSA, 300,000 SNP), SOCCS (HumanHap300/HumanHap240S, 510,000 SNP), SABRE (Illumina Human Core Bead Chip, 330,000 SNP).

EPIC-Potsdam cohort study included 27,000 participants at the ages from 35 to 65, who were selected randomly from the population of the city of Potsdam (Germany) in the years

from 1994 to 1998 (Boeing et al., 1999). PainOmics (Allegrì et al., 2016) was the case-control study aimed at finding potential biomarkers for dorsalgia and therapeutic targets for its management. The sample of 3400 participants including the residents of Italy, Belgium, England, and Croatia was composed in the years from 2014 to 2016.

Loci the associations of which were confirmed using independent samples (except for *KREMEN1*) in (Sharapov et al., 2020)

SNP	Position	Gene	EA/RA	EAF	Trait	BETA	SE	p	N
rs186127900	1:25318225	<i>RUNX3</i>	G/T	0.99	FA2G2S2	1.24	0.19	1.16E-10	1245
rs1257220	2:135015347	<i>MGAT5</i>	A/G	0.26	G4total, A4total	0.22	0.02	6.11E-20	4343
rs4839604	3:142960273	<i>SLC9A9</i>	C/T	0.80	FBS2/FS2	-0.20	0.03	3.87E-11	3592
rs17775791	3:186722362	<i>ST6GAL1</i>	C/T	0.26	FG1S1/(FG1+FG1S1)	-0.49	0.02	8.60E-97	4343
rs3115663	6:31601843	<i>PRRC2A</i>	C/T	0.18	M9	-0.15	0.03	1.63E-07	4343
rs6421315	7:50355207	<i>IKZF1</i>	C/G	0.37	A2[6]BG1n	-0.23	0.02	1.19E-27	4802
rs13297246	9:33128617	<i>B4GALT1</i>	A/G	0.17	FA2G2n	0.31	0.03	1.28E-24	4051
rs3967200	11:126232385	<i>ST3GAL4</i>	C/T	0.86	G4S3/G4S4	0.63	0.03	1.20E-106	4802
rs7928758	11:134265967	<i>B3GAT1</i>	G/T	0.15	A4G4S3	-0.36	0.03	6.43E-27	3592
rs735396	12:121438844	<i>HNF1A</i>	C/T	0.35	G3Fa/G3total	-0.21	0.02	4.91E-20	4343
rs11621121	14:65822493	<i>FUT8</i>	C/T	0.42	FG3/G3total	-0.31	0.02	8.94E-45	4187
rs35590487	14:105989599	<i>IGH, TMEM121</i>	C/T	0.75	FA2[3]G1n	-0.20	0.03	1.38E-09	2469
rs3760776	19:5839746	<i>FUT6</i>	A/G	0.09	G3Fa/G3total	-0.48	0.05	3.85E-23	2469
rs9624334	22:24166256	<i>SMARCB1, DERL3, CHCHD10</i>	C/G	0.17	FA2[6]BG1n	-0.31	0.03	7.15E-26	4051
rs909674	22:39859169	<i>MGAT3</i>	A/C	0.70	FBn	-0.22	0.02	1.88E-20	4343
rs140053014	22:29550678	<i>KREMEN1</i>	Ins/Del	0.98	FA2[3]G1n	-0.68	0.23	0.0027	459

Note. EA/RA – effect allele/reference allele; EAF – effect allele frequency; BETA/SE – effector allele effect on a trait and its standard error.

The Scottish project SOCCS (Theodoratou et al., 2016; Vučković et al., 2016) was the case-control study aimed at investigating the risk factors of colorectal cancer. The data on 2000 colorectal cancer patients and 2100 control subjects were collected in the research. SABRE is the population study initiated in 1988 (Tillin et al., 2012). Overall, the data on 4800 participants aged from 40 to 69 residing in West London (Great Britain) were collected.

To prioritize new protein glycosylation regulator genes in the confirmed loci and pose hypotheses on potential mechanisms at work in these loci, the authors of (Sharapov et al., 2019) used a combination of quantitative genetics and bioinformatics methods and approaches as follows.

1. Gene prioritization based on the results of eQTL colocalization analysis. Colocalization methods, particularly the SMR/HEIDI method (Zhu et al., 2016) used by the authors, made it possible to identify genes, the changes in the expression of which (at the mRNA level) mediated the association between SNPs and the studied trait.
2. Gene prioritization based on the determination of possible functional consequences of nucleotide substitutions with high SNP linkage disequilibrium associated with N-glycome traits. The VEP (McLaren et al., 2016), FATHMM-XF (Rogers et al., 2018), and FATHMM-InDel (Ferlaino et al., 2017) methods were used to select SNPs,

where substitutions changed the primary amino acid sequence of a protein and/or were recognized as pathogenic. Genes with sequences affected by said substitutions were prioritized as candidate genes.

3. Gene prioritization based on their involvement in various biological pathways. The DEPICT method (Pers et al., 2015) prioritized genes and biological pathways based on the results of enrichment analysis (overrepresentation of genes pertaining to specific biological pathways in the associated loci), which in turn was performed based on the pre-calculated probability of involvement of a specific locus in a particular gene network and/or biological pathway.

If the methods above failed to achieve gene prioritization for a certain locus, then the gene closest to the SNP with the most significant association in the locus was selected.

Candidate genes involved in N-glycosylation of blood plasma proteins

As a result of *in silico* studies within the investigation of the total N-glycome of blood plasma (Sharapov et al., 2019), 20 candidate genes were prioritized for 15 loci (Fig. 4).

The detailed description of these genes and the hypotheses on their possible roles in N-glycosylation regulation of human blood plasma proteins are presented in this section.

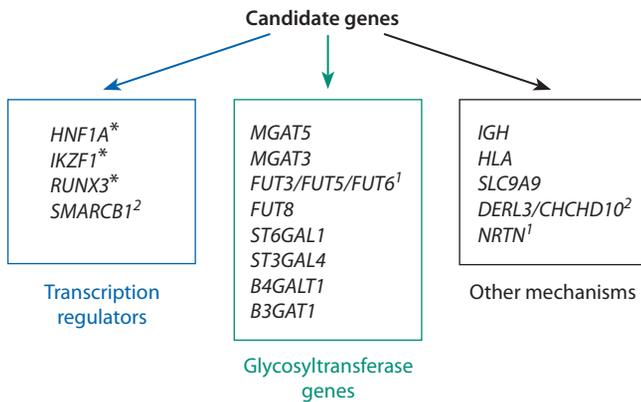


Fig. 4. Candidate genes regulating N-glycosylation levels of human blood plasma proteins suggested in (Sharapov et al., 2019).

The asterisks indicate the genes experimentally confirmed to be involved in N-glycosylation regulation. The superscripts indicate prioritization of candidate genes within the locus.

The genes coding for glycosyltransferase enzymes involved in N-glycan biosynthesis emerge as candidate genes in 8 loci (*MGAT5*, *MGAT3*, *FUT3/FUT5/FUT6*, *FUT8*, *ST6GAL1*, *ST3GAL4*, *B4GALT1*, *B3GAT1*) out of 15.

MGAT5 coding for GnT-V enzyme, i. e. alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase, is the candidate gene in the locus on the second chromosome, 125 Mbp. This enzyme transports the N-acetylglucosamine residue to the mannose of N-glycan, which produces a tri- or tetraantennary N-glycan. The locus with *MGAT5* showed an association with glycomic traits reflecting tri- and tetraantennary glycan levels (Sharapov et al., 2019).

MGAT3 coding for N-acetylglucosaminyltransferase GnT-III, i. e. beta-1,4-mannosylglycoprotein 4-beta-N-acetylglucosaminyltransferase, is the candidate gene in the locus on the 22nd chromosome, 39 Mbp. This enzyme transports the N-acetyl glucosamine residue to the mannose of N-glycan so as to produce backbone bisection. The pleiotropic effect of this locus on both N-glycan levels and *MGAT3* expression in CD19+ cells (B-lymphocytes) was demonstrated (Sharapov et al., 2019; Klarić et al., 2020).

FUT8 coding for Fuc-TVIII enzyme, i. e. alpha-(1, 6)-fucosyltransferase, is the candidate gene in the locus on the 14th chromosome, 66 Mbp. This enzyme transports fucose residue to N-acetylglucosamine of the N-glycan backbone, and through that is responsible for N-glycan backbone fucosylation. It is worth noting that loci *FUT8* and *MGAT3* showed association with traits FBS2/(FS2+FBS2) and FBS2/FS2 reflecting the presence of backbone bisection in biantennary glycans with backbone fucosylation (Sharapov et al., 2019), which is consistent with the known phenomenon of interference of Fuc-TVIII and GnT-III enzyme activities (Brockhausen, Schachter, 1996).

FUT6, *FUT5*, *FUT3*, and *NRTN* are candidate genes in the locus on the 19th chromosome, 5.8 Mbp. *NRTN* codes for a neurotrophic factor regulating neuron survival and function-

ing. *FUT6* and *FUT3/FUT5* code for Fuc-TVI and Fuc-TIII enzymes, i. e. fucosyltransferase 6 and 3, respectively, transporting fucose residue from the GDP-fucose to the N-acetylglucosamine by forming an alpha-1,3(4)-glycosidic bond. These enzymes are responsible for antennary fucosylation of N-glycans. In (Sharapov et al., 2019), it was shown that this locus is associated with antennary fucosylation in tri- and tetraantennary glycans. It is of note that rs17855739 SNP is located in *FUT6*. This SNP codes for G>A substitution (A allele frequency in human populations is about 12 %, according to the TopMED database), which leads to the replacement of negatively charged glutamic acid with positively charged lysine at position 247 (p.Glu247Lys). This substitution is located in the catalytic domain of Fuc-TVI enzyme and causes enzyme inactivation, and therefore this variant may have functional effect on glycosylation of human blood plasma proteins. It should be mentioned that *FUT3*, *FUT5*, and *FUT6* descend from a common ancestral gene as a result of two duplications (Dupuy et al., 2002). In addition, *FUT5* expression at the transcription and translation level in a human organism is much weaker compared to *FUT3* and *FUT6* (Taniguchi et al., 2014).

ST6GAL1 is the candidate gene in the locus on the third chromosome, 186 Mbp. *ST6GAL1* codes for alpha-2,6-sialyltransferase enzyme catalyzing the formation of the alpha-2,6-glycosidic bond between N-acetylneuraminic acid and N-acetylglucosamine bound to galactose of N-glycan. The *ST6GAL1* locus showed the association with the levels of mono- and disialylated N-glycans and their precursors (Sharapov et al., 2019).

ST3GAL4 is the candidate gene in the locus on the 11th chromosome, 126 Mbp. *ST3GAL4* codes for alpha-2,3-sialyltransferase enzyme transporting the N-acetylneuraminic acid residue. This locus showed an association with the levels of various sialylated N-glycans (Sharapov et al., 2019).

B4GALT1 is the candidate gene in the locus on the 9th chromosome, 33 Mbp. *B4GALT1* codes for galactosyltransferase enzyme binding galactose to various substrates, including N-acetyl glucosamine. The *B4GALT1* locus was associated with the levels of galactosylated biantennary N-glycans and their precursors (Sharapov et al., 2019). It is also known that a series of mutations in *B4GALT1* leads to a congenital disorder of glycosylation (Staretz-Chacham et al., 2020).

B3GAT1 coding for galactosylgalactosylxylosylprotein-3-beta-glucuronosyltransferase 1 enzyme is the candidate gene in the locus on the 11th chromosome, 134 Mbp. This enzyme catalyzes the transport of glucuronic acid in HNK-1 epitope biosynthesis. This epitope is expressed on lymphocytes, but its presence on blood plasma proteins remained undiscovered for some time. The association of this locus with N-glycan levels in blood plasma proteins was first shown in (Huffman et al., 2011). The presence of glucuronic acid in N-glycome of blood plasma, which can explain the association of the locus, was discovered in (Sharapov et al., 2019).

Candidate genes in seven other loci are not glycosyltransferase genes. Three genes, *SMARCB1*, *DERL3*, and *CHCHD10*, were prioritized in the locus on the 22nd chromosome, 39 Mbp. The strongest association signal in the locus is observed in the coding sequence of *SMARCB1* gene. *SMARCB1* codes for the protein of hSWI/SNF complex acting as a chromatin remodeler. *SMARCB1* gene product plays a major part in carcinogenesis inhibition, cell proliferation and differentiation (Pottier et al., 2007).

DERL3 codes for the enzyme involved in the degradation of luminal glycoproteins with incorrect tertiary structure in the endoplasmic reticulum (Oda et al., 2006). The pathogenic variant rs3177243 is also found in this locus, in the coding sequence of *DERL3* gene.

CHCHD10 codes for mitochondrial protein observed in fibrils of mitochondrial cristae. It was shown that genetic association of this locus with N-glycan levels in proteins may be mediated by the effect of nucleotide substitutions on *CHCHD10* expression in blood cells (Sharapov et al., 2019). The direct involvement of mitochondrial proteins in glycosylation processes remained undiscovered before 2017, when the paper showing the role of mitochondrial fragmentation and the number of ER-mitochondria contacts in representation of sialylated glycans on the surface of glioblastoma cells, which in turn affected glioblastoma cell recognition by lymphocytes, was published (Martinvalet, 2018).

The locus on the 14th chromosome, 105 Mbp, contains the *IGH* gene cluster coding for heavy chains of immunoglobulins. IgG is the most common N-glycoprotein in human blood plasma (Clerc et al., 2016), and its constitutive N-glycosylation site is located in the heavy chain.

SLC9A9 is the candidate gene in the locus on the third chromosome, 142 Mbp. *SLC9A9* codes for the Na⁺/H pump, presumably regulating the pH level in the Golgi apparatus (GA). Protein glycosylation occurs in the GA, and, according to the available data, it is a pH-sensitive process (Kellokumpu, 2019). The processes in the GA affect the synthesis of heterodimeric complexes responsible for glycosylation (Hassinen et al., 2011). It was shown in (Rivinoja et al., 2009) that a pH increase in the GA may disrupt terminal N-glycosylation (including sialylation) due to incorrect localization of glycosyltransferases. In accordance with this hypothesis, the *SLC9A9* locus showed an association with tetra-sialylated N-glycan levels in (Huffman et al., 2011) and with sialylated N-glycan levels in (Sharapov et al., 2019).

HNF1A is the candidate gene in the locus on the 12th chromosome, 121 Mbp. A detailed functional study into this locus in (Lauc et al., 2010a) showed that *HNF1A* coding for the hepatocyte transcription factor regulates the expression of most fucosyltransferase encoding genes, *FUT3*, *FUT5*, *FUT6*, *FUT8*, *FUT10*, and *FUT11*, in the HepG2 cell line obtained from liver cells. The same study demonstrated that *HNF1A* regulates the expression of genes encoding the key GDP-fucose synthetase enzymes, and GDP-fucose acts as a substrate for fucosyltransferases. This implies that *HNF1A* plays a major part in glycan fucosylation processes.

IKZF1 is the candidate gene for the locus on the 7th chromosome, 50 Mbp. It was shown earlier (Lauc et al., 2013) that this locus was associated with IgG glycosylation, and *IKZF1* was suggested as the candidate gene for the locus. *IKZF1* encodes the DNA-binding protein Ikaros, a transcription regulator involved in chromatin remodeling. It is of note that the *IKZF1* locus showed the association with levels of N-glycans with backbone fucosylation in blood plasma proteins, with which the *FUT8* locus was associated (Sharapov et al., 2019). *IKZF1* is considered as an important lymphocyte differentiation regulator (Sellars et al., 2009; Marke et al., 2018).

Since IgG-secreting cells are lymphocyte derivatives, *IKZF1* gene was selected as the candidate gene in the locus, and the hypothesis on its role in regulation of backbone fucosylation in IgG N-glycans through *FUT8* expression regulation was posed (Sharapov et al., 2019). In addition, it was experimentally shown in (Klarić et al., 2020) that *IKZF1* knockdown in *MATAT6* IgG-secreting cells leads to more than tripled *FUT8* expression and increased fucosylation level in the secreted IgG.

RUNX3 is the candidate gene in the locus on the first chromosome, 25 Mbp. This gene codes for Runt domain-containing protein, a transcription factor, which, similarly to *IKZF1* (Sellars et al., 2009), plays a major part in B-lymphocyte maturation and differentiation.

The candidate genes for the *HLA* locus (human major histocompatibility complex) on the sixth chromosome, 25–32 Mbp, are not presented due to a high chance of false positive. The *HLA* locus is unique in terms of quantitative genetics of multifactorial human traits (Kennedy et al., 2017). This locus shows the highest gene density in the human genome; it also demonstrates the highest degree of polymorphism at the nucleotide level; locus alleles show high linkage disequilibrium throughout the whole locus spanning 8 Mbp.

Gene regulatory network of N-glycosylation of human blood plasma proteins

The recent studies into N-glycome of blood plasma (Sharapov et al., 2019, 2020) have demonstrated a significant association of 15 loci with 116 out of 117 glycan traits. In total, significant association has been shown by 214 locus-trait pairs. These data were used in (Sharapov et al., 2019) to reconstruct the gene regulatory network of N-glycan levels in blood plasma proteins (Fig. 5). This network visualizes the association between the discovered loci and N-glycan levels in blood plasma proteins.

To build the network, glycomic traits were classified into four groups based on the tissue secreting N-glycoproteins into blood plasma. The first group included the traits reflecting N-glycan levels in immunoglobulins (IgA, IgG, IgD, IgE, IgM) secreted by lymphocytic series cells, i.e. B-lymphocytes, plasmoblasts, and plasmocytes. The second group were the traits reflecting N-glycan levels in proteins (transferrin, haptoglobin, etc.) primarily secreted by hepatocytes, i.e. liver cells. The third group were the

and *IGH/TMEM121* genes. These loci are associated with N-glycans linked to immunoglobulins secreted into blood stream by lymphocytary series cells. In addition, it was shown in GWAS studies of N-glycan levels in IgG that these loci are associated with IgG N-glycosylation (Shen et al., 2017; Klarić et al., 2020). Since IgG is the most common glycoprotein in blood plasma, it can be hypothesized that candidate genes from this network regulate N-glycosylation processes in B-lymphocytes and their descendants. The role of candidate genes from this network should likely be tested in antibody-producing cells and cells close to them.

The role of transcription factor *IKZF1* in *FUT8* expression regulation in the lymphoid line *GM12878* was proved in (Klarić et al., 2020). Furthermore, *IKZF1* knockdown resulted in increased fucosylated protein level, which proves the role of transcription factor *IKZF1* in protein fucosylation regulation as a result of *in vitro* experiment.

The GWAS approach was used to identify a total of 16 loci, and associations of 15 of them were confirmed in independent samples. An *in silico* study was performed for 15 confirmed loci, and 20 candidate genes were suggested. As a result of *in vitro* experiments, the role of transcription factor *IKZF1* in protein fucosylation regulation and the role of *HNFLA* in fucosyltransferase expression regulation were proved. The role of transcription factor *RUNX3* in N-glycosylation regulation was confirmed by targeted genome editing (using CRISPR-dCas9 system) in cell lines VPR-dCas9 and KRAB-dCas9 HEK-293F secreting IgG into the environment. Comparison of the IgG N-glycosylation profile with the non-modified control cell line showed that increased *RUNX3* gene expression leads to a significant reduction of galactosylated structures with a further increase in agalactosylated structures (Mijakovac et al., 2022).

Conclusion

The results of GWAS studies of N-glycan levels in blood plasma proteins confirm the understanding of N-glycosylation of human blood plasma proteins as a complex process controlled by genes involved in various biological pathways and expressed in various tissues. The candidate genes suggested as a result of a large-scale *in silico* investigation (Sharapov et al., 2019) of the confirmed loci make it possible to pose functional hypotheses on the mechanisms underlying the effect of the discovered loci on N-glycosylation of blood plasma proteins. These hypotheses will be of use in the planning of *in vitro* and *in vivo* molecular genetic studies of glycome and its role in pathogenesis of socially and economically important human diseases. The results of the performed *in vitro* experiments solidify the scientific credence of functional hypotheses with regard to candidate genes suggested using the GWAS approach.

There are several development trends for human population glycogenomics. Larger-scale GWAS studies into N-glycan levels using larger samples will be performed. New functional genomic data applicable to studying N-glycosylation processes will be available, which, combined with the GWAS results, will make it possible to identify more loci

and potential N-glycosylation regulators. The application of the GWAS approach in glycosylation regulation studies is currently restricted to the analysis of the total N-glycome of blood plasma and N-glycome of IgG and transferrin. The development of N-glycome profiling technologies will expand the variety of proteins, the individual N-glycosylation profiles of which will be studied. On the other hand, high-throughput technologies for N-glycome profiling in other human tissues are likely to emerge. The advancements listed above will make it possible to better understand N-glycosylation regulation in human proteins and through that determine the role of glycosylation in pathogenesis of glycome-associated diseases and boost the development of new methods for prediction, prophylaxis, diagnostics and management of these diseases.

References

- Akmačić I.T., Ugrina I., Štambuk J., Gudelj I., Vučković F., Lauc G., Pučić-Baković M. High-throughput glycomics: optimization of sample preparation. *Biochemistry (Mosc.)*. 2015;80(7):934-942. DOI 10.1134/S0006297915070123.
- Allegrì M., De Gregori M., Minella C.E., Klersy C., Wang W., Sim M., Gieger C., Manz J., Pemberton I.K., MacDougall J., Williams F.M., Van Zundert J., Buyse K., Lauc G., Gudelj I., Primorac D., Skelin A., Aulchenko Y.S., Karssen L.C., Kapural L., Rauck R., Fanelli G., PainOMICS Group "Omics" biomarkers associated with chronic low back pain: protocol of a retrospective longitudinal study. *BMJ Open*. 2016;6(10):e012070. DOI 10.1136/bmjopen-2016-012070.
- Anthony R.M., Wermeling F., Ravetch J.V. Novel roles for the IgG Fc glycan. *Ann. N. Y. Acad. Sci.* 2012;1253(1):170-180. DOI 10.1111/j.1749-6632.2011.06305.x.
- Boeing H., Korfmann A., Bergmann M.M. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. *Ann. Nutr. Metab.* 1999;43(4):205-215. DOI 10.1159/000012787.
- Brockhausen I., Schachter H. Glycosyltransferases involved in N- and O-glycan biosynthesis. In: Gabius H.-J., Gabius S. (Eds.). *Glycosciences*. Weinheim: Chapman & Hall, 1997;79-113. DOI 10.1002/9783527614738.ch5.
- Bulik-Sullivan B., Finucane H.K., Anttila V., Gusev A., Day F.R., Loh P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L., Perry J.R.B., Patterson N., Robinson E.B., Daly M.J., Price A.L., Neale B.M. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 2015;47(11):1236-1241. DOI 10.1038/ng.3406.
- Chauhan J.S., Rao A., Raghava G.P.S. In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. *PLoS One*. 2013;8(6):e67008. DOI 10.1371/journal.pone.0067008.
- Clerc F., Novokmet M., Dotz V., Reiding K.R., de Haan N., Kammeijer G.S.M., Dalebout H., Bladergroen M.R., Vukovic F., Rapp E., IBD-BIOM Consortium, Targan S.R., Barron G., Manetti N., Lattiano A., McGovern D.P.B., Annese V., Lauc G., Wuhrer M. Plasma N-glycan signatures are associated with features of inflammatory bowel diseases. *Gastroenterology*. 2018;155(3):829-843. DOI 10.1053/j.gastro.2018.05.030.
- Clerc F., Reiding K.R., Jansen B.C., Kammeijer G.S.M., Bondt A., Wuhrer M. Human plasma protein N-glycosylation. *Glycoconj. J.* 2016;33(3):309-343. DOI 10.1007/s10719-015-9626-2.
- Cobb B.A. The history of IgG glycosylation and where we are now. *Glycobiology*. 2020;30(4):202-213. DOI 10.1093/glycob/cwz065.
- Connelly M.A., Gruppen E.G., Otvos J.D., Dullaart R.P.F. Inflammatory glycoproteins in cardiometabolic disorders, autoimmune diseases and cancer. *Clin. Chim. Acta.* 2016;459:177-186. DOI 10.1016/j.cca.2016.06.012.

- Craveur P., Rebehmed J., de Brevern A.G. PTM-SD: a database of structurally resolved and annotated posttranslational modifications in proteins. *Database (Oxford)*. 2014;2014:bau041. DOI 10.1093/database/bau041.
- de Haan N., Pučić-Baković M., Novokmet M., Falck D., Lageveen-Kammeijer G., Razdorov G., Vučković F., Trbojević-Akmačić I., Gornik O., Hanić M., Wuhrer M., Lauc G., The Human Glycome Project. Developments and perspectives in high-throughput protein glycomics: enabling the analysis of thousands of samples. *Glycobiology*. 2022;32(8):651-663. DOI 10.1093/glycob/cwac026.
- Dotz V., Wuhrer M. N-glycome signatures in human plasma: associations with physiology and major diseases. *FEBS Lett*. 2019;593(21):2966-2976. DOI 10.1002/1873-3468.13598.
- Dupuy F., Germot A., Marendia M., Oriol R., Blancher A., Julien R., Maftah A. α 1,4-fucosyltransferase activity: A significant function in the primate lineage has appeared twice independently. *Mol. Biol. Evol.* 2002;19(6):815-824. DOI 10.1093/oxfordjournals.molbev.a004138.
- Egorova K.S., Smirnova N.S., Toukach P.V. CSDB_GT, a curated glycosyltransferase database with close-to-full coverage on three most studied nonanimal species. *Glycobiology*. 2021;31(5):524-529. DOI 10.1093/glycob/cwaa107.
- Ferlaino M., Rogers M.F., Shihab H.A., Mort M., Cooper D.N., Gaunt T.R., Campbell C. An integrative approach to predicting the functional effects of small indels in non-coding regions of the human genome. *BMC Bioinformatics*. 2017;18(1):442. DOI 10.1186/s12859-017-1862-y.
- Fuster M.M., Esko J.D. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer*. 2005;5(7):526-542. DOI 10.1038/nrc1649.
- Gagneux P., Aebi M., Varki A. Evolution of glycan diversity. In: Varki A., Cummings R.D., Esko J.D., Stanley P., Hart G.W., Aebi M., Darvill A.G., Kinoshita T., Packer N.H., Prestegard J.H., Schnaar R.L., Seeberger P.H. (Eds.). 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2015–2017. Chapter 20. PMID 28876829. DOI 10.1101/glycobiology.3e.020.
- Gudelj I., Lauc G., Pezer M. Immunoglobulin G glycosylation in aging and diseases. *Cell. Immunol.* 2018a;333:65-79. DOI 10.1016/j.cellimm.2018.07.009.
- Gudelj I., Salo P.P., Trbojević-Akmačić I., Albers M., Primorac D., Perola M., Lauc G. Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up. *Biochim. Biophys. Acta Mol. Basis Dis.* 2018b;1864(6 Pt. A):2034-2039. DOI 10.1016/j.bbadis.2018.03.018.
- Harvey D.J., Merry A.H., Royle L., Campbell M.P., Dwek R.A., Rudd P.M. Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics*. 2009;9(15):3796-3801. DOI 10.1002/pmic.200900096.
- Hassinen A., Pujol F.M., Kokkonen N., Pieters C., Kihlström M., Korhonen K., Kellokumpu S. Functional organization of Golgi N- and O-glycosylation pathways involves pH-dependent complex formation that is impaired in cancer cells. *J. Biol. Chem.* 2011;286(44):38329-38340. DOI 10.1074/jbc.M111.277681.
- Hemani G., Zheng J., Elsworth B., Wade K.H., Haberland V., Baird D., Laurin C., Burgess S., Bowden J., Langdon R., Tan V.Y., Yarmolinsky J., Shihab H.A., Timpson N.J., Evans D.M., Relton C., Martin R.M., Davey Smith G., Gaunt T.R., Haycock P.C. The MR-Base platform supports systematic causal inference across the human phenotype. *Elife*. 2018;7:e34408. DOI 10.7554/elife.34408.
- Huffman J.E., Knezevic A., Vitart V., Kattla J., Adamczyk B., Novokmet M., Igl W., Pucic M., Zgaga L., Johannson Å., Redzic I., Gornik O., Zemunik T., Polasek O., Kolcic I., Pehlic M., Koelerman C.A.M., Campbell S., Wild S.H., Hastie N.D., Campbell H., Gyllensten U., Wuhrer M., Wilson J.F., Hayward C., Rudan I., Rudd P.M., Wright A.F., Lauc G. Polymorphisms in *B3GAT1*, *SLC9A9* and *MGAT5* are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum. Mol. Genet.* 2011;20(24):5000-5011. DOI 10.1093/hmg/ddr414.
- Huffman J.E., Pučić-Baković M., Klarić L., Hennig R., Selman M.H.J., Vučković F., Novokmet M., Krištić J., Borowiak M., Muth T., Polašek O., Razdorov G., Gornik O., Plomp R., Theodoratou E., Wright A.F., Rudan I., Hayward C., Campbell H., Deelder A.M., Reichl U., Aulchenko Y.S., Rapp E., Wuhrer M., Lauc G. Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol. Cell. Proteomics*. 2014;13(6):1598-1610. DOI 10.1074/mcp.M113.037465.
- Jain S., Gautam V., Naseem S. Acute-phase proteins: As diagnostic tool. *J. Pharm. Bioallied. Sci.* 2011;3(1):118-127. DOI 10.4103/0975-7406.76489.
- Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353-D361. DOI 10.1093/nar/gkw1092.
- Kellokumpu S. Golgi pH, ion and redox homeostasis: How much do they really matter? *Front. Cell Dev. Biol.* 2019;7:93. DOI 10.3389/fcell.2019.00093.
- Kennedy A.E., Ozbek U., Dorak M.T. What has GWAS done for HLA and disease associations? *Int. J. Immunogenet.* 2017;44(5):195-211. DOI 10.1111/iji.12332.
- Keser T., Gornik I., Vučković F., Selak N., Pavić T., Lukić E., Gudelj I., Gašparović H., Biočina B., Tilin T., Wennerström A., Männistö S., Salomaa V., Havulinna A., Wang W., Wilson J.F., Charutvedti N., Perola M., Campbell H., Lauc G., Gornik O. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. *Diabetologia*. 2017;60(12):2352-2360. DOI 10.1007/s00125-017-4426-9.
- Khoury G.A., Baliban R.C., Floudas C.A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* 2011;1:90. DOI 10.1038/srep00090.
- Klarić L., Tsepilov Y.A., Stanton C.M., Mangino M., Sikka T.T., Esko T., Pakhomov E., ... Wilson J.F., Zoldoš V., Vitart V., Specor T., Aulchenko Y.S., Lauc G., Hayward C. Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Sci. Adv.* 2020;6(8):eaax0301. DOI 10.1126/sciadv.aax0301.
- Knezević A., Polasek O., Gornik O., Rudan I., Campbell H., Hayward C., Wright A., Kolcic I., O'Donoghue N., Bones J., Rudd P.M., Lauc G. Variability, heritability and environmental determinants of human plasma N-glycome. *J. Proteome Res.* 2009;8(2):694-701. DOI 10.1021/pr800737u.
- Kukuruzinska M.A., Lennon K. Protein N-glycosylation: molecular genetics and functional significance. *Crit. Rev. Oral Biol. Med.* 1998;9(4):415-448. DOI 10.1177/10454411980090040301.
- Landini A., Trbojević-Akmačić I., Navarro P., Tsepilov Y.A., Sharapov S.Z., Vučković F., Polašek O., Hayward C., Petrović T., Vilaj M., Aulchenko Y.S., Lauc G., Wilson J.F., Klarić L. Genetic regulation of post-translational modification of two distinct proteins. *Nat. Commun.* 2022;13(1):1586. DOI 10.1038/s41467-022-29189-5.
- Lauc G., Essafi A., Huffman J.E., Hayward C., Knezević A., Kattla J.J., Polašek O., ... Gyllensten U., Wilson J.F., Wright A.F., Hastie N.D., Campbell H., Rudd P.M., Rudan I. Genomics meets glycomics—the first GWAS study of human N-glycome identifies HNF1 α as a master regulator of plasma protein fucosylation. *PLoS Genet.* 2010a;6(12):e1001256. DOI 10.1371/journal.pgen.1001256.
- Lauc G., Huffman J.E., Pučić M., Zgaga L., Adamczyk B., Mužinić A., Novokmet M., ... Wuhrer M., Wright A.F., Rudd P.M., Hayward C., Aulchenko Y., Campbell H., Rudan I. Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet.* 2013;9(1):e1003225. DOI 10.1371/journal.pgen.1003225.

- Lauc G., Pezer M., Rudan I., Campbell H. Mechanisms of disease: The human *N*-glycome. *Biochim. Biophys. Acta*. 2016;1860(8):1574-1582. DOI 10.1016/j.bbagen.2015.10.016.
- Lauc G., Rudan I., Campbell H., Rudd P.M. Complex genetic regulation of protein glycosylation. *Mol. Biosyst.* 2010b;6(2):329-335. DOI 10.1039/b910377e.
- Lombard V., Golaconda Ramulu H., Drula E., Coutinho P.M., Henrisat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42(D1):D490-D495. DOI 10.1093/nar/gkt1178.
- Marke R., van Leeuwen F.N., Scheijen B. The many faces of IKZF1 in B-cell precursor acute lymphoblastic leukemia. *Haematologica*. 2018;103(4):565-574. DOI 10.3324/haematol.2017.185603.
- Martinvalet D. The role of the mitochondria and the endoplasmic reticulum contact sites in the development of the immune responses. *Cell Death Dis.* 2018;9(3):336. DOI 10.1038/s41419-017-0237-7.
- McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S., Thormann A., Flicek P., Cunningham F. The ensemble variant effect predictor. *Genome Biol.* 2016;17(1):122. DOI 10.1186/s13059-016-0974-4.
- Mehta A., Herrera H., Block T. Glycosylation and liver cancer. *Adv. Cancer Res.* 2015;126:257-279. DOI 10.1016/bs.acr.2014.11.005.
- Mijakovac A., Miškec K., Kristić J., Vičić Bočkor V., Tadić V., Bošković M., Lauc G., Zoldoš V., Vojta A. A transient expression system with stably integrated CRISPR-dCas9 fusions for regulation of genes involved in immunoglobulin G glycosylation. *CRISPR J.* 2022;5(2):237-253. DOI 10.1089/crispr.2021.0089.
- Mizushima T., Yagi H., Takemoto E., Shibata-Koyama M., Isoda Y., Iida S., Masuda K., Satoh M., Kato K. Structural basis for improved efficacy of therapeutic antibodies on defucosylation of their Fc glycans. *Genes Cells.* 2011;16(11):1071-1080. DOI 10.1111/j.1365-2443.2011.01552.x.
- Moayyeri A., Hammond C.J., Hart D.J., Spector T.D. The UK adult twin registry (TwinsUK resource). *Twin Res. Hum. Genet.* 2013; 16(1):144-149. DOI 10.1017/thg.2012.89.
- Mohanty S., Chaudhary B., Zoetewey D. Structural insight into the mechanism of *N*-linked glycosylation by oligosaccharyltransferase. *Biomolecules.* 2020;10(4):624. DOI 10.3390/biom10040624.
- Moremen K.W., Tiemeyer M., Nairn A.V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* 2012;13(7):448-462. DOI 10.1038/nrm3383.
- Mulloy B., Dell A., Stanley P., H. Prestegard J. Structural analysis of glycans. In: Varki A., Cummings R.D., Esko J.D., Stanley P., Hart G.W., Aebi M., Darvill A.G., Kinoshita T., Packer N.H., Prestegard J.H., Schnaar R.L., Seeberger P.H. (Eds.). *Essentials of Glycobiology*. [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2015. Chapter 50, 2017. DOI 10.1101/glycobiology.3e.050.
- Nairn A.V., Aoki K., dela Rosa M., Porterfield M., Lim J.-M., Kulik M., Pierce J.M., Wells L., Dalton S., Tiemeyer M., Moremen K.W. Regulation of glycan structures in murine embryonic stem cells. *J. Biol. Chem.* 2012;287(45):37835-37856. DOI 10.1074/jbc.m112.405233.
- Nairn A.V., York W.S., Harris K., Hall E.M., Pierce J.M., Moremen K.W. Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J. Biol. Chem.* 2008;283(25):17298-17313. DOI 10.1074/jbc.M801964200.
- Nikolac Perkovic M., Pucic Bakovic M., Kristic J., Novokmet M., Huffman J.E., Vitart V., Hayward C., Rudan I., Wilson J.F., Campbell H., Polasek O., Lauc G., Pivac N. The association between galactosylation of immunoglobulin G and body mass index. *Prog. Neuropsychopharmacol. Biol. Psychiatry.* 2014;48:20-25. DOI 10.1016/j.pnpbp.2013.08.014.
- Oda Y., Okada T., Yoshida H., Kaufman R.J., Nagata K., Mori K. Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. *J. Cell Biol.* 2006;172(3):383-393. DOI 10.1083/jcb.200507057.
- Ohtsubo K., Marth J.D. Glycosylation in cellular mechanisms of health and disease. *Cell.* 2006;126(5):855-867. DOI 10.1016/j.cell.2006.08.019.
- Pasaniuc B., Price A.L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* 2017;18(2):117-127. DOI 10.1038/nrg.2016.142.
- Peipp M., Lammerts van Bueren J.J., Schneider-Merck T., Bleeker W.W.K., Dechant M., Beyer T., Repp R., van Berkel P.H.C., Vink T., van de Winkel J.G.J., Parren P.W.H.I., Valerius T. Antibody fucosylation differentially impacts cytotoxicity mediated by NK and PMN effector cells. *Blood.* 2008;112(6):2390-2399. DOI 10.1182/blood-2008-03-144600.
- Pers T.H., Karjalainen J.M., Chan Y., Westra H.-J., Wood A.R., Yang J., Lui J.C., Vedantam S., Gustafsson S., Esko T., Frayling T., Speliotes E.K., Boehnke M., Raychaudhuri S., Fehrmann R.S.N., Hirschhorn J.N., Franke L. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* 2015;6(1):5890. DOI 10.1038/ncomms6890.
- Poole J., Day C.J., von Itzstein M., Paton J.C., Jennings M.P. Glycointeractions in bacterial pathogenesis. *Nat. Rev. Microbiol.* 2018; 16(7):440-452. DOI 10.1038/s41579-018-0007-2.
- Pottier N., Cheok M.H., Yang W., Assem M., Tracey L., Obenaus J.C., Panetta J.C., Relling M.V., Evans W.E. Expression of SMARCB1 modulates steroid sensitivity in human lymphoblastoid cells: identification of a promoter SNP that alters PARP1 binding and SMARCB1 expression. *Hum. Mol. Genet.* 2007;16(19):2261-2271. DOI 10.1093/hmg/ddm178.
- Reiding K.R., Bondt A., Hennig R., Gardner R.A., O'Flaherty R., Trbojević-Akmačić I., Shubhakar A., Hazes J.M.W., Reichl U., Fernandes D.L., Pučić-Baković M., Rapp E., Spencer D.I.R., Dolhain R.J.E.M., Rudd P.M., Lauc G., Wührer M. High-throughput serum *N*-glycomics: method comparison and application to study rheumatoid arthritis and pregnancy-associated changes. *Mol. Cell. Proteomics.* 2019;18(1):3-15. DOI 10.1074/mcp.RA117.000454.
- Reily C., Stewart T.J., Renfrow M.B., Novak J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* 2019;15(6):346-366. DOI 10.1038/s41581-019-0129-4.
- Rivinoja A., Hassinen A., Kokkonen N., Kauppila A., Kellokumpu S. Elevated Golgi pH impairs terminal *N*-glycosylation by inducing mislocalization of Golgi glycosyltransferases. *J. Cell. Physiol.* 2009; 220(1):144-154. DOI 10.1002/jcp.21744.
- Rogers M., Shihab H.A., Mort M., Cooper D., Gaunt T.R., Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics.* 2018;34(3):511-513. DOI 10.1093/bioinformatics/btx536.
- Russell A.C., Šimurina M., Garcia M.T., Novokmet M., Wang Y., Rudan I., Campbell H., Lauc G., Thomas M.G., Wang W. The *N*-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. *Glycobiology.* 2017;27(5):501-510. DOI 10.1093/glycob/cwx022.
- Saito M., Ishii A. T3Gal-V (GM3 synthase, SAT-I). In: *Handbook of Glycosyltransferases and Related Genes*. Tokyo: Springer, 2002; 289-294. DOI 10.1007/978-4-431-67877-9_39.
- Saldova R., Asadi Shehni A., Haakensen V.D., Steinfeld I., Hilliard M., Kifer I., Helland A., Yakhini Z., Børresen-Dale A.-L., Rudd P.M. Association of *N*-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J. Proteome Res.* 2014;13(5):2314-2327. DOI 10.1021/pr401092y.
- Sellars M., Reina-San-Martin B., Kastner P., Chan S. Ikaros controls isotype selection during immunoglobulin class switch recombination. *J. Exp. Med.* 2009;206(5):1073-1087. DOI 10.1084/jem.2008.2311.
- Shadrina A.S., Zlobin A.S., Zaytseva O.O., Klarić L., Sharapov S.Z., Pakhomov E., Perola M., Esko T., Hayward C., Wilson J.F., Lauc G., Aulchenko Y.S., Tsepilov Y.A. Multivariate genome-wide analysis

- of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function. *Hum. Mol. Genet.* 2021;30(13):1259-1270. DOI 10.1093/hmg/ddab072.
- Sharapov S.Z., Shadrina A.S., Tsepilov Y.A., Elgaeva E.E., Tiys E.S., Feoktistova S.G., Zaytseva O.O., ... Dagostino C., Gieger C., Allegri M., Williams F., Schulze M.B., Lauc G., Aulchenko Y.S. Replication of fifteen loci involved in human plasma protein N-glycosylation in 4,802 samples from four cohorts. *Glycobiology.* 2020; 31(2):82-88. DOI 10.1093/glycob/cwaa053.
- Sharapov S.Z., Tsepilov Y.A., Klaric L., Mangino M., Thareja G., Shadrina A.S., Simurina M., ... Louis E., Georges M., Suhre K., Spector T., Williams F.M.K., Lauc G., Aulchenko Y.S. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum. Mol. Genet.* 2019;28(12):2062-2077. DOI 10.1093/hmg/ddz054.
- Shen X., Klarić L., Sharapov S., Mangino M., Ning Z., Wu D., Trbojević-Akmačić I., Pučić-Baković M., Rudan I., Polašek O., Hayward C., Spector T.D., Wilson J.F., Lauc G., Aulchenko Y.S. Multi-variate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nat. Commun.* 2017;8(1):447. DOI 10.1038/s41467-017-00453-3.
- Skropeta D. The effect of individual N-glycans on enzyme activity. *Bioorg. Med. Chem.* 2009;17(7):2645-2653. DOI 10.1016/j.bmc.2009.02.037.
- Spector T.D., Williams F.M.K. The UK adult twin registry (TwinsUK). *Twin Res. Hum. Genet.* 2006;9(6):899-906. DOI 10.1375/183242706779462462.
- Staley J.R., Blackshaw J., Kamat M.A., Ellis S., Surendran P., Sun B.B., Paul D.S., Freitag D., Burgess S., Danesh J., Young R., Butterworth A.S. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics.* 2016;32(20):3207-3209. DOI 10.1093/bioinformatics/btw373.
- Staretz-Chacham O., Noyman I., Wormser O., Abu Quider A., Hazan G., Morag I., Hadar N., Raymond K., Birk O.S., Ferreira C.R., Koifman A. B4GALNT1-congenital disorders of glycosylation: Expansion of the phenotypic and molecular spectrum and review of the literature. *Clin. Genet.* 2020;97(6):920-926. DOI 10.1111/cge.13735.
- Taniguchi N., Honke K., Fukuda M., Narimatsu H., Yamaguchi Y., Angata T. (Eds.). *Handbook of Glycosyltransferases and Related Genes.* Tokyo: Springer, 2014. DOI 10.1007/978-4-431-54240-7.
- Taniguchi N., Kizuka Y. Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. *Adv. Cancer Res.* 2015;126:11-51. DOI 10.1016/bs.acr.2014.11.001.
- Thanabalasingham G., Huffman J.E., Kattla J.J., Novokmet M., Rudan I., Gloyn A.L., Hayward C., ... Hastie N.D., Campbell H., McCarthy M.I., Rudd P.M., Owen K.R., Lauc G., Wright A.F. Mutations in HNF1A result in marked alterations of plasma glycan profile. *Diabetes.* 2013;62(4):1329-1337. DOI 10.2337/db12-0880.
- Theodoratou E., Thaqi K., Agakov F., Timofeeva M.N., Štambuk J., Pučić-Baković M., Vučković F., Orchard P., Agakova A., Din F.V.N., Brown E., Rudd P.M., Farrington S.M., Dunlop M.G., Campbell H., Lauc G. Glycosylation of plasma IgG in colorectal cancer prognosis. *Sci. Rep.* 2016;6:28098. DOI 10.1038/srep28098.
- Tillin T., Forouhi N.G., McKeigue P.M., Chaturvedi N., SABRE Study Group. Southall and Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. *Int. J. Epidemiol.* 2012;41(1):33-42. DOI 10.1093/ije/dyq175.
- Trbojević Akmačić I., Ventham N.T., Theodoratou E., Vučković F., Kennedy N.A., Krištić J., Nimmo E.R., Kalla R., Drummond H., Štambuk J., Dunlop M.G., Novokmet M., Aulchenko Y., Gornik O., Campbell H., Pučić Baković M., Satsangi J., Lauc G. Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm. Bowel Dis.* 2015;21(6):1237-1247. DOI 10.1097/mib.0000000000000372.
- Uhlén M., Fagerberg L., Hallström B.M., Lindskog C., Oksvold P., Mardinoglu A., Sivertsson Å., ... Forsberg M., Persson L., Johansson F., Zwahlen M., von Heijne G., Nielsen J., Pontén F. Proteomics. Tissue-based map of the human proteome. *Science.* 2015; 347(6220):1260419. DOI 10.1126/science.1260419.
- Uhlén M., Karlsson M.J., Hober A., Svensson A.-S., Scheffel J., Kotol D., Zhong W., ... Voldborg B.G., Tegel H., Hober S., Forsström B., Schwenk J.M., Fagerberg L., Sivertsson Å. The human secretome. *Sci. Signal.* 2019;12(609):eaa0274. DOI 10.1126/scisignal.aaz0274.
- Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology.* 1993;3(2):97-130. DOI 10.1093/glycob/3.2.97.
- Varki A., Cummings R.D., Esko J.D., Stanley P., Hart G.W., Aebi M., Darvill A.G., Kinoshita T., Packer N.H., Prestegard J.H., Schnaar R.L., Seeberger P.H. (Eds.). *Essentials of Glycobiology.* [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2017.
- Varki A., Kornfeld S. Historical background and overview. In: Varki A., Cummings R.D., Esko J.D., Stanley P., Hart G.W., Aebi M., Darvill A.G., Kinoshita T., Packer N.H., Prestegard J.H., Schnaar R.L., Seeberger P.H. (Eds.). *Essentials of Glycobiology.* [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2017.
- Vilaj M., Lauc G., Trbojević-Akmačić I. Evaluation of different PNGase F enzymes in immunoglobulin G and total plasma N-glycans analysis. *Glycobiology.* 2021;31(1):2-7. DOI 10.1093/glycob/cwaa047.
- Visscher P.M., Brown M.A., McCarthy M.I., Yang J. Five years of GWAS discovery. *Am. J. Hum. Genet.* 2012;90(1):7-24. DOI 10.1016/j.ajhg.2011.11.029.
- Visscher P.M., Wray N.R., Zhang Q., Sklar P., McCarthy M.I., Brown M.A., Yang J. 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* 2017;101(1):5-22. DOI 10.1016/j.ajhg.2017.06.005.
- Vreeker G.C.M., Nicolardi S., Bladergroen M.R., van der Plas C.J., Mesker W.E., Tollenaar R.A.E.M., van der Burgt Y.E.M., Wuhrer M. Automated plasma glycomics with linkage-specific sialic acid esterification and ultrahigh resolution MS. *Anal. Chem.* 2018;90(20): 11955-11961. DOI 10.1021/acs.analchem.8b02391.
- Vučković F., Theodoratou E., Thaqi K., Timofeeva M., Vojta A., Štambuk J., Pučić-Baković M., Rudd P.M., Đerek L., Servis D., Wennerström A., Farrington S.M., Perola M., Aulchenko Y., Dunlop M.G., Campbell H., Lauc G. IgG glycome in colorectal cancer. *Clin. Cancer Res.* 2016;22(12):3078-3086. DOI 10.1158/1078-0432.CCR-15-1867.
- Wahl A., van den Akker E., Klaric L., Štambuk J., Benedetti E., Plomp R., Razdorov G., Trbojević-Akmačić I., Deelen J., van Heemst D., Slagboom P.E., Vučković F., Grallert H., Krumsiek J., Strauch K., Peters A., Meitinger T., Hayward C., Wuhrer M., Beekman M., Lauc G., Gieger C. Genome-wide association study on immunoglobulin G glycosylation patterns. *Front. Immunol.* 2018;9: 277. DOI 10.3389/fimmu.2018.00277.
- Wang Y., Klarić L., Yu X., Thaqi K., Dong J., Novokmet M., Wilson J., Polasek O., Liu Y., Krištić J., Ge S., Pučić-Baković M., Wu L., Zhou Y., Ugrina I., Song M., Zhang J., Guo X., Zeng Q., Rudan I., Campbell H., Aulchenko Y., Lauc G., Wang W. The association between glycosylation of immunoglobulin G and hypertension. *Medicine (Baltimore).* 2016;95(17):e3379. DOI 10.1097/md.0000000000003379.
- Winkler T.W., Day F.R., Croteau-Chonka D.C., Wood A.R., Locke A.E., Mägi R., Ferreira T., Fall T., Graff M., Justice A.E., Luan J., Gustafsson S., Randall J.C., Vedantam S., Workalemahu T., Kilpeläi-

- nen T.O., Scherag A., Esko T., Kutalik Z., Heid I.M., Loos R.J.F., Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* 2014;9(5):1192-1212. DOI 10.1038/nprot.2014.071.
- Yang J., Ferreira T., Morris A.P., Medland S.E., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, Madden P.A.F., Heath A.C., Martin N.G., Montgomery G.W., Weedon M.N., Loos R.J., Frayling T.M., McCarthy M.I., Hirschhorn J.N., Goddard M.E., Visscher P.M. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* 2012;44(4):369-375, S1-3. DOI 10.1038/ng.2213.
- Zaytseva O.O., Freidin M.B., Keser T., Štambuk J., Ugrina I., Šimurina M., Vilaj M., Štambuk T., Trbojević-Akmačić I., Pučić-Baković M., Lauc G., Williams F.M.K., Novokmet M. Heritability of human plasma N-glycome. *J. Proteome Res.* 2020;19(1):85-91. DOI 10.1021/acs.jproteome.9b00348.
- Zhu Z., Zhang F., Hu H., Bakshi A., Robinson M.R., Powell J.E., Montgomery G.W., Goddard M.E., Wray N.R., Visscher P.M., Yang J. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 2016;48(5):481-487. DOI 10.1038/ng.3538.

ORCID ID

S.Zh. Sharapov orcid.org/0000-0003-0279-4900
A.N. Timoshchuk orcid.org/0000-0001-8529-4498
Y.S. Aulchenko orcid.org/0000-0002-7899-1575

Acknowledgements. This article is an output of a research project implemented as part of the Research Program at the MSU Institute for Artificial Intelligence. The work of Y.S.A. was supported in part by the state-funded project No. FWNR-2022-0020.

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

Received November 17, 2022. Revised January 20, 2023. Accepted January 23, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Species identification of spider mites (Tetranychidae: Tetranychinae): a review of methods

A.V. Razuvaeva¹, E.G. Ulyanova², E.S. Skolotneva¹✉, I.V. Andreeva²

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

² Siberian Federal Scientific Centre of Agro-BioTechnologies of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk Region, Russia

✉ skolotnevaES@bionet.nsc.ru

Abstract. Spider mites (Acari: Tetranychidae) are dangerous pests of agricultural and ornamental crops, the most economically significant of them belonging to the genera *Tetranychus*, *Eutetranychus*, *Oligonychus* and *Panonychus*. The expansion of the distribution areas, the increased harmfulness and dangerous status of certain species in the family Tetranychidae and their invasion of new regions pose a serious threat to the phytosanitary status of agro- and biocenoses. Various approaches to acarofauna species diagnosis determine a rather diverse range of currently existing methods generally described in this review. Identification of spider mites by morphological traits, which is currently considered the main method, is complicated due to the complexity of preparing biomaterials for diagnosis and a limited number of diagnostic signs. In this regard, biochemical and molecular genetic methods such as allozyme analysis, DNA barcoding, restriction fragment length polymorphism (PCR-RFLP), selection of species-specific primers and real-time PCR are becoming important. In the review, close attention is paid to the successful use of these methods for species discrimination in the mites of the subfamily Tetranychinae. For some species, e.g., the two-spotted spider mite (*Tetranychus urticae*), a range of identification methods has been developed – from allozyme analysis to loop isothermal amplification (LAMP), while for many other species a much smaller variety of approaches is available. The greatest accuracy in the identification of spider mites can be achieved using a combination of several methods, e.g., examination of morphological features and one of the molecular approaches (DNA barcoding, PCR-RFLP, etc.). This review may be useful to specialists who are in search of an effective system for spider mite species identification as well as when developing new test systems relevant to specific plant crops or a specific region.

Key words: spider mites; species identification; allozyme analysis; MALDI-TOF MS; DNA barcoding; PCR-RFLP; ITS; COI.

For citation: Razuvaeva A.V., Ulyanova E.G., Skolotneva E.S., Andreeva I.V. Species identification of spider mites (Tetranychidae: Tetranychinae): a review of methods. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(3):240-249. DOI 10.18699/VJGB-23-30

Видовая идентификация паутиных клещей (Tetranychidae: Tetranychinae): обзор методов

А.В. Разуваева¹, Е.Г. Ульянова², Е.С. Сколотнева¹✉, И.В. Андреева²

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

² Сибирский федеральный научный центр агробиотехнологий Российской академии наук, р. п. Краснообск, Новосибирская область, Россия

✉ skolotnevaES@bionet.nsc.ru

Аннотация. Паутиные клещи (Acari: Tetranychidae) являются опасными вредителями сельскохозяйственных и декоративных культур; наиболее экономически значимые из них относятся к родам *Tetranychus*, *Eutetranychus*, *Oligonychus* и *Panonychus*. Расширение ареалов распространения, усиление вредоносности и статуса по степени опасности отдельных видов тетранихид, инвазия опасных вредителей в новые регионы представляют серьезную угрозу для фитосанитарного состояния агро- и биоценозов. Различные подходы к видовой диагностике акарофауны определяют достаточно разнообразный спектр существующих в настоящее время методов, общая информация по которым приведена в данном обзоре. Идентификация паутиных клещей по морфологическим признакам, считающаяся основным методом, осложнена вследствие трудоемкости подготовки биоматериала для диагностики и ограниченного числа диагностических признаков. В связи с этим важное значение приобретают биохимические и молекулярно-генетические методы, такие как аллозимный анализ, ДНК-штрихкодирование, полиморфизм длины рестрикционного фрагмента (ПЦР-ПДФ), подбор видоспецифичных праймеров, ПЦР в реальном времени и изотермическая амплификация. В обзоре пристальное внимание уделено успехам применения этих методов для видовой дискриминации клещей подсемейства Tetranychinae. Для некоторых видов, например обыкновенного паутинового клеща (*Tetranychus urticae*), разработан спектр методов идентификации – от аллозимного анализа до петлевой изотермической амплификации (LAMP), тогда как для многих других видов доступно гораздо меньшее разнообразие подходов. Наибольшей точности

в определении паутиных клещей можно добиться, используя комбинацию нескольких методов, например осмотр морфологических признаков и один из молекулярных подходов (ДНК-штрихкодирование, ПЦР-ПДРФ и др.). Обзор может быть полезен специалистам, находящимся в поиске эффективной системы для видовой дискриминации паутиных клещей, а также при разработке новых тест-систем, актуальных для конкретных культур растений или определенного региона.

Ключевые слова: паутиные клещи; молекулярная идентификация; аллозимный анализ; MALDI-TOF MS; ДНК-штрихкодирование; ПЦР-ПДРФ; ITS; COI.

Introduction

Herbivorous mites cause significant damage to agricultural and ornamental crops. The harmfulness of the mites is manifested in a decrease in yield and deterioration in the quality of crop production; leads to a decrease in drought resistance and winter hardiness as well as to the loss of decorative properties of cultivated plants (Devi et al., 2019; Chaires-Grijalva et al., 2021; Obasa et al., 2022; Ulyanova et al., 2022). Changes in weather and climate conditions in almost all regions of the world, including the Russian Federation, contribute to a wider spread and mass reproduction of pests, transformation of their geographical habitats, changes in the population dynamics and trophic relationships with the host plants (Musolin, Saulich, 2012; Zeynalov, 2017; Koshkin et al., 2021), which fully applies to herbivorous acarofauna (Volkova, Matveikina, 2016; Zeynalov, Orel, 2021), e.g., expansion of habitats, increased harmfulness, and the emergence of populations resistant to acaricides have recently been observed in spider mites (Acari: Tetranychidae), one of the main groups of phytophagous mites (Balykina et al., 2017; Zeynalov, Orel, 2021). The phytosanitary situation is complicated by the high risk of introducing quarantine species with the plant material imported as part of international trade for the purpose of growing, propagating and selling flower and other plant products (Rak, Litvinova, 2010; Petrov et al., 2016; Kamayev, 2018; Kamayev, Mironova, 2018; Vásquez, Colmenárez, 2020).

The Tetranychidae family is subdivided into two subfamilies, Bryobiinae and Tetranychinae, and includes at least 71 genera and more than 1250 described species, 100 of which are dangerous pests (Migeon et al., 2010). The most common species of the Tetranychidae family belong to genera *Tetranychus*, *Eutetranychus*, *Oligonychus*, and *Panonychus* (Ben-David et al., 2007). Among them, the two-spotted spider mite (*Tetranychus urticae* Koch.) and European red mite (*Panonychus ulmi* Koch.) are regarded as the most harmful species (Ben-David et al., 2007). *T. urticae* is ubiquitous and damages a wide range of crops as well as ornamental woody and herbaceous plants from various botanical families. The habitats of the two-spotted spider mite and other representatives of this family in open-ground agro- and biocenoses have been gradually expanding, covering more and more territories of Russia (Kamayev, Karpun, 2020; Ulyanova et al., 2022). The harmfulness of the spider mites that damage coniferous plants has been increasing as well. Thus in the south

of Western Siberia, the spruce spider mite (*Oligonychus ununguis* Jacobi) damaging the European and Siberian spruces and Siberian fir used for planting greenery in urban areas has become a threat (Ulyanova et al., 2022). The risk of importing quarantine pests with planting material has also increased, e.g., the sugi spider mite (*Oligonychus hondoensis* Ehara), invasive for this territory, was found in the Krasnodar Region (Kamayev, Karpun, 2020).

Correct identification of spider mite species is of great scientific and practical importance for the study of their population dynamics and timely control of their numbers in agro- and biocenoses as well as for the elimination of international plant quarantine-based trade barriers (Li et al., 2015). Currently, several methods are in use to diagnose the members of the Tetranychidae family, including identification by the morphological characteristics of adult as well as biochemical (protein-based) and molecular (DNA-based) methods. The objective of this review is to consider the modern methods and approaches used to identify the most common species of spider mites (Tetranychidae: Tetranychinae).

Morphological methods

At present, the identification of herbivorous mites is carried out mainly by the traditional methods based on visual examination of morphological features. The foundations of the morphological method in determining spider mite were laid in the former USSR by prominent scientists V.I. Mitrofanov, I.Z. Livshits and Z.I. Strunkova who hugely contributed to the establishment of a whole area of scientific research devoted to species Tetranychidae family diagnosis (Mitrofanov et al., 1987). The approach was further developed in the research and applied works by S.Ya. Popov et al. (Popov, 2013), A.K. Akhatov (2016) et al., many of which still remain relevant.

Determination of the mite's genus and species is usually based on the shape and size of male genitalia (Morphological Identification..., 2014). However, their microscopic size, slight differences in diagnostic features in the species belonging to the same genus, and the laboriousness of preparing the biomaterial for analysis significantly complicate their morphological identification (Konoplev et al., 2017). Another drawback of this approach is the impossibility of determining the species based on other stages of their development (eggs, larvae, nymphs), since diagnostic differences exist only in adults. Moreover, for some closely related species, morphological identification is almost

impossible, e. g., genus *Amphitetranychus* includes three species and only one of them, *A. viennensis*, can be recognized by the shape of the aedeagus, while *A. quercivorus* and *A. savenkoae* are difficult to separate based on this trait (Arabuli et al., 2019).

The slides are prepared using modified Faure–Berlese media, of which Hoyer’s medium (Walter, Krantz, 2009) is the most common for fixing herbivorous mites and consists of 50 ml of distilled water, 30 ml of gum arabic, 200 ml of chloral hydrate, 20 ml of glycerin. Mites are placed in a drop of medium on a glass slide and covered with a coverslip. The slide is heated at temperatures of 40–60 °C with exposure time varying from 24 hours to 5–10 days, or 3 hours at 70–85 °C (Kamayev, 2019), which contributes to the clarification and straightening of the mite, so its proper diagnosis can be performed. If slides are to be stored for a long time, one is recommended to dry them additionally in a thermostat at a temperature of 40–45 °C for 5–7 days. After drying, the edges of the coverslips are filled with varnish, so the slides can be stored indefinitely. Mite species are diagnosed in transmitted light using a phase-contrast microscope at a 10 to 1000-time magnification.

In general, the morphological method requires considerable time and is very demanding in terms of qualification and experience of those who apply it. In some cases, morphological identification is supplemented by ecological and behavioral responses of the species, as well as by information on the host plant on which the species has been found, which can facilitate species identification (Akhatov, 2016). Some features of the life cycle, such as the diapausing phase, wintering sites, diapause exit time, concentration on certain organs of host plants, and the specificity of damage symptoms, can also play an important role in the species diagnosis.

Species crossing

Genetic incompatibility when crossing closely related species of tetranychid mites is considered one of the best criteria for their discrimination. The reproductive barriers can be caused either by the morphological features (the size of the aedeagus knob exceeding the size of the female epigyne) or by the death of zygotes, which affects the viability of eggs or manifests itself in a different sex ratio in the offspring. For instance, the experiments of Russian researchers on crossing between species *Tetranychus atlanticus*, *T. urticae*, and *T. sawzdargi* showed their complete genetic isolation (Popov, 2013). The reproductive isolation of two morphologically related species of *Amphitetranychus* spp., manifesting itself in the absence of female offspring, was confirmed in the reciprocal crossing of *A. savenkoae* and *A. quercivorus* (Arabuli et al., 2019).

According to a number of scientists (Popov, 2013; Arabuli et al., 2019), a combination of the morphological method with the crossing results for closely related or distant species is a reliable criterion for spider mite species

diagnosis. Unfortunately, this method is only applicable for scientific purposes, and cannot be used for practical applications.

Biochemical methods

One of the biochemical methods for identifying species of spider mites is **allozyme analysis** that is protein electrophoresis enabling for efficient detection of enzyme polymorphisms (Navajas, Fenton, 2000). To perform the analysis, a single individual is homogenized, its protein extract is isolated and subjected to electrophoresis. In the electric field, proteins are separated according to their size and net electrical charge. After electrophoresis, the gel is histochemically stained to detect specific enzymes. Based on the number and arrangement of the stained fractions, one can make judgments about the alleles encoding a given enzyme in each individual (Kutlunina, Ermoshin, 2017). For correct identification, it is important to choose the enzyme system in which there will be no variability in stripe patterns within a species (polymorphic alleles) and common stripes across species (Gotoh et al., 2007).

The analysis has been widely used to identify insects (Turak, Hales, 1994) and ixodid ticks (Lampo et al., 1997). It has also been applied for spider mite identification, using such enzymes as esterase, phosphoglucosemerase, and malate dehydrogenase (Enohara, Amano, 1996; Goka, Takafuji, 1998; Gotoh et al., 2004, 2007; Arabuli et al., 2019). Such spider mite species as *Panonychus citri* and *P. mori* were first identified using allozyme analysis of esterases (Osakabe, 1987). Later it was shown that the same method could be used to identify three more species of the *Panonychus* mites endemic for Japan (Gotoh, 1992) (see the Table). Using esterase zymograms, Gotoh et al. (2007) were able to distinguish females of four *Tetranychus* species, and based on the analysis of phosphoglucomutase isoenzymes – all 13 Japanese species of genus *Tetranychus* (see the Table). The allozyme analysis of esterases was also used for species diagnosis of the representatives of genus *Amphitetranychus* (Arabuli et al., 2019). The obtained zymograms were species-specific and made it possible to confidentially distinguish all three species (see the Table).

Thus, allozyme analysis is an effective tool for identifying spider mite species and has a much higher resolution than the morphological method. However, it does not evaluate all possible allele variants present in populations (Altukhov, 2003).

Another biochemical method proposed for the identification of spider mites is **matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)** (Kajiwara et al., 2016) that has been a well-established method for identifying microbial species (Singhal et al., 2015). In recent years, this method has also been used to investigate plant pests and insects, such as nematodes, *Drosophila* fruit flies, mosquitoes, etc. (Ahmad et al., 2012).

In this method, species identification is achieved via comparing the spectra of the studied samples with the samples from the spectral database of previously identified organisms. The databases provided by instrument manufacturers currently do not contain reference spectra for plant pests and parasites, but the instrument software usually provides an option to integrate your own reference spectra into existing databases and subsequently perform automated comparisons (Murugaiyan, Roesler, 2017). To create the reference spectra, it is necessary to use the samples whose species have been identified by, e.g., morphological methods.

One of the method's important advantages if compared to other identification techniques is the fast sample preparation and short testing time (Murugaiyan, Roesler, 2017). For diagnosis, it is sufficient having just one individual, whose protein extract must first be isolated using 70 % formic acid and acetonitrile, or the mite can be directly located on the metal target of the device, where formic acid and acetonitrile are added, followed by a matrix solution (Kajiwara et al., 2016). The method's main disadvantage is the high cost of the device, but the price of one analysis is very low and includes just the cost of the matrix solution and the calibration standard (provided that there is a reusable steel plate for samples).

Kajiwara et al. (2016) showed that closely related species of *Tetranychus* spider mites (see the Table) had different mass spectra allowing for their identification. The identification was performed by comparing the three main spectral peaks within a m/z range from 2000 to 10 000 corresponding to ribosomal proteins. In adults of both sexes, all three peaks were differentiated, which made it possible to attribute them to the same species. In the case of nymphs, only two of the three main peaks could be identified. Minor peaks were either sex- or developmental stage-specific and were not used for identification (Kajiwara et al., 2016).

In general, MALDI-TOF MS is a promising method for the identification of spider mites, but its application is still limited and requires further research. In particular, there may be spectral differences between geographically different populations, which has been shown for some insects (see Murugaiyan, Roesler, 2017 and references to this article), and it is also assumed that the spectra of one species can be influenced by its host plant (Kajiwara et al., 2016).

Molecular methods

For the purposes of mite species diagnosis, the molecular methods based on DNA sequencing are used along with biochemical ones. In these methods, the DNA fragments amplified by PCR serve as biological barcodes to identify the species a sample belongs to (Hajibabaei et al., 2007). Both nuclear and mitochondrial genome sequences can be used as such barcodes. Since more and less conserved regions alternate within a ribosomal DNA cluster (genes 18S, 5.8S, 28S rRNA and internal transcribed spacers 1, 2

(ITS1, ITS2), respectively), it becomes possible to select different marker sequences to separate taxa of different ranks. The 18S and 28S rDNA gene sequences can be used to compare phylogenetically distant taxa (Matsuda et al., 2014), and the ITS regions are effective for distinguishing between species and even populations (Hillis, Dixon, 1991; Hurtado et al., 2008). Among the mitochondrial genes, the cytochrome *c* oxidase subunit I (*COI*) gene sequence is used to identify species and analyze phylogenetic relationships. In addition to the Folmer fragment (Barcode of Life (BOLD), Folmer et al., 1994) being the 5'-terminal sequence of the *COI* gene, other gene sequences are used for barcoding spider mites (Hinomoto et al., 2007; Ros, Breeuwer, 2007; Matsuda et al., 2013; İnak et al., 2022).

In (Ben-David et al., 2007), the authors used the ITS2 sequence as a barcode to distinguish 16 species of spider mites of the Tetranychidae family endemic for Israel (see the Table), and the barcode sequence of each type was uniquely distinguishable from all others. The effectiveness of the DNA barcoding for species identification was also confirmed by *COI* gene fragment-based identification of the spider mite species collected from agricultural fields in Vietnam (Hinomoto et al., 2007) (see the Table).

The results also demonstrated the method's limitations, e.g., some samples could not be classified due to lack of sequence information in the databases. Other species, such as *T. urticae* and *T. turkestanii*, *T. neocaledonicus* and *T. gloveryi*, could not be uniquely identified by their DNA sequences. The impossibility to distinguish between *T. urticae* and *T. turkestanii* based on the *COI* gene sequence was also confirmed in the mites collected in different regions of Russia and Turkey (Konoplev et al., 2017; İnak et al., 2022), which can be overcome by further accumulation of data on the sequences of different DNA regions from morphologically identified samples. Matsuda et al. (2013) identified the mite species of genus *Tetranychus* by sequencing the ITS and *COI* genes. The authors concluded that 10 out of 13 species of the Japanese mites of genus *Tetranychus* can be identified using the ITS sequence, while the *COI* gene sequence made it possible to identify all 13 species (see the Table).

In contrast to genus *Tetranychus*, where some species cannot be distinguished based on DNA sequences, this method proved to be effective for genus *Oligonychus* (Matsuda et al., 2012), so all 17 Japanese species were successfully distinguished based on any of the sequences (*COI*, ITS, and 28S rDNA), including those that were difficult to differentiate by morphological characters, such as *O. castaneae* and *O. coffeae* (see the Table). Molecular analysis of the *COI* gene sequence of the mites of genus *Amphitetranychus* showed that the sequence is effective for the identification of all three species (Arabuli et al., 2019) (see the Table).

Thus, the DNA barcoding is an effective tool for the identification of individual genera of spider mites (such

Systems developed for identification of spider mite species

Diagnosis method	Species	Allozyme analysis					MALDI-TOF MS				DNA barcoding				PCR RFLP				Species-specific primers			TaqMan PCR		LAMP
		Osakabe, 1987	Gotoh, 1992	Gotoh et al., 2007	Arabuli et al., 2019	Kajiwara et al., 2016	Ben-David et al., 2007	Hinomoto et al., 2007	Matsuda et al., 2012	Matsuda et al., 2013	Arabuli et al., 2019	Osakabe et al., 2008	Arimoto et al., 2013	Hurtado et al., 2008	Ovalle et al., 2020	Khaing et al., 2015	Shim et al., 2016	Zélé et al., 2018	Sinaie et al., 2018	Li et al., 2015	Chen et al., 2020	Sinaie et al., 2019		
Tetranychus	<i>urticae</i>			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>ludeni</i>			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>kanzawai</i>			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>phaselus</i>			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>evansi</i>				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>truncatus</i>			+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>piercei</i>			+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>okinawanus</i>			+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>turkestani</i>						+						+	+					+					
	<i>pueraricola</i>			+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>parakanzawai</i>			+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>ezoensis</i>			+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>neocaledonicus</i>			+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>takafujii</i>			+							+	+	+	+	+	+	+	+	+	+	+	+	+	
	<i>lambi</i>												+	+	+	+	+	+	+	+	+	+	+	
	<i>malaysiensis</i>												+	+	+	+	+	+	+	+	+	+	+	
	<i>pacificus</i>												+	+	+	+	+	+	+	+	+	+	+	
	<i>merganser</i>												+	+	+	+	+	+	+	+	+	+	+	
<i>misumaiensis</i>								+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
Panonychus	<i>citri</i>	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>ulmi</i>		+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>mori</i>	+						+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>thelytokus</i>		+																					
	<i>bambusicola</i>		+																					
	<i>caglei</i>																							
Oligonychus	<i>osmanthi</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>coffea</i>								+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>gotohi</i>								+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>perseae</i>						+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>afraziaticus</i>						+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>mangiferus</i>						+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>clavatus</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>pustulosus</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>karamatus</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>hondoensis</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>tsudomei</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>ilicis</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>camelliae</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>ununguis</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>perditus</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>castaneae</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>amiensis</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>orthius</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>modestus</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>biharensis</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
<i>rubicundus</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
Mononychellus	<i>caribbeanae</i>													+	+	+	+	+	+	+	+	+		
	<i>mcgregori</i>													+	+	+	+	+	+	+	+	+		
	<i>tanajoa</i>													+	+	+	+	+	+	+	+	+		
Amphitetranychus	<i>viennensis</i>				+					+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>quercivorus</i>				+					+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>savencoae</i>				+					+	+	+	+	+	+	+	+	+	+	+	+	+		
Schizotetranychus	<i>asparagi</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+			
Eutetranychus	<i>palmatus</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
	<i>orientalis</i>							+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		

Note. The "+" sign denotes a species available for identification.

as *Oligonychus* and *Amphitetranynchus*). However, some closely related species (e. g., from genus *Tetranychus*) may be indistinguishable from each other by their barcode sequences, as it has been shown for *T. kanzawai*, *T. parakanzawai*, and *T. ezoensis* having similar ITS sequences (Matsuda et al., 2013); and *T. urticae* and *T. turkestanii* having similar *COI* gene sequences (Hinamoto et al., 2007; Konoplev et al., 2017). For these species, other marker genes should be selected.

The desire to reduce the time spent for analysis has led to the development of faster methods for identifying spider mites, one of these being **PCR-restriction fragment length polymorphism (PCR-RFLP)**. In this method, fragments of genomic DNA (nuclear or mitochondrial), amplified by PCR, are subjected to hydrolysis by restriction endonucleases. The restriction products are subjected to gel electrophoresis, so a conclusion can be made about the presence or absence of a restriction site in this sample and, as a result, a species can be identified.

PCR-RFLP is widely used for species identification of various organisms (Ratcliffe et al., 2003; Han et al., 2004; Alam et al., 2007). In spider mites, the method was first applied to distinguish between three species of genus *Panonychus* and *T. urticae* (Osakabe, Sakagami, 1994) (see the Table). Later, Gotoh et al. (1998) were able to differentiate *T. urticae* and *T. pueraricola* by applying PCR-RFLP to the ITS2 region. In (Hurtado et al., 2008), the authors applied the method to the ITS fragment (ITS1, 5.8S rDNA, ITS2) to identify five species of spider mites that inhabited Spanish citrus orchards (see the Table). The most harmful was the common spider mite (*T. urticae*) that made it necessary to quickly distinguish it among the others, which was achieved through the *RsaI* enzyme patterning the restriction fragments, while hydrolysis with two other enzymes made it possible to identify the other four species (Hurtado et al., 2008). The PCR-RFLP applied to ITS regions is used in the Japanese Imported Plant Quarantine Department (Arimoto et al., 2013; Li et al., 2015) and enables one to detect all 14 spider mite species found in Japan, including five exotic ones (Osakabe et al., 2002, 2008; Arimoto et al., 2013) (see the Table). Meanwhile, the PCR-RFLP applied to the *COI* region was successfully used to identify four spider mite species thriving on cassava in Colombia (Ovalle et al., 2020) (see the Table).

PCR-RFLP is a relatively inexpensive and effective method for spider mite identification, used in many countries around the world (Hurtado et al., 2008; Arimoto et al., 2013; Ovalle et al., 2020). However, it can only be applied to those species it has been developed for, e. g., to identify mites in an area whose species composition is well established. When new species are added to the system, their RFLP pattern should be analyzed and, if the pattern matches the existing ones, the system should be modified by selecting other diagnostic restrictases. The RFLP pattern can also vary in different populations of the same species

due to nucleotide substitutions at the restriction site (Arimoto et al., 2013).

Another approach for rapid mite identification has been selection of **species-specific primers**. The DNA regions where such primers are defined should be common for the species being determined, but the sequences themselves should differ between species of the same genus (Shim et al., 2016). In this case, both whole pairs of species-specific primers can be used, as well as a combination of one universal primer and the second primer being unique for a particular species. Such primers increase the accuracy and speed of PCR-based identification of spider mite species, but their development is not a trivial task. Khaing et al. (2014) succeeded in selecting pairs of species-specific primers for the ITS2 region for four species of spider mites belonging to genus *Panonychus* (see the Table). Later, species-specific primers were developed for the spider mites of genus *Tetranychus* (see the Table) common in Korea and very similar morphologically (Shim et al., 2016).

Combining several species-specific primers in a single tube allows applying the **multiplex approach** to identify several species at once. In this case, the melting temperature of the primers should be high, and the resulting amplicons – short, but different in length. Additionally, multiplex PCR requires selecting amplification conditions, such as the ratio of primer pairs and the elongation time (Zélé et al., 2018). Zélé et al. (2018) designed and successfully multiplexed primers complementary to the ribosomal locus in a single reaction to identify the most common spider mites found in southwestern Europe (see the Table). Multiplex PCR has also been used to discriminate between the two main spider mite species, *T. urticae* and *T. turkestanii* found in Iranian greenhouses (Sinaie et al., 2018).

Using **real-time PCR** for species identification significantly reduces the time for analysis because this method does not require gel electrophoresis since the accumulation of the PCR product is monitored directly during the reaction using optical sensors built into the cycler. Two types of labels are used to detect a PCR product: intercalating agents (e. g., SYBR Green) or modified oligonucleotides containing fluorophores (DNA probes) (Bikbulatova et al., 2012). In addition, real-time PCR analysis combines primer annealing and elongation steps, resulting in a shorter reaction time than conventional PCR (Li et al., 2015).

Li et al. (2015) developed the TaqMan PCR detection system to distinguish *T. urticae* among other closely related species that has proved to be highly specific and reliable. TaqMan is one of hybridization DNA probes – an oligonucleotide complementary to the amplified internal region of the DNA fragment, labeled at the ends with fluorophores – being a reporter and a quencher, respectively. When they are on the same probe, the quencher absorbs the signal from the reporter. During amplification, the polymerase moving along the DNA destroys the probe, so the reporter and quencher move away from each other, and the

reporter's fluorescence becomes noticeable (Bikbulatova et al., 2012). The species-specific primers and probe were designed for *T. urticae*'s ITS1 sequence, since it contained more intraspecific polymorphisms than the ITS2 sequence widely used for *Tetranychus* species phylogeny and identification (Li et al., 2015).

Another team (Chen et al., 2020) developed TaqMan species-specific probes for the identification of spider mites thriving in the cotton fields of Australia (see the Table), and selected the conditions for their use in one test tube (multiplex approach). An important feature of this approach was the possibility of extending it to other types of spider mites with minor modification. Since it used a pair of primers universal for spider mites for amplification, it was necessary, when adding another species to the test system, to develop and add a species-specific probe for it and select reaction conditions (Chen et al., 2020). The authors used DNA probes for three species of mites in one reaction and considered it was possible to increase their number to five (add or replace with DNA probes for the required species) to diagnose up to five different mite species with one PCR.

Another method that can be used for species identification purposes is **loop mediated isothermal amplification (LAMP)** (Tomita et al., 2008). Unlike the classical PCR, LAMP uses a different thermostable polymerase with high displacement capacity that can itself displace the second strand without thermal denaturation. That is why LAMP reaction takes place at the same temperature (60–65 °C) and does not require an DNA amplifier. Another feature of LAMP is that not two, but four or six primers are used during the reaction, which determines its high specificity (Notomi et al., 2000). On the other hand, selecting such primers for LAMP is a rather laborious task. The reaction's product can be detected in several ways, including the naked eye after adding a fluorescent dye to confirm the reaction has taken place (Tomita et al., 2008). Thus, this method not only does not require special equipment or specially trained specialists, but is also suitable for large-scale field studies (Hsieh et al., 2012; Sinaie et al., 2019).

As a well-established technique, LAMP has been used in a wide range of applications, including the identification of plant pathogens and insect species (Ahuja, Somvanshi, 2021; Dermauw et al., 2022). For the identification of the common spider mite, a highly sensitive method combining PCR and LAMP (PCR-LAMP) has been developed (Sinaie et al., 2019). It includes the standard PCR performed prior to isothermal amplification to reduce the false negatives and increase the sensitivity of the LAMP assay. The authors have shown that PCR-LAMP is a fast and reliable method for *T. urticae* biomaterial detection.

Conclusion

Our review of currently available methods for diagnosing the most significant species of herbivorous mites (Tetranychidae: Tetranychinae) has demonstrated that a wide

variety of approaches have been accumulated that can be used by choice, depending on the goals of species identification. The morphological methods, despite the complexity of required preliminary operations to prepare mites for analysis, still remain the main ones in determining this group of phytophages. However, the new approaches developed in recent decades and based on biochemical and molecular markers open up even greater opportunities for faster and more accurate identification of species. For instance, some species of spider mites such as *T. urticae* can be properly diagnosed using one of the molecular approaches (DNA barcoding, PCR-RFLP, TaqMan PCR, etc.). At the same time, some other species have not yet been diagnosed based on their DNA sequences. In such cases, the use of several identification methods at once is justified. Such an integrated approach was applied to the diagnosis of the species of genus *Amphitetranychus* (Arabuli et al., 2019) whose morphological characteristics were supplemented by biochemical, and molecular ones and verified in crossing experiments. It is important to note that several approaches can be implemented on the same individual, provided that DNA extraction methods are used without destroying the sample (Khaing et al., 2013; Shim et al., 2016) or a whole mite is used as a PCR matrix (direct PCR) (Sakamoto, Gotoh, 2017), and for subsequent morphological characterization. It has been shown that, after such manipulations, specimens retain their morphological features intact, including dorsal setae and aedeagus (Sakamoto, Gotoh, 2017).

References

- Ahmad F., Babalola O.O., Tak H.I. Potential of MALDI-TOF mass spectrometry as a rapid detection technique in plant pathology: identification of plant-associated microorganisms. *Anal. Bioanal. Chem.* 2012;40(4):1247-1255. DOI 10.1007/s00216-012-6091-7.
- Ahuja A., Somvanshi V.S. Diagnosis of plant-parasitic nematodes using loop-mediated isothermal amplification (LAMP): a review. *Crop Prot.* 2021;147:105459. DOI 10.1016/j.cropro.2020.105459.
- Akhatov A.K. Guide to the Identification of Acarids and Insects in Vegetable Greenhouses. Moscow: KMK Publ., 2016. (in Russian)
- Alam M.T., Das M.K., Dev V., Ansari M.A., Sharma Y.D. PCR-RFLP method for the identification of four members of the *Anopheles annularis* group of mosquitoes (Diptera: Culicidae). *Trans. R. Soc. Trop. Med. Hyg.* 2007;101(3):239-244. DOI 10.1016/j.trstmh.2006.03.007.
- Altukhov Yu.P. Genetic Processes in Populations. Moscow: Akademkniga Publ., 2003. (in Russian)
- Arabuli T., Negm M.W., Matsuda T., Kitashima Y., Abramishvili T., Akimov I.A., Zhovnerchuk O.V., Popov S.Ya., Gotoh T. Morphological identification of *Amphitetranychus* species (Acari: Tetranychidae) with crossbreeding, esterase zymograms and DNA barcode data. *PLoS One.* 2019;14(9):e0221951. DOI 10.1371/journal.pone.0221951.
- Arimoto M., Satoh M., Uesugi R., Osakabe M. PCR-RFLP analysis for identification of *Tetranychus* spider mite species (Acari: Tetranychidae). *J. Econ. Entomol.* 2013;106(2):661-668. DOI 10.1603/EC12440.
- Balykina E.B., Yagodinskaya L.P., Rybareva T.S., Balitsky N.V. Influence of acaricides on the structure of a phytophagous acarid complex.

- Bulleten Gosudarstvennogo Nikitskogo Botanicheskogo Sada = Bulletin of the Nikita Botanical Garden*. 2017;123:58-64. (in Russian)
- Ben-David T., Melamed S., Gerson U., Morin S. ITS2 sequences as barcodes for identifying and analyzing spider mites (Acari: Tetranychidae). *Exp. Appl. Acarol.* 2007;41(3):169-181. DOI 10.1007/s10493-007-9058-1.
- Bikbulatova S.M., Chemeris D.A., Nikonorov Yu.M., Mashkov O.I., Garafutdinov R.R., Chemeris A.V., Vakhitov V.A. Methods for detecting polymerase chain reaction in real-time mode. *Vestnik Bashkirskogo Universiteta = Herald of the Bashkir State University*. 2012;17(1):59-67. (in Russian)
- Chaires-Grijalva M.P., Serrano-Domínguez A.K., Coronado-Blanco J.M. Mites associated with maize in Mexico. *Rev. Mex. Cienc. Agríc.* 2021;12(8):1497-1510. DOI 10.29312/remexca.v12i8.2718.
- Chen Y., Woolley L., Nguyen D., Gupta R., Chandler G.T., Nehl D., Herron G.A. Development and use of a single real-time PCR assay to identify the three spider mite species *Tetranychus urticae*, *Tetranychus lambi* and *Tetranychus ludeni* (Acari: Tetranychidae). *Austral. Entomol.* 2020;59(3):630-638. DOI 10.1111/aen.12457.
- Dermauw W., Moerkerecke Y.V., Ebrahimi N., Casteels H., Bonte J., Witters J. A loop-mediated isothermal amplification (LAMP) assay for rapid identification of *Ceratitidis capitata* and related species. *Curr. Res. Insect Sci.* 2022;2:100029. DOI 10.1016/j.cris.2022.100029.
- Devi M., Challa N., Mahesh G. Important mite pests of temperate and sub-tropical crops: a review. *J. Entomol. Zool. Stud.* 2019;7(4):1378-1384.
- Enohara K., Amano H. Simple method for discriminating six common species of red *Tetranychus* spider mites (Acari: Tetranychidae) in Japan. *Jpn. J. Appl. Entomol. Zool.* 1996;40(4):311-315. DOI 10.1303/jjaez.40.311. (in Japanese)
- 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.
- Goka K., Takafuji A. Electrophoretic detection of enzyme variation among Japanese red-coloured spider mites of the genus *Tetranychus* (Acari: Tetranychidae). *Exp. Appl. Acarol.* 1998;22:167-176. DOI 10.1023/A:1006006129339.
- Gotoh T. Separation of *Panonychus ulmi*, *P. thelytokus* and *P. bambusicola* (Acari: Tetranychidae) by esterase zymograms. *Appl. Entomol. Zool.* 1992;27(4):598-601. DOI 10.1303/aez.27.598.
- Gotoh T., Gutierrez J., Navajas M. Molecular comparison of the sibling species *Tetranychus pueraricola* Ehara et Gotoh and *T. urticae* Koch (Acari: Tetranychidae). *Entomol. Sci.* 1998;1(1):55-57.
- Gotoh T., Kitashima Y., Adachi I. Geographic variation of esterase and malate dehydrogenases in two spider mite species, *Panonychus osmanthi* and *P. citri* (Acari: Tetranychidae) in Japan. *Int. J. Acarol.* 2004;30(1):45-54. DOI 10.1080/01647950408684368.
- Gotoh T., Kitashima Y., Goka K. *Tetranychus* mite species identification using esterase and phosphoglucumutase zymograms. *Appl. Entomol. Zool.* 2007;42(4):579-585. DOI 10.1303/aez.2007.579.
- Hajibabaei M., Singer G.A.C., Hebert P.D.N., Hickey D.A. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* 2007;23(4):167-172. DOI 10.1016/j.tig.2007.02.001.
- Han H., Cho M.R., Jeon H.Y., Lim C.K., Jang H.I. PCR-RFLP identification of three major *Meloidogyne* species in Korea. *J. Asia Pac. Entomol.* 2004;7(2):171-175. DOI 10.1016/S1226-8615(08)60212-5.
- Hillis D.M., Dixon M.T. Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 1991;66(4):411-453. DOI 10.1086/417338.
- Hinomoto N., Tran D.P., Pham A.T., Le T.B.N., Tajima R., Ohashi K., Osakabe M., Takafuji A. Identification of spider mites (Acari: Tetranychidae) by DNA sequences: a case study in Northern Vietnam. *Int. J. Acarol.* 2007;33(1):53-60. DOI 10.1080/01647950708684501.
- Hsieh C.-H., Wang H.-Y., Chen Y.-F., Ko C.-C. Loop-mediated isothermal amplification for rapid identification of biotypes B and Q of the globally invasive pest *Bemisia tabaci*, and studying population dynamics. *Pest Manag. Sci.* 2012;68(8):1206-1213. DOI 10.1002/ps.3298.
- Hurtado M.A., Ansaloni T., Cros-Arteil S., Jacas J.A., Navajas M. Sequence analysis of the ribosomal internal transcribed spacers region in spider mites (Prostigmata: Tetranychidae) occurring in citrus orchards in Eastern Spain: use for species discrimination. *Ann. Appl. Biol.* 2008;153(2):167-174. DOI 10.1111/j.1744-7348.2008.00250.x.
- İnak E., Çobanoğlu S., Auger P., Migeon A. Molecular identification and phylogenetic analysis of spider mites (Prostigmata: Tetranychidae) of Turkey. *Exp. Appl. Acarol.* 2022;87:195-205. DOI 10.1007/s10493-022-00728-5.
- Kajiwara H., Hinomoto N., Gotoh T. Mass fingerprint analysis of spider mites (Acari) by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for rapid discrimination. *Rapid Commun. Mass Spectrom.* 2016;30(8):1037-1042. DOI 10.1002/rcm.7534.
- Kamayev I.O. *Oligonychus* (Acari: Tetranychidae) mites occurring in conifer planting material. In: Lectures in memoriam O.A. Kataev. Dendrobiont Invertebrates and Fungi and Their Role in Forest Ecosystems. Vol. 1. Insects and other invertebrates. Proceed. of the int. conf. St. Petersburg, Oct. 22–25, 2018. St. Petersburg, 2018;44. (in Russian)
- Kamayev I.O. Approach to diagnostics of spider mites (Acari: Tetranychidae) in phytosanitary practice. In: Monitoring and Biological Control Methods of Woody Plant Pests and Pathogens: From theory to practice. Proceed. of Second int. conf. Moscow, April 22–26, 2019. Moscow; Krasnoyarsk, 2019;82. (in Russian)
- Kamayev I.O., Karpun N.N. A new data on spider mites (Acari: Trombidiformes: Tetranychidae) inhabiting ornamental plants on the Black Sea coast of Krasnodar Region, Russia. *Kavkazskiy Entomologicheskij Bulletin = Caucasian Entomological Bulletin*. 2020;16(2):295-298. DOI 10.23885/181433262020162-295298. (in Russian)
- Kamayev I.O., Mironova M.K. The phytosanitary risk of herbivorous mites (Arachnida: Acariiformes). *Karantin Rasteniy. Nauka i Praktika = Plant Quarantine: Science and Policy*. 2018;3(25):13-20. (in Russian)
- Khaing T.M., Lee J.-H., Lee W.-G., Lee K.-Y. A new record of *Amphitetanychus quercivorus* (Acari: Tetranychidae) in Korea and molecular comparison with *A. viennensis*. *J. Asia Pac. Entomol.* 2013;16(2):155-160. DOI 10.1016/j.aspen.2012.12.003.
- Khaing T.M., Shim J.-K., Lee K.-Y. Molecular identification and phylogenetic analysis of economically important acaroid mites (Acari: Astigmata: Acaroidea) in Korea. *Entomol. Res.* 2014;44(6):331-337. DOI 10.1111/1748-5967.12085.
- Khaing T.M., Shim J.-K., Lee K.-Y. Molecular identification of four *Panonychus* species (Acari: Tetranychidae) in Korea, including new records of *P. caglei* and *P. mori*. *Entomol. Res.* 2015;45:345-353. DOI 10.1111/1748-5967.12135.
- Konoplev N.D., Ignatov A.N., Popov S.Ya. To question of opportunity of *Tetranychus urticae* and *T. atlanticus* (Acari: Tetranychidae) identification using mtDNA COI gene. *Vestnik Zashchity Rasteniy = Plant Protection News*. 2017;4(94):43-47. (in Russian)
- Koshkin E.I., Andreeva I.V., Guseinov G.G., Guseinov K.G., Dzhali-lov F.S.-U., Mityushev I.M. Specific aspects of the plant and phytophage interactions in agroecosystems under climate change. *Agro-*

- khimiya = Agrochemistry*. 2021;1:79-96. DOI 10.31857/S0002188121010063. (in Russian)
- Kutlunina N.A., Ermoshin A.A. *Molecular Methods in Plant Studies*. Yekaterinburg: Urals State Univ. Publ., 2017. (in Russian)
- Lampo M., Rangel Y., Mata A. Genetic markers for the identification of two tick species, *Amblyomma dissimile* and *Amblyomma rotundatum*. *J. Parasitol.* 1997;83(3):382-386. DOI 10.2307/3284398.
- Li D., Fan Q.-H., Waite D.W., Gunawardana D., George S., Kumarsinghe L. Development and validation of a real-time PCR assay for rapid detection of two-spotted spider mite, *Tetranychus urticae* (Acari: Tetranychidae). *PLoS One*. 2015;10(7):e0131887. DOI 10.1371/journal.pone.0131887.
- Matsuda T., Fukumoto C., Hinomoto N., Gotoh T. DNA-based identification of spider mites: molecular evidence for cryptic species of the genus *Tetranychus* (Acari: Tetranychidae). *J. Econ. Entomol.* 2013;106(1):463-472. DOI 10.1603/EC12328.
- Matsuda T., Hinomoto N., Singh R.N., Gotoh T. Molecular-based identification and phylogeny of *Oligonychus* species (Acari: Tetranychidae). *J. Econ. Entomol.* 2012;105(3):1043-1050. DOI 10.1603/EC11404.
- Matsuda T., Morishita M., Hinomoto N., Gotoh T. Phylogenetic analysis of the spider mite sub-family Tetranychinae (Acari: Tetranychidae) based on the mitochondrial COI gene and the 18S and the 5' end of the 28S rRNA genes indicates that several genera are polyphyletic. *PLoS One*. 2014;9(10):e108672. DOI 10.1371/journal.pone.0108672.
- Migeon A., Nouguié E., Dorkeld F. Spider Mites Web: a comprehensive database for the Tetranychidae. In: Sabelis M., Bruin J. (Eds.) *Trends in Acarology*. Dordrecht: Springer, 2010;557-560. DOI 10.1007/978-90-481-9837-5_96.
- Mitrofanov V.I., Strunkova Z.I., Livshits I.Z. *Guide to Spider Mites in the Fauna of the USSR and Neighboring Countries*. Dushanbe: Donish Publ., 1987. (in Russian)
- Morphological Identification of Spider Mites (Tetranychidae) Affecting Imported Fruits. The Secretariat of the North American Plant Protection Organization, Canada, 2014.
- Murugaiyan J., Roesler U. MALDI-TOF MS profiling-advances in species identification of pests, parasites, and vectors. *Front. Cell Infect. Microbiol.* 2017;7:184. DOI 10.3389/fcimb.2017.00184.
- Musolin D.L., Saulich A.K. Responses of insects to the current climate change: from physiology and behavior to range shifts. *Entomological Review*. 2012;92(7):715-740. DOI 10.1134/S0013873812070019.
- Navajas M., Fenton B. The application of molecular markers in the study of diversity in acarology: a review. *Exp. Appl. Acarol.* 2000;24(10-11):751-774. DOI 10.1023/a:1006497906793.
- Notomi T., Okayama H., Masubuchi H., Yonekawa T., Watanabe K., Amino N., Hase T. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 2000;28(12):e63. DOI 10.1093/nar/28.12.e63.
- Obasa K., Alabi O.J., Sétamou M. Wheat curl mite: a new source of the Eriophyoid mite in wheat fields identified. *PhytoFrontiers*. 2022;2(4):342-346. DOI 10.1094/PHYTOFR-04-22-0038-SC.
- Osakabe Mh. Difference of esterase isozymes between non-diapausing and diapausing strains of the citrus red mite, *Panonychus citri* (McGregor) (Acarina: Tetranychidae). *Appl. Entomol. Zool.* 1987;22(4):577-584. DOI 10.1303/aez.22.577.
- Osakabe Mh., Hirose T., Satô M. Discrimination of four Japanese *Tetranychus* species (Acari: Tetranychidae) using PCR-RFLP of the inter-transcribed spacer region of nuclear ribosomal DNA. *Appl. Entomol. Zool.* 2002;37(3):399-407. DOI 10.1303/aez.2002.399.
- Osakabe Mh., Kotsubo Yu., Tajima R., Hinomoto N. Restriction fragment length polymorphism catalog for molecular identification of Japanese *Tetranychus* spider mites (Acari: Tetranychidae). *J. Econ. Entomol.* 2008;101(4):1167-1175. DOI 10.1093/jee/101.4.1167.
- Osakabe Mh., Sakagami Y. RFLP analysis of ribosomal DNA in sibling species of spider mite, genus *Panonychus* (Acari: Tetranychidae). *Insect Mol. Biol.* 1994;3(1):63-66. DOI 10.1111/j.1365-2583.1994.tb00152.x.
- Ovalle T.M., Vásquez-Ordóñez A.A., Jimenez J., Parsa S., Cuellar W.J., Lopez-Lavalle L.A.B. A simple PCR-based method for the rapid and accurate identification of spider mites (Tetranychidae) on cassava. *Sci. Rep.* 2020;10(1):19496. DOI 10.1038/s41598-020-75743-w.
- Petrov D.L., Zhorov D.G., Sautkin F.V. Leaf gall mite *Aceria erinea* (Nalepa, 1891) (Acariformes: Eriophyidae) – a new invasive species of pests of walnut (*Juglans regia* L.) in Belarus. *Vestnik Beloruskogo Gosudarstvennogo Universiteta = Vestnik BGU*. 2016;2:75-77. (in Russian)
- Popov S.Ya. *The Ecological Aspect of Controlling the Harm of Insect and Mite Populations: A collection of articles*. Moscow: RGAU-MSKha Publ., 2013. (in Russian)
- Rak N.S., Litvinova S.V. Hazardous organisms migration and acclimatization during the plant introduction in the Far North Russia greenhouses. *Hortus Botanicus*. 2010;5:2. (in Russian)
- Ratcliffe S.T., Webb D.W., Weinzievr R.A., Robertson H.M. PCR-RFLP identification of Diptera (Calliphoridae, Muscidae and Sarcophagidae) – a generally applicable method. *J. Forensic Sci.* 2003;48(4):783-785. DOI 10.1520/JFS2002136.
- Ros V.I.D., Breeuwer J.A.J. Spider mite (Acari: Tetranychidae) mitochondrial COI phylogeny reviewed: host plant relationships, phylogeography, reproductive parasites and barcoding. *Exp. Appl. Acarol.* 2007;42(4):239-262. DOI 10.1007/s10493-007-9092-z.
- Sakamoto H., Gotoh T. Non-destructive direct polymerase chain reaction (direct PCR) greatly facilitates molecular identification of spider mites (Acari: Tetranychidae). *Appl. Entomol. Zool.* 2017;52:661-665. DOI 10.1007/s13355-017-0512-1.
- Shim J.-K., Khaing T.M., Seo H.-E., Ahn J.-Y., Jung D.-O., Lee J.-H., Lee K.-Y. Development of species-specific primers for rapid diagnosis of *Tetranychus urticae*, *T. kanzawai*, *T. phaselus* and *T. truncatus* (Acari: Tetranychidae). *Entomol. Res.* 2016;46(2):162-169. DOI 10.1111/1748-5967.12154.
- Sinaie S., Sadeghi Namaghi H., Fekrat L. A multiplex PCR assay for simultaneous discrimination of two predominant spider mites of the genus *Tetranychus* (Acari: Tetranychidae) in greenhouses of Iran. *J. Agr. Sci. Tech.* 2018;20(4):733-744.
- Sinaie S., Sadeghi-Namaghi H., Fekrat L. Loop-mediated isothermal amplification combined with PCR for specific identification of injurious mite, *Tetranychus urticae* (Trombidiformes: Tetranychidae). *Biologia*. 2019;74(5):477-485. DOI 10.2478/s11756-018-00187-7.
- Singhal N., Kumar M., Kanaujia P.K., Viridi J.S. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Front. Microbiol.* 2015;6:791. DOI 10.3389/fmicb.2015.00791.
- Tomita N., Mori Y., Kanda H., Notomi T. Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nat. Protoc.* 2008;3(5):877-882. DOI 10.1038/nprot.2008.57.
- Turak E., Hales D.F. An allozyme method for identifying individual aphids of morphologically similar taxa (Hemiptera: Aphididae). *J. Aust. Ent. Soc.* 1994;33(1):57-59. DOI 10.1111/j.1440-6055.1994.tb00920.x.
- Ulyanova E., Andreeva I., Shatalova E. The species composition of herbivorous acarofauna in the south of Western Siberia. *Persian J. Acarol.* 2022;11(2):365-370. DOI 10.22073/pja.v11i2.72456.
- Vásquez C., Colmenárez Y. Invasive mite species in the Americas: bioecology and impact. In: Haouas D., Hufnagel L. (Eds.) *Pests Control and Acarology*. London: IntechOpen, 2020. DOI 10.5772/intechopen.86127.

- Volkova M.V., Matveikina E.A. Structural changes in the phytophagous mite complex in the commercial vineyards of the Crimea. *Plodovodstvo i Vinogradarstvo Yuga Rossii = Fruit Growing and Viticulture of South Russia*. 2016;38(2):130-139. (in Russian)
- Walter D.E., Krantz G.W. Collecting, rearing and preparing specimens. In: Krantz G.W., Walter D.E. (Eds.) *A Manual of Acarology*. Texas Tech Univ. Press, 2009;91-92.
- Zeynalov A.S. Ecological and phytosanitary consequences of climate change in the plantings of fruit cultures. *Uspekhi Sovremennoy Nauki = Modern Science Success*. 2017;2(9):94-100. (in Russian)
- Zeynalov A.S., Orel D.S. Change in species composition, bioecology and harmfulness of main applian phytophages in the Central Region of the Non-Black Earth Zone of Russia under the influence of climate factors. *Vestnik Kazanskogo GAU = Vestnik of the Kazan State Agrarian University*. 2021;16(1):15-21. DOI 10.12737/2073-0462-2021-15-21. (in Russian)
- Zélé F., Weill M., Magalhães S. Identification of spider-mite species and their endosymbionts using multiplex PCR. *Exp. Appl. Acarol.* 2018;74(2):123-138. DOI 10.1007/s10493-018-0224-4.

ORCID ID

A.V. Razuvaeva orcid.org/0000-0002-0746-3660
E.G. Ulyanova orcid.org/0000-0002-9154-5238
E.S. Skolotneva orcid.org/0000-0001-8047-5695
I.V. Andreeva orcid.org/0000-0001-5627-8667

Acknowledgements. This research was funded by the Ministry of Science and Higher Education of the Russian Federation, Grant No. FWNR-2022-0007.

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

Received July 14, 2022. Revised December 25, 2022. Accepted January 7, 2023.

LIM-kinase 1 effects on memory abilities and male courtship song in *Drosophila* depend on the neuronal type

A.V. Zhuravlev¹ , E.S. Zalomaeva^{1, 2}, E.S. Egozova², A.D. Emelin², V.V. Sokurova², E.A. Nikitina^{1, 2}, E.V. Savvateeva-Popova¹

¹ Pavlov Institute of Physiology of the Russian Academy of Sciences, St. Petersburg, Russia

² Herzen State Pedagogical University of Russia, St. Petersburg, Russia

 beneor@mail.ru

Abstract. The signal pathway of actin remodeling, including LIM-kinase 1 (LIMK1) and its substrate cofilin, regulates multiple processes in neurons of vertebrates and invertebrates. *Drosophila melanogaster* is widely used as a model object for studying mechanisms of memory formation, storage, retrieval and forgetting. Previously, active forgetting in *Drosophila* was investigated in the standard Pavlovian olfactory conditioning paradigm. The role of specific dopaminergic neurons (DAN) and components of the actin remodeling pathway in different forms of forgetting was shown. In our research, we investigated the role of LIMK1 in *Drosophila* memory and forgetting in the conditioned courtship suppression paradigm (CCSP). In the *Drosophila* brain, LIMK1 and p-cofilin levels appeared to be low in specific neuropil structures, including the mushroom body (MB) lobes and the central complex. At the same time, LIMK1 was observed in cell bodies, such as DAN clusters regulating memory formation in CCSP. We applied GAL4 × UAS binary system to induce *limk1* RNA interference in different types of neurons. The hybrid strain with *limk1* interference in MB lobes and glia showed an increase in 3-h short-term memory (STM), without significant effects on long-term memory. *limk1* interference in cholinergic neurons (CHN) impaired STM, while its interference in DAN and serotonergic neurons (SRN) also dramatically impaired the flies' learning ability. By contrast, *limk1* interference in *fruitless* neurons (FRN) resulted in increased 15–60 min STM, indicating a possible LIMK1 role in active forgetting. Males with *limk1* interference in CHN and FRN also showed the opposite trends of courtship song parameters changes. Thus, LIMK1 effects on the *Drosophila* male memory and courtship song appeared to depend on the neuronal type or brain structure.

Key words: *Drosophila*; LIMK1; conditioned courtship suppression paradigm; memory; forgetting; dopaminergic neurons; cholinergic neurons; *fruitless*; male courtship song.

For citation: Zhuravlev A.V., Zalomaeva E.S., Egozova E.S., Emelin A.D., Sokurova V.V., Nikitina E.A., Savvateeva-Popova E.V. LIM-kinase 1 effects on memory abilities and male courtship song in *Drosophila* depend on the neuronal type. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2023;27(3):250-263. DOI 10.18699/VJGB-23-31

Влияние LIM-киназы 1 на память и брачную песню самцов дрозофилы зависит от типа нейронов

A.V. Журавлев¹ , E.C. Заломаева^{1, 2}, E.C. Егозова², A.Д. Емелин², В.В. Сокурова², E.A. Никитина^{1, 2}, E.B. Савватеева-Попова¹

¹ Институт физиологии им. И.П. Павлова Российской академии наук, Санкт-Петербург, Россия

² Российский государственный педагогический университет им. А.И. Герцена, Санкт-Петербург, Россия

 beneor@mail.ru

Аннотация. Сигнальный каскад ремоделирования актина, в состав которого входят LIM-киназа 1 (LIMK1) и ее субстрат кофилин, участвует в регуляции различных процессов в нейронах позвоночных и беспозвоночных животных. *Drosophila melanogaster* широко используется как модельный объект для изучения механизмов формирования, сохранения и воспроизведения памяти, а также забывания. Ранее активное забывание у дрозофилы исследовали с помощью классического павловского ольфакторного обучения. Было показано, что в разных формах забывания участвуют специфические дофаминергические нейроны и компоненты актинового каскада. В данной работе мы оценивали роль LIMK1 в процессах памяти и забывания у дрозофилы в парадигме условно-рефлекторного подавления ухаживания. В мозге дрозофилы уровень LIMK1 и фосфокофилина избирательно снижен в отдельных структурах нейропиля, включая лопасти грибовидных тел и центральный комплекс. В то же время LIMK1 присутствует в телах нервных клеток, таких как кластеры дофаминергических нейронов, регулирующие формирование памяти при условно-рефлекторном подавлении ухаживания. С использованием системы бинарного скрещивания GAL4 × UAS мы инициировали РНК-интерференцию *limk1* в различных типах нервных клеток. У гибридных линий с интерференцией *limk1* в лопастях грибовидных тел и глии наблюдалось усиление 3-часовой краткосрочной памяти, без видимого влияния на долгосрочную память. Интерференция *limk1* в холинергических нейронах приводила к снижению краткосрочной памяти, в дофаминергических и серотонинергических нейронах ее результатом было также существенное нарушение способности мух к обучению. Напротив, интерференция *limk1* в нейронах *fruitless* усиливала 15–60-минутную краткосрочную память, что указывает на возможную роль LIMK1 в процессах активного забывания. У самцов с интерференцией *limk1* в

холинергических и *fruitless* нейронах также были отмечены разнонаправленные изменения параметров брачной песни. Таким образом, эффекты LIMK1 на память и брачную песню самцов дрозофилы определяются типом нервных клеток или структурой мозга.

Ключевые слова: дрозофила; LIMK1; условно-рефлекторное подавление ухаживания; память; забывание; дофаминергические нейроны; холинергические нейроны; *fruitless*; брачная песня самца.

Introduction

Memory formation and forgetting serve as the basis of behavioral plasticity. Whereas memory is a specific process of information acquisition, storage and retrieval by the nervous system, active forgetting is defined as “a mechanism or series of mechanisms to remove memories that become unused” (Davis, Zhong, 2017). Associative memory formation and active forgetting occur in both mammals and invertebrates, including *Drosophila melanogaster* (Medina, 2018), which is a well-known object of classical genetics. Having a short life cycle and relatively simple nervous system, the fruit fly makes it easy to perform genetic analysis of the molecular basis of behavioral and cognitive processes.

There are several experimental techniques to form associative memory in *Drosophila*, including short-term memory (STM) and protein synthesis-dependent long-term memory (LTM). The most widely used technique is classical Pavlovian learning with negative electroshock reinforcement, or olfactory aversive learning (OAVL), which revealed genes responsible for different types of memory (Tully et al., 1994). More natural is conditioned courtship suppression paradigm (CCSP) (Siegel, Hall, 1979; Kamyshev et al., 1999). GAL4 × UAS binary expression system (Duffy, 2002) is used to study the effects of specific genes on memory processes. The fine neural organization of the mushroom bodies (MB), a principle structure responsible for associative olfactory learning in *Drosophila*, was evaluated in detail. The MB output neurons (MBON) are the main effectors of MB, whereas specific clusters of dopaminergic neurons (DAN) regulate the activity of MB – MBON synaptic contacts (Aso et al., 2009, 2014a, b). Among them are aSP13 DAN of the protocerebral anterior medial cluster (PAM), which innervate $\gamma 5$ area of MB, playing a crucial role in CCSP learning and memory (Keleman et al., 2012).

The molecular and neural mechanisms of active forgetting implicate the activity of DAN and Rac1-dependent signal pathways (Medina, 2018). Small GTPases of the Rho family, including Rho and Rac, regulate neuronal actin polymerization during the *Drosophila* nervous system development. Rho via its effector ROCK or Rac/Cdc42 via its effector Pak activate LIM-kinase 1 (LIMK1), which phosphorylates *Drosophila* cofilin (twinstar) protein, blocking its actin-depolymerization activity and inhibiting axon growth. Rac also acts through Pak-independent pathway to antagonize LIMK1 and promote axon growth (Ng, Luo, 2004). In addition to its role in neurogenesis, Rac is crucial for both interference-induced and passive forgetting in OAVL paradigm. PAK/LIMK1/cofilin pathway probably acts downstream Rac1 (Shuai et al., 2010). Forgetting specific types of memory depends on different signal proteins (Zhang et al., 2016; Gao et al., 2019).

Forgetting in OAVL paradigm is caused by several DAN of the protocerebral posterior lateral 1 (PPL1) cluster, which innervates some MB structures, such as pedunculus, lower and upper stalk. Memory acquisition and forgetting are regu-

lated by different dopamine receptors, dDA1 and DAMB respectively (Berry et al., 2012). Coincidence of conditioned and unconditioned stimuli creates a memory trace in MBON- $\gamma 2\alpha'1$, probably inhibiting the MB > MBON- $\gamma 2\alpha'1$ synapses. The unconditioned stimulus alone activates DAN- $\gamma 2\alpha'1$, which in turn disinhibit MB > MBON- $\gamma 2\alpha'1$ synapses and cause forgetting (Berry et al., 2018). DAN that innervate the MB $\alpha\alpha'$ tip induce the interference-based forgetting through the scaffold protein Scribble, binding together Rac1, PAK3 and cofilin (Cervantes-Sandoval et al., 2016).

Whereas multiple data prove the importance of DAN and actin-remodeling signal pathway for forgetting in OAVL paradigm, there is virtually no data for molecular mechanisms of memory decay in CCSP. Effects of LIMK1-dependent signal cascade on CCSP learning and memory were firstly shown for the temperature-sensitive mutant *agn^{ts3}*, with LIMK1 increase in the adult brain compared to the wild type *Canton-S* (CS). Temperature rise leads to a decrease in *agn^{ts3}* LIMK1 level, simultaneously restoring its learning ability and 3 h memory, which are drastically impaired in the norm (Medvedeva et al., 2008). *agn^{ts3}* has multiple polymorphisms within and near *limk1* gene, as well as a changed profile of microRNA expression, and can serve as a model object for Williams syndrome (Nikitina et al., 2014; Savvatееva-Popova et al., 2017). The temporal profile of STM learning index (LI) was assayed in CCSP for *agn^{ts3}*, as well as for the wild-type strains with *limk1* polymorphisms, CS and *Oregon R*. Only CS was able to learn and store memory up to 24 h (Zalomaeva et al., 2021).

The behavioral effects of LIMK1 changes in *agn^{ts3}* do not give information about specific cell types, where LIMK1 can be involved in learning and memory. In this study, we performed the analysis of memory decay for several *Drosophila* GAL4 × UAS strains with neuronal type-specific *limk1* RNA interference. LIMK1 distribution in the *Drosophila* brain structures was studied in detail using confocal microscopy. The effect of *limk1* interference on fly memory ability depended on both neural type and memory form. LIMK1 also appeared to be involved in regulation of male courtship song: *limk1* interference in different neuronal types specifically affected some song parameters.

Materials and methods

***Drosophila* strains.** Fly strains were provided by the Research Center “Biocollection of Pavlov Institute of Physiology RAS for the study of integrative mechanisms of the nervous and visceral systems”. The strain numbers (#) are given in accordance with the Research Center and Bloomington *Drosophila* Stock Center, USA (Cook et al., 2010). The following strains were used:

1. *Canton-S* (CS) – the wild-type strain with *limk1* polymorphisms.
2. *agn^{ts3}* – temperature-sensitive mutation on CS genetic background with *limk1* polymorphisms, characterized by learning and memory defects.

3. Strains expressing GAL4 in specific neuronal types:
#6794: w[*]; P{w[+mC]=nrv2-GAL4.S}8 P{w[+mC]=UAS-GFP.S65T}eg[T10]. GAL4 and green fluorescent protein (GFP) are expressed in nervous system under *Nrv2* regulatory element;
#6793: w[*]; P{w[+mC]=ChAT-GAL4.7.4}19B P{w[+mC]=UAS-GFP.S65T}Myo31DF[T2]. GAL4 and GFP are expressed in cholinergic neurons (CHN) under *ChAT* and *VAcHT* regulatory elements;
#7009: w[1118]; P{w[+mC]=Ddc-GAL4.L}Lmpt[4.36]. GAL4 is expressed in dopaminergic (DAN) and serotonergic (SRN) neurons;
#30027: w[1118]; P{w[+mW.hs]=GawB}fru[NP0021]. GAL4 is expressed in *fruitless* neurons regulating mating behavior.
4. Act-GAL4: w[1118]; P{w[+mC]=} 25FO1/CyO, y[+]; *Canton-S* background. GAL4 is expressed in the whole body under actin promoter.
5. Strains with UAS-dependent *limk1* suppression:
#26294: y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.JF02063}attP2. The strain expresses interfering RNA against *limk1* (RNAi) under UAS (*limk1-KD*, knockdown).
#36303: y[1] v[1]; P{y[+t7.7]=CaryP}attP2. The control strain with genetic background identical to that for #26294, but lacking RNAi (*limk1*^{“+”}).
6. Strains with GFP gene regulated by UAS:
#32186: w[*]; P{y[+t7.7] w[+mC]=10XUAS-IVS-mCD8::GFP}attP40.
#32202: w[*]; P{y[+t7.7] w[+mC]=10XUAS-IVS-GFP-WPRE}attP2.

To induce *limk1* RNA interference in specific neuronal types, a strain carrying GAL4 activator expressed under tissue-specific promoter was crossed to UAS strain #26294. The cross product of a GAL4 strain and #36303 strain served as a control.

Flies were raised on the standard yeast–raisin medium at 25±0.5 °C and a 12:12 daily illumination cycle. For behavioral tests, experimental males were collected without anesthesia and kept individually. 5–6-day-old males were used in experiments. Females (CS) were collected as virgins and brought together with CS males for fertilization in CCSP one day before experiment.

Antibodies. Primary antibodies: Rat anti-LIMK1 multi-specific monoclonal (Enzo Life Sciences, ALX-803-343-C100); mouse anti-*Drosophila* cysteine string protein (CSP); rabbit anti-*Drosophila* tyrosine hydroxylase (TH) (Abcam, ab128249); rabbit anti-GFP (Abcam, ab290).

Secondary antibodies: Goat anti-mouse Alexa Fluor 488 (Invitrogen, A32723), donkey-anti-rat Alexa Fluor 594 (ThermoFisherScientific, A-21209), goat anti-rabbit Alexa Fluor 633 (Invitrogen, A21071).

RNA extraction and RT-PCR analysis of *limk1* expression. The level of *limk1* expression was assayed using semi-quantitative PCR in complex with reverse transcription (RT-PCR). Flies were anesthetized by freezing. 10 male flies or 70 male heads were homogenized in 300 µl TRI reagent (MRC, TR 118). Total RNA was extracted from homogenates according to the manufacturer’s protocol. The quality of RNA was checked by 1.5 % agarose gel electrophoresis. 1 µg

RNA was reverse-transcribed by MMLV reverse transcriptase (Evrogen, #SK022S) according to the manufacturers’ protocol, using random hexamer primers and RNase inhibitor (Syntol, #E-055). Semi-quantitative PCR was performed on a StepOne Plus (Applied Biosystems, Inc., USA) using qPCRmix HS SYBR+LowROX (Evrogen, #PK156L) containing direct and reverse primers (0.5 mM each). Baseline and cycle threshold values were determined by automatic analysis using StepOne software v2.3 (Applied Biosystems, USA). *rpl32* transcript was used as an internal control. The predesigned *limk1* primers (PP12636 in FlyPrimerBank, <http://www.flyrnai.org/flyprimerbank>) were used to bind all five *limk1* cDNA isoforms, both premature and mature forms, as primers do not span the exon-intron borders. The relative *limk1* transcript level was calculated using the comparative $\Delta\Delta C_t$ method. The number of biological replicates (independent RNA extractions with reverse transcription) was 3–5, the number of technical replicates was 3.

The primer sequences were the following:

rpl32:

Forward: 5'-TATGCTAAGCTGTCGCACAAATGGC-3'

Reverse: 5'-GTTCTGCATGAGCAGGACCTCCA-3'

limk1:

Forward: 5'-GTGAACGGCACACCAGTTAGT-3'

Reverse: 5'-ACTTGCACCGGATCATGCTC-3'

PCR parameters:

1. 1 cycle: 95 °C – 5 min.
2. 45 cycles. 95 °C – 20 s, 60 °C – 20 s, 72 °C – 20 s, 77 °C – 15 s (detection).
3. Melting curve: 95 °C – 15 s, 60 °C – 1 min, 60–95 °C ($\Delta 0.3$ °C, 15 s).

Immunofluorescent staining of *Drosophila* brains.

5–6-day-old imago males were anesthetized by freezing. The brains were prepared in PBS buffer (pH 7.5) using needle-sharp tweezers (Merck, T4412), fixed in 4 % paraformaldehyde in PBS for 1 h at RT and stained according to (Thapa et al., 2019), without a freezing stage. Antibodies were diluted in PBT (0.2 % Tween 20, 5 % BSA in PBS) as 1:200, for anti-CSP – 1:20. Previously, for better staining of brains, we increased the time of incubation with primary antibodies up to 5 days (Zhuravlev et al., 2020). Here, the incubation was performed at 4 °C for 3 days (with primary antibodies) or overnight (with secondary antibodies). Brains were mounted with Vectashield mounting medium containing DAPI (Vector laboratories, H-1200-10).

Protein distribution analysis in the brain by confocal microscopy.

Brains were scanned frontally using laser scanning confocal microscopy (LSM 710 Carl Zeiss; Confocal microscopy Resource Center; Pavlov Institute of Physiology of the Russian Academy of Sciences, St. Petersburg, Russia). Scanning was performed using X63 objective at different depths (z-step 2 µm). Images were analyzed using Fiji software. The brain structures were visually mapped using the *Drosophila* brain online atlas (Virtual Fly Brain). To measure the average level of LIMK1 inside the brain structures, the average signal intensity was measured in three small square areas (~10 × 10 µm) within each of the structures. The average values were obtained and normalized to the average structure intensity for the given brain. Colocalization Threshold analysis

was performed to measure co-localization of LIMK1 with neurospecific markers. To prepare figures, auto contrast function was used for each optical slice.

Learning and courtship suppression tests in *Drosophila* males. Flies learning and STM were estimated in CCSP, as described in (Zhuravlev et al., 2022). In the case of long term memory (LTM), learning was performed by placing flies in food-containing glasses (20 mm diam., ~20 mm high) for 5 h (Kamyshev et al., 1999). Courtship index (CI) and learning index (LI) were estimated at the following time points after learning: for short-term memory (STM) analysis: 0 min (learning), 3 h; for STM decay analysis: 15, 30, 60 min, 24 h; for LTM analysis: 0 min, 2 days, 8 days. In all groups, naive males (without mating experience) served as a control to calculate LI:

$$LI = [(CI_N - CI_T) / CI_T] \times 100 \% = (1 - CI_N / CI_T) \times 100 \%,$$

where CI_N is the middle CI for naive males, and CI_T is the middle CI for males after training. The naive and trained males were the same age. The decrease in LI compared to LI (0 min) was considered a time-dependent memory decay. The decrease in LI for a mutant strain compared to that for the wild-type strain *CS* was considered a strain-specific impairment of learning or memory.

Courtship song analysis. The 5-day-old imago male courtship song was recorded as in (Savvateeva-Popova et al., 2008). A naive male of the studied line and a fertilized female (*CS*) were placed together in a Perplex chamber with a latticed bottom on top of a microphone. The chamber was placed in a foam box in a soundproof room. The sounds were recorded for 5 min using Audacity software (Mazzoni, Dannenberg, 2020). The sound signals were filtered to exclude noises, obtaining signals within 100–800 Hz. The level of noise was decreased using a standard Audacity plugin. The software *Drosophila* courtship song analysis (DCSA) (Iliadi et al., 2009) was used to automatically detect pulse and sine song components.

The results of analysis were manually edited. The mean values of the song parameters were calculated for each fly. The following parameters were estimated: pulse song index (PInd, % of the total time), pulse song initiation frequency (PFR; 100/s), sine song index (SInd, % of the total time), sine song frequency (SFR, 100/s), interpulse interval (IPI, ms), period of song pulse train (Per, s), intertrain interval (ITI, ms), train duration (TrainDur), pulse number in train (PulseN), sine song duration (SDur, ms), sine song amplitude (SAmp, C.U.), IPI variance (Var(IPI), ms^2). Per is the time between the starts of the neighboring trains. ITI is the time between the end of the previous and the start of the next train.

Statistical analysis. Analysis of LIMK1 mRNA level was performed using two-sided *t*-test, Social Science Statistic online resource ($p < 0.05$). Analysis of LI and courtship song parameters was performed using two-sided randomization test at significance level α of 0.05 ($n = 20$), using *Drosophila* Courtship Lite software (Nikolai Kamyshev, 2006, nkamster@gmail.com), with 10000 iterations. The program is freely available from the author upon request. Randomization test was reported to be better for LI comparison than *t*-test or some nonparametric tests (Kamyshev et al., 1999). Court-

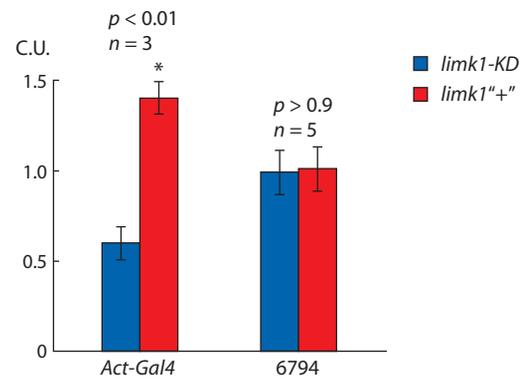


Fig. 1. Comparative levels of *limk1* RNA in the UAS > GAL4 hybrids with and without *limk1* RNA interference.

X axis: GAL4 strains. Y axis, conditional units (C.U., expression levels are normalized to the average value). Statistical differences: * between *limk1-KD* > GAL4 and *limk1'+* > GAL4 (two-sided *t*-test; p and n are shown above the charts). Standard error is shown.

ship song parameters were also analyzed using two-sided Mann–Whitney U-test. Python 3 scripts were used to draw the box plots charts.

Results

limk1 RNA level in *Drosophila* UAS × GAL4 hybrids

To check that GAL4 really induces *limk1* RNA interference in 26294 strain, we compared *limk1* RNA level in the UAS (f) > GAL4 (m) hybrids. Females with and without transgenic RNAi for *limk1* suppression (*limk1-KD* and *limk1'+*, respectively) were crossed to *Act-GAL4* males, expressing GAL4 in the whole body. The level of total *limk1* RNA was approximately 2-fold lower in the hybrid with *limk1* interference. These data confirmed the efficiency of RNAi-dependent *limk1* suppression in 26294 strain (*limk1-KD*) upon its activation by GAL4. At the same time, there were no differences for *limk1-KD* > 6794 and *limk1'+* > 6794, where RNA expression was measured in heads and was regulated by neuronal type-specific GAL4 (Fig. 1). Thus, *limk1* RNA differences after neural type-specific *limk1* RNA interference might be local or too low to be detected in the whole *Drosophila* heads.

LIMK1 distribution in the *Drosophila* brain

When studying LIMK1 distribution, we focused on the central part of the *Drosophila* brain, without the optic lobes (OL), mainly at the level of the superior medial protocerebrum (SMP) and gamma-lobes (γ L) of MB. Here the PAM clusters of DAN are located (Mao, Davis, 2009), responsible for the *Drosophila* courtship learning and memory (Keleman et al., 2012). Additionally, the area including the central complex (CC) and calyx (Cal) surrounded by Kenyon cells (KC) was studied. The CSP-positive neuropil structures and tyrosine hydroxylase (TH)-positive DAN cell bodies and processes served as landmarks in describing the LIMK1 distribution. The following description is given for the wild-type strain *CS* (see the Table and Suppl. Material 1)¹.

¹ Supplementary Materials 1–6 are available in the online version of the paper: <https://vavilovj-icg.ru/download/pict-2023-27/appx9.pdf>

LIMK1 and TH distribution in the *Canton-S* brain (visual analysis)

Z	N	Neuropil structures (CSP)	TH-positive structures	LIMK1 level
I	0	AL and SMP; the part of γ L is observed as a shading (CSP lever is lower)	Two bilaterally symmetrical clusters of PAM cell bodies, the processes of which form two semicircular tracts (#1), enveloping SMP below	AL and SMP: high; thin tissue layers resembling the neuropil glia: high; cell bodies surrounding neuropil: relatively low; cell nuclei: low
II	3	γ L of MB, subdivided into three unequal parts; the putative γ 5 area is marked with an asterisk (*); AOT lateral to SMP	Semicircular tracts (#1) move laterally; two tracts (#2) above SMP; granularity (#3) above γ L. TH level is low in γ L	γ L: lower than in SMP; the thin tissue layers surrounding AL and SMP: high
III	7	γ L is half hidden by β 'L, with the horizontal β L below it. The beginning of α/α 'L is seen	Granularity (#3) above and laterally the β 'L area; tracts (#2) join at the central line (#4). TH level is low in β L	All MB lobes: low
IV	11	α L and α 'L. The part of γ L (#7) lateral to α L	Granularity (#5) at the area of β 'L tips, where the PAM processes converge	Area #5: low; granularity along the central axis of the brain (#6): high; γ L (#7): very low
V	14	The tips of α L and α 'L; the part of EB	Area 5 is subdivided into two (#8), corresponding to β 'L tips. Commissure (#9) above (#8)	Area #8: low; EB: low
VI	16	EB; α L and γ L join to the pedunculus (Ped) of MB	Commissure #9 above EB	Ped and EB: low; the semicircular granularity (#10) above and laterally EB: high; commissure (#11) between EB and the esophagus (ES): high
VII	20	CC (EB, FB and No)	Commissure #9 continues laterally	CC and Ped: low; tissue layers surrounding neuropil structures: high
VIII	26	CC (FB), W, Ped	Tracts around FB	Commissure (#12) above CC and semicircular structures (#13) surrounding wedge: high
IX	29	CC is no longer visible	The cell bodies and processes of DAN clusters around Cal are partly visible	The great commissure (#GC) above the ES: high; glia-like layers among neuropil: high
X	36	Cal, KC and PB; Ped below Cal – a small shading (#14)	DAN clusters around Cal: PPL1 and PPM2	Cal and PB: high; KC: relatively low; glia-like layers (#15) in the bottom part near ES: high

Note. The depth of the studied zone (Z) is given for the brain optical slices, from the PAM cell bodies to CC (I–V) or from CC to Cal (VI–X). N is the number of the brain optic slice for a given zone (step 2 μ m). For different brains, there may be slight differences in depth of the given N.

The brain structures (here and below): α , α ', β , β ' and γ L – the corresponding lobes of MB, AL – the antennal lobes, AOT – the anterior optic tubercle, Cal – the calyx of MB, CC – the central complex, EB – the ellipsoid body of CC, ES – the esophagus, FB – the fan-shaped body of CC, KC – the Kenyon cells, MB – the mushroom body, MBI – the median bundle, No – the noduli of CC, PB – the protocerebral bridge, Ped – the pedunculus of MB, SEG – the subesophageal ganglion, SMP – the superior medial protocerebrum, W – wedge. DAN clusters, according to (Mao, Davis, 2009): PAM – the protocerebral anterior medial cluster, PPL – the protocerebral posterior lateral clusters, PPM – the protocerebral posterior medial clusters.

The distribution of DAN clusters corresponded to that described in (Mao, Davis, 2009). PAM clusters were clearly observed near SMP, with processes extending towards the central part of the brain. The processes formed glomerular structures around the MB horizontal lobes (γ , β and β 'L), probably being the synaptic endings innervating the correspond areas. The structure #3 was located above γ L, the structures #5 and 8 – in the β 'L area, the commissure #9 was seen in the central part of the brain. TH signal was relatively low in β L (see Suppl. Material 1). PPL1, PPM2 and other DAN clusters were observed around Cal, sending their processes to the different brain areas (Fig. 2, a).

LIMK1 was concentrated in the neuropil structures of the anterior part of the brain, such as SMP and AL. The LIMK1-positive granularity was observed inside SMP, between the β 'L tips (#8) and around the ellipsoid body (EB) of CC (#10, 11). LIMK1 level was also high in thin tissue layers adjacent to neuropil and some neural tracts, such as #12 around the great commissure, #13 around wedge (W) and #15 near esophagus (ES), morphologically resembling glia (Hartenstein,

2011). LIMK1 signal was lower in cell bodies of the neurons surrounding AL (ALCB), probably being the cell bodies of the projection neurons, as well as in KC surrounding Cal. Here, LIMK1 was mainly concentrated in the cytoplasm, beyond the nuclei. LIMK1 level was significantly decreased in all the MB lobes and pedunculus (Ped), as well as in the CC structures, whereas in Cal and the protocerebral bridge (PB) it was relatively high (see Suppl. Material 1). LIMK1 and TH colocalization was observed in SMP, AL, Cal, the TH-positive cells and processes, and in glomerular densities, such as #3, 5 and 6 (see Fig. 2, b).

To check that the antibody specifically binds to LIMK1, the distribution of LIMK1 main product p-cofilin was assayed in CS. The pattern of p-cofilin distribution was generally similar to that for LIMK1 (Suppl. Material 2). The level of p-cofilin was low in MB (including Cal) and CC (mostly EB, as well as in the case of LIMK1). In contrast to LIMK1, p-cofilin was mainly concentrated in the cell nuclei in the peripheral area of the fly brain, such as Kenyon cells around Cal, as well as in PB, subesophageal ganglion (SEG), and cell bodies surrounding

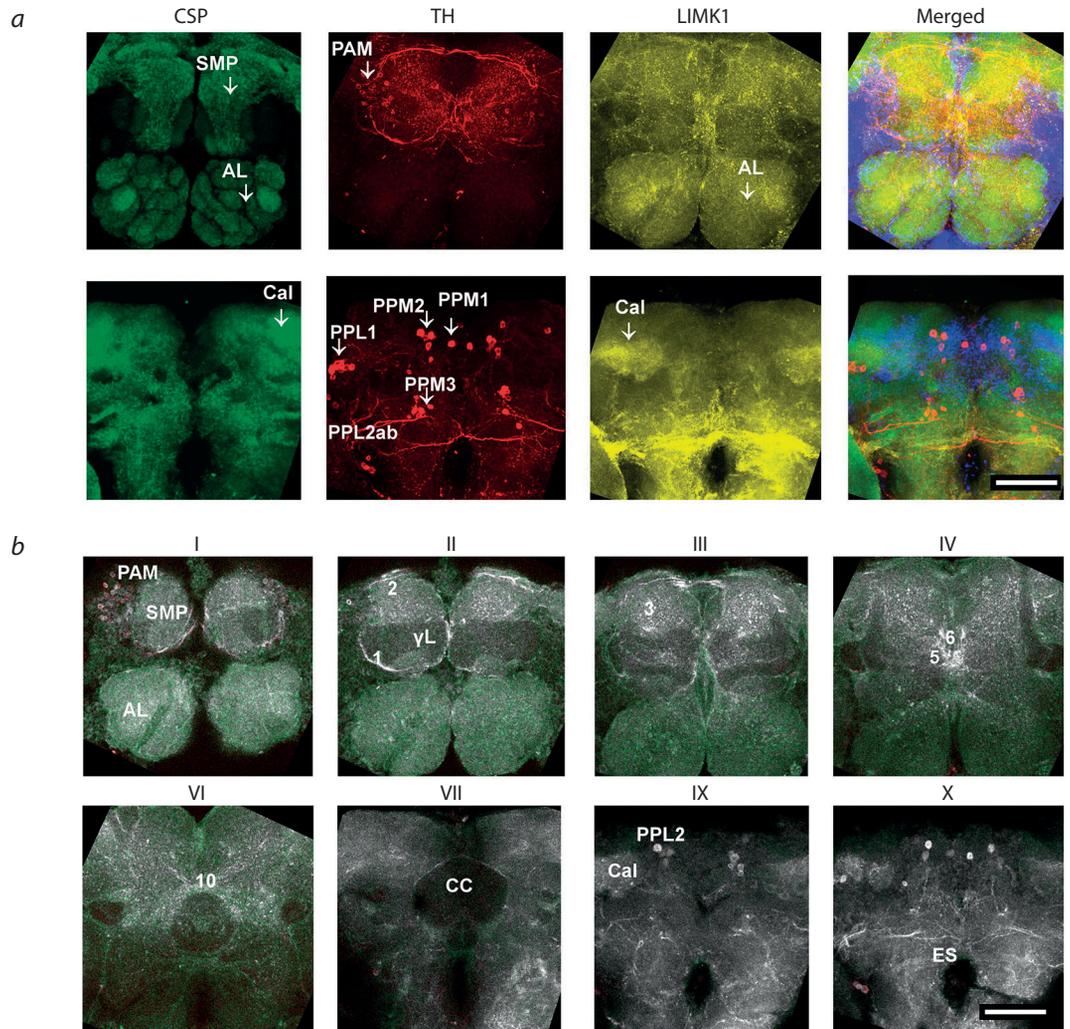


Fig. 2. LIMK1 spatial distribution in the CS brain.

a, Z-projection. AL level (above) and Cal level (below). Color scheme: green – CSP, red – TH, yellow – LIMK1, blue – DAPI. *b*, Co-localization of TH and LIMK1 in different optic zones. Color scheme: green – LIMK1, red – TH, white – the areas of LIMK1 – TH co-localization. The scale bar size is 50 μm. See the Table for abbreviations and description of the optic zones.

AL. p-Cofilin was also localized in diffuse layers within the brain structures, such as EB, probably formed by glia. The p-cofilin-enriched cells were found in SEG, forming structures with two glomerular branches (*), and around CC structures, fan-shaped body (FB) and EB (**).

Several GAL4 activators were used to initiate *limk1* RNA interference. Both 6793 and 6794 strains specifically express the green fluorescent protein (GFP) under GAL4 promoter. In strain 6794, GAL4 was reported to be expressed in OL, thoracic ganglion, different nerves, and cortex glia (Sun et al., 1999; Okamoto, Nishimura, 2015). In 6794 > *limk1-KD* hybrid, GAL4-driven GFP expression was detected in glia-like cells, surrounding the neuropil structures, such as AL, SMP, CC and its parts, as well as in the MB lobes, Ped and some KC (Suppl. Material 3, *a*). GFP level was higher in αL and βL compared to α'L, β'L and γL. The signal was lower in Cal and virtually absent in most neuropil structures, such as AL, SMP, CC and others. Thus, *limk1* interference should occur in Cal and some glia cells, where the levels of both GAL4 and LIMK1 were relatively high. Similar distribu-

tion was observed in the control 6794 > *limk1*^{+/+} strain (see Suppl. Material 3, *b*). In the strain 6793, GAL4 is expressed in cholinergic neurons (CHN), with GFP signal in OL, AL with the surrounding interneurons, the parts of CC, the great commissure (GC), Cal and the mechanosensory area of SEG (Salvatterra, Kitamoto, 2001). In the strains 6793 > *limk1-KD* and 6793 > *limk1*^{+/+} (see Suppl. Material 3, *c, d*), we observed a strong GFP signal in cell bodies surrounding SMP and AL, as well as in some KC, several neuropil structures (AL, αβL, EB, FB), and GC. In all the studied strains, LIMK1 distribution character appeared to be similar to that in CS.

To check that GAL4 is active in 7009 and 30027 strains, we crossed them to strains expressing GFP under GAL4 promoter. In 7009 > 32186 hybrid, we observed a prominent GFP signal in cell clusters near SMP, morphologically similar to the TH-positive PAM clusters. Some cells might be SRN, but they constitute the minority of the observed neurons in this area (Albin et al., 2015; Kasture et al., 2018). The processes of PAM neurons extended to the horizontal MB lobes, including γL, and the densely innervated β'L tip (#5) connected by

a commissure, and to a much lesser extent the β L tip (Suppl. Material 4, a). EB was surrounded by the GFP-positive processes extending from different parts of the brain. The GFP-positive DAN around Cal were also observed. Hence, GAL4 activator of 7009 should suppress *limk1* inside DAN, including PAM neurons, which regulate memory storage in CCSP. The *fruitless*-positive neurons (FRN) are responsible for mating behavior. In 30027 > 32202 hybrid, we observed GFP in some KC, in the cell bodies located near SMP and AL, and glomerular structure, forming a ring-like structure around Ped (see Suppl. Material 4, b). Similar structures were described in (Yu et al., 2010). The distribution of LIMK1 in the hybrid strains with and without *limk1* knockdown in the above neurons was similar to that in CS (Suppl. Material 5).

The normalized intensity of LIMK1 signal was calculated for several brain structures (Fig. 3). The LIMK1 relative levels in specific structures were very similar for the CS brain and the average brain of all the strains. The biggest differences were observed for the TH-positive glomerular structure #6 (TH+(6)), which is possibly responsible for memory formation in CCSP. In the average brain, ALCB had the normalized LIMK1 level about 1. Compared to them, AL, SMP and TH+(6) structures had the higher LIMK1 level, whereas the MB lobes, EB and Ped had the lower LIMK1 level. In *agn^{ts3}*, AL and ALCB had the higher LIMK1 level compared to CS, whereas most of the rest studies structures had lower LIMK1 level. This corresponds to more contrast LIMK1 staining in *agn^{ts3}* relative to CS (Suppl. Material 6). There were no prominent differences after *limk1* knockout, except for several structures with minor changes. The interstrain differences might be local or beyond the resolution of the method.

3 h STM differs in hybrids with and without *limk1* interference

3 h STM was estimated for *limk1-KD* (f) > 6794 (m) and the control *limk1^{ts3}* (f) > 6794 (m) hybrids. In both cases, we observed the decrease of courtship index (CI) after learning, with its partial recovery after 3 h. The box plot height was minimal for CS and rather big for UAS × GAL4 hybrids, showing that the value of courtship suppression significantly varied for individual flies. All strains were capable to learn in CCSP, with learning index LI (0 h) immediately after training of about 60–70 % (Fig. 4, a). The CS LI was still high after 3 h, indicating STM preservation, in agreement with (Zhuravlev et al., 2022). The strain with *limk1* interference also preserved STM: although its LI (3 h) was only about 20 %, it did not statistically differ from that for CS, as well as from LI (0 h). In the control strain, LI (3 h) decreased compared to LI (0 h) and did not differ from zero, indicating the impaired STM. Thus, while 3 h memory storage or retrieval was impaired in the control strain, *limk1* interference seems to improve 3 h STM. At the same time, it did not affect the impaired 8 day LTM, with only minor positive effect on 2 day LTM (see Fig. 4, b).

Neuron type-specific *limk1* interference differentially affects STM dynamics

To investigate the dynamics of STM decay in different strains with *limk1* interference, we performed LI analysis immediately and 24 h after learning (Fig. 5). To exclude the possible effect

of eye color on learning and memory abilities, we applied GAL4 (f) > UAS (m) crossing scheme, where both the strain with *limk1* knockdown and the control strain had the same wild-type eye color. For 6794 activator (MB and glia), the control strain showed nearly the same LI within 24 h, whereas the strain with *limk1* interference demonstrated a steeper forgetting curve. Hence, 6794 > *limk1-KD* showed high LI after learning, but seemed to increase the speed of memory forgetting on the interval 0–30 min. *limk1* knockdown in CHN (6793) was associated with significant decrease of LI within 60 min after training.

For DAN and SRN activator (7009), both the strain with *limk1* interference and the control strain showed nearly the same dynamics of STM decay, except for 30–60 min period. *limk1* interference was associated with a dramatic defect on learning: LI did not differ from zero. LI (24 h) was negative in both hybrids, possibly being the effect of sensitization: males did not suppress the courtship activity but courted more actively some time after training. For FRN activator (30027), the effect of *limk1* knockdown was the opposite to that of MB/glia and CHN activators: *limk1* interference decreased the speed of forgetting, and LI (30 min) did not differ from zero. Thus, the effect of *limk1* interference on STM dynamics appeared to depend on the neuronal type.

LIMK1 interference in CHN and FRN neurons differentially affects courtship song parameters

Finally, we studied the influence of *limk1* interference on the male courtship song parameters. The hybrids with CHN and FRN drivers were studied (Fig. 6). There were no interstrain differences in interpulse interval (IPI), the species- and population-specific parameter (Ritchie et al., 1994), and IPI variance (Var(IPI)), the marker of neurodegenerative processes (Savvateeva-Popova et al., 2003). *limk1* interference in CHN (6793) decreased the pulse song index and frequency (PInd, PFr), increasing the mean period (Per), intertrain interval (ITI), train duration (TrainDur), sine song duration (SDur) and train pulse number (PulseN). On the contrary, in the strain with FRN activator (30027), *limk1* interference resulted in PFr increase, as well as Per, ITI, SInd and SDur decrease. *limk1* knockdown by two different activators had the opposite effects on PInd, PFr, Per, ITI and SDur, leveling the initial differences between SInd, TrainDur and PulseN. Thus, *limk1* interference in CHN seemed to decrease the rate of switching from the singing mode to silence mode and back, resulting in longer trains and ITI, while *limk1* interference in FRN neurons generally had the opposite effect.

Discussion

In mammals, LIMK1- and cofilin-dependent actin remodeling is widely involved in regulation of synaptic processes, such as exocytosis, receptor trafficking and remodeling of dendritic spines. These processes underlay long-term potentiation (LTP), long-term depression (LTD) and different forms of memory. LIMK1 also affects gene expression through CREB and LTM formation. Deregulation of LIMK1-dependent actin remodeling is involved in multiple pathologies, such as Alzheimer's and Parkinson's diseases, Williams syndrome, schizophrenia, and autism (Ben Zablah et al., 2021).

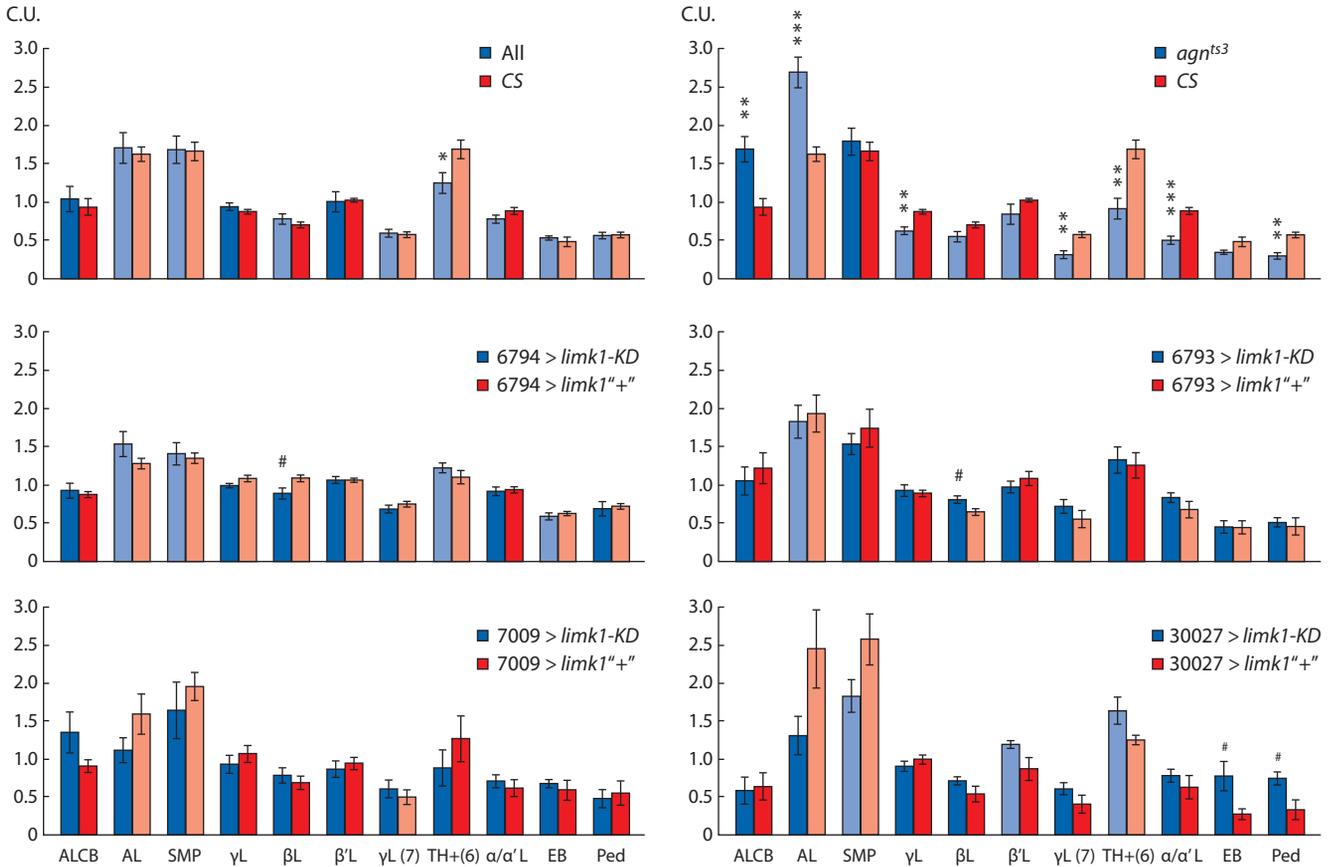


Fig. 3. LIMK1 relative level in different brain structures.

The normalized levels are shown in conventional units (C.U.). All – the averaged brain of the all studied strains. Statistical differences (two-sided *t*-test): * from CS ($p < 0.05, 0.01, 0.001$ for one, two and three asterisks, respectively); # from the control hybrid without *limk1* interference; blue > cyan and red > pink color change – the difference from ALCBB ($p < 0.05$; blue and red – no difference). $N = 47-61$ (All), $5-6$ (CS), $4-6$ (*agn^{ts3}*), $7-8$ ($6794 > limk1-KD$), $7-10$ ($6794 > limk1'+''$), $5-6$ ($6793 > limk1-KD$), $6-7$ ($6793 > limk1'+''$), $3-5$ ($7009 > limk1-KD$), $3-5$ ($7009 > limk1'+''$), 4 ($30027 > limk1-KD$), $3-4$ ($30027 > limk1'+''$). Standard error is shown.

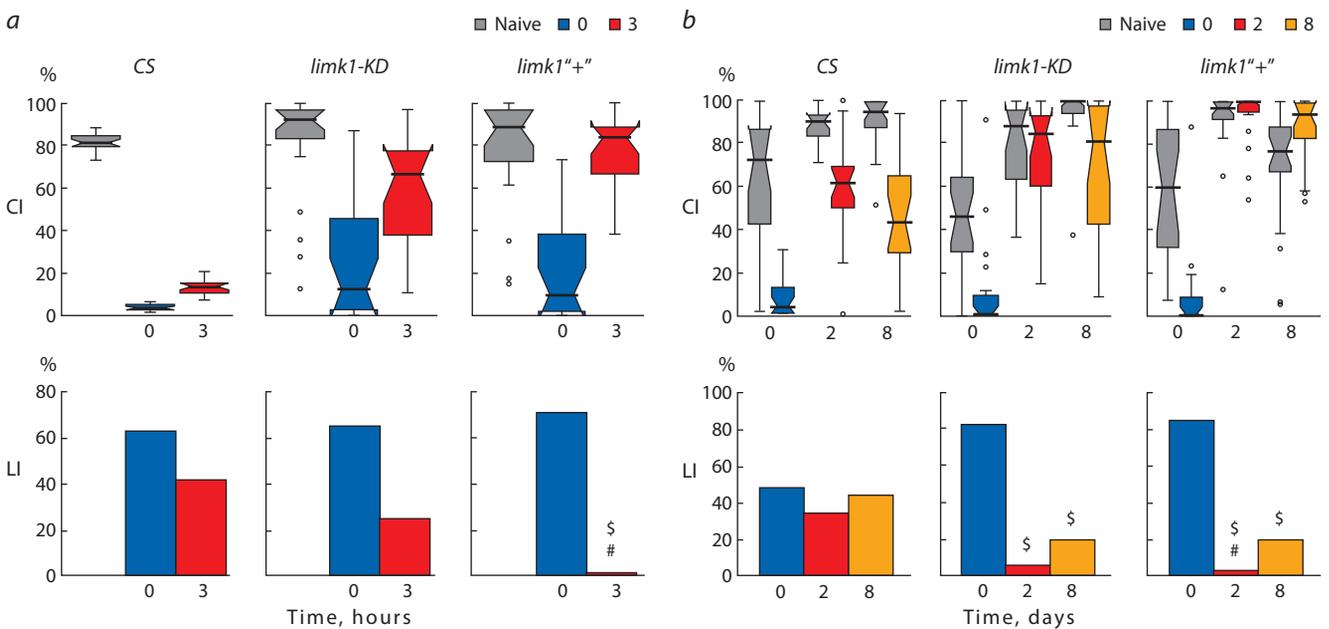


Fig. 4. The memory abilities of the *Drosophila* strain with *limk1* suppression by 6794 activator.

a, STM. *b*, LTM. CI – courtship index; box plot and whisker chart, median is shown by the bold black line. LI – learning index. Statistical differences: # from CS, \$ from LI (0 h/0 day) (two-sided randomization test; $p < 0.05, n = 20$).

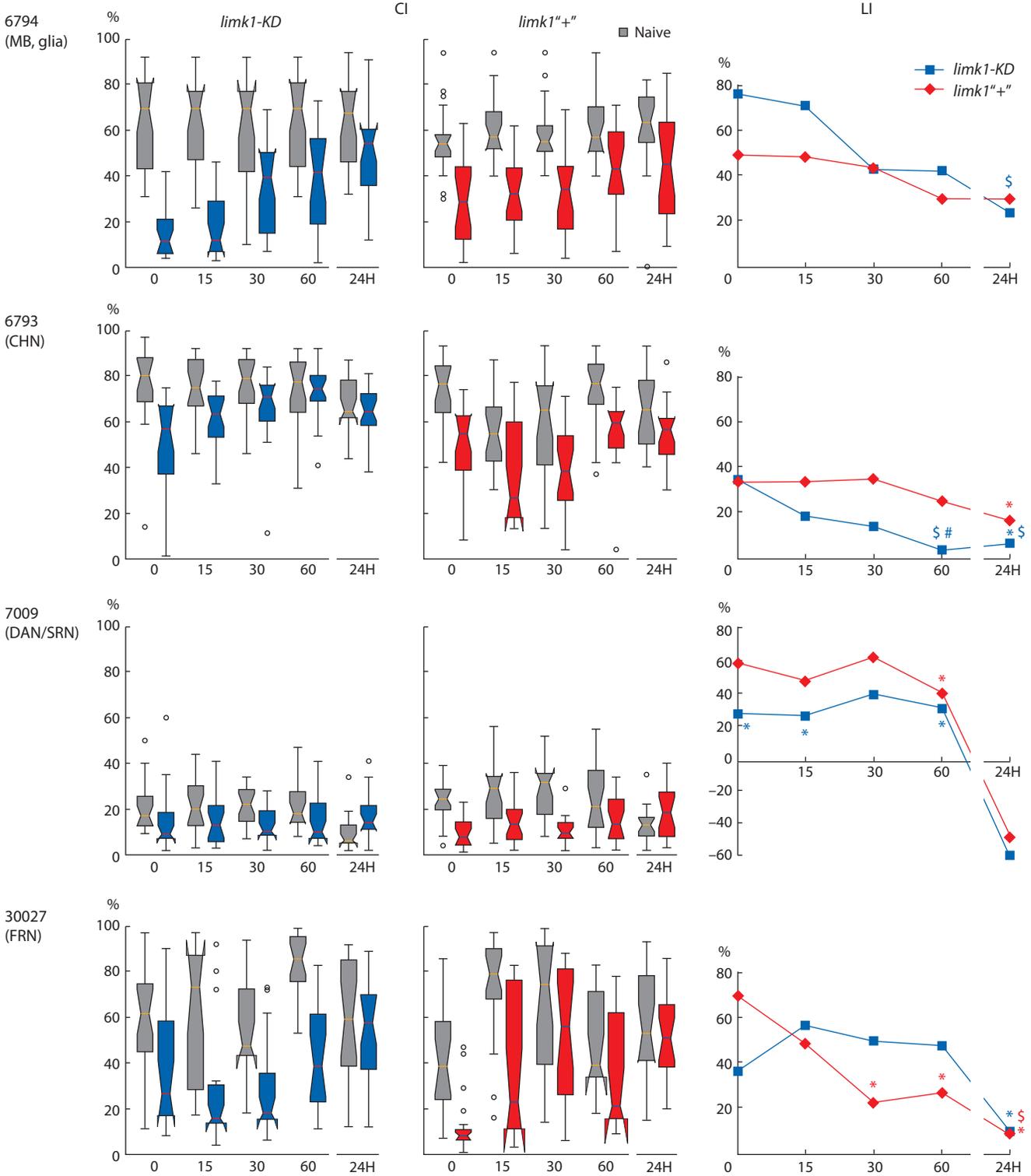


Fig. 5. STM dynamics in *Drosophila* strains with neuron type-specific *limk1* suppression.

CI – courtship index, LI – learning index. X axis: time (min; 24H – 24 hours). Statistical differences: # from CS, \$ from LI (0 min), * no difference from zero (two-sided randomization test; $p < 0.05$, $n = 20$).

In mature neurons, actin is enriched in both pre- and post-synaptic structures, such as dendritic spines, regulated by Rho signaling pathway. The action of LIMK1 on actin polymerization and memory processes is rather complex, being dependent on the mode of LIMK1 regulation (transient or

long-term overexpression) and cofilin level. While the active cofilin destabilizes fibrillar actin, in high concentrations it increases actin polymerization and nucleates actin filaments in dendritic spines during long-term potentiation (reviewed in Cuberos et al., 2015). Thus, it is hard to predict the integral

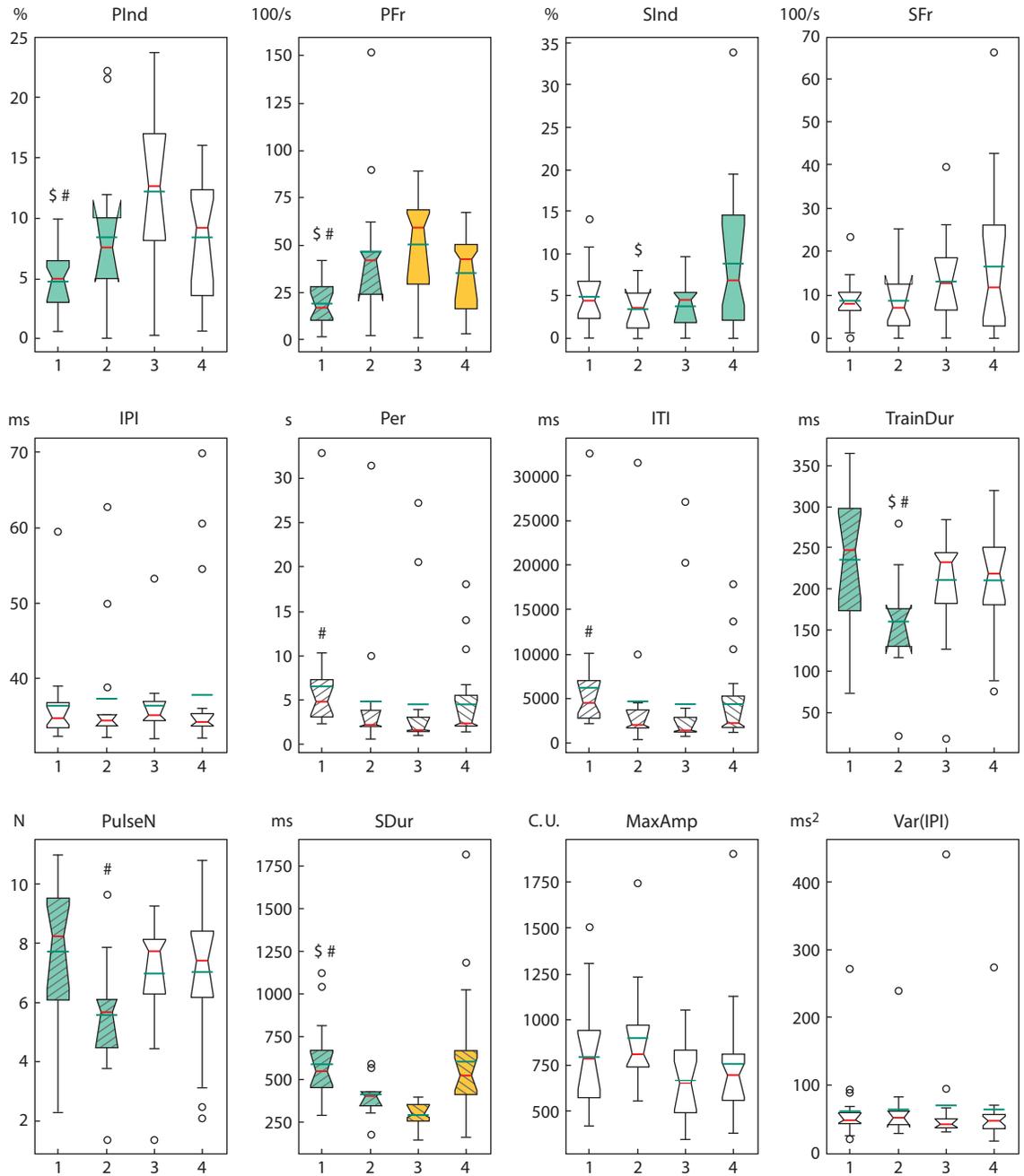


Fig. 6. Courtship song parameters in *Drosophila* hybrids with neuron type-specific *limk1* interference.

Hybrid numbers: 1. 6793 > *limk1*-KD; 2. 6793 > *limk1*"+"(CHN); 3. 30027 > *limk1*-KD; 4. 30027 > *limk1*"+"(FRN). Statistical differences: hybrids with *limk1* interference vs control – the pairs with difference are shown by color (two-sided randomization test, $p < 0.05$) or hatching (two-sided Mann–Whitney test, $p < 0.05$); hybrids with CHN driver vs hybrids with FRN driver: \$ two-sided randomization test; # two-sided Mann–Whitney test. For hybrids 1, 2, 3, 4, respectively: $n = 20, 15, 17, 22$ for pulse song parameters; $n = 18, 13, 15, 20$ for sine song parameters. Median is shown by the red line, the mean value is shown by the green line.

effect of LIMK1 and cofilin activity on memory processes. It was crucial to check the behavioral effects of *Drosophila limk1* suppression in specific types of neurons.

Using paraffin section staining, both LIMK1 and its product p-cofilin were shown to be homogeneously distributed in the brain neuropil, with maximum level in CC (Lopatina et al., 2007). Our results of the confocal microscopy analysis gave a different picture of LIMK1 distribution, quite similar for all the studied strains. We have shown a specific LIMK1 decrease

in MB, which is responsible for associative learning, as well as in CC, involved in higher movement control (Strauss, Heisenberg, 1993). This puts under question the role of LIMK1 in the aforementioned processes. However, LIMK1 was present in the cell bodies and processes of DAN, which interact with MB and CC (Mao, Davis, 2009) and regulate memory and forgetting. The observed p-cofilin distribution resembled this for LIMK1: its level was low in the MB lobes and CC. The low level of p-cofilin in the MB lobes had been previously shown

(Abe et al., 2014). In contrast to LIMK1, p-cofilin level was low in Cal formed by PN and KC terminals and high in cell nuclei. The latter corresponds to its functioning in the cell, as cofilin phosphorylation is necessary for its translocation into the nucleus (Abe et al., 1993).

The effectiveness of *limk1* suppression at the RNA level was confirmed using Act-GAL4 activator in the whole *Drosophila* body. GAL4 was also active in specific brain areas of the corresponding strains. However, we failed to quantitatively check the changes of *limk1* expression in *Drosophila* UAS × GAL4 hybrids with neuronal-specific GAL4 expression. The decrease in LIMK1 level might be local or too small. *limk1* interference might also induce the compensatory activation of LIMK1 translation.

To study *limk1* knockdown effects on memory, we used CCSP modification applied by (Kamyshev et al., 1999): training was performed with the mated female. In this case, courtship learning results from the rise of sensitivity to the antiaphrodisiac cis-vaccenyl acetate (cVA) due to unsuccessful courtship. cVA is not required for learning, being necessary for memory performance. aSP13 DAN, which innervate the *fru*-positive tip of γ L, are necessary and sufficient for courtship learning (Keleman et al., 2012). 24 h memory consolidation requires the prolonged aSP13 stimulation and Orb2 dimerization in some γ neurons (Krüttner et al., 2015). α/β neurons are involved in LTM processes (Redt-Clouet et al., 2012; Jones et al., 2018). Hence, other DAN innervating α/β L of MB, including PAM and PPL1 cells (Aso et al., 2014a), may also be involved in LTM.

The behavioral differences were observed after *limk1* interference, e. g., the restoration of 3 h STM for *limk1-KD* > 6794 strain. GAL4 drivers themselves affected memory abilities, which were generally decreased compared to *CS*. The drivers also significantly affected the forgetting curves. Thus, we studied the effects of *limk1* interference relative to a control strain with the same GAL4 driver. We applied two different crossing schemes – UAS (f) > 6794 (m) (for 3 h STM and LTM analysis) and reverse (in the other behavioral experiments). In the first case, the control UAS > *limk1*“+” hybrid had bright eyes due to *v[1]* recessive allele, in contrast to UAS > *limk1-KD* hybrid with the wild type dark red eyes. The observed 3 h STM differences are unlikely to be associated with the differences in eyes pigmentation, as *v[1]* flies showed a normal 3 h STM and 2 day LTM in CCSP (Nikitina et al., 2021), while both forms of memory were impaired in the control strain. However, memory retention depends on parent affect, with some paternal epigenetic factors affecting STM strength (Medvedeva et al., 2021). For 6794 > *limk1-KD* strain, we did not see STM difference from the control strain, though learning ability slightly increased after *limk1* knockdown (see Fig. 5). Thus, when studying LIMK1 effects on learning and memory, it is necessary to consider the crossing direction.

Acetylcholine is the major excitatory neurotransmitter in *Drosophila*. Among CNH are: PN forming synapses on KC of MB (Yasuyama et al., 2002), the MB intrinsic neurons that are responsible for olfactory memory, expressing ChAT and VAcT (Barnstedt et al., 2016), and the α/β core neurons required for LTM consolidation (Yi et al., 2013). In the hybrids with 6793 driver (GAL4 expressed in CHN), GFP level was

specifically high in α/β L compared to the other MB lobes. Here, *limk1* interference resulted in faster STM forgetting. This contradicts the cofilin role in active forgetting shown in OAVL, where cofilin was proposed to be regulated by LIMK1 (Shuai, Zhong, 2010). The involvement of LIMK1 and cofilin in forgetting may occur locally, within specific neuronal populations or synaptic terminals. At the same time, LIMK1 may be crucial for memory storage and retrieval in CCSP. The glutamatergic MBON M6 neurons serve for STM output: aSP13 DAN prolongs potentiation of γ L – M6 synapses (Zhao et al., 2018). Some cholinergic MBON appeared to regulate the *Drosophila* visual appetitive memory (Aso et al., 2014b). As the extrinsic MB cells responsible for CCSP memory were similar to those used for appetitive memory (Montague, Baker, 2016), the decrease in 60 min STM might occur due to *limk1* suppression in some of these neurons.

The hybrids with 7009 driver (DAN and SRN) showed generally low CI values and negative LI values 24 h after learning. Males of these strains had pale pink eyes because of defects of eyes pigmentation, due to non-complete *w[1118]* rescue. *w[1118]* males demonstrated low courtship activity and success, presumably due to some defects of sexual development and maturation (Xiao et al., 2017). However, the control 7009 > *limk1*“+” strain had normal LI up to 60 min after training. *limk1* knockdown by 7009 driver was associated with dramatic defects of learning and memory: LI just after training did not statistically differ from zero. Thus, LIMK1-dependent signaling in DAN and SRN seems to be important for learning and memory in CCSP.

limk1 knockdown by 30027 driver (FRN) decreased the forgetting rate in the time interval 30–60 min. This corresponds to the role of actin-remodeling pathway in forgetting in OAVL paradigm (Shuai et al., 2010). LI of the control strain did not differ from zero starting from 30 min after learning, while *limk1* knockdown increased it. In males, FRN are responsible for courtship behavior. There are ~1500 FRN in the *Drosophila* brain, including sensory organs, lateral horn, lateral protocerebrum, SMP arch and motor control centers. Together they provide multisensory integration to regulate the male courtship process (Yu et al., 2010; Liu et al., 2019). Some CHN and DAN are also *Fru*-positive, such as ~300 γ L neurons and aSP13 DAN located in SMP, which regulates courtship learning and memory. The activity of *fru* gene was reported to decrease upon LTM formation in CCSP (Winbush et al., 2012). Hence, suppression of some FRN activity may be associated with memory prolongation and consolidation.

In addition to memory processes, *limk1* interference affected some parameters of the male courtship song. As well as for courtship memory, we observed the opposite functional effects of *limk1* knockdown in FRN and CHN. FRN of the P1 class initiate *Drosophila* courtship behavior and trigger courtship song. pIP10 neurons possibly convey the P1 signal to thoracic dPR1 and vPR6 neurons, proposed to be the parts of a central pattern generator (CPG), which defines the time and shape of the pulse song. vPR6 possibly encode IPI (von Philipsborn et al., 2011). Pulse and sine CPG either contain FRN or interact with them. As sine and pulse song normally do not overlap, the mutually inhibitory mechanisms must exist, switching

between quasilinear and relaxation modes of oscillation for sine and pulse song, respectively. Some descending interneurons may control the type of the song, while the others trigger singing or terminate the song (Clyne, Miesenböck, 2008).

Indeed, we observed the opposite changes of PInd/PFr and SInd/SDur upon *limk1* interference in CHN and FRN, moving the balance toward the sine and pulse song, respectively. The increase in PFr after *limk1* knockdown in FRN might indicate the negative role of LIMK1-dependent signaling on activity of the pulse CPG or the upstream brain centers, which switch them from active to silent mode. CC is important for control of stability of pacemakers, which regulate the rhythmic structure of courtship song. PB destruction leads to sound signal distortions, FB and EB destruction additionally decreases sine and pulse trains (Popov et al., 2004). CC includes a large number of neuronal types, such as CHN, DAN, SRN, and others. CHN are present in FB, EB, No and PB (Kahsai, Winther, 2011), similarly to what we observed in our research. Hence, they probably play some role in regulation of male singing. The opposite effects of *limk1* interference in CHN and FRN may indicate a specific role of LIMK1 in courtship controlling network, whereas the other parts of the brain possibly have a total antagonistic effects on its activity. Alternatively, CHN and FRN *fru* neurons may differ in some aspects of regulation of LIMK1-dependent signaling pathway.

Conclusion

In summary, we have shown that effects of *limk1* interference in *Drosophila* male courtship memory and song depend on both the neuronal type and specific behavioral parameter. *limk1* interference in CHN and FRN had generally opposite effects, whereas its suppression in DAN and SRN impaired the flies' ability to learn. Using activator strains with a narrower pattern of GAL4 expression would help to better localize the brain structures, where LIMK1 regulates memory and forgetting in CCSP. Among such putative structures are γ L and aSP13 DAN innervating γ 5 area, as well as other DAN participating in memory formation, consolidation and retrieval. Studying the behavioral consequences of *limk1* overexpression in different brain areas will complement the estimates of the effects of its suppression. The above investigation should also focus on LIMK1 partner proteins, such as cofilin in its active and phosphorylated form.

References

- Abe H., Nagaoka R., Obinata T. Cytoplasmic localization and nuclear transport of cofilin in cultured myotubes. *Exp. Cell Res.* 1993; 206(1):1-10. DOI 10.1006/excr.1993.1113.
- Abe T., Yamazaki D., Murakami S., Hiroi M., Nitta Y., Maeyama Y., Tabata T. The NAV2 homolog Sickie regulates F-actin-mediated axonal growth in *Drosophila* mushroom body neurons via the non-canonical Rac-Cofilin pathway. *Development.* 2014;141(24):4716-4728. DOI 10.1242/dev.113308.
- Albin S.D., Kaun K.R., Knapp J.M., Chung P., Heberlein U., Simpson J.H. A subset of serotonergic neurons evokes hunger in adult *Drosophila*. *Curr. Biol.* 2015;25(18):2435-2440. DOI 10.1016/j.cub.2015.08.005.
- Aso Y., Grübel K., Busch S., Friedrich A.B., Siwanowicz I., Tanimoto H. The mushroom body of adult *Drosophila* characterized by GAL4 drivers. *J. Neurogenet.* 2009;23(1-2):156-172. DOI 10.1080/01677060802471718.
- Aso Y., Hattori D., Yu Y., Johnston R.M., Iyer N.A., Ngo T.B., Dionne H., Abbott L.F., Axel R., Tanimoto H., Rubin G.M. The neuronal architecture of the mushroom body provides a logic for associative learning. *eLife.* 2014a;3:e04577. DOI 10.7554/eLife.04577.
- Aso Y., Sitaraman D., Ichinose T., Kaun K.R., Vogt K., Belliard-Guérin G., Plaçais P., Robie A.A., Yamagata N., Schnaitmann C., Rowell W.J., Johnston R.M., Ngo T.B., Chen N., Korff W., Nita-bach M.N., Heberlein U., Preat T., Branson K.M., Tanimoto H., Rubin G.M. Mushroom body output neurons encode valence and guide memory-based action selection in *Drosophila*. *eLife.* 2014b; 3:e04580. DOI 10.7554/eLife.04580.
- Barnstedt O., Oswald D., Felsenberg J., Brain R., Moszynski J.P., Talbot C.B., Perrat P.N., Waddell S. Memory-relevant mushroom body output synapses are cholinergic. *Neuron.* 2016;89(6):1237-1247. DOI 10.1016/j.neuron.2016.02.015.
- Ben Zablah Y., Zhang H., Gugustea R., Jia Z. LIM-kinases in synaptic plasticity, memory, and brain diseases. *Cells.* 2021;10(8):2079-2103. DOI 10.3390/cells10082079.
- Berry J.A., Cervantes-Sandoval I., Nicholas E.P., Davis R.L. Dopamine is required for learning and forgetting in *Drosophila*. *Neuron.* 2012;74(3):530-542. DOI 10.1016/j.neuron.2012.04.007.
- Berry J.A., Phan A., Davis R.L. Dopamine neurons mediate learning and forgetting through bidirectional modulation of a memory trace. *Cell Rep.* 2018;25(3):651-662. DOI 10.1016/j.celrep.2018.09.051.
- Cervantes-Sandoval I., Chakraborty M., MacMullen C., Davis R.L. Scribble scaffolds a signalosome for active forgetting. *Neuron.* 2016;90(6):1230-1242. DOI 10.1016/j.neuron.2016.05.010.
- Clyne J.D., Miesenböck G. Sex-specific control and tuning of the pattern generator for courtship song in *Drosophila*. *Cell.* 2008;133(2): 354-363. DOI 10.1016/j.cell.2008.01.050.
- Cook K.R., Parks A.L., Jacobus L.M., Kaufman T.C., Matthews K.A. New research resources at the Bloomington *Drosophila* Stock Center. *Fly (Austin).* 2010;4(1):88-91. DOI 10.4161/fly.4.1.11230.
- Cuberos H., Vallée B., Vourc'h P., Tastet J., Andres C.R., Bénédetti H. Roles of LIM kinases in central nervous system function and dysfunction. *FEBS Lett.* 2015;589(24 Pt.B):3795-3806. DOI 10.1016/j.febslet.2015.10.032.
- Davis R.L., Zhong Y. The biology of forgetting – a perspective. *Neuron.* 2017;95(3):490-503. DOI 10.1016/j.neuron.2017.05.039.
- Duffy J.B. GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis.* 2002;34(1-2):1-15. DOI 10.1002/gene.10150.
- Gao Y., Shuai Y., Zhang X., Peng Y., Wang L., He J., Zhong Y., Li Q. Genetic dissection of active forgetting in labile and consolidated memories in *Drosophila*. *Proc. Natl. Acad. Sci. USA.* 2019;116(42): 21191-21197. DOI 10.1073/pnas.1903763116.
- Hartenstein V. Morphological diversity and development of glia in *Drosophila*. *Glia.* 2011;59(9):1237-1252. DOI 10.1002/glia.21162.
- Iliadi K.G., Kamyshev N.G., Popov A.V., Iliadi N.N., Rashkovets-kaya E.L., Nevo E., Korol A.B. Peculiarities of the courtship song in the *Drosophila melanogaster* populations adapted to gradient of microecological conditions. *J. Evol. Biochem. Physiol.* 2009;45(5): 579-588. DOI 10.1134/S0022093009050041.
- Jones S.G., Nixon K.C.J., Chubak M.C., Kramer J.M. Mushroom body specific transcriptome analysis reveals dynamic regulation of learning and memory genes after acquisition of long-term courtship memory in *Drosophila*. *G3: Genes Genomes Genetics (Bethesda).* 2018;8(11):3433-3446. DOI 10.1534/g3.118.200560.
- Kahsai L., Winther A.M. Chemical neuroanatomy of the *Drosophila* central complex: distribution of multiple neuropeptides in relation to neurotransmitters. *J. Comp. Neurol.* 2011;519(2):290-315. DOI 10.1002/cne.22520.
- Kamyshev N.G., Iliadi K.G., Bragina J.V. *Drosophila* conditioned courtship: two ways of testing memory. *Learn. Mem.* 1999;6(1): 1-20.
- Kasture A.S., Hummel T., Susic S., Freissmuth M. Big lessons from tiny flies: *Drosophila melanogaster* as a model to explore dysfunction of dopaminergic and serotonergic neurotransmitter systems. *Int. J. Mol. Sci.* 2018;19(6):1788-1805. DOI 10.3390/ijms19061788.

- Keleman K., Vrontou E., Krüttner S., Yu J.Y., Kurtovic-Kozaric A., Dickson B.J. Dopamine neurons modulate pheromone responses in *Drosophila* courtship learning. *Nature*. 2012;489(7414):145-149. DOI 10.1038/nature11345.
- Krüttner S., Traummüller L., Dag U., Jandrasits K., Stepien B., Iyer N., Fradkin L.G., Noordermeer J.N., Mensh B.D., Keleman K. Synaptic Orb2A bridges memory acquisition and late memory consolidation in *Drosophila*. *Cell Rep*. 2015;11(12):1953-1965. DOI 10.1016/j.celrep.2015.05.037.
- Liu W., Ganguly A., Huang J., Wang Y., Ni J.D., Gurav A.S., Aguilar M.A., Montell C. Neuropeptide F regulates courtship in *Drosophila* through a male-specific neuronal circuit. *eLife*. 2019;8:1-29. DOI 10.7554/eLife.49574.
- Lopatina N.G., Zachepilo T.G., Chesnokova E.G., Savvateeva-Popova E.V. Mutations in structural genes of tryptophan metabolic enzymes of the kynurenine pathway modulate some units of the L-glutamate receptor – actin cytoskeleton signaling cascade. *Russ. J. Genet*. 2007;43(10):1168-1172. DOI 10.1134/S1022795407100110.
- Mao Z., Davis R.L. Eight different types of dopaminergic neurons innervate the *Drosophila* mushroom body neuropil: anatomical and physiological heterogeneity. *Front. Neural Circuits*. 2009;3:1-17. DOI 10.3389/neuro.04.005.2009.
- Mazzoni D., Dannenberg R. Audacity®. Audacity Team, 2020. URL: <https://audacityteam.org>. Last access: 05.12.2021.
- Medina J.H. Neural, cellular and molecular mechanisms of active forgetting. *Front. Syst. Neurosci*. 2018;12:1-10. DOI 10.3389/fnsys.2018.00003.
- Medvedeva A.V., Molotkov D.A., Nikitina E.A., Popov A.V., Karagodin D.A., Baricheva E.M., Savvateeva-Popova E.V. Systemic regulation of genetic and cytogenetic processes by a signal cascade of actin remodeling: locus *agnostic* in *Drosophila*. *Russ. J. Genet*. 2008;44(6):669-681. DOI 10.1134/S1022795408060069.
- Medvedeva A.V., Tokmatcheva E.V., Kaminskaya A.N., Vasileva S.A., Nikitina E.A., Zhuravlev A.V., Zakharov G.A., Zatsepina O.G., Savvateeva-Popova E.V. Parent-of-origin effects on nuclear chromatin organization and behavior in a *Drosophila* model for Williams–Beuren Syndrome. *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov Journal of Genetics and Breeding*. 2021;25(5):472-485. DOI 10.18699/VJ21.054.
- Montague S.A., Baker B.S. Memory elicited by courtship conditioning requires mushroom body neuronal subsets similar to those utilized in appetitive memory. *PLoS One*. 2016;11(10):e0164516. DOI 10.1371/journal.pone.0164516.
- Ng J., Luo L. Rho GTPases regulate axon growth through convergent and divergent signaling pathways. *Neuron*. 2004;44(5):779-793. DOI 10.1016/j.neuron.2004.11.014.
- Nikitina E.A., Medvedeva A.V., Zakharov G.A., Savvateeva-Popova E.V. The *Drosophila* *agnostic* locus: involvement in the formation of cognitive defects in Williams syndrome. *Acta Naturae*. 2014;6(2):53-61. DOI 10.32607/20758251-2014-6-2-53-61.
- Nikitina E.A., Zhuravlev A.V., Savvateeva-Popova E.V. Effect of impaired kynurenine synthesis on memory in *Drosophila*. *Integrativnaya Fiziologiya = Integrative Physiology*. 2021;2(1):49-60. DOI 10.33910/2687-1270-2021-2-1-49-60. (in Russian)
- Okamoto N., Nishimura T. Signaling from glia and cholinergic neurons controls nutrient-dependent production of an insulin-like peptide for *Drosophila* body growth. *Dev. Cell*. 2015;35(3):295-310. DOI 10.1016/j.devcel.2015.10.003.
- Popov A.V., Peresleni A.I., Savvateeva-Popova E.V., Vol'f R., Heisenberg M. The role of the mushroom bodies and of the central complex of *Drosophila melanogaster* brain in the organization of courtship behavior and communicative sound production. *J. Evol. Biochem. Phys.* 2004;40:641-652. DOI 10.1007/s10893-004-0005-z.
- Redt-Clouet C., Trannoy S., Boulanger A., Tokmatcheva E., Savvateeva-Popova E., Parmentier M.L., Preat T., Dura J.M. Mushroom body neuronal remodeling is necessary for short-term but not for long-term courtship memory in *Drosophila*. *Eur. J. Neurosci*. 2012;35(11):1684-1691. DOI 10.1111/j.1460-9568.2012.08103.x.
- Ritchie M.G., Yate V.H., Kyriacou C.P. Genetic variability of the inter-pulse interval of courtship song among some European populations of *Drosophila melanogaster*. *Heredity (Edinb)*. 1994;72(Pt.5):459-464. DOI 10.1038/hdy.1994.64.
- Salvaterra P.M., Kitamoto T. *Drosophila* cholinergic neurons and processes visualized with Gal4/UAS-GFP. *Brain Res. Gene Expr. Patterns*. 2001;1(1):73-82. DOI 10.1016/S1567-133X(01)00011-4.
- Savvateeva-Popova E.V., Popov A.V., Grossman A., Nikitina E.A., Medvedeva A.V., Peresleni A.I., Molotkov D.A., Kamyshev N.G., Pyatkov K.I., Zatsepina O.G., Schostak N., Zelentsova E.S., Pavlova G., Pantelev D., Riederer P., Evgen'ev M.B. Non-coding RNA as a trigger of neuropathologic disorder phenotypes in transgenic *Drosophila*. *J. Neural Transm. (Vienna)*. 2008;115(12):1629-1642. DOI 10.1007/s00702-008-0078-8.
- Savvateeva-Popova E.V., Popov A.V., Heinemann T., Riederer P. *Drosophila* mutants of the kynurenine pathway as a model for ageing studies. *Adv. Exp. Med. Biol*. 2003;527:713-722. DOI 10.1007/978-1-4615-0135-0_84.
- Savvateeva-Popova E.V., Zhuravlev A.V., Brázda V., Zakharov G.A., Kaminskaya A.N., Medvedeva A.V., Nikitina E.A., Tokmatcheva E.V., Dolgaya J.F., Kulikova D.A., Zatsepina O.G., Funikov S.Y., Ryazansky S.S., Evgen'ev M.B. *Drosophila* model for the analysis of genesis of LIM-kinase 1-dependent Williams–Beuren syndrome cognitive phenotypes: INDELS, transposable elements of the Tc1/mariner superfamily and microRNAs. *Front. Genet*. 2017;8:123. DOI 10.3389/fgene.2017.00123.
- Shuai Y., Lu B., Hu Y., Wang L., Sun K., Zhong Y. Forgetting is regulated through Rac activity in *Drosophila*. *Cell*. 2010;140(4):579-589. DOI 10.1016/j.cell.2009.12.044.
- Shuai Y., Zhong Y. Forgetting and small G protein Rac. *Protein Cell*. 2010;1(6):503-506. DOI 10.1007/s13238-010-0077-z.
- Siegel R.W., Hall J.C. Conditioned responses in courtship behavior of normal and mutant *Drosophila*. *Proc. Natl. Acad. Sci. USA*. 1979;76(7):3430-3434. DOI 10.1073/pnas.76.7.343.
- Strauss R., Heisenberg M. A higher control center of locomotor behavior in the *Drosophila* brain. *J. Neurosci*. 1993;13(5):1852-1861. DOI 10.1523/JNEUROSCI.13-05-01852.1993.
- Sun B., Xu P., Salvaterra P.M. Dynamic visualization of nervous system in live *Drosophila*. *Proc. Natl. Acad. Sci. USA*. 1999;96(18):10438-10443. DOI 10.1073/pnas.96.18.10438.
- Thapa A., Sullivan S.M., Nguyen M.Q., Buckley D., Ngo V.T., Dada A.O., Blankinship E., Cloud V., Mohan R.D. Brief freezing steps lead to robust immunofluorescence in the *Drosophila* nervous system. *Biotechniques*. 2019;67(6):299-305. DOI 10.2144/btn-2018-0067.
- Tully T., Preat T., Boynton S.C., Vecchio M.D. Genetic dissection of consolidated memory in *Drosophila*. *Cell*. 1994;79(1):35-47. DOI 10.1016/0092-8674(94)90398-0.
- Virtual Fly Brain. URL: <https://v2.virtualflybrain.org>. Last access: 01.12.2021.
- von Philipsborn A.C., Liu T., Yu J.Y., Masser C., Bidaye S.S., Dickson B.J. Neuronal control of *Drosophila* courtship song. *Neuron*. 2011;69(3):509-522. DOI 10.1016/j.neuron.2011.01.011.
- Winbush A., Reed D., Chang P.L., Nuzhdin S.V., Lyons L.C., Arbeitman M.N. Identification of gene expression changes associated with long-term memory of courtship rejection in *Drosophila* males. *G3: Genes Genomes Genetics (Bethesda)*. 2012;2(11):1437-1445. DOI 10.1534/g3.112.004119.
- Xiao C., Qiu S., Robertson R.M. The white gene controls copulation success in *Drosophila melanogaster*. *Sci. Rep*. 2017;7(1):7712-7725. DOI 10.1038/s41598-017-08155-y.
- Yasuyama K., Meinertzhagen I.A., Schürmann F.W. Synaptic organization of the mushroom body calyx in *Drosophila melanogaster*. *J. Comp. Neurol*. 2002;445(3):211-226. DOI 10.1002/cne.10155.
- Yi W., Zhang Y., Tian Y., Guo J., Li Y., Guo A. A subset of cholinergic mushroom body neurons requires Go signaling to regulate sleep in *Drosophila*. *Sleep*. 2013;36:1809-1821. DOI 10.5665/sleep.3206.

- Yu J.Y., Kanai M.I., Demir E., Jefferis G.S., Dickson B.J. Cellular organization of the neural circuit that drives *Drosophila* courtship behavior. *Curr. Biol.* 2010;20(18):1602-1614. DOI 10.1016/j.cub.2010.08.025.
- Zalomaeva E.S., Falina V.S., Medvedeva A.V., Nikitina E.A., Savvateeva-Popova E.V. Learning and forgetting in *Drosophila melanogaster* in *limk1* gene polymorphism. *Integrativnyaya Fiziologiya = Integrative Physiology.* 2021;2(3):318-327. DOI 10.33910/2687-1270-2021-2-3-318-327. (in Russian)
- Zhang X., Li Q., Wang L., Liu Z.J., Zhong Y. Cdc42-dependent forgetting regulates repetition effect in prolonging memory retention. *Cell Rep.* 2016;16(3):817-825. DOI 10.1016/j.celrep.2016.06.041.
- Zhao X., Lenek D., Dag U., Dickson B.J., Keleman K. Persistent activity in a recurrent circuit underlies courtship memory in *Drosophila*. *eLife.* 2018;7:1-16. DOI 10.7554/eLife.31425.
- Zhuravlev A.V., Shchegolev B.F., Zakharov G.A., Ivanova P.N., Nikitina E.A., Savvateeva-Popova E.V. 3-Hydroxykynurenine as a potential ligand for Hsp70 proteins and its effects on *Drosophila* memory after heat shock. *Mol. Neurobiol.* 2022;59(3):1862-1871. DOI 10.1007/s12035-021-02704-3.
- Zhuravlev A.V., Vetrovoy O.V., Ivanova P.N., Savvateeva-Popova E.V. 3-Hydroxykynurenine in regulation of *Drosophila* behavior: the novel mechanisms for cardinal phenotype manifestations. *Front. Physiol.* 2020;11:971. DOI 10.3389/fphys.2020.00971.

ORCID ID

A.V. Zhuravlev orcid.org/0000-0003-2673-4283
E.S. Zalomaeva orcid.org/0000-0002-6005-3433
E.S. Egozova orcid.org/0000-0002-0055-3778

A.D. Emelin orcid.org/0000-0002-8498-9017
V.V. Sokurova orcid.org/0000-0003-0821-9974
E.A. Nikitina orcid.org/0000-0003-1897-8392
E.V. Savvateeva-Popova orcid.org/0000-0002-6925-4370

Acknowledgements. This research was funded by the Russian Foundation for Basic Research (No. 20-015-00300 A). We thank Prof. E. Buchner (Institute of Clinical Neurobiology, Würzburg) for *Drosophila* anti-CSP antibodies. We also thank the Bloomington *Drosophila* Stock Center (NIH P40OD018537) at Indiana University (USA) and TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing us with transgenic GAL4 and UAS fly stocks. This study was supported by the State Program 47 GP "Scientific and Technological Development of the Russian Federation" (2019–2030), theme 0134-2019-0004.

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

Received June 9, 2022. Revised August 19, 2022. Accepted August 24, 2022.

Original Russian text <https://vavilovj-icg.ru/>

On the nature of picobirnaviruses

A.Yu. Kashnikov , N.V. Epifanova, N.A. Novikova

I.N. Blokhina Nizhny Novgorod Research Institute of Epidemiology and Microbiology, Nizhny Novgorod, Russia
 a.kashn@yandex.ru

Abstract. The picobirnaviruses (*Picobirnaviridae*, *Picobirnavirus*, PBVs) are currently thought to be animal viruses, as they are usually found in animal stool samples. However, no animal model or cell culture for their propagation has yet been found. In 2018, a hypothetical assumption about PBVs belonging to prokaryotic viruses was put forward and experimentally substantiated. This hypothesis is based on the presence of Shine–Dalgarno sequences in the genome of all PBVs before three reading frames (ORF) at the ribosomal binding site, with which the prokaryotic genome is saturated, while in the eukaryotic genome such regions occur with low frequency. The genome saturation with the Shine–Dalgarno sequences, as well as the preservation of this saturation in the progeny, according to scientists, allows us to attribute PBVs to prokaryotic viruses. On the other hand, there is a possibility that PBVs belong to viruses of eukaryotic hosts – fungi or invertebrates, since PBV-like sequences similar to the genome of fungal viruses from the families of mitoviruses and partitiviruses have been identified. In this regard, the idea arose that, in terms of reproduction mode, PBVs resemble fungal viruses. The divergence of views on the true PBV host(s) has sparked discussions among scientists and required further research to elucidate their nature. The review highlights the results of the search for a PBV host. The reasons for the occurrence of atypical sequences among the PBV genome sequences that use an alternative mitochondrial code of lower eukaryotes (fungi and invertebrates) for the translation of viral RNA-dependent RNA polymerase (RdRp) instead of the standard genetic code are analyzed. The purpose of the review was to collect arguments in support of the hypothesis about the phage nature of PBVs and to find the most realistic explanation of the reasons for identifying non-standard genomic sequences for PBVs. Based on the hypothesis about the genealogical relationship of PBVs with RNA viruses from other families with similar segmented genomes, such as *Reoviridae*, *Cystoviridae*, *Totiviridae* and *Partitiviridae*, virologists support the assumption of a decisive role in the origin of atypical PBV-like reassortment strains between PBVs and viruses of the listed families. The collected arguments given in this review indicate a high probability of a phage nature of PBVs. The data presented in the review show that the belonging of PBV-like progeny to prokaryotic or eukaryotic viruses is determined not only by its genome saturation level with a prokaryotic motif, standard or mitochondrial genetic code. The primary structure of the gene encoding the viral capsid protein responsible for the presence or absence of specific proteolytic properties of the virus that determine its ability for independent horizontal transmission into new cells may also be a decisive factor.

Key words: picobirnavirus; genome segment; host cell; mitochondrial genetic code; phylogenetic tree; reassortment.

For citation: Kashnikov A.Yu., Epifanova N.V., Novikova N.A. On the nature of picobirnaviruses. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2023;27(3):264-275. DOI 10.18699/VJGB-23-32

О природе пикобирнавирусов

А.Ю. Кашников , Н.В. Епифанова, Н.А. Новикова

Нижегородский научно-исследовательский институт эпидемиологии и микробиологии им. академика И.Н. Блохиной Роспотребнадзора, Нижний Новгород, Россия
 a.kashn@yandex.ru

Аннотация. Считается, что пикобирнавирусы (*Picobirnaviridae*, *Picobirnavirus*, ПБВ) являются вирусами животных, так как их обычно находят в образцах стула животных. Однако до сих пор не найдена модель животного или клеточная культура для их размножения. В 2018 г. было выдвинуто и экспериментально обосновано гипотетическое предположение о принадлежности ПБВ к вирусам прокариот. Эта гипотеза основана на присутствии в геноме всех ПБВ перед тремя открытыми рамками считывания (ORF) в сайте связывания с рибосомой последовательностей Шайна–Дальгарно, которыми насыщен геном прокариот, тогда как в геноме эукариот такие участки встречаются с низкой частотой. Насыщенность генома последовательностями Шайна–Дальгарно, а также сохранение такой насыщенности у потомства, по мнению ученых, позволяет отнести ПБВ к вирусам прокариот. В то же время существует вероятность принадлежности ПБВ к вирусам эукариотических хозяев – грибов или беспозвоночных, поскольку были обнаружены ПБВ-подобные последовательности, сходные с геномом вирусов грибов из семейств митовирусов и партитивирусов. В связи с этим возникло представление, что по способу репродукции ПБВ напоминают вирусы грибов. Расхождение во взглядах на истинного хозяина (хозяев) ПБВ вызвало дискуссию среди ученых и потребовало дальнейших исследований для выяснения их природы.

В обзоре освещены результаты исследований по поиску хозяина ПБВ. Проанализированы причины появления среди множества характерных для ПБВ последовательностей генома атипичных последовательностей, использующих для трансляции вирусной РНК-зависимой РНК-полимеразы (RdRp) вместо стандартного генетического кода альтернативный митохондриальный код низших эукариот (грибов и беспозвоночных). Цель обзора состояла в сборе аргументов в поддержку гипотезы, полагающей фаговую природу ПБВ и поиск наиболее реалистичного объяснения причин выявления нестандартных для ПБВ геномных последовательностей. Опираясь на гипотезу о генеалогическом родстве ПБВ с РНК-вирусами из других семейств со сходным сегментированным геномом, такими как *Reoviridae*, *Cystoviridae*, *Totiviridae* и *Partitiviridae*, вирусологи поддерживают предположение о решающей роли в происхождении атипичных ПБВ-подобных штаммов реассортации между ПБВ и вирусами перечисленных семейств. Собранные аргументы свидетельствуют о большой вероятности фаговой природы у ПБВ. Представленные в обзоре данные свидетельствуют, что принадлежность ПБВ-подобного потомства к вирусам прокариот или эукариот определяется не только степенью насыщения его генома прокариотическим мотивом, стандартным или митохондриальным генетическим кодом. Решающим фактором может являться также первичная структура гена, кодирующего белок вирусного капсида, отвечающего за наличие или отсутствие специфических протеолитических свойств у вируса, определяющих его способность к самостоятельному горизонтальному распространению в новые клетки.

Ключевые слова: пикобирнавирус; сегмент генома; клетка-хозяин; митохондриальный генетический код; филогенетическое дерево; реассортация.

Introduction

Picobirnaviruses (PBVs) are small, nonenveloped bisegmented double-stranded RNA viruses that have been detected in a wide variety of animal species including invertebrates and in environmental samples. Since PBVs are ubiquitous in faeces/gut contents of humans and other animals with or without diarrhea, they were considered opportunistic enteric pathogens of mammals and avian species. However, an animal cell culture or a gnotobiotic animal for propagation of this virus has not yet been identified. This fact led some researchers to doubt that picobirnaviruses belong to eukaryotic viruses. Indian scientists Krishnamurthy, Wang (2018) have analyzed a large number of full-size (almost full-size) genomic sequences of PBVs found in faeces of humans and animals as well as environmental samples. This analysis revealed prokaryotic motifs (regions) in the PBV genome located before the open reading frames at the ribosomal binding site. Based on the data obtained, a hypothesis was put forward and experimentally substantiated that PBVs belong to prokaryotic viruses – bacteriophages (Krishnamurthy, Wang, 2018), which was later supported by a number of other studies (Adriaenssens et al., 2018; Boros et al., 2018; Kleymann et al., 2020).

On the other hand, after the discovery of new PBV-like nucleotide sequences encoding RNA-dependent RNA polymerase, but using an alternative (non-standard) mitochondrial genetic code (of molds or invertebrates) for translation, it was suggested that PBVs could be fungal viruses with a reproduction mode resembling that of mitoviruses (Yinda et al., 2019; Kleymann et al., 2020).

These contradictions, which have caused a discussion in the scientific community on the nature of the true host(s) of PBVs, can be resolved by a hypothesis put forward in 2018 by Wolf et al. (2018) and supported by other researchers (Chauhan et al., 2021). This hypothesis explains the origin of abnormal strains of PBVs by the tendency of these viruses to abrupt genetic modification following the reassortment of genome segments described earlier (Conceição-Neto et al., 2016) and the acquisition of the ability of the bacteriophage to reproduce in the cells of the organism of another taxonomic group – the lower eukaryote.

This review analyzes the available publications on modern ideas about the origin and evolution of PBVs, as well as a discussion on the prokaryotic or eukaryotic nature of their true hosts. The purpose of the review was to collect arguments to support the hypothesis about the phage nature of PBVs and to search for the most realistic explanation of the reasons for the identification of genomic sequences that are nonstandard for viruses.

Characteristics of PBVs

Picobirnaviruses (PBVs) are small, nonenveloped particles 33–37 nm in diameter with icosahedral type of symmetry ($T = 2$) belonging to the only genus *Picobirnavirus* within the family *Picobirnaviridae* of the order *Diplohnvirales*. Double-stranded (ds) RNA-genome of PBVs consists of two segments 2525 and 1745 bp in length (Fig. 1). Information about the structure of the virion and genome of PBV, the area of prevalence, connection with diarrhea, the level of excretion, the opportunistic (conditionally pathogenic) and zoonotic nature of the virus, the wide tissue tropism and genetic variability is given in previously published reviews (Ganesh et al., 2014; Kashnikov et al., 2020; Ghosh, Malik, 2021). The methods of amplification, PCR diagnostics and genome sequencing of these viruses are also described there.

Using phylogenetic analysis based on the nucleotide sequence of the RNA-dependent RNA polymerase (RdRp) gene located in the segment 2 of the genome, researchers divide PBVs into five genogroups: GI, GII (Rosen et al., 2000), GIII (Smits et al., 2014), GIV and GV (Li et al., 2015), among which there are genetically variable clusters. Genogroups GI and GII are more common in the PBV cluster detected in vertebrates and humans. Genogroup GIII has been identified in invertebrates (Shi et al., 2016) while genogroups GIV and GV have been identified in fungal and prokaryotic host cells (Knox et al., 2018). All five PBV genogroups identified in one host (marmot) are shown in the phylogram (Fig. 2) in the study (Luo et al., 2018). The main PBV genogroups are genogroups GI and GII, of which PBVs of genogroup GI are the most common (Shi et al., 2016; Kumar et al., 2020; Ghosh, Malik, 2021).

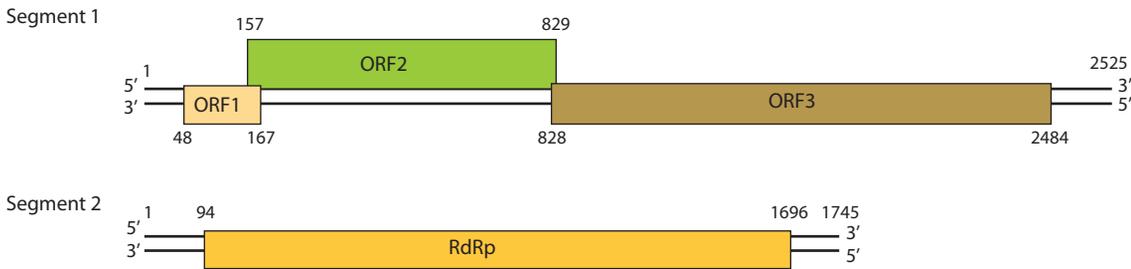


Fig. 1. Genomic organization of the human PBV (strain Hy005102) belonging to the genogroup I (Ghosh, Malik, 2021).

Segment 1 of Hy005102 genome consists of three open reading frames (ORF) ORF1, ORF2, ORF3. Reading frame ORF3 encodes the precursor of the virus capsid protein. Segment 2 has one ORF encoding viral RNA-dependent RNA polymerase (RdRp). ORF1 and ORF2 products are not identified.

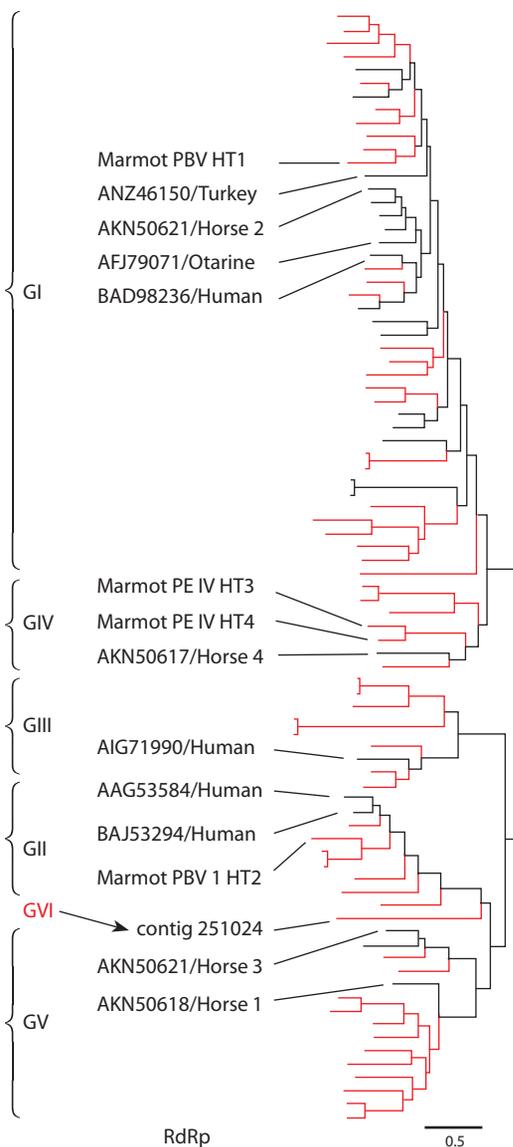


Fig. 2. Phylogenetic tree showing the presence of five proposed PBV genogroups identified in marmot is built on the basis of the nucleotide sequences of the complete RdRp gene.

Sequences of picobirnaviruses obtained from marmot are shown in red. Sequences of picobirnaviruses obtained from other hosts are shown in black (Luo et al., 2018, with modifications).

It has been established that viruses belonging to the same RdRp-genogroup can be detected in suspected hosts belonging to different species. It has also been shown that PBVs of different genogroups are detected in the host of the same species (Ganesh et al., 2014; Malik et al., 2014; Woo et al., 2014, 2019; Li et al., 2015; Gallagher et al., 2017; Navarro et al., 2017; Boros et al., 2018; Duraisamy et al., 2018; Ghosh et al., 2018; Yinda et al., 2019; Joycelyn et al., 2020; Kleymann et al., 2020; Ghosh, Malik, 2021). At the same time, the true host of viruses has not yet been identified. Among the higher eukaryotes, they did not succeed in identifying either a cell culture or gnotobiotic animals for the virus propagation (Ganesh et al., 2014; Malik et al., 2014; Delmas et al., 2019; Kleymann et al., 2020).

In the future, as the PBV studies continued, researchers began to doubt the fact that the cells of higher eukaryotes could be the hosts of these viruses (Adriaenssens et al., 2018; Boros et al., 2018; Krishnamurthy, Wang, 2018). Recently it has been discovered that PBVs differ from dsRNA viruses of higher eukaryotes (*Reoviridae*) not only in the architecture of the capsid, but also in presumably being able to infect prokaryotic cells (Knox et al., 2018; Krishnamurthy, Wang, 2018).

Hypothesis about the phage nature of PBV and associated doubts

Prior to the hypothesis of Krishnamurthy and Wang (2018), PBVs were thought to be eukaryotic viruses because they were identified in a wide variety of animal species, including invertebrates. Since PBVs are ubiquitous in the gut contents of humans and other animals with or without diarrhea, they were considered opportunistic enteric pathogens. But intestinal virom in animals contains not only eukaryotic, but also prokaryotic viruses, which usually make up the largest share of it (Yinda et al., 2018). It was logical to assume that PBVs are not present in the gut cells, but in the gut contents and can be prokaryotic viruses of the gut microbiome (Adriaenssens et al., 2018; Kunz et al., 2018; Delmas et al., 2019; Bell et al., 2020; Guajardo-Leiva et al., 2020; Ghosh, Malik, 2021).

In this case, the level of virus identification should correspond to the number of bacteria in which they multiply. In particular, the study of Kleymann et al. (2020) reported high rates of identification of GI PBVs (35.36 %, 29/82) in Indian mongoose stool samples on the island of St. Kitts (one of the

Lesser Antilles). The percentage of PBV identification on the island of St. Kitts could mean the concentration of host bacteria in this area. Moreover, there was a difference in the frequency of PBV identification in mongooses from the urban and wild habitats, 33.33 % (19/57) and 40.00 % (10/25) respectively, which could indicate a difference in bacterial load in these places.

The suggestion that the PBVs may be prokaryotic viruses appeared when researchers began to analyze the diversity of full-sized (almost full-sized) sequences of their genome. With the help of next-generation sequencing technologies (NGS) and polymerase chain reaction (PCR) using specific primers, they succeeded in identifying some features of the PBV genome that indicate that these viruses may actually be prokaryotic or fungal viruses (Shi et al., 2016; Adriaenssens et al., 2018; Boros et al., 2018; Krishnamurthy, Wang, 2018; Wolf et al., 2018; Yinda et al., 2018; Delmas et al., 2019; Kleymann et al., 2020; Ghosh, Malik, 2021).

In 2018, while analyzing different reference genomes of RNA-containing viruses individually and at the family level, Indian scientists Krishnamurthy and Wang have discovered conservative regions in the PBV genome called Shine–Dalgarno sequences or SD-sequences (Krishnamurthy, Wang, 2018). Such regions are present in the genomes of all prokaryotic and eukaryotic viruses and usually consist of six nucleotides – AGGAGG. They are located in the 5'-untranslated region before the open reading frames (ORFs) at a distance of 1 to 18 nucleotides (spacer region) to the start codon (AUG) initiating the translation of the viral genome products (Krishnamurthy, Wang, 2018; Ghosh, Malik, 2021). Functional SD-sequences are ribosome binding sites and promote the translation of viral proteins.

However, genome enrichment with SD-sequences was observed only in families of viruses that infect prokaryotes, but not in families infecting eukaryotes. This observation made it possible for Krishnamurthy and Wang to suggest that the high frequency of appearance of SD-sequences in the viral genomes may be a defining feature of the prokaryotic type of virus, and any viral families the genomes of which are enriched with such SD-regions are prokaryotic viral families. Among the viruses infecting prokaryotes, for example, some bacteriophages of the family *Cystoviridae* have a high content of SD-sequences, the genome of which consists of several fragments of dsRNA (Mindich, 1988; Boros et al., 2018).

In the PBV genome, SD-sequences were present before all ORFs in segments 1 and 2. The level of enrichment with SD-regions in PBVs is higher than in any known prokaryotic virus family and this level was constant (in 100 % of the studied genes), while not in all prokaryotic viruses it is maintained in the virus replication process. Such a high level of preservation of prokaryotic regions in the PBV genome should correlate with the level of their preservation in the genome of bacteria of a certain type from the spectrum of hosts that PBVs can infect. This level varies in different viral families (Krishnamurthy, Wang, 2018). Preservation of the level of enrichment with prokaryotic regions in the genomes of prokaryotic viruses depends on whether the host bacterium itself retains them in its genes. It is known that not all bacterial species preserve the prokaryotic Shine–Dalgarno sequence to the same extent.

For example, in the genome of bacteria belonging to the type *Firmicutes*, the prokaryotic motif is preserved in more than 80 % of genes, while in *Bacteroides*, less than 10 % (Omotajo et al., 2015). It is known that different families of bacterial RNA-viruses can consist of evolutionarily related viruses capable of infecting one type of bacteria. From this fact it follows that PBVs can infect bacteria within the type *Firmicutes* that most corresponds to the level of preservation of the prokaryotic motif in genes (more than 80 %) to PBVs and is most common in the fecal microbiota (Sekelja et al., 2011).

The hypothesis about the phage nature of PBVs put forward by Krishnamurthy and Wang (2018) is confirmed by the results obtained by other researchers (Adriaenssens et al., 2018; Boros et al., 2018; Kleymann et al., 2020). In particular, the study of Boros et al. (2018) revealed in the genome of chicken PBVs SD-regions that were located in segment 1 before the three ORFs and in segment 2 before ORFs above the initiation codons. Using 6xHis tagging and western blotting of genomic segment 1 of PBVs containing the SD-motif, these researchers have succeeded in showing *in vivo* the possibility of its expression and functionalization in *Escherichia coli* (Boros et al., 2018). The results obtained, according to the authors, serve as proof of the existence of a bacterial culture for the reproduction of PBVs.

The assumption that PBVs represent a new family of RNA bacteriophages with a high level of genomic diversity was also confirmed in the work of Adriaenssens et al. (2018). The authors of this work found a hexameric prokaryotic AGGAGG motif in 100 % of the genomic sequences of PBVs, while eukaryotic viruses from different families had SD-regions with low frequency and mainly consisted of tetramers (AGGA, GGAG, GAGG) (Adriaenssens et al., 2018). The review of Ghosh, Malik (2021) presents a number of conservative prokaryotic sequences (motives) found before all ORFs in segments 1 and 2 of PBV and PBV-like genomes, indicating the location and with access numbers in GenBank (Ghosh, Malik, 2021).

However, despite the practical results obtained, indicating the possible PBV affiliation to prokaryotic viruses, many authors believe that it is premature to talk about the final proof of the phage nature of PBVs (Ramesh et al., 2021). The host in which PBVs would successfully reproduce has not yet been identified. Given the fact that the gut microbiome consists of several hundreds of mostly uncultivable bacteria, the identification of true bacterial or archeal host(s) of PBVs (if any) will be challenging (Boros et al., 2018).

In addition, ORFs encoding the RdRp gene were found in the genome of some PBV strains, in which, during translation, instead of the expected standard genetic code, an alternative code of fungi and invertebrates was used (Shi et al., 2016; Yinda et al., 2018, 2019; Kleymann et al., 2020). So an assumption was made that the PBV hosts may be cells of lower eukaryotes.

PBV-like strains with nonstandard genetic code

As is known, the gene sequences in the viral genomes have their characteristic conserved regions – motifs by which viruses are identified. Motives characteristic of PBV and PBV-like genomes include the prokaryotic region Shine–Dalgarno

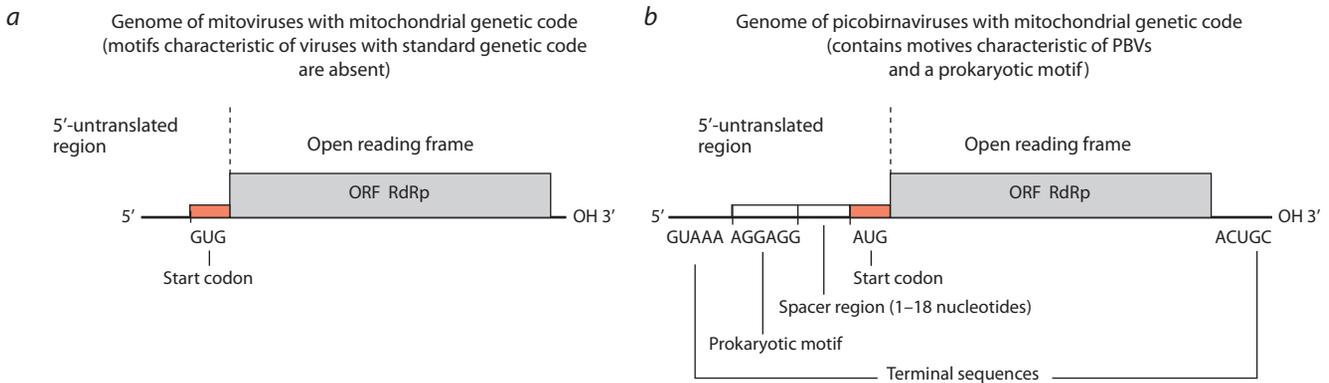


Fig. 3. Mitovirus (a) and picobirnavirus (b) genomes.

no AGGAGG (Adriaenssens et al., 2018; Boros et al., 2018; Ghosh et al., 2018; Krishnamurthy, Wang, 2018; Yinda et al., 2018, 2019; Guajardo-Leiva et al., 2020; Joycelyn et al., 2020; Kleymann et al., 2020), terminal sequences 5'-GUAAA and 3'-ACUGC (Ghosh et al., 2018; Delmas et al., 2019; Woo et al., 2019; Kleymann et al., 2020), and three regions represented in amino acid sequences DFXKFD, SGSGGT and GDD (Kleymann et al., 2020).

When translating the RdRp gene of most picobirnaviruses, the standard genetic code is used. However, recently, new PBV-like RdRp gene sequences with an alternative (non-standard) mitochondrial genetic code have been identified in human (Woo et al., 2019), mongoose (Kleymann et al., 2020), bat (Yinda et al., 2018) and invertebrate (Shi et al., 2016) fecal samples. The mitochondrial code is characteristic of viruses of mold fungi and invertebrates. In particular, five PBV-like genomic sequences consisting of dsRNA with mitochondrial code were isolated from the gut contents of bats (Yinda et al., 2018). In four of them (P11-300, P11-378, P14-90 and P15-218) they failed to identify the presence of the capsid gene. These PBV-like genomes contained only RdRp gene sequence with the mitochondrial genetic code of the mold. The absence of the capsid protein gene in them resembled the genome of mitoviruses from the family *Mitoviridae*, which are known to infect the mitochondria of unicellular mold fungi (Fig. 3) (Hillman, Cai, 2013; Shi et al., 2016).

The mitovirus genome as well as four abnormal PBV-like genomes consists of dsRNA and encodes only RdRp. Mitoviruses lack a capsid, replicate in mitochondria and use the genetic mitochondrial code of mold fungi for the RdRp translation. So an assumption was made that mold fungi can also be PBV hosts (Yinda et al., 2018; Kleymann et al., 2020).

However, unlike mitoviruses, the noncapsid PBV-like strains identified by Yinda et al. (2018) in a single segment of the RdRp genome contained conserved regions characteristic of PBVs. In the genome of the fifth PBV-like strain (P16-366), in addition to the conservative features of the RdRp gene characteristic of PBVs, a sequence encoding the capsid protein was found. This strain was clustered on the phylogenetic tree along with GII PBVs. However, it used an alternative genetic code and was very similar in terms of genome organization to fungal viruses of the family *Partitiviridae* (Duquerroy

et al., 2009). These families have a minimalistic genome consisting of two dsRNA segments encoding RdRp and the capsid protein, respectively, which in these families is clearly homologous (Wolf et al., 2018).

Similarly, researchers Kleymann et al. (2020) isolated the M17A strain among typical PBV-like strains from mongoose, the RdRp gene of which retained all conservative for PBVs motives, but had an alternative mitochondrial mold code, and the capsid sequence in the genome of this strain was absent. Assumption that PBVs can be fungal or invertebrate viruses, as Ghosh, Malik (2021) rightly noted, has further complicated the discussion about the true PBV hosts.

Discussion on the origin of abnormal PBV-like strains

During the discussion about the true PBV hosts, researchers attempted to interpret the causes of the appearance of PBV-like strains (Shackelton et al., 2008; Wolf et al., 2018; Yinda et al., 2018; Shahi et al., 2019; Ghosh, Malik, 2021). In particular, Yinda et al. (2018) suggested that PBV-like strains found in the gut contents of different eukaryotic hosts, with a genome resembling the mitovirus genome, may have a reproduction mode similar to mitoviruses. Following the assumption of Yinda et al. (2018), the capsid protein gene is not needed with this reproduction mode, since mitoviruses do not use the pathway of the independent horizontal transmission from cell to cell, but are transmitted vertically from mother to daughter cells (during division) or horizontally (by merging hyphae). Similarly, when assembling new PBV-like particles, simplified structures can be formed with the adaptation to the existence in a fungal cell characteristic of mitoviruses and lost capsid protein gene. Such an interpretation of the appearance of noncapsid RNA viruses is consistent with one of the trends in the evolution of RNA viruses associated with the loss of their structural module by their genome, noticed by Wolf et al. (2018).

The trend in the evolution of RNA viruses associated with the loss of the genome segment with the capsid gene is demonstrated by the pedigree diagram of eukaryotic viruses with the RNA genome, shown in Figure 4. This scheme also explains the origin of capsidated RNA viruses of unicellular eukaryotes.

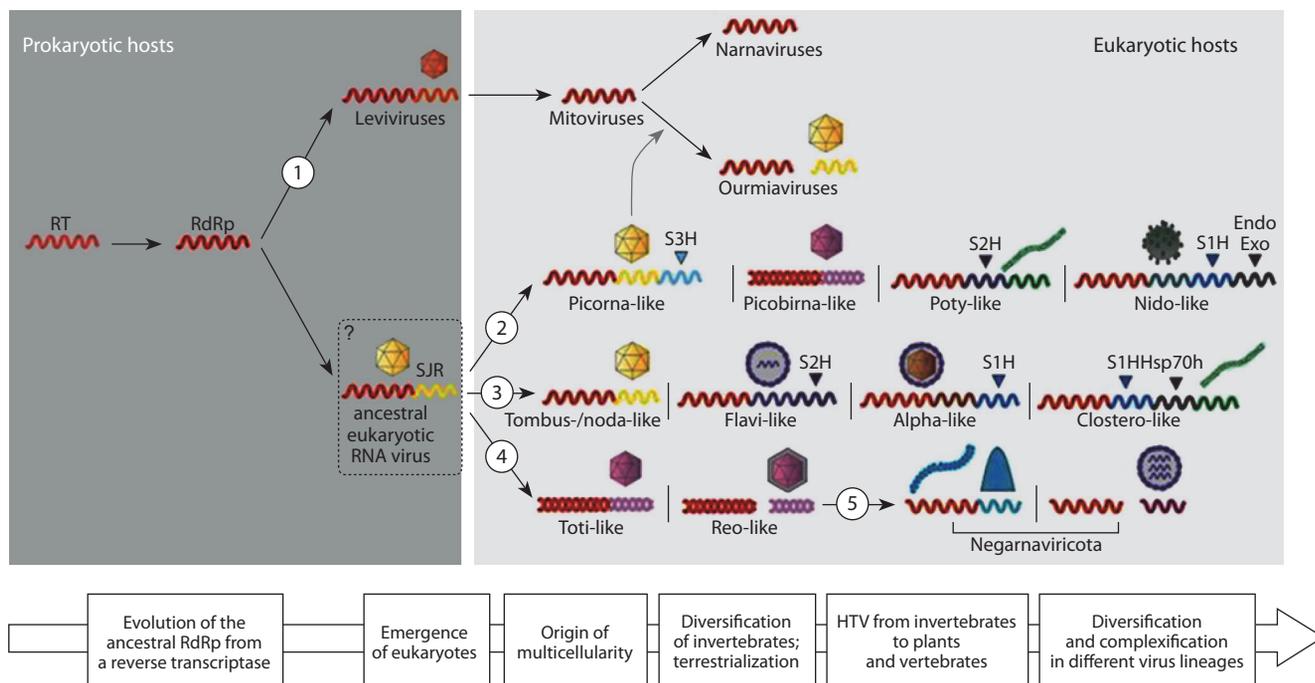


Fig. 4. A rough scheme of the key steps of RNA virus evolution (Wolf et al., 2018).

The evolution of capsid-free viruses of lower eukaryotes (fungi and invertebrates) follows the path of genome reduction due to the loss of the capsid protein gene of the probable ancestor – levivirus infecting prokaryotes (the vector of evolution under the number 1 in Figure 4, leading to the appearance at the end of the family *Narnaviridae*). In accordance with the scheme, the ancestor of mitoviruses from the family *Mitoviridae* descended from the family *Leviviridae*, previously containing the capsid protein gene. An illustrative example of genome reduction is hypoviruses – capsid-free viruses with dsRNA, the structure and phylogenetic analysis of the genome of which showed that they originated from potyviruses that lost the capsid protein (Dawe, Nuss, 2013; Chauhan et al., 2021) and the capsid protein gene, entering the cells of unicellular eukaryotes. Replication of mitoviruses occurs only in mitochondria, where their capsid-free (‘naked’) RNA genomes (replicons) are replicated. With the loss of the structural gene, mitoviruses have lost the ability for independent horizontal transmission (Shackelton, Holmes, 2008).

Similarly, the origin of the group of capsid-free PBV-like strains isolated by Yinda et al. (2018) and Kleymann et al. (2020) from Cameroonian bats (P11-300, P11-378, P14-90 and P15-218) and from a mongoose (strain M17A) can be explained by the loss of a genome segment encoding the capsid protein by the parent strain.

On the other hand, a capsid-free RNA virus (mitovirus) in the process of evolution could gain a capsid (the final direction of vector 1, leading to the occurrence of the family *Ourmiaviridae* due to the fusion of its ‘naked’ RNA replicon with the replicon of the capsid protein of +eukaryotic RNA virus, possibly having the eukaryotic virus as an ancestor (evolution vector 2 in Figure 4). Encapsulated strains of RNA viruses,

and probably PBV-like analogues, could have occurred in this evolutionary way.

Researchers explain shuffling or loss of genome segments in RNA viruses by another trend in their evolution – the ability to reassort – to redistribute gene modules (RdRp and capsid protein) between closely related virus families with similar genes. Reassortment modification of the genome has also been observed in PBVs (Conceição-Neto et al., 2016). The creative role of reassortment between families of RNA viruses with similar genes explains the origin of the PBV-like strain P16-366 found by Yinda et al. (2018) in bats. This strain contained, together with the gene sequence RdRp (with a non-standard mitochondrial code), the sequence encoding the capsid protein. The proposed genetic similarity between families of viruses with a bisegmented dsRNA genome is based on the hypothesis of a reassortment mechanism for the evolution of these viruses.

On the relationship of PBVs with families of viruses with the dsRNA genome

The phylogenetic relationship between the families of RNA viruses with a segmented genome is based on the information about the primary structure of the RdRp gene. This gene is universal and, despite a much greater distribution among eukaryotic viruses, is present in almost all RNA viruses (including capsid-less RNA replicons) with the exception of some satellite viruses (Dolja, Koonin, 2018).

The universality of the RdRp gene indicates the possibility of its common origin in RNA viruses. In 2018, Wolf et al. put forward a hypothesis suggesting the presence of phylogenetic relationships between families of RNA-containing viruses, and a phylogenetic tree was constructed (Fig. 5), the topo-

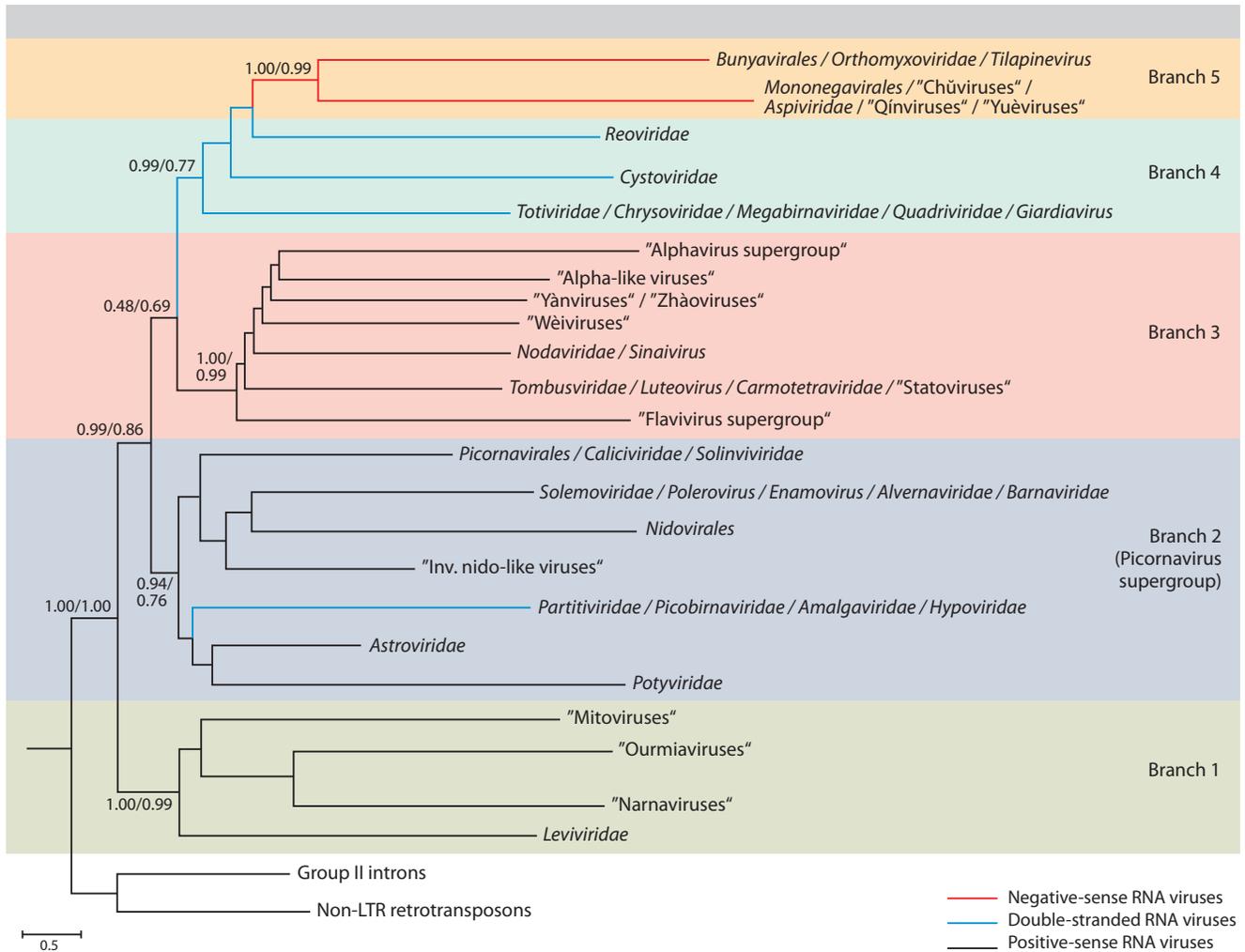


Fig. 5. Phylogenetic tree of viruses with the RNA-genome built on the basis of gene sequences of RNA-dependent RNA polymerase and reverse transcriptase (Wolf et al., 2018).

Five main branches are shown, among which three branches 1, 2 and 4 directly relate to the origin of PBV-like strains.

logy of which demonstrates a possible relationship between the families.

The phylogenetic tree constructed on the basis of a set of phylogenetic data gives an idea of the possible origin of PBVs. This tree confirms the evolution scenario in which the last common ancestors of virus lineages with a dsRNA genome were simple viruses the segmented genome of which contained two genes (RdRp and capsid protein). According to the authors of the hypothesis, all viruses with a dsRNA genome in branches 2 and 4 of the phylogenetic tree have similar capsid proteins that can be combined with the genome RdRp from different viruses with +RNA genome.

Viruses with the dsRNA genome – *Partitiviridae*, *Picobirnaviridae*, *Cystoviridae*, *Reoviridae* and *Totiviridae* form separate lineages in two branches of the phylogenetic tree. Viruses of the family *Picobirnaviridae* are located on the same line with the families *Hypoviridae*, *Amalgaviridae* and *Partitiviridae*. The most phylogenetically close families forming one cluster are *Picobirnaviridae* and *Partitiviridae*. The close location of these families demonstrates a high de-

gree of relationship between them. These families are united by a similar organization of virions and homologous capsid proteins (Duquerroy et al., 2009; Wolf et al., 2018). The difference between these viruses lies in the fact that the surface PBV capsid proteins, unlike those of partitiviruses, have perforation activity, which determines the ability of the virus to penetrate into the cell. In addition, due to the ability of PBVs for horizontal transmission, two segments of its genome are combined into one capsid during assembly, and in partitiviruses, the RdRp gene and the capsid protein gene use separate capsids (Vainio et al., 2018).

The PBV capsid protein gene is distantly related to the capsid protein genes of other viruses with the dsRNA genome (*Totiviridae*, *Reoviridae* and *Cystoviridae*), which make up three parallel evolutionary lineages in branch 4 of the phylogenetic tree (Wolf et al., 2018). Significant homology of the genes encoding the capsid protein of viruses of the families *Reoviridae*, *Totiviridae*, *Cystoviridae*, *Picobirnaviridae* and *Partitiviridae* was noted earlier (El Omari et al., 2013; Lvov et al., 2013; Luque et al., 2014).

Three evolutionary lineages of branch 4 of the phylogenetic tree are formed by families of viruses infecting both prokaryotes (*Cystoviridae*) and eukaryotes (*Reoviridae*, *Totiviridae*). The location of these viral families on one branch, according to the authors of the phylogenetic tree, does not exclude the possibility of the origin of eukaryotic +RNA viruses from their prokaryotic analogues. They admit that mitoviruses (replicating in the mitochondria of mold cells) originated from an ancestor common with leviviruses – a prokaryotic RNA virus parasitizing in enterobacteria. The proof of this is the evolutionary relationship between cystoviruses (bacterial viruses) and reoviruses (Poranen, Bamford, 2012; El Omari et al., 2013).

According to supporters of the creative role of reassortment, the origin of the PBV-like strain P16-366 could result from a reassortment between an as yet undiscovered PBV relative from branch 2, which has passed to reproduction in the mitochondria of fungal cells, and one of the viruses with the dsRNA genome of branch 4 (see Fig. 5). Moreover, it was noted that genetic restrictions on the ability to create reassortants during coinfection with viruses of families forming branch 4 of the phylogenetic tree may be less strict for the prokaryotic virus family *Cystoviridae*. The appearance of the encapsulated PBV-like strain P16-366 occurred due to the unification of the segment with the gene RdRp of the +RNA virus of branch 2 (possibly a naked RNA replicon) using the mitochondrial code, with a fragment of the capsid gene of the dsRNA virus of branch 4.

Another direction in the evolution of viruses with a bisegmented dsRNA genome is the acquisition of partition – the packaging of genome segments into one (monopartite) or separate (bi-multipartite) particles. This trend is interesting because it gives us some insight into the nature of the hosts of PBV-like strains (prokaryotic or eukaryotic). For example, bi-partition is observed only in fungal viruses with a dsRNA genome and in plant viruses with a ssDNA genome (*Begomoviruses*) (Nibert et al., 2013). In bacterial viruses in general and with a segmented dsRNA genome, in particular (*Cystoviridae*), the packaging of genome segments into individual particles is not observed. Partition is related to the virus transmission mode (independent or non-independent). For example, partitiviruses parasitizing in fungal cells do not have the ability for independent (horizontal) transmission and they are bipartite, while PBVs transmitting independently are monopartite. The horizontal (independent) transmission pathway allows them to transfer both segments of the genome into a new cell with a high probability. The ability for independent (horizontal) transmission into new cells can be considered as one of the main criteria for determining the true host of the detected virus.

Presence of a molecular apparatus for penetration into the cell is an important criterion determining the nature of the PBV host

During familiarization with the studies solving the question of the nature of PBV and PBV-like strains, we came to understand that the belonging of these viruses to the bacterial viruses, higher or lower eukaryotes can be determined not only by the characteristic saturation level of their genome with

a prokaryotic motif, standard or mitochondrial genetic code. This affiliation should no less be determined by the presence or absence of specific (in relation to the host cell) proteolytic activity of the capsid protein, which determines the possibility of independent penetration of the virus into the cell. The ability for independent horizontal transmission is characteristic of animal viruses and phages, while it is absent in viruses of lower eukaryotes (partitiviruses, mitoviruses). PBV capsid proteins have perforating activity (Duquerroy et al., 2009). This makes PBVs capable of independent penetration into cells and, therefore, can characterize them as bacterial viruses.

Thus, we can conclude that in solving the question of the true PBV host in which they can reproduce, in addition to conservative motifs characteristic of the PBV genome, the determining factor is the presence of a mechanism of specific horizontal penetration into the cell – capsid proteins with specific perforating activity. If these proteins are specific to the receptors of a prokaryotic cell, for example, *Firmicutes* cells (Adriaenssens et al., 2018), then we are dealing with a phage, if they are specific to the receptors of a eukaryotic cell – with a virus of higher eukaryotes, and if there are no capsid proteins at all, then such a virus can be attributed to viruses of lower eukaryotes (not capable of independent transmission).

Assumption about the origin of atypical PBV-like strains does not contradict the hypothesis of the phage nature of PBVs

The existence of atypical PBV-like strains cannot constitute a refutation of supposed phage nature of PBVs for a number of reasons. According to the hypothesis of Wolf et al. (2018), there is a related relationship between the families of viruses with the dsRNA genome in the structure of the phylogenetic tree, as evidenced by the presence of homologous genes and a common ancestor related to prokaryotic RNA viruses. Moreover, the families of eukaryotic and prokaryotic RNA viruses may be related, as evidenced by the evolutionary relationship between cystoviruses and reoviruses (Poranen, Bamford, 2012; El Omari et al., 2013). This supports the assumption of the possibility of the exchange of homologous segments of the genome between prokaryotic viruses and related viruses of lower eukaryotes, when both are in the mold fungus cell.

It has recently become known that bacterial viruses can not only infect bacterial cells, but also pass through the epithelial cells of eukaryotes using the mechanism of phage transcytosis. From epithelial cells through the blood or lymph, phages can enter various organs and tissues of animals. However, phages can penetrate inside eukaryotic cell only with a bacterial cell infected by them (Nguyen et al., 2017). The ubiquitous isolation of PBVs from environments where bacteria occur means that these viruses may not be intracellular eukaryotic viruses, but prokaryotic viruses of the gut microbiome (Kashnikov et al., 2020; Ghosh, Malik, 2021). The assumption that PBVs can infect prokaryotic cells is confirmed by the presence of a conservative prokaryotic Shine–Dalgarno sequence in their genome (Adriaenssens et al., 2018; Boros et al., 2018; Krishnamurthy, Wang, 2018).

The identification of atypical PBV-like sequences with a non-standard mitochondrial genetic code of fungi and invertebrates can serve as evidence that PBVs are capable of

reproducing their own kind in the cells of lower eukaryotes, undergoing genetic changes when changing hosts (Yinda et al., 2018; Kleymann et al., 2020; Ghosh, Malik, 2021).

Therefore, it can be assumed that PBVs can get into eukaryotic cell (whether it is a cell of a fungus, an invertebrate or another host), where they will meet eukaryotic or prokaryotic virus with similar genome. Viruses like mitoviruses, partiti-viruses, or cystoviruses may be among the PBV reassortment partners. After rearranging the genome segments of the PBVs with partner viruses, PBV-like reassortants described by Yinda et al. (2018) and Kleymann et al. (2020) can appear. Then, according to the assumptions of the authors of the reassortation interaction hypothesis, depending on the presence or absence of prokaryotic motives and motives characteristic of PBVs, as well as depending on the genetic code used by the genes of a new virus (standard or mitochondrial) will determine not only the degree of its relationship with PBV, but also its prokaryotic or eukaryotic nature.

The hypothesis explaining the formation of atypical PBV strains through the exchange of homologous genome segments between related viral families does not exclude the possibility of their appearance using satellite relationships between PBVs and helper viruses from 2nd or 4th branches of the phylogenetic tree. Viruses from the families *Partitiviridae*, *Totiviridae*, *Reoviridae* are known as helper viruses, which require some satellite dsRNAs for their reproduction (Lvov et al., 2013). Assuming that PBV is a satellite that uses an RdRp enzyme for reproduction in the fungal cells similar to mitovirus (from branch 2), it is possible to explain the appearance of non-capsid PBV-like strains with mitochondrial genetic code P11-300, P11-378, P14-90 and P15-218, isolated by Yinda et al. (2018) or M17A isolated by Kleymann et al. (2020).

By using as a helper a virus from branches 2 or 4 (similar to *Partitiviridae*, *Reoviridae*, *Totiviridae* or *Cystoviridae*) the appearance of the encapsulated PBV-like strain P16-366 can be explained. In other words, in this case, the ability of the progeny resulting from the interaction of the satellite virus with the helper virus to reproduce in a bacterial cell or in the mitochondria of mold will depend on the nature of the helper virus. And this does not contradict the supposed prokaryotic nature of PBVs, but only means the possibility of its change in the reassortment process.

Arguments in support of the hypothesis that PBV belongs to prokaryotic viruses

PBVs are found wherever bacteria are found – in environmental samples, in lower eukaryotes (fungi and invertebrates), in the gut contents of higher eukaryotes (including reptiles). This means that PBVs may not be intracellular eukaryotic viruses but prokaryotic viruses of the gut microbiome (Ghosh, Malik, 2021).

Like prokaryotic cystoviruses, PBV genomes contain and preserve in a saturated state (in both genome segments in each reading frame) functional binding sites for bacterial-type ribosomes (Shine–Dalgarno sequences), while in eukaryotic viruses the genome is not saturated with them (Boros et al., 2018; Krishnamurthy, Wang, 2018).

The identification of related PBV strains in different animal species may mean that the PBV hosts are a specific type of bacteria, possibly *Firmicutes*, found in the gut of different

vertebrates and invertebrates. The level of preservation of prokaryotic sites in the genome of these bacteria (more than 80 % of the genes) corresponds to that of PBVs (Krishnamurthy, Wang, 2018).

PBV identification in the gut, respiratory organs of animals (cattle, humans, monkeys, pigs), in blood and respiratory samples (Lee, Bent, 2014; Blanco-Picazo et al., 2020) does not refute the assumption that bacterial cells can be their hosts. Phages cannot directly infect cells of different organs of higher eukaryotes, but they can get into these organs by non-specific translocation through gut epithelium with blood flow (Nguyen et al., 2017) or with the help of bacteria in which they multiply (Dabrowska et al., 2005). This is how phages penetrate into the blood, lymph, organs and even the brain.

The presence of a capsid protein with perforating activity (the ability to penetrate through the cell membrane), as in viruses capable of infecting animal cells (Duquerroy et al., 2009), does not contradict the fact that PBVs can be prokaryotic RNA viruses. It is known that representatives of the family of bacterial RNA viruses also have a molecular apparatus for penetrating into bacterial cells using transcytosis (Reed et al., 2013; Nguyen et al., 2017).

The acquisition of immunity to PBVs by infected animals also does not mean that PBVs can be considered eukaryotic viruses, since it has been established that host immune responses can also occur against bacterial viruses (Dabrowska et al., 2005; Górski et al., 2006). Possibly, PBVs cause an immune response to infection not of the host cells, but of the bacterial cells that make up its microbiome, which does not exclude the possibility that PBVs are prokaryotic viruses.

Like the phages, which are virulent and moderate, in an infected organism, PBVs can be active (excreted) and inactive (temporarily not excreted), while infected animals will be asymptomatic carriers (Ganesh et al., 2014; Malik et al., 2014; Kumar et al., 2020; Ghosh, Malik, 2021).

The identification of PBV-like strains with the RdRp gene using an alternative mitochondrial genetic code of eukaryotes (mold, invertebrates) for translation is also not a refutation of the assumption that PBVs belong to prokaryotic viruses. According to the hypothesis of Wolf et al. (2018), they are the result of reassortment between the PBVs and the families of RNA viruses with a similar genome. The appearance of atypical PBV-like strains only indicates the possibility of an evolutionary transition from the prokaryotic nature of the virus to the eukaryotic nature and back as a result of rearrangement of genomic segments, and is explained by a change in the degree of saturation of the genome of the new virus with prokaryotic regions, which may change its nature. The possibility of satellite relationships between PBVs and RNA viruses of lower eukaryotes such as *Partitiviridae*, *Totiviridae* and *Reoviridae*, which are known as helper viruses, is not excluded. There are known cases of satellite relationships between prokaryotic and lower eukaryotic viruses infecting one single-celled host, allowing phages to reproduce new progeny in a eukaryotic cell (Gogarten, Townsend, 2005; Claverie, Abergel, 2009; Thannesberger et al., 2017).

The expression and functioning of PBV segment 1 *in vivo* in *Escherichia coli* (Boros et al., 2018) was carried out, which confirms the existence of a bacterial culture for the propagation of PBVs.

Conclusion

The arguments given convincingly show that PBVs can indeed be prokaryotic viruses, since they are found wherever bacteria are found – in environmental samples, in the gut contents of vertebrates, in the cells of fungi and invertebrates.

Similar to prokaryotic cystoviruses, PBV genomes contain and retain functional binding sites for bacterial-type ribosomes in a saturated state, while in eukaryotic viruses the genome is not saturated with these motifs. The high level of preservation of prokaryotic sites in the PBV genome suggests that they belong to a new family of RNA viruses that infect a certain type of bacteria with a high content of SD-sequences in the genome. Such host bacteria of PBVs can be *Firmicutes* found in the gut of vertebrates and invertebrates, which have a similar level of prokaryotic motifs characteristic of bacteria. The high frequency of the presence of SD-sequences in the PBV genome can be considered one of the main criteria for identifying new virus families for affiliation to a prokaryotic or eukaryotic host.

But in order for a virus to be finally classified as a phage, its genome must contain a structural gene encoding a protein with specific proteolytic properties that determine the ability of the virus to independently penetrate into a bacterial cell. A protein with proteolytic properties is present in the PBV capsid, which indicates the ability of this virus for horizontal transmission (independently) from cell to cell and from one host to another. At the same time, the identification of PBVs in the gut of various hosts may suggest that their horizontal transmission can be carried out by bacteria and, therefore, these viruses themselves are phages. The identification of PBVs in various tissues of the body can also be explained by the fact that, being phages, they are capable of nonspecific translocation through the walls of gut epithelial cells.

The selection of a culture for the virus propagation is necessary for the final determination of its nature. To date, the belonging of PBVs to viruses of higher eukaryotes has not been proven, since it was not possible to isolate them from eukaryotic cell cultures. To establish definitively whether PBVs are prokaryotic viruses, it is necessary to direct efforts to select a host for their reproduction among prokaryotic cells obtained from the gut microbiome of mutants and conditions for the cultivation of these cells.

And, finally, the presence of probable relatives among the families of RNA viruses with similar genes with which PBVs can exchange genome segments, replicating atypical genetic variants, does not contradict the correctness of the hypothesis about the phage nature of PBVs, but indicates their potential ability to adapt to new conditions of existence, allowing them to infect eukaryotic or prokaryotic host cells.

References

- Adriaenssens E.M., Farkas K., Harrison C., Jones D.L., Allison H.E., McCarthy A.J. Viromic analysis of wastewater input to a river catchment reveals a diverse assemblage of RNA viruses. *mSystems*. 2018; 3(3):1-18. DOI 10.1128/mSystems.00025-18.
- Bell N., Khamrin P., Kumthip K., Rojjanadumrongkul K., Nantachit N., Maneekarn N. Molecular detection and characterization of picobirnavirus in environmental water in Thailand. *Clin. Lab.* 2020;66: 85-88. DOI 10.7754/CLIN.LAB.2019.191013.
- Blanco-Picazo P., Fernández-Orth D., Brown-Jaque M., Miró E., Espinal P., Rodríguez-Rubio L., Muniesa M., Navarro F. Unravelling the consequences of the bacteriophages in human samples. *Sci. Rep.* 2020;10:6737. DOI 10.1038/s41598-020-63432-7.
- Boros Á., Polgár B., Pankovics P., Fenyvesi H., Engelmann P., Phan T.G., Delwart E., Reuter G. Multiple divergent picobirnaviruses with functional prokaryotic Shine–Dalgarno ribosome binding sites present in cloacal sample of a diarrheic chicken. *Virology*. 2018;525:62-72. DOI 10.1016/j.virol.2018.09.008.
- Chauhan R., Awasthi S., Narayan R.P. Chapter 11 – Evolution and diversity of plant RNA viruses. In: Rajarshi Kumar Gaur, S.M. Paul Khurana, Pradeep Sharma, Thomas Hohn (Eds.) *Plant Virus-Host Interaction: Molecular Approaches and Viral Evolution*. Second Edn. Acad. Press, 2021;303-318. DOI 10.1016/B978-0-12-821629-3.00020-8.
- Claverie J.M., Abergel C. Mimivirus and its virophage. *Ann. Rev. Genet.* 2009;43:49-66. DOI 10.1146/annurev-genet-102108-134255.
- Conceição-Neto N., Mesquita J.R., Zeller M., Yinda C.K., Álvares F., Roque S., Petrucci-Fonseca F., Godinho R., Heylen E., Van Ranst M., Matthijnsens J. Reassortment among picobirnaviruses found in wolves. *Arch. Virol.* 2016;161:2859-2862. DOI 10.1007/s00705-016-2987-4.
- Dabrowska K., Swiata-Jelen K., Opolski A., Weber-Dabrowska B., Gorski A. Bacteriophage penetration in vertebrates. *J. Appl. Microbiol.* 2005;98:7-13. DOI 10.1111/j.1365-2672.2004.02422.x.
- Dawe A.L., Nuss D.L. Hypovirus molecular biology: from Koch's postulates to host self-recognition genes that restrict virus transmission. *Adv. Virus Res.* 2013;86:109-147.
- Delmas B., Attoui H., Ghosh S., Malik Y.S., Mundt E., Vakharia V.N. ICTV virus taxonomy profile: *Picobirnaviridae*. *J. Gen. Virol.* 2019; 100:133-134. DOI 10.1099/jgv.0.001186.
- Dolja V.V., Koonin E.V. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res.* 2018;244:36-52. DOI 10.1016/j.virusres.2017.10.020.
- Duquerry S., Da Costa B., Henry C., Vigouroux A., Libersou S., Lepault J., Navaza J., Delmas B., Rey F.A. The picobirnavirus crystal structure provides functional insights into virion assembly and cell entry. *EMBO J.* 2009;28:1655-1665. DOI 10.1038/emboj.2009.109.
- Duraisamy R., Akiana J., Davoust B., Mediannikov O., Michelle C., Robert C., Parra H.-J., Raoult D., Biagini P., Desnues C. Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses. *Virus Genes*. 2018;54: 256-271. DOI 10.1007/s11262-018-1543-6.
- El Omari K., Sutton G., Ravantti J.J., Zhang H., Walter T.S., Grimes J.M., Bamford D.H., Stuart D.I., Mancini E.J. Plate tectonics of virus shell assembly and reorganization in phage $\phi 8$, a distant relative of mammalian reoviruses. *Structure*. 2013;21:1384-1395. DOI 10.1016/j.str.2013.06.017.
- Gallagher C.A., Navarro R., Cruz K., Aung M.S., Ng A., Bajak E., Beierschmitt A., Lawrence M., Dore K.M., Ketzis J., Malik Y.S., Kobayashi N., Ghosh S. Detection of picobirnaviruses in vervet monkeys (*Chlorocebus sabaeus*): molecular characterization of complete genomic segment-2. *Virus Res.* 2017;230:13-18. DOI 10.1016/j.virusres.2016.12.021.
- Ganesh B., Masachessi G., Mladenova Z. Animal picobirnavirus. *Virus Disease*. 2014;25:223-238. DOI 10.1007/s13337-014-0207-y.
- Ghosh S., Malik Y.S. The true host/s of picobirnaviruses. *Front. Vet. Sci.* 2021;7:1-9. DOI 10.3389/fvets.2020.615293.
- Ghosh S., Shiokawa K., Aung M.S., Malik Y.S., Kobayashi N. High detection rates of picobirnaviruses in free roaming rats (*Rattus* spp.): molecular characterization of complete gene segment-2. *Infect. Genet. Evol.* 2018;65:131-135. DOI 10.1016/j.meegid.2018.07.024.
- Gogarten J.P., Townsend J.P. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 2005;3(9):679-687. DOI 10.1038/nrmicro1204.

- Górski A., Kniotek M., Perkowska-Ptasńska A., Mróz A., Przerwa A., Gorczyca W., Dabrowska K., Weber-Dabrowska B., Nowaczyk M. Bacteriophages and transplantation tolerance. *Transplant. Proc.* 2006;38:331-333. DOI 10.1016/j.transproceed.2005.12.073.
- Guajardo-Leiva S., Chnaiderman J., Gaggero A., Díez B. Metagenomic insights into the sewage RNA virosphere of a large city. *Viruses.* 2020;12:1050. DOI 10.3390/v12091050.
- Hillman B.I., Cai G. Chapter Six – The family *Narnaviridae*: simplest of RNA viruses. *Adv. Virus Res.* 2013;86:149-176. DOI 10.1016/B978-0-12-394315-6.00006-4.
- Joycelyn S.J., Ng A., Kleymann A., Malik Y.S., Kobayashi N., Ghosh S. High detection rates and genetic diversity of picobirnaviruses (PBVs) in pigs on St. Kitts Island: identification of a porcine PBV strain closely related to simian and human PBVs. *Infect. Genet. Evol.* 2020;84:104383. DOI 10.1016/j.meegid.2020.104383.
- Kashnikov A.Yu., Epifanova N.V., Novikova N.A. Picobirnaviruses: prevalence, genetic diversity, detection methods. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2020;24(6):661-672. DOI 10.18699/VJ20.660.
- Kleymann A., Becker A.A.M.J., Malik Y.S., Kobayashi N., Ghosh S. Detection and molecular characterization of picobirnaviruses (PBVs) in the mongoose: identification of a novel PBV using an alternative genetic code. *Viruses.* 2020;12(1):99. DOI 10.3390/v12010099.
- Knox M.A., Gedye K.R., Hayman D.T.S. The challenges of analysing highly diverse picobirnavirus sequence data. *Viruses.* 2018;10:685. DOI 10.3390/v10120685.
- Krishnamurthy S.R., Wang D. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. *Virology.* 2018;516:108-114. DOI 10.1016/j.virol.2018.01.006.
- Kumar N., Mascarenhas J.D.A.P., Ghosh S., Maschessi G., da Silva Bandeira R., Nates S.V., Dhama K., Singh R.K., Malik Y.S. Picobirnavirus. In: Malik Y.S., Singh R.K., Dhama K. (Eds.) *Animal-Origin Viral Zoonoses*. Singapore: Springer, 2020;291-312.
- Kunz A.F., Possatti F., de Freitas J.A., Alfieri A.A., Takiuchi E. High detection rate and genetic diversity of picobirnavirus in a sheep flock in Brazil. *Virus Res.* 2018;255:10-13. DOI 10.1016/j.virusres.2018.06.016.
- Lee C.K., Bent S.J. Uncovering the hidden villain within the human respiratory microbiome. *Diagnosis.* 2014;1:203-212. DOI 10.1515/dx-2014-0039.
- Li L., Giannitti F., Low J., Keyes C., Ullmann L.S., Deng X., Aleman M., Pesavento P.A., Pusterla N., Delwart E. Exploring the virome of diseased horses. *J. Gen. Virol.* 2015;96:2721-2733. DOI 10.1099/vir.0.000199.
- Luo X.L., Lu S., Jin D., Yang J., Wu S.S., Xu J. *Marmota himalayana* in the Qinghai-Tibetan plateau as a special host for bi-segmented and unsegmented picobirnaviruses. *Emerg. Microbes Infect.* 2018;7(1):20. DOI 10.1038/s41426-018-0020-6.
- Luque D., Gómez-Blanco J., Garriga D., Brilot A.F., González J.M., Havens W.M., Carrascosa J.L., Trus B.L., Verdager N., Ghabrial S.A., Castón J.R. Cryo-EM near-atomic structure of a dsRNA fungal virus shows ancient structural motifs preserved in the dsRNA viral lineage. *Proc. Natl. Acad. Sci. USA.* 2014;111:7641-7646. DOI 10.1073/pnas.1404330111.
- Lvov D.K., Alkhovskiy S.V., Shchelkanov M.Yu. Satellites. In: *Guide to Virology. Viruses and viral infections of humans and animals*. Moscow, 2013;350-352. (in Russian)
- Malik Y.S., Kumar N., Sharma K., Dhama K., Shabbir M.Z., Ganesh B., Kobayashi N., Banyai K. Epidemiology, phylogeny, and evolution of emerging enteric picobirnaviruses of animal origin and their relationship to human strains. *Biomed. Res. Int.* 2014;2014:780752. DOI 10.1155/2014/780752.
- Mindich L. Bacteriophage Ø6: a unique virus having a lipid-containing membrane and a genome composed of three dsRNA segments. *Adv. Virus Res.* 1988;35:137-173. DOI 10.1016/S0065-3527(08)60710-1.
- Navarro R., Yibin C., Nair R., Peda A., Aung M.S., Ketzis J., Malik Y.S., Kobayashi N., Ghosh S. Molecular characterization of complete genomic segment-2 of picobirnavirus strains detected in a cat and a dog. *Infect. Genet. Evol.* 2017;54:200-204. DOI 10.1016/j.meegid.2017.07.006.
- Nguyen S., Baker K., Padman B.S., Patwa R., Dunstan R.A., Weston T.A., Schlosser K., Bailey B., Lithgow T., Lazarou M., Luque A., Rohwer F., Blumberg R.S., Barr J.J. Bacteriophage transcytosis provides a mechanism to cross epithelial cell layers. *mBio.* 2017;8:e01874-17. DOI 10.1128/mBio.01874-17.
- Nibert M.L., Tang J., Xie J., Collier A.M., Ghabrial S.A., Baker T.S., Tao Y.J. Chapter Three – 3D structures of fungal partitiviruses. *Adv. Virus Res.* 2013;86:59-85. DOI 10.1016/B978-0-12-394315-6.00003-9.
- Omotajo D., Tate T., Cho H., Choudhary M. Distribution and diversity of ribosome binding sites in prokaryotic genomes. *BMC Genomics.* 2015;16(1):604. DOI 10.1186/s12864-015-1808-6.
- Poranen M.M., Bamford D.H. Assembly of large icosahedral double-stranded RNA viruses. *Adv. Exp. Med. Biol.* 2012;726:379-402. DOI 10.1007/978-1-4614-0980-9_17.
- Ramesh A., Bailey E.S., Ahyon V., Langelier C., Phelps M., Neff N., Sit R., Tato C., DeRisi J.L., Greer A.G., Gray G.C. Metagenomic characterization of swine slurry in a North American swine farm operation. *Sci. Rep.* 2021;11:16994. DOI 10.1038/s41598-021-95804-y.
- Reed C.A., Langlais C., Wang I.-N., Young R. A₂ expression and assembly regulates lysis in Qβ infections. *Microbiology.* 2013;159:507-514. DOI 10.1099/mic.0.064790-0.
- Rosen B.I., Fang Z.-Y., Glass R.I., Monroe S.S. Cloning of human picobirnavirus genomic segments and development of an RT-PCR detection assay. *Virology.* 2000;277(2):316-329. DOI 10.1006/viro.2000.0594.
- Sekelja M., Berget I., Naes T., Rudi K. Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *ISME J.* 2011;5(3):519-531. DOI 10.1038/ismej.2010.129.
- Shackelton L.A., Holmes E.C. The role of alternative genetic codes in viral evolution and emergence. *J. Theor. Biol.* 2008;254:128-134. DOI 10.1016/j.jtbi.2008.05.024.
- Shahi S., Eusebio-Cope A., Kondo H., Hillman B.I., Suzuki N. Investigation of host range of and host defense against a mitochondrially replicating mitovirus. *J. Virol.* 2019;93:e01503-18. DOI 10.1128/JVI.01503.
- Shi M., Lin X.D., Tian J.H., Chen L.J., Chen X., Li C.X., Qin X.C., Li J., Cao J.P., Eden J.S., Buchmann J., Wang W., Xu J., Holmes E.C., Zhang Y.-Z. Redefining the invertebrate RNA virosphere. *Nature.* 2016;540:539-543. DOI 10.1038/nature20167.
- Smits S.L., Schapendonk C.M., van Beek J., Vennema H., Schürch A.C., Schipper D., Bodewes R., Haagmans B.L., Osterhaus A.D., Koopmans M.P. New viruses in idiopathic human diarrhea cases, the Netherlands. *Emerg. Infect. Dis.* 2014;20(7):1218-1222. DOI 10.3201/eid2007.140190.
- Thannesberger J., Hellinger H.J., Klymiuk I., Kastner M.T., Rieder F.J.J., Schneider M., Fister S., Lion T., Kosulin K., Laengle J., Bergmann M., Ratte T., Steininger C. Viruses comprise an extensive pool of mobile genetic elements in eukaryote cell cultures and human clinical samples. *FASEB J.* 2017;31:1987-2000. DOI 10.1096/fj.201601168R.
- Vainio E.J., Chiba S., Ghabrial S.A., Maiss E., Roossinck M., Sabanadzovic S., Suzuki N., Xie J., Nibert M. ICTV Virus Taxonomy Profile: *Partitiviridae*. *J. Gen. Virol.* 2018;99(1):17-18. DOI 10.1099/jgv.0.000985.
- Wolf Y.I., Kazlauskas D., Iranzo J., Lucía-Sanz A., Kuhn J.H., Krupovic M., Dolja V.V., Koonin E.V. Origins and evolution of the global RNA virome. *mBio.* 2018;9:e02329-18. DOI 10.1128/mBio.02329-18.
- Woo P.C.Y., Lau S.K.P., Teng J.L.L., Tsang A.K.L., Joseph M., Wong E.Y.M., Tang Y., Sivakumar S., Bai R., Wernery R., Wer-

- nery U., Yuen K.-Y. Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. *Virology*. 2014;471-473:117-125. DOI 10.1016/j.virol.2014.09.020.
- Woo P.C.Y., Teng J.L.L., Bai R., Tang Y., Wong A.Y.P., Li K.S.M., Lam C.S.F., Fan R.Y.Y., Lau S.K.P., Yuen K.-Y. Novel picobirnaviruses in respiratory and alimentary tracts of cattle and monkeys with large intra- and inter-host diversity. *Viruses*. 2019;11:574. DOI 10.3390/v11060574.
- Yinda C.K., Ghogomu S.M., Conceição-Neto N., Beller L., Deboutte W., Vanhulle E., Maes P., Van Ranst M., Matthijssens J. Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobimaviruses using an alternative genetic code. *Virus Evol.* 2018;4:vey008. DOI 10.1093/ve/vey008.
- Yinda C.K., Vanhulle E., Conceição-Neto N., Beller L., Deboutte W., Shi C., Ghogomu S.M., Maes P., Ranst M.V., Matthijssens J. Gut virome analysis of Cameroonians reveals high diversity of enteric viruses, including potential interspecies transmitted viruses. *mSphere*. 2019;4(1):e00585-18. DOI 10.1128/mSphere.00585-18.

ORCID ID

A.Yu. Kashnikov orcid.org/0000-0003-1033-7347
N.V. Epifanova orcid.org/0000-0001-7679-8029
N.A. Novikova orcid.org/0000-0002-3710-6648

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

Received May 26, 2022. Revised July 21, 2022. Accepted July 22, 2022.

Original Russian text <https://vavilovj-icg.ru/>

Detection of gene clusters for biodegradation of alkanes and aromatic compounds in the *Rhodococcus qingshengii* VKM Ac-2784D genome

Yu.A. Markova¹, I.S. Petrushin^{1, 2}✉, L.A. Belovezhets³

¹ Siberian Institute of Plant Physiology and Biochemistry of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

² Irkutsk State University, Irkutsk, Russia

³ A.E. Favorsky Irkutsk Institute of Chemistry of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

✉ ivan.kiel@gmail.com

Abstract. Bacterial species of the genus *Rhodococcus* are known to be efficient degraders of hydrocarbons in contaminated soil. They are also employed for bioremediation of polluted environments. These bacteria are widely met in soil, water and living organisms. Previously, we have isolated the *Rhodococcus qingshengii* strain VKM Ac-2784D from the rhizosphere of couch grass growing on oil-contaminated soil. This strain can effectively degrade oil and some model compounds (naphthalene, anthracene and phenanthrene). The results of phylogenetic analysis show that this strain belongs to the species *R. qingshengii*. To understand the catabolic properties of this strain, we have studied its gene clusters possessing such properties. The alkane destruction genes are represented by two clusters and five separate *alkB* genes. The destruction of aromatic compounds involves two stages, namely central and peripheral. The *R. qingshengii* VKM Ac-2784D genome contains four out of eight known central metabolic pathways for the destruction of aromatic compounds. The structure of the gene clusters is similar to that of the known strains *R. jostii* RHA1 and *R. ruber* Chol-4. The peripheral pathways include the genes encoding proteins for benzoic acid destruction. The presence of biphenyl 2,3-dioxygenases as well as gene clusters of benzoate and 2-hydroxypentandienoate pathways suggests that *R. qingshengii* VKM Ac-2784D could degrade polychlorinated biphenyls. The biodegradation ability can be enhanced by biosurfactants, which are known to be synthesized by *Rhodococcus*. The *R. qingshengii* VKM Ac-2784D genome contains the *otsA*, *otsB*, *treY*, *treZ* genes. The bioinformatics data are supported by the previous biochemical experiments that allow a mixture of species with a wide variation of metabolic pathways to be obtained.

Key words: biodegradation; *Rhodococcus*; oil destruction; genomics.

For citation: Markova Yu.A., Petrushin I.S., Belovezhets L.A. Detection of gene clusters for biodegradation of alkanes and aromatic compounds in the *Rhodococcus qingshengii* VKM Ac-2784D genome. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2023;27(3):276-282. DOI 10.18699/VJGB-23-33

Обнаружение генных кластеров биодеструкции алканов и ароматических соединений в геноме *Rhodococcus qingshengii* VKM Ac-2784D

Ю.А. Маркова¹, И.С. Петрушин^{1, 2}✉, Л.А. Беловежец³

¹ Сибирский институт физиологии и биохимии растений Сибирского отделения Российской академии наук, Иркутск, Россия

² Иркутский государственный университет, Иркутск, Россия

³ Иркутский институт химии им. А.Е. Фаворского Сибирского отделения Российской академии наук, Иркутск, Россия

✉ ivan.kiel@gmail.com

Аннотация. Многие представители рода *Rhodococcus* известны как активные биодеструкторы компонентов нефти (в том числе алканов) и ароматических соединений. Пристальное внимание, которое стали уделять родококкам в последнее время, является следствием их высокого катаболического потенциала. Ранее нами выделен штамм *Rhodococcus qingshengii* VKM Ac-2784D из ризосферы пырея, произрастающего на нефтезагрязненной почве. По результатам филогенетического анализа этот штамм может быть отнесен к виду *Rhodococcus qingshengii*. На сегодняшний день расшифрованы пути и идентифицированы гены деструкции многих загрязнителей. Для оценки способности рассматриваемого штамма к деградации нефти и нефтепродуктов мы исследовали генные кластеры, ассоциированные с такой способностью. Ферменты деструкции алканов представлены двумя кластерами и пятью отдельно расположенными генами *alkB*. Деструкция ароматических соединений состоит из двух этапов: периферического и центрального. Геном *R. qingshengii* VKM Ac-2784D содержит четыре из восьми известных центральных путей деструкции ароматических соединений. Структура генных кластеров

сходна с описанными в литературе штаммами *R. jostii* RHA1 и *R. ruber* Chol-4. Периферические пути представлены кластерами генов, кодирующих белки деструкции бензойной кислоты, ген бифенил 2,3-диоксигеназы и два гена бифенил-2,3-диол 1,2-диоксигеназы, участвующих в катаболизме бифенила. Присутствие генов бифенил 2,3-диоксигеназы, кластера генов бензоатного и 2-гидроксипентадиеноатного путей указывает на способность исследуемого штамма деструктировать полихлорированные бифенилы. Усилить активность биодеструкции загрязнителей помогают сурфактанты, улучшающие доступность разлагаемых веществ. Для некоторых видов *Rhodococcus* описаны биосурфактанты на основе трегалозы. В геноме *R. qingshengii* VKM Ac-2784D присутствуют гены, кодирующие белки биосинтеза сурфактантов – *otsA*, *otsB*, *treY*, *treZ*. Данные биоинформационного анализа согласуются с результатами ранее проведенных биохимических исследований. Знание метаболического потенциала отдельного организма, полученное в результате анализа генома, позволит создавать адаптированные к заданным условиям смеси бактериальных штаммов, объединяющие микроорганизмы с разным спектром метаболических путей деструкции.

Ключевые слова: биодеструкция; *Rhodococcus*; нефтедеструкция; геномика.

Introduction

Rhodococcus is a genus of Gram-positive bacteria of the *Actinobacteria* phylum, which are widespread in nature. These bacteria have been isolated from soil, water and living organisms. Some species of *Rhodococcus* are known to be pathogens, e. g. *R. hoagii* (former *R. equi*), which causes zoonotic infection, and *R. fascians* phytopathogen (Garrido-Sanz et al., 2020).

Over the last decade, rhodococci have attracted considerable interest owing to their high catabolic properties. These bacteria are able to degrade various pollutants (polyaromatic hydrocarbons, PAHs; dioxines, dioxin-like polychlorinated biphenyls, etc.) and, therefore, can be employed for bioremediation of soils (Martínková et al., 2009).

The application of rhodococci in bioactive mixtures requires an understanding of which pollutants are capable of destroying a microorganism. This goal can be achieved by the joint use of bioinformatic and biochemical approaches. Currently, most of the genes and pathways of pollutants degradation are known.

To date, numerous gene clusters of rhodococci genomes, which encode oil-degrading enzymes, have been reported (Zampolli et al., 2019). The key components of alkane degradation are alkane-monooxygenase (gene *alkB*, soluble di-iron monooxygenase, SDIMO) and cytochrome (CYP153, member of superfamily P450). Enzymes *AlkB* and CYP153 are usually involved in the oxidation of liquid alkanes, while SDIMO exerts an action on short-chain alkanes (<C8) (Coleman et al., 2011).

The destruction of aromatic compounds involves central and peripheral pathways. The latter are diverse: more than twenty such pathways are known in the literature. Owing to these pathways, rhodococci can oxidize polyaromatic hydrocarbons (PAHs), biphenyls, steroids or phthalates to common intermediates, which are further oxidized by central destruction routes. Eight central destruction pathways are described for *Rhodococcus* species: β -keto adipate, phenylacetate, 2-hydroxypentadienoate, gentisate, homogentisate, hydroxyquinol, homoprotocatechuate, and a pathway with an unknown substrate (Guevara et al., 2019). Genes encoding proteins of peripheral pathways are usually located in plasmids, while genes of central pathways are localized in a chromosome.

Biosurfactants play an auxiliary role in the degradation of hydrophobic compounds. These surface-active agents (SAAs) improve bioavailability of the degrading substances. Rhodo-

cocci can synthesize the biosurfactants from trehalose and mycolic acids (Kuyukina, Ivshina, 2010).

In the present work, we tried to determine the genes encoding the above metabolic pathways in the *Rhodococcus qingshengii* VKM Ac-2784D genome as well as to evaluate the ability of this strain to degradation of oil components.

Materials and methods

The *Rhodococcus qingshengii* VKM Ac-2784D strain was isolated from the rhizosphere of couch grass (*Elytrigia repens*) growing on oil-contaminated soil near Tyret village, Irkutsk region, Russia (Belovezhets et al., 2017). It was found that this strain can oxidize both polyaromatic hydrocarbons (PAHs) and alkanes (Belovezhets et al., 2017).

To study genomic features and biodegradation potential, the whole genome was sequenced, assembled, annotated and submitted to NCBI GenBank (accession no. CP064920, Petrushin et al., 2021). The annotated *R. qingshengii* VKM Ac-2784D genome contained 5775 genes encoding 5716 protein-coding sequences, 53 tRNAs, 3 noncoding RNAs (ncRNAs), and 3 rRNAs.

The sequence of the 16S rRNA subunit was assembled from the initial sequencing data using MATAM software (Pericard et al., 2018). To compare genomic features and define phylogenetic relationship of the *R. qingshengii* VKM Ac-2784D strain, the phylogenetic tree was built using 16S rRNA sequences of strains, described in recent reviews of *Rhodococcus* systematics (Gürtler, Seviour, 2010; Sangal et al., 2019). The phylogenetic tree was built with MEGA X using the Tamura–Nei model and other settings set to default (Kumar et al., 2018). For close related species, average nucleotide identity distance matrix and phylogenetic tree were built using pyani v. 0.2.11 software, mode “ANI_m” (<https://github.com/widdowquinn/pyani>) with default settings. These distances were calculated by the MUMmer algorithm (Deloger et al., 2009) using whole genomes divided to fragments as described in (Richter, Rosselló-Móra, 2009).

Genes related to the metabolism of alkanes, biphenyls and other pollutants were located using BLAST search against whole genome of the *R. qingshengii* VKM Ac-2784D strain with known sequences of functional genes, described in previous studies. All figures, gene annotations were performed using UGENE software (Okonechnikov et al., 2012). Metabolic pathway models were analyzed in RAST SEED (Overbeek et al., 2014) online service.

Results

Phylogenetic relationship of *R. qingshengii* VKM Ac-2784D

To identify phylogenetic relationship of the *R. qingshengii* VKM Ac-2784D strain, a traditional approach based on the construction of a phylogenetic tree according to 16S rRNA sequences was used (Fig. 1). In Figure 1, accession numbers in NCBI GenBank are given for each strain with its species name. Several dozens of species belong to *Rhodococcus*. To make the picture of the tree clearer, only the species that are close to *R. qingshengii* VKM Ac-2784D are shown. Two closest species are *R. qingshengii* and *R. erythropolis*. Both of them are well-known degraders of oil components (Táncsics et al., 2015).

The selection of related strains with oil-degradation activity

Although more than 50 rhodococci are known, we focused our attention on the species capable of degrading oil components, and the whole genome of which was disclosed. After literature analysis, 30 *Rhodococcus* genomes were selected for further study. For these species, the whole-genome and phylogenetic tree were built on the basis of average nucleo-

tide identity distance matrix (presented as heatmap) (Fig. 2). This approach permits to consider the whole genomic data and not only differences in 16S rRNA fragments. Despite the fact that species of some strains were not determined, it was suggested that *Rhodococcus* sp. BH4, like *R. qingshengii* VKM Ac-2784D, is referred to *R. qingshengii*. It should be noted that *R. erythropolis* and *R. rhodochrous* ATCC 17895 species differ considerably from *R. qingshengii*. Interestingly enough, one of the first known oil-degrading strains *R. jostii* RHA1 is very close to *Rhodococcus* sp. DK17.

Gene clusters of alkanes degradation

It was previously reported that the core gene of alkane degradation, *alkB*, is usually co-located with genes of rubredoxin (*alkG1*, *alkG2* or *rubA1*, *rubA2*) or rubredoxin reductase (*alkT* or *rubB*) (Whyte et al., 2002). In the present work, it was shown that *R. qingshengii* VKM Ac-2784D genome has two clusters containing genes of alkane-monooxygenase and rubredoxins (Fig. 3). Also, five separately located genes of alkane-monooxygenase were found (three in chromosome and two in plasmid). The *R. qingshengii* VKM Ac-2784D genome includes 14 genes of cytochrome P450 (11 in chromosome and 3 in plasmid) and has no SDIMO encoding genes.

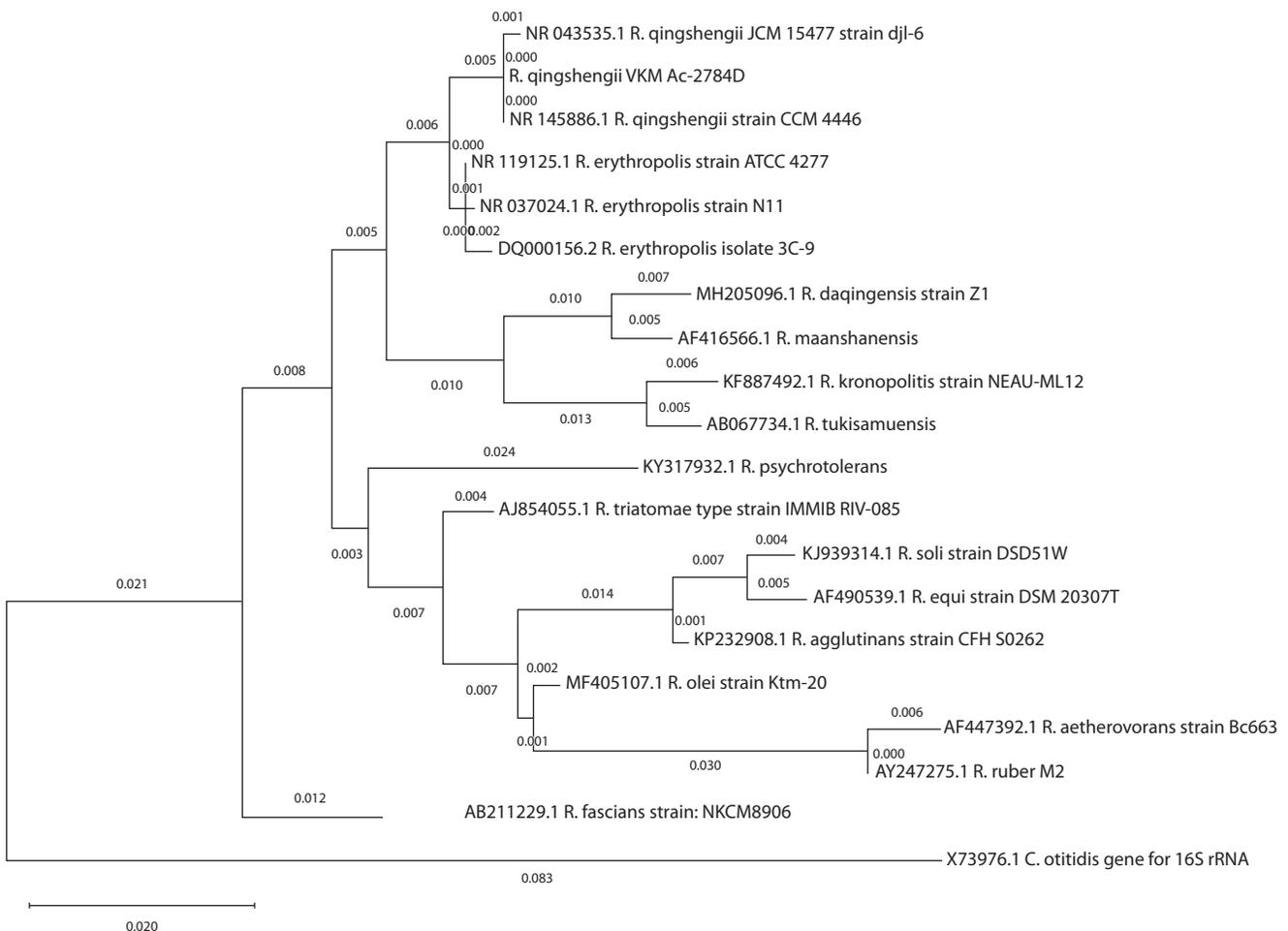


Fig. 1. Phylogenetic tree of *R. qingshengii* VKM Ac-2784D and closer species.

The tree was constructed using 16S rRNA sequences of strains. Accession numbers in NCBI GenBank are given for each strain with its species name.

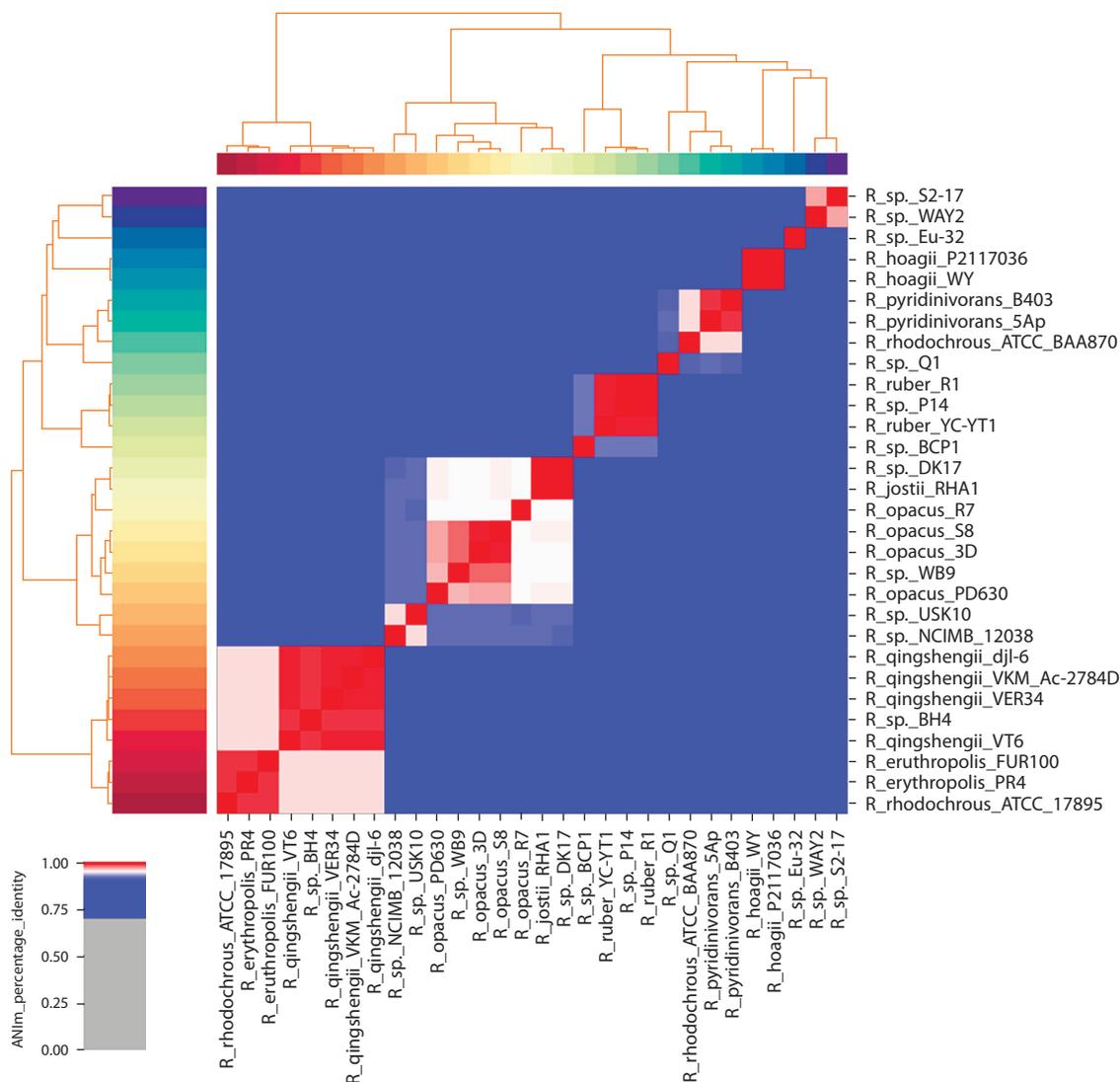


Fig. 2. Heatmap and the phylogenetic tree of *R. qingshengii* VKM Ac-2784D and related species possessing oil-degrading activity. Scale ANIm_percentage_identity shows the similarity percentage of a given pair of the genomes (from 0 to 1).

Central pathways of PAHs destruction

Four central pathways of destruction were revealed for *R. qingshengii* VKM Ac-2784D: β -ketoacid, phenylacetate, 2-hydroxypentadienoate, homogentisate. The structure of the gene clusters (Fig. 4) is similar to that of the previously described strains: *R. jostii* RHA1, *R. ruber* Chol-4 (Navarro-Llorens et al., 2005; Yam et al., 2010; Gibu et al., 2019; Guevara et al., 2019).

Peripheral pathways of PAHs destruction

Generally, the aromatic ring is cleaved by dioxygenase systems Rieske 2Fe-2S, which are involved in the degradation of biphenyls, ethylbenzene and naphthalene (gene clusters including the *bph*, *etb* and *nah* genes). These systems possess a wide range of substrate specificity and can be present in the *Rhodococcus* genome simultaneously (Shumkova et al., 2015). The *R. qingshengii* VKM Ac-2784D genome contains 88 probable oxygenases: 25 dioxygenases and 63 monooxygenases.

Several gene clusters in the *R. qingshengii* VKM Ac-2784D genome belong to peripheral pathways. Among them are 1,2-dioxygenase, 2,3-dioxygenase, benzoic acid degradation gene as well as two genes of biphenyl-2,3-diol, which participate in biphenyl catabolism. The structure of the *bphABCDK* gene cluster is shown in Figure 5. Moreover, the genes encoding dioxygenases, responsible for the intradiol and extradiol aromatic ring-cleavage, were found. At the same time, the *R. qingshengii* VKM Ac-2784D genome contains no *tmo* gene cluster, which takes part in transformation of toluene to *n*-cresol.

Biosurfactants synthesis

For some *Rhodococcus* species, gene clusters related to the trehalose-based synthesis of biosurfactants were documented. The pioneering works dedicated to synthetic approaches to such surfactants dealt with the pathogenic bacteria *Mycobacterium tuberculosis* (De Smet et al., 2000). Similar approaches were proposed for *Rhodococcus* species (Retamal-Morales et

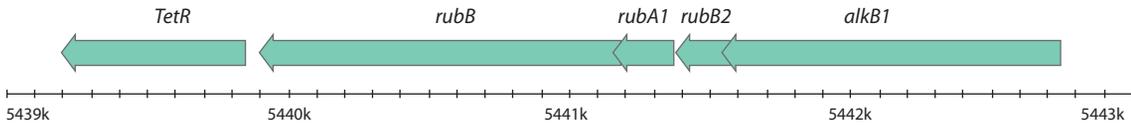


Fig. 3. The structure of gene cluster encoding alkane-monoxygenase, rubredoxins and rubredoxin reductase in the *R. qingshengii* VKM Ac-2784D chromosome.

Here and in Figures 4 and 5: arrows show transcription direction.

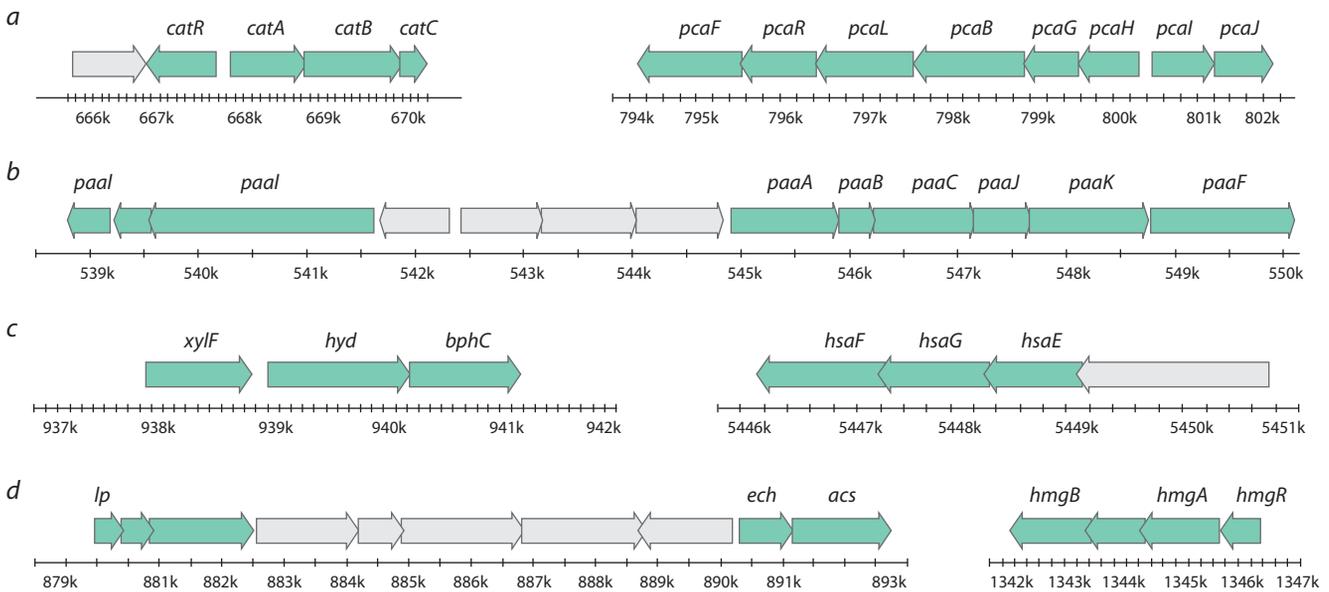


Fig. 4. Gene clusters structure of four central pathways in the *R. qingshengii* VKM Ac-2784D genome: β -ketoadipate (a), phenylacetate (b), 2-hydroxypentadienoate (c), homogentisate (d).

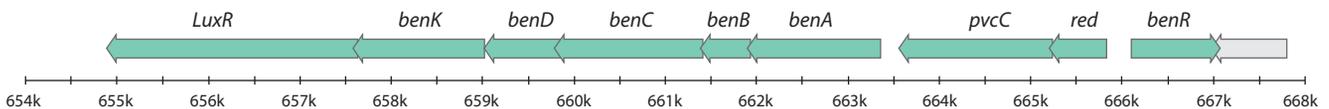


Fig. 5. Structure of gene clusters encoding peripheral pathways of PAHs degradation in the *R. qingshengii* VKM Ac-2784D genome: benzoic acid degradation.

al., 2018). The *R. qingshengii* VKM Ac-2784D genome contains genes encoding proteins of the biosurfactants synthesis: *otsA*, *otsB*, *treY*, *treZ*.

Discussion

The ability of the *R. qingshengii* VKM Ac-2784D strain to degrade oil components, PAHs and some other pollutants is supported by the results of our previous experimental studies (Tretyakova et al., 2019a, b; Belovezhets et al., 2020). It was shown that this strain mainly degraded the alkane fraction of oil, in particular C_{14} – C_{24} alkanes (Belovezhets et al., 2021a). In addition, PAHs turned out to be also efficient in experiments with degradation of model oil compounds (Belovezhets et al., 2021b). Such metabolic activity can be explained by the synthesis of biosurfactants. The strain under study appeared to be an effective producer of extracellular and cell-binded forms of bio-SAAs. The highest amount of the synthesized

extracellular biosurfactants was almost 1.5 g/L (substrate optimization was not carried out) (Belovezhets et al., 2021b).

To determine the mechanisms of PAH and alkane degradation, the whole genome sequencing and analysis of the gene clusters related to the corresponding metabolic pathways were performed.

Nowadays genome sequencing has become a routine procedure for studying gene properties. Gene annotation of a major part of genes (related to catabolic potential) is performed automatically with NCBI GenBank Prokaryotic Genome Annotation Pipeline (PGAP). These annotations allow some metabolic gene clusters to be determined at the stage of preliminary bioinformatic analysis. Further, BLAST search was employed to find particular functional genes from previous studies (Navarro-Llorens et al., 2005; Yam et al., 2010; Guevara et al., 2019). It was revealed that the *R. qingshengii* VKM Ac-2784D genome contains gene clusters 4 out

of 8 known central pathways for PAHs destruction. Alkanes degradation enzymes are represented by two gene clusters containing genes of alkane-monooxygenase and rubredoxins and five separately located *alkB* genes (three in chromosome and two in plasmid). The presence of biphenyl 2,3-dioxygenase genes as well as gene clusters of benzoate and 2-hydroxypentandienoate indicates that *R. qingshengii* VKM Ac-2784D can degrade polychlorinated biphenyls.

The detailed data on the studied genes and their presence in genomes of the strains are given in Supplementary Material¹. For the *R. qingshengii* VKM Ac-2784D genome, the loci names of functional genes, references to the experimental works and degree of similarity for each gene are indicated. It is found that 7 out of 29 strains are most similar to *R. qingshengii* VKM Ac-2784D. These are *R. ruber* YC-YT1, *R. qingshengii* VER34, *R. qingshengii* VT6, *R. qingshengii* djl-6-2, *Rhodococcus* sp. PR4, *Rhodococcus* sp. BH4, *R. rhodochrous* ATCC 17895. The names of these strains are marked in gray in Supplementary Material.

The bioinformatics data are supported by the results of previous biochemical studies. As a result, it is shown that the catabolic properties of *R. qingshengii* VKM Ac-2784D permit to apply this strain for the destruction of various pollutants.

Conclusion

Nowadays there many genes and pathways related to the degradation of the various pollutants are identified. Genome analysis allows to uncover the metabolic potential of the particular microorganism, and obtain the specialized microbial mixtures with wide bioactive degradation spectrum. To understand the catabolic potential of these bacteria to degrade oil and its components we study its gene clusters, associated with such abilities. The structure of gene clusters is the same of known for strains *R. jostii* RHA1 and *R. ruber* Chol-4. Alkanes destruction genes grouped by two clusters and five separate genes *alkB*. The biodegradation activity may be enhanced by the biosurfactants, which are known to be synthesized by *Rhodococcus*. This knowledge can help to create the mixture of species with wide variation of metabolic pathways, but to evaluate the effectiveness of such mixtures further experiments are required.

References

- Belovezhets L.A., Makarova L.A., Tretyakova M.S., Markova Yu.A., Dudareva L.V., Semenova N.V. Possible pathways for destruction of polyaromatic hydrocarbons by some oil-degrading bacteria isolated from plant endosphere and rhizosphere. *Appl. Biochem. Microbiol.* 2017;53(1):68-72. DOI 10.1134/S0003683817010069.
- Belovezhets L.A., Markova Yu.A., Tretyakova M.S., Klyba L.V., Sanzheeva E.R. Destruction of an oil paraffin fraction by microorganisms. *Chem. Technol. Fuels Oils.* 2021a;56(6):919-925. DOI 10.1007/s10553-021-01208-z.
- Belovezhets L.A., Tretyakova M.S., Markova Yu.A. Physicochemical properties of biosurfactants produced by oil destructor microorganisms. *Chem. Sustain. Dev.* 2021b;29(1):20-25. DOI 10.15372/csd2021273.
- Belovezhets L.A., Tretyakova M.S., Markova Yu.A., Levchuk A.A. Use of rhizosphere microorganisms for bioremediation of oil contaminated soils. *IOP Conf. Ser.: Earth Environ. Sci.* 2020;408:012086. DOI 10.1088/1755-1315/408/1/012086.

- Coleman N.V., Yau S., Wilson N.L., Nolan L.M., Migocki M.D., Ly M., Crosssett B., Holmes A.J. Untangling the multiple monooxygenases of *Mycobacterium chubuense* strain NBB4, a versatile hydrocarbon degrader. *Environ. Microbiol. Rep.* 2011;3(3):297-307. DOI 10.1111/j.1758-2229.2010.00225.x.
- Deloger M., El Karoui M., Petit M.A. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J. Bacteriol.* 2009;191(1):91-99. DOI 10.1128/JB.01202-08.
- De Smet K.A.L., Weston A., Brown I.N., Young D.B., Robertson B.D. Three pathways for trehalose biosynthesis in mycobacteria. *Microbiology.* 2000;146(1):199-208. DOI 10.1099/00221287-146-1-199.
- Garrido-Sanz D., Redondo-Nieto M., Martin M., Rivilla R. Comparative genomics of the *Rhodococcus* genus shows wide distribution of biodegradation traits. *Microorganisms.* 2020;8(5):774. DOI 10.3390/microorganisms8050774.
- Gibu N., Kasai D., Ikawa T., Akiyama E., Fukuda M. Characterization and transcriptional regulation of *n*-alkane hydroxylase gene cluster of *Rhodococcus jostii* RHA1. *Microorganisms.* 2019;7(11):479. DOI 10.3390/microorganisms7110479.
- Guevara G., Lopez M.-C., Alonso S., Perera J., Navarro-Llorens M. New insights into the genome of *Rhodococcus ruber* strain Chol-4. *BMC Genomics.* 2019;20(1):332. DOI 10.1186/s12864-0195677-2.
- Gürtler V., Seviour R.J. Systematics of members of the genus *Rhodococcus* (Zopf 1891) Emend Goodfellow et al. 1998. In: Alvarez H. (Ed.) *Biology of Rhodococcus*. Microbiology Monographs. Vol. 16. Berlin; Heidelberg: Springer, 2010;1-28. DOI 10.1007/978-3-642-12937-7_1.
- Kumar S., Stecher G., Knyaz C., Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 2018;35(6):1547-1549. DOI 10.1093/molbev/msy096.
- Kuyukina M.S., Ivshina I.B. Application of *Rhodococcus* in bioremediation of contaminated environments. In: Alvarez H. (Ed.) *Biology of Rhodococcus*. Microbiology Monographs. Vol. 16. Berlin; Heidelberg: Springer, 2010;231-262. DOI 10.1007/978-3-642-12937-7_9.
- Martínková L., Uhnakova B., Nesvera J., Kren V. Biodegradation potential of the genus *Rhodococcus*. *Environ. Int.* 2009;35(1):162-177. DOI 10.1016/j.envint.2008.07.018.
- Navarro-Llorens J.M., Patrauchan M.A., Stewart G.R., Davies J.E., Eltis L.D., Mohn W.W. Phenylacetate catabolism in *Rhodococcus* sp. strain RHA1: a central pathway for degradation of aromatic compounds. *J. Bacteriol.* 2005;187(13):4497-4504. DOI 10.1128/JB.187.13.4497-4504.2005.
- Okonechnikov K., Golosova O., Fursov M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012;28(8):1166-1167. DOI 10.1093/bioinformatics/bts091.
- Overbeek R., Olson R., Pusch G., Olsen G., Davis J., Disz T., Edwards R., Gerdes S., Parrello B., Shukla M., Vonstein V., Wattam A., Xia F., Stevens R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014;42(D1):D206-D214. DOI 10.1093/nar/gkt1226.
- Pericard P., Dufresne Y., Couderc L., Blanquart S., Touzet H. MATAM: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes. *Bioinformatics.* 2018;34(4):585-591. DOI 10.1093/bioinformatics/btx644.
- Petruhin I.S., Markova Yu.A., Karepova M.D., Zaytseva Y.V., Belovezhets L.A. Complete genome sequence of *Rhodococcus qingshengii* strain VKM Ac-2784D, isolated from *Elytrigia repens* rhizosphere. *Microbiol. Resour. Announc.* 2021;10(11):e00107-21. DOI 10.1128/mra.00107-21.
- Retamal-Morales G., Heine T., Tischler J.S., Erler B., Groning J.A.D., Kaschaber S.R., Schlomann M., Levican G., Tischler D. Draft genome sequence of *Rhodococcus erythropolis* B7g, a biosurfactant producing actinobacterium. *J. Biotechnol.* 2018;280:38-41. DOI 10.1016/J.JBIOTECH.2018.06.001.
- Richter M., Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA.* 2009;106(45):19126-19131. DOI 10.1073/pnas.0906412106.

¹ Supplementary Material is available in the online version of the paper: <https://vavilovj-icg.ru/download/pict-2023-27/appx10.pdf>

- Sangal V., Goodfellow M., Jones A.L., Seviour R.J., Sutcliffe I.C. Refined systematics of the genus *Rhodococcus* based on whole genome analyses. In: Alvarez H. (Ed.) *Biology of Rhodococcus*. Microbiology Monographs. Vol. 16. Cham: Springer, 2019;1-21. DOI 10.1007/978-3-030-11461-9_1.
- Shumkova E.S., Egorova D.O., Boronnikova S.V., Plotnikova E.G. Polymorphism of the *bphA* genes in bacteria destructing biphenyl/chlorinated biphenils. *Mol. Biol.* 2015;49(4):569-580. DOI 10.1134/S0026893315040159.
- Táncsics A., Benedek T., Szoboszlay S., Veres P.G., Farkas M., Mathe I., Marialigeti K., Kukolya J., Lanyi S., Kriszt B. The detection and phylogenetic analysis of the alkane 1-monoxygenase gene of members of the genus *Rhodococcus*. *Syst. Appl. Microbiol.* 2015;38(1):1-7. DOI 10.1016/j.syapm.2014.10.010.
- Tretyakova M.S., Ivanova M.V., Belovezhets L.A., Markova Yu.A. Possible use of oil-degrading microorganisms for protection of plants growing under conditions of oil pollution. *IOP Conf. Ser.: Earth Environ. Sci.* 2019a;315:072046. DOI 10.1088/1755-1315/315/7/072046.
- Tretyakova M.S., Markova Yu.A., Belovezhets L.A. Microbial preparation for bioremediation of soil contaminated with oil and oil products. 2019b. Available at: <https://patents.google.com/patent/RU2705290C1/en>.
- Whyte L.G., Smits T.H.M., Labbé D., Witholt B., Greer C.W., van Beilen J.B. Gene cloning and characterization of multiple alkane hydroxylase systems in *Rhodococcus* strains Q15 and NRRL B-16531. *Appl. Environ. Microbiol.* 2002;68(12):5933-5942. DOI 10.1128/AEM.68.12.5933-5942.2002.
- Yam K.C., van der Geize R., Eltis L.D. Catabolism of aromatic compounds and steroids by *Rhodococcus*. In: Alvarez H. (Ed.) *Biology of Rhodococcus*. Microbiology Monographs. Vol. 16. Berlin; Heidelberg: Springer, 2010;133-169. DOI 10.1007/978-3-642-12937-7_6.
- Zampolli J., Zeaiter Z., Di Canito A., Di Gennaro P. Genome analysis and -omics approaches provide new insights into the biodegradation potential of *Rhodococcus*. *Appl. Microbiol. Biotechnol.* 2019; 103:1069-1080. DOI 10.1007/s00253-018-9539-7.

ORCID ID

Yu.A. Markova orcid.org/0000-0001-7767-4204
I.S. Petrushin orcid.org/0000-0002-8788-5352
L.A. Belovezhets orcid.org/0000-0001-5922-3397

Acknowledgements. This study was carried out within the framework of basic budget funding No. 122041100050-6 (FWSS-2022-0004). The bioinformatics data analysis was performed in part on the equipment of the Bioinformatics Shared Access Center, the Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (ICG SB RAS).

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

Received May 16, 2022. Revised September 5, 2022. Accepted September 6, 2022.

Original Russian text <https://vavilovj-icg.ru/>

Compilation and functional classification of telomere length-associated genes in humans and other animal species

E.V. Ignatieva , N.S. Yudin, D.M. Larkin 

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

 eignat@bionet.nsc.ru; dmlarkin@gmail.com

Abstract. Telomeres are the terminal regions of chromosomes that ensure their stability while cell division. Telomere shortening initiates cellular senescence, which can lead to degeneration and atrophy of tissues, so the process is associated with a reduction in life expectancy and predisposition to a number of diseases. An accelerated rate of telomere attrition can serve as a predictor of life expectancy and health status of an individual. Telomere length is a complex phenotypic trait that is determined by many factors, including the genetic ones. Numerous studies (including genome-wide association studies, GWAS) indicate the polygenic nature of telomere length control. The objective of the present study was to characterize the genetic basis of the telomere length regulation using the GWAS data obtained during the studies of various human and other animal populations. To do so, a compilation of the genes associated with telomere length in GWAS experiments was collected, which included information on 270 human genes, as well as 23, 22, and 9 genes identified in the cattle, sparrow, and nematode, respectively. Among them were two orthologous genes encoding a shelterin protein (*POT1* in humans and *pot-2* in *C. elegans*). Functional analysis has shown that telomere length can be influenced by genetic variants in the genes encoding: (1) structural components of telomerase; (2) the protein components of telomeric regions (shelterin and CST complexes); (3) the proteins involved in telomerase biogenesis and regulating its activity; (4) the proteins that regulate the functional activity of the shelterin components; (5) the proteins involved in telomere replication and/or capping; (6) the proteins involved in the alternative telomere lengthening; (7) the proteins that respond to DNA damage and are responsible for DNA repair; (8) RNA-exosome components. The human genes identified by several research groups in populations of different ethnic origins are the genes encoding telomerase components such as *TERC* and *TERT* as well as *STN1* encoding the CST complex component. Apparently, the polymorphic loci affecting the functions of these genes may be the most reliable susceptibility markers for telomere-related diseases. The systematized data about the genes and their functions can serve as a basis for the development of prognostic criteria for telomere length-associated diseases in humans. Information about the genes and processes that control telomere length can be used for marker-assisted and genomic selection in the farm animals, aimed at increasing the duration of their productive lifetime. Key words: telomere length; candidate genes; genome-wide association study; functional classification.

For citation: Ignatieva E.V., Yudin N.S., Larkin D.M. Compilation and functional classification of telomere length-associated genes in humans and other animal species. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(3):283-292. DOI 10.18699/VJGB-23-34

Компиляция и функциональная классификация генов, ассоциированных с длиной теломер, у человека и других видов животных

Е.В. Игнатьева , Н.С. Юдин, Д.М. Ларкин 

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
 eignat@bionet.nsc.ru; dmlarkin@gmail.com

Аннотация. Теломеры – это концевые участки хромосом, обеспечивающие их стабильность в ходе клеточного деления. Укорочение теломер инициирует процесс старения клеток, что может приводить к дегенерации и атрофии тканей. Укорочение теломер связано с сокращением продолжительности жизни и с предрасположенностью к ряду заболеваний, поэтому данный показатель может быть использован в качестве предиктора продолжительности жизни и состояния здоровья отдельного индивида. Длина теломер – сложный фенотипический признак, который определяется многими факторами, в том числе генетическими. Многочисленные исследования (включая полногеномный анализ ассоциаций, ПГАА) свидетельствуют о полигенном характере контроля длины теломер. Цель работы – охарактеризовать генетические основы регуляции длины теломер на основе данных ПГАА, полученных при исследовании различных популяционных выборок человека и других животных. Для этого авторами была собрана компиляция генов, ассоциированных с длиной теломер по дан-

ным ПГАА, которая включала сведения о 270 генах человека, а также 23, 22 и 9 генах, выявленных у крупного рогатого скота, домового воробья и нематоды соответственно. Среди них присутствовали два гена-ортолога, кодирующих белок шелтеринового комплекса (*POT1* у человека и *pot-2* у *C. elegans*). Функциональный анализ показал, что на длину теломер могут влиять генетические варианты в генах, кодирующих: 1) структурные компоненты теломеразы; 2) белковые компоненты теломерных участков хромосом (шелтериновый комплекс и CST комплекс); 3) белки, участвующие в биогенезе теломеразы и регулирующие ее активность; 4) белки, регулирующие функциональную активность компонентов шелтеринового комплекса; 5) белки, участвующие в репликации и/или кэпировании теломер; 6) белки, контролируемые альтернативный путь удлинения теломер; 7) белки, реагирующие на повреждения ДНК и отвечающие за репарацию; 8) компоненты РНК экзосом. В работе выявлены гены человека, идентифицированные несколькими исследовательскими группами в популяциях различного этнического происхождения. Это гены, кодирующие компоненты теломеразы (*TERC* и *TERT*), а также ген *STN1*, кодирующий белок CST комплекса. По-видимому, полиморфные локусы, затрагивающие функции этих генов, могут быть наиболее надежными маркерами предрасположенности к заболеваниям, связанным с длиной теломер. Систематизированные нами данные о генах и их функциях будут полезны при разработке прогностических критериев заболеваний человека, для которых показана связь с длиной теломер. Сведения о генах и процессах, контролируемых длину теломер, могут быть востребованы для маркер-ориентированной и геномной селекции сельскохозяйственных животных, направленной на повышение продолжительности их хозяйственного использования.

Ключевые слова: длина теломер; гены-кандидаты; полногеномный анализ ассоциаций; функциональный анализ.

Introduction

Telomeres are the terminal regions of chromosomes that ensure their stability and represented by evolutionary conserved tandemly repeated DNA sequences (e.g., a hexanucleotide TTAGGG repeat in vertebrates) of several kb in length (Podlevsky, 2008; Monaghan, Ozanne, 2018). For example, their lengths in humans at birth are 10–15 kb (Jafri et al., 2016). 3' terminal end of a telomere is a single-stranded guanine-rich DNA region (150–200 nucleotides), whose end interacts with the double-stranded region to form the so-called T-loop at the telomere end. T-loop formation and stabilization are ensured by a shelterin complex (Fig. 1). This structure prevents recognition of a chromosome terminal region by repair proteins (de Lange, 2018).

DNA polymerase is unable to fully replicate the 3'-end of a linear DNA during cell division, which leads to a loss of 50–200 nucleotides of the telomeric sequence at each cell division (Fan et al., 2021). Telomere shortening can also be facilitated by other factors and processes (Suppl. Material 1)¹, such as oxidative stress, inflammation, UV irradiation, effects of toxic agents, DNA replication errors, etc. (Aviv, Shay, 2018; Monaghan, Ozanne, 2018). These factors are likely to produce different effects depending on cell types and the organism's development stage and species (Monaghan, Ozanne, 2018).

Telomere shortening initiates cellular senescence. Activation of DNA damage response signaling pathways results in a cell cycle arrest, which may in turn lead to apoptosis and eventually – to progressive tissue degeneration (Jafri et al., 2016; Aviv, Shay, 2018; Monaghan, Ozanne, 2018). The data collected at the cellular level *in vitro* and in model organisms lay the foundation for using the telomere length as a predictor of life expectancy and health status of an individual. Indeed, studies in humans (Crocco et al., 2021), mice (Vera et al., 2012), sheep (Wilbourn et al., 2018), cattle (Seeker et al.,

2021), wild birds (Bichet et al., 2020), and other animals have shown that shorter telomere length may be associated with reduced life expectancy. The studies in humans discovered an association between the telomere length and cardiovascular diseases, cancer, diabetes, inflammation, and other pathological states (Kong et al., 2013; Jafri et al., 2016; Aviv, Shay, 2018).

Telomere shortening is prevented by telomerase, a specialized ribonucleoprotein complex acting as a reverse transcriptase. In humans, telomerase is active in almost all the cancer cells studied (Jafri et al., 2016), in blastocyst, in most somatic tissues at 16–20 weeks of development (except for brain cells), and ovary and sperm cells at all ontogenetic stages (except for mature spermatozooids and oocytes) (Wright et al., 1996).

Telomerase activity is controlled by the proteins regulating expression of telomerase components, their movement to various cell compartments, processing, and assembly as well as by the proteins maintaining stability of the telomerase complex or, on the contrary, activating its degradation (Egan, Collins, 2012; Tseng et al., 2015; Schrupfová, Fajkus, 2020). The main stages of telomerase biogenesis are presented in Figure 2. The examples of proteins affecting the telomerase activity are presented in Suppl. Material 2. In addition, the telomerase activity is also affected by the shelterin (Diotti, Loayza, 2011; de Lange, 2018) and CST complex (Fig. 3) (Chen et al., 2012).

Telomere length is a complex phenotypic trait determined by multiple factors including genetic ones. The meta-analysis of heritability data for this trait performed in eighteen vertebrate species showed the averaged heritability index of 45 %. The studies showed the value of 52 % in humans, 42 % – in Holstein cattle breed, 35 % – in hamadryas baboons, and 5 % – in sheep (Chik et al., 2022).

The problem of genetic basis of telomere length regulation is of interest for many researchers. The information on telo-

¹ Supplementary Materials 1–11 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Ignatieva_Engl_27_3.pdf.

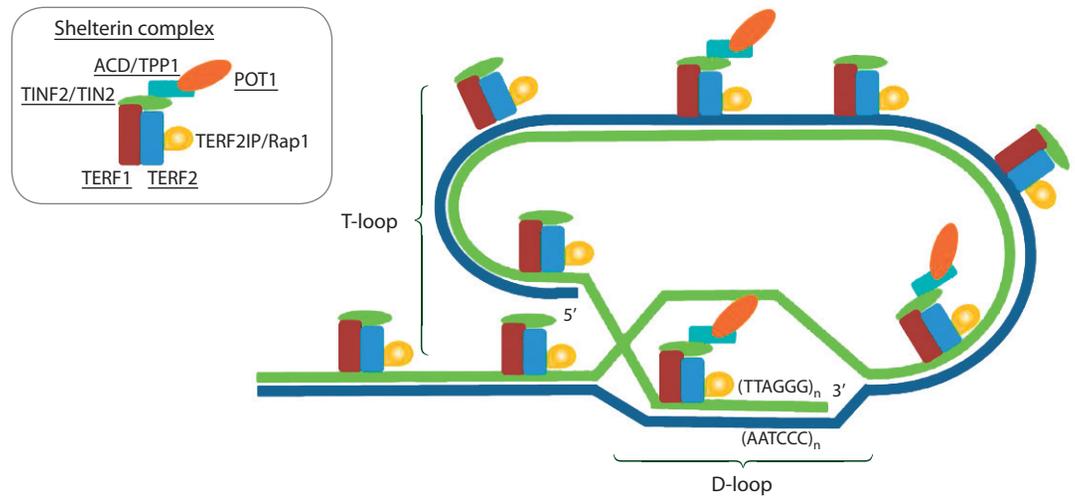


Fig. 1. Structure of a chromosome's telomeric region.

Telomeric DNA is presented as a T-loop reconstructed following the black-and-white illustration from Fan and co-workers (2021); nucleotide sequences in the DNA strands are not shown (Fan et al., 2021). The top left is simplified illustration of the relative positions of shelterin subunits, following the description from Jafri and co-workers (2016). Since the ACD/TPP1 and POT1 are much less abundant in nuclei (de Lange, 2018), some shelterin complexes are depicted without these subunits. D-loop is a structure, where two strands of the double-stranded DNA are separated, and one of them connects with the third DNA strand (a single-stranded 3'-end of the telomeric DNA region). The names of proteins corresponding to the human genes associated with telomere length according to the GWAS data are underlined.

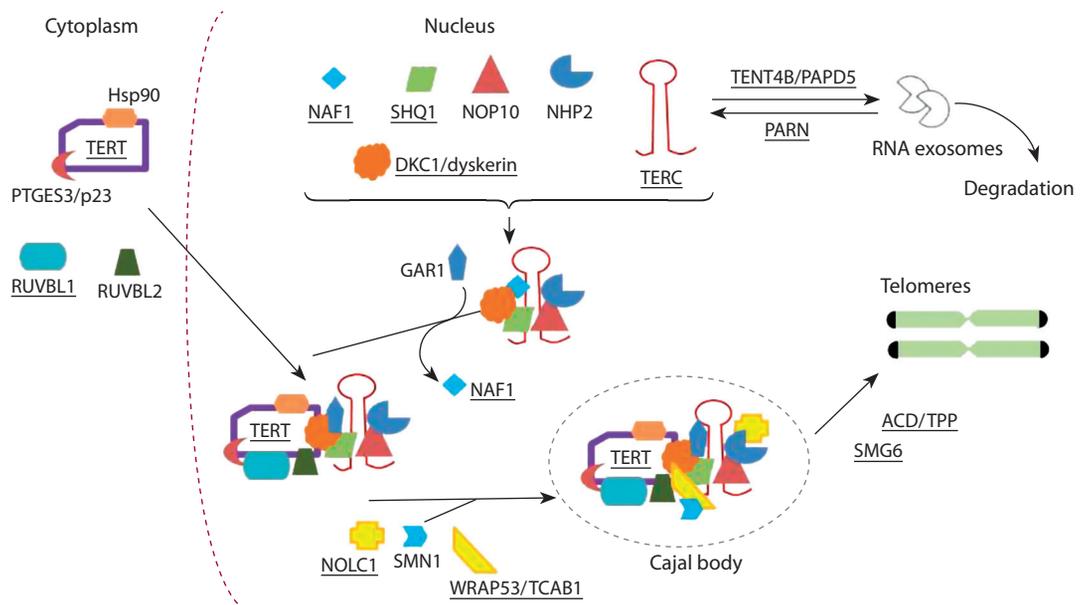


Fig. 2. Simplified presentation of the main stages of telomerase biogenesis.

The names of proteins and RNAs corresponding to the human genes associated with telomere length (as per GWAS results) are underlined. The scheme is based on the data on protein functions from the research articles cited in Suppl. Material 2.

merase components and proteins involved in telomere length regulation (including 20 proteins identified in mammals) can be found in The Telomerase Database (<http://telomerase.asu.edu/>) (Podlevsky et al., 2008). Joyce and co-workers (2018) presented a set of 80 human genes with telomere-related functions (Joyce et al., 2018).

The GWAS data also indicate a polygenic nature of telomere length heritability. For instance, the GWAS Catalog ([https://](https://www.ebi.ac.uk/gwas/)

www.ebi.ac.uk/gwas/) cites 99 human genes that either include or neighbor the telomere length-associated allelic variants. One of the largest GWA studies presents the data on 138 human genomic loci, whose allelic variants are associated with telomere length (Codd et al., 2021).

In addition to the telomere length association data gathered using GWA studies in various human population samples, the data obtained in other animal species also appear to be

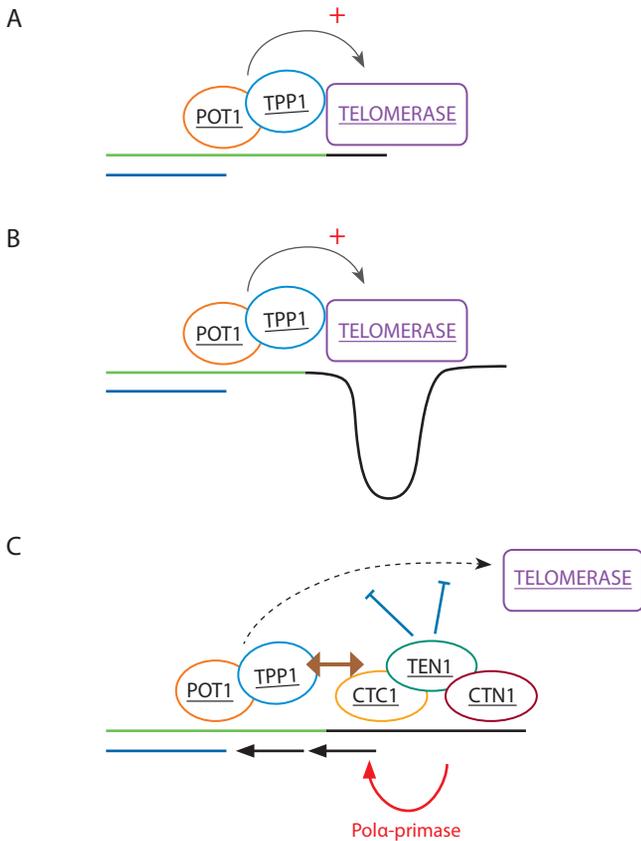


Fig. 3. The role of CST proteins in telomerase activity regulation at the late S/G2 phase.

The first step of the five-step mechanism described by Chen et al. (2012) is the recruitment of telomerase and additional ACD/TPP1 and POT1 by shelterin complex (Step 1, *Recruitment*, not shown in the figure). Then, telomerase starts extending the single-stranded region of the DNA molecule (Step 2, *Extension I*) (shown in Panel A, the newly synthesized DNA region is represented by a black line). After that (Step 3, *Extension II*) the single-stranded region of the DNA molecule is further extended (see Panel B). CST proteins interact with the newly synthesized single-stranded DNA region (~60 nucleotides) hindering the stimulating effect of ACD/TPP1 and POT1 on telomerase (Step 4, *Termination*) and initiating C-strand synthesis by DNA polymerase alpha-primase (Polα-primase) (Step 5, *Fill-in*). Steps 4 and 5 are presented in Panel C. The names of proteins corresponding to the human genes associated with telomere length (as per GWAS results) are underlined.

of interest. However, these studies are rather scarce and are available only for Holstein–Friesian cattle (Ilska-Warner et al., 2019), house sparrow nestlings (Pepke et al., 2021), and *C. elegans* (Cook et al., 2016).

The objective of this review was to characterize the genetic basis of telomere length regulation using the GWAS data collected in various human populations and to compare them with the results of similar experiments in other animal species. For this purpose, (1) the data on genes identified in GWAS telomere length experiments were systematized; (2) functional annotation of genes was performed, and the set of biological processes affecting telomere length was identified.

Materials and methods

The data on telomere length-associated genes were obtained from the papers available in the PubMed database ([https://](https://pubmed.ncbi.nlm.nih.gov/)

pubmed.ncbi.nlm.nih.gov/) using such keywords as ‘telomere length’ and ‘GWAS’. Functional annotation of genes was performed using information obtained from the papers presenting GWAS data, PubMed, The Telomerase Database (<http://telomerase.asu.edu/>) queries, and the DAVID knowledgebase (<https://david.ncifcrf.gov/>) (Sherman et al., 2022).

Results and discussion

Human genes identified through GWA studies

PubMed queries produced 18 scientific papers presenting the results of identifying telomere length-associated polymorphic loci in human genome based on GWAS data. These papers were analyzed, and the data on 270 telomere length-associated genes were collected (Suppl. Material 3). Most genes (262) were identified in European-ancestry population samples, the data on 15 genes were obtained from the studies in Southeast Asian population samples (natives of China, Bangladesh, and India), five genes were identified as a result of trans-ethnic meta-analysis of Singaporean Chinese and European ancestry data (Dorajoo et al., 2019), and one gene was found in African Americans (Zeiger et al., 2018).

The data on functional significance in the context of telomere length regulation were presented by the authors of GWA studies for 52 genes out of 270 (see Suppl. Material 3). The fact that the data on gene significance in the context of telomere length regulation were unavailable for a number of loci reflects the capabilities and limitations of GWAS methodology. Most loci identified by GWAS and associated with the trait of interest are located in intergenic regions. As a rule, in these cases, the set of candidate genes includes the nearest genes, whose functional significance is often difficult to interpret. To identify the mechanisms and genes, through which intergenic variants affect the studied traits, additional experiments are required. For example, it was shown that T-to-C substitution of rs1421085 in the intron of *FTO* gene affects the expression of *IRX3* and *IRX5*, whose transcription start sites are far away (~520 and ~1160 kb) from rs1421085 (Claussnitzer et al., 2015).

Main functional groups of human telomere length-associated genes

A functional classification was performed for a set of 52 human genes for which there was information about their functional significance in the context of telomere length regulation (Suppl. Material 4). As a result, several functional groups of genes have been identified (Fig. 4):

Genes encoding telomerase components: TERC is the telomerase RNA component acting as a matrix for DNA strand extension at the telomere end and TERT is a reverse transcriptase enzyme subunit (Egan, Collins, 2012; Tseng et al., 2015).

Genes encoding shelterin proteins: components of this complex (TERF1/TRF1, TERF2/TRF2, POT1, TERF2IP/Rap1/DRIP5, TIN2/TIN2, and ACD/TPP1/TINT1) can bind to both double-stranded and single-stranded telomeric DNA regions (see Fig. 1), stabilize them, protect them from exo-

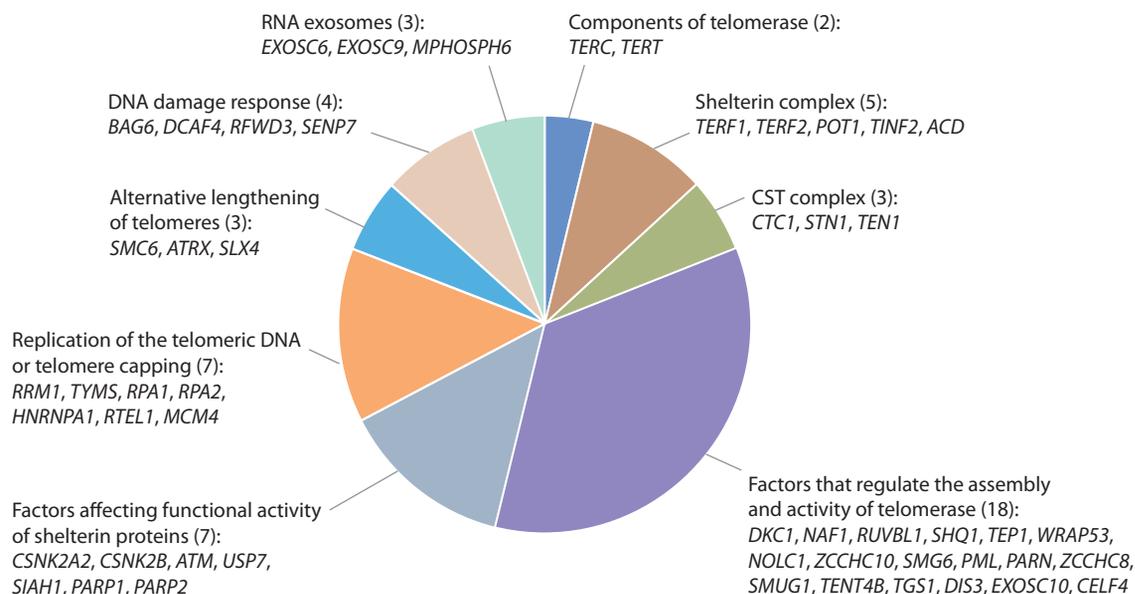


Fig. 4. Functional groups of human telomere length-associated genes.

The classification is presented for 52 genes, whose role in telomere length regulation is characterized in Suppl. Material 4. The numbers given in parentheses indicate the numbers of genes in groups.

nucleases, reduce telomerase access, and inhibit the proteins activated by damaged DNA and involved in double-stranded break repair (Diotti, Loayza, 2011; de Lange, 2018). The GWAS data on telomere length association were obtained for the genes coding for five out of six shelterin proteins (TERF1, TERF2, POT1, TINF2 and ACD/TPP1/TINT1) (see Fig. 4, Suppl. Material 4).

Genes encoding CST proteins: CTC1, STN1, TEN1. CST complex acts as a telomerase negative regulator at the late S to G2 phase of the cell cycle (see Fig. 3) (Chen et al., 2012).

Genes encoding proteins involved in telomerase biogenesis and regulating its activity. One of these genes, *ZCCHC10*, encodes a protein regulating telomerase synthesis at transcriptional level: *ZCCHC10* suppresses *TERT* transcription (Ohira et al., 2019). Processing and assembly of a telomerase RNA subunit involves DKC1, NAF1, and SHQ1 (Egan, Collins, 2012), ribonuclease PARN, exoribonuclease DIS3, the component of a nuclear exosome targeting (NEXT) complex, *ZCCHC8* (Tseng et al., 2015), *SMUG1* (Kroustallaki et al., 2019), and *CELF4/BRUNOL4* (Mangino et al., 2009). Non-canonical polymerase *TENT4B/PAPD5* (Nagpal et al., 2020), trimethylguanosine synthetase *TGS1* (Chen et al., 2020), and *EXOSC10* RNA exosome component (Stuparević et al., 2021) cause a decrease in the level of active TERC. The assembly of telomerase nucleoprotein complex involves ATPase *RUVBL1/pontin* (Jafri et al., 2016) and telomerase-associated protein *TEP1* (Codd et al., 2021). Two proteins (*WRAP53/WDR79/TCAB1* and *NOLC1/NOOP140*) provide telomerase accumulation in Cajal bodies, the small nuclear organelles where processing of small nuclear and nucleolar RNAs and assembly or ribonucleoprotein complexes occur (Bizarro et al., 2019; Schrupfová, Fajkus, 2020). Telomerase activity

is modulated by activator protein *SMG6/EST1A*, which also binds to a single-stranded DNA (Snow et al., 2003), and *PML* protein, whose isoform *PML-IV* suppresses telomerase activity (Oh et al., 2009).

Genes encoding proteins regulating functional activity of shelterin proteins. *CSNK2A2* and *CSNK2B* are the subunits of casein kinase which phosphorylates *TERF1*, increasing its binding to telomeres (Saxena et al., 2014; Li et al., 2020). *ATM* serine/threonine kinase, on the contrary, decreases *TERF1* binding to the telomeric DNA (Li et al., 2020). Peptidase *USP7* and ubiquitin ligase *SIAH1* activate proteasomal degradation of *POT1* and *TERF2*, respectively (Codd et al., 2021). ADP ribosylases *PARP1* and *PARP2* reduce the DNA binding activity of *TERF2* (Dorajoo et al., 2019; Codd et al., 2021).

Genes encoding proteins involved in telomere replication and/or capping: (1) enzymes *RRM1* and *TYMS* involved in synthesis of deoxynucleoside triphosphates (dNTP) and thymidylates required for DNA synthesis (Dorajoo et al., 2019; Nersisyan et al., 2019); (2) helicases *RTEL1* and *MCM4* (Codd et al., 2013, 2021); (3) *RPA1* and *RPA2*, the subunits of the RPA complex capable to unfold G-quadruplex structures that may block DNA replication (Codd et al., 2021); (4) *HNRNPA1* promoting telomere capping after DNA replication (Codd et al., 2021).

Genes encoding proteins affecting the alternative telomere lengthening pathway. This telomerase-independent mechanism (ALT or Alternative Lengthening of Telomere, see the description in Suppl. Material 5) includes the recombination between telomeric regions of two DNA molecules (Sobinoff, Pickett, 2017, 2020). Three genes identified in GWA studies were attributed to this group (see Fig. 4, Suppl.

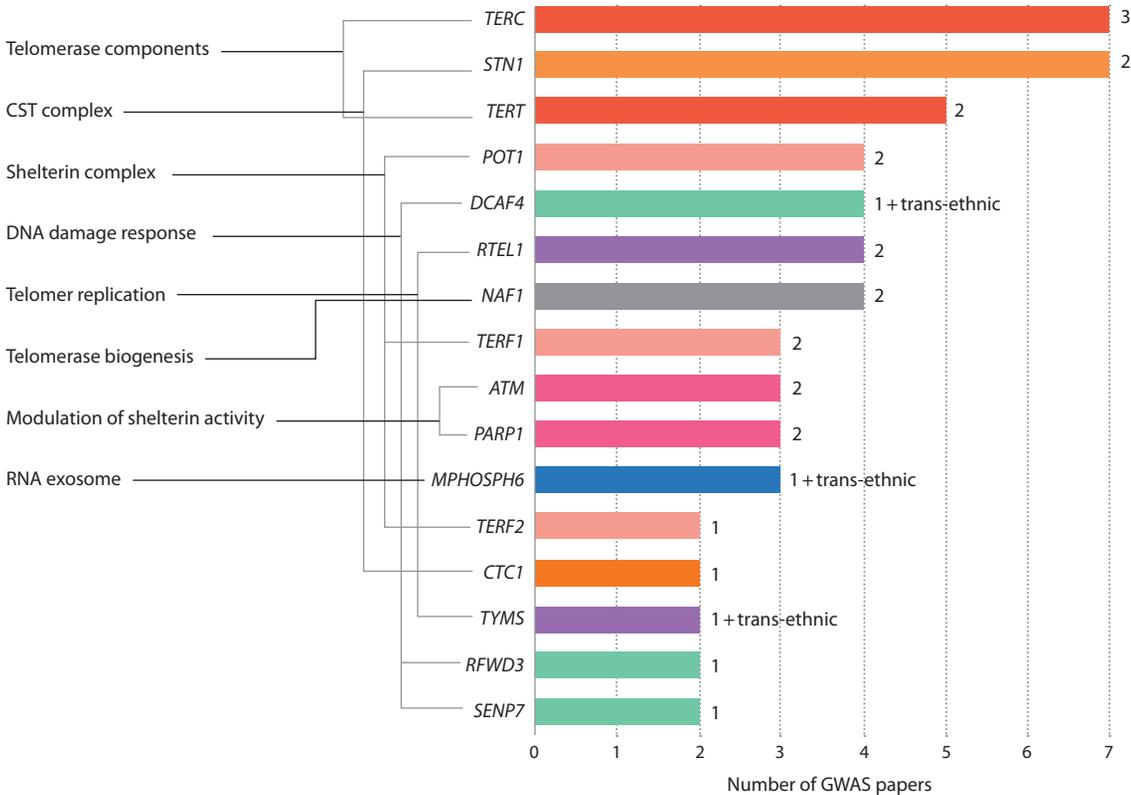


Fig. 5. Genes revealed in at least two GWAS papers.

Colors of columns indicate functional groups the genes belong to. The numbers on the right of columns represent the number of ethnically diverse GWAS population samples, in which the genes were identified. Trans-ethnic is the group consisting of Singaporean Chinese and Europeans.

Material 4). These genes encode SMC6 which activates ALT (Potts, Yu, 2007) and its two inhibiting proteins: ATRX with chromatin remodeling activity and SLX4 endonuclease (Sobinoff, Pickett, 2017).

Genes encoding DNA damage response proteins: (1) peptidase *SENP7* (Li et al., 2020); (2) chaperone protein *BAG6* (Li et al., 2020); (3) *DCAF4* interacting with *CUL4-DDB1* ligase (Mangino et al., 2015); (4) *RFWD3* interacting with RPA protein (replication protein A) (Li et al., 2020).

Genes encoding subunits of RNA exosomes: *EXOSC6*, *EXOSC9* (Codd et al., 2021) and *MPHOSPH6* (Dorajoo et al., 2019). These proteins are functionally significant, because it is known that *TERC* may be subjected to 3'-processing, and the RNA-exosomes are involved in this process (Tseng et al., 2015).

Human candidate genes identified in more than one study

As mentioned above, we have analyzed 18 papers on identifying telomere length-associated human genome loci based on GWA studies and collected the data on 270 such genes (see Suppl. Material 3). Notably, only 16 genes were identified in at least two studies (Fig. 5).

The most frequently identified genes were the ones encoding both telomerase components (*TERC* and *TERT*) and *STN1* encoding a component of the CST complex (revealed in 7, 5, and 7 studies, respectively). Three genes *POT1*, *TERF1*,

and *TERF2* encoding components of the shelterin complex were mentioned in 4, 3, and 2 publications, respectively. Three more genes *DCAF4*, *RTEL1*, and *NAF1* controlling the DNA damage response, telomere replication, and telomerase biogenesis were identified in four studies. *ATM*, *PARP1*, *MPHOSPH6*, *RFWD3*, *SENP7*, and *TYMS* were identified in 3 or 2 papers.

Most of 16 genes listed above were identified in population samples of different ethnic origin: (1) *TERC* in three ethnic groups, namely Europeans, Bangladeshis, and Singaporean Chinese; (2) *DCAF4*, *MPHOSPH6*, and *TYMS* in Europeans and as a result of the trans-ethnic meta-analyses (Singaporean Chinese+Europeans); (3) *TERT*, *STN1*, *POT1*, *RTEL1*, *NAF1*, *TERF1*, *ATM*, *PARP1* in two ethnic groups, namely Europeans and Singaporean Chinese.

Identification of the genes related to telomere length regulation according to DAVID

Using DAVID, we found the terms from the GOTERM_BP_DIRECT dictionary that were significantly (FDR < 0.05) associated with the list of 270 human genes presented in Suppl. Material 3. Sixteen terms indicating biological processes that directly control telomere length are presented in Suppl. Material 6, and the remaining fifteen terms are listed in Suppl. Material 7. There were 30 genes associated with the terms from the first group (see Suppl. Material 6), with two of them

(*SIRT6* and *TP53*) previously not recognized as biologically interpretable (these genes were presented in (Codd et al., 2021) without comment on their functional significance in the context of telomere length regulation). The analysis of scientific papers showed that proteins encoded by both genes can function in the subtelomeric regions of chromosomes (Tennen et al., 2021; Tutton et al., 2016), which means they could be indirectly involved in telomere length regulation.

Then, the genes associated with the second group of GOTERM_BP_DIRECT terms identified at FDR < 0.05 (see Suppl. Material 7) were analyzed. Among them, 29 genes were found that had no biological interpretation (highlighted in red in Suppl. Material 7, and listed in Suppl. Material 8). This group of 29 genes included the above mentioned *SIRT6* and *TP53*, as well as *BRCA1*, *SAMHD1*, and *BRCC3* associated with the maximum number of GO terms (six, four, and four, respectively). Apparently, the genes from the list thus obtained may also be of interest in the context of telomere length regulation.

Telomere length-associated genes found in other animal species

Genome-wide search for telomere length-associated loci and genes was carried out in three animal species: cattle (*Bos taurus*), sparrows (*Passer domesticus*), and nematodes (*Caenorhabditis elegans*).

A GWA study to investigate the species *Bos taurus* was carried out on 702 animals of the Holstein–Friesian breed (Ilska-Warner et al., 2019). The study of the DNA isolated from the whole blood of cows sampled at birth showed six telomere length-associated polymorphic loci, and three additional loci were identified when analyzing the DNA from blood samples at the first lactation. An analysis of the quantitative trait loci (QTL) corresponding to the identified genetic variants revealed 14 candidate genes at birth and 9 at the first lactation (see the Table and Suppl. Material 9). The authors were unable to find any data on direct involvement of the identified genes in processes associated with telomere length regulation.

NUP93 nucleoporin gene encoding a nuclear pore component was considered a potential regulator, because it was shown earlier in yeast that nucleoporins facilitated silencing of genes in proximity of telomeric regions (Van de Vosse et al., 2013).

The recently published results of GWA study (Pepke et al., 2021) in house sparrow (*Passer domesticus*) nestlings made it possible to identify 22 candidate genes (see the Table and Suppl. Material 10). According to the authors, the genes of interest in the context of telomere length regulation seem to be as follows: (1) *WNT9B* encoding a protein component of Wnt/β-catenin signaling pathway due to β-catenin involvement in *Tert* activation in embryonic stem cells of mice; (2) *CDCA4*, *GH*, and *GHRHR* regulating cell proliferation, apoptosis, and body growth; (3) *RHOF* involved in cytoskeletal organization; (4) *RNF34* (E3 ubiquitin-protein ligase RNF34) regulating ubiquitination; (5) *AQP1* due to involvement of aquaporin protein in transport of nitrogen oxide and active forms of oxygen, which increases oxidative stress which can in turn affect telomerase activity; (6) *SCN4A*, because its expression in human stem cells correlates with telomere length.

Our analysis showed that none of the candidate genes identified in cattle (23 genes) and house sparrows (22 genes) (see the Table) had orthologs among the 270 genes identified based on GWA studies in humans and presented in Suppl. Material 3.

The study in *C. elegans* (Cook et al., 2016) produced 9 candidate genes (see the Table and Suppl. Material 11). One out of nine genes, *pot-2*, is orthologous to *POT1* encoding a shelterin complex component in humans. The authors assume that another gene *ZK1127.4* may also be involved in telomere length regulation, because BCCIP encoded by an orthologous human gene interacts with *BRCA2* involved in DNA replication.

In general, when comparing sets of candidate genes identified in humans and three other animal species, almost no orthologous genes are detected, which may be due to species-specific features of telomere length regulation, some peculiarities of regulation at various ontogenetic stages, and differences in sampled tissues or gender of the studied individuals.

Telomere length-associated animal genes (see Suppl. Materials 9–11 for additional data)

Species	Method / DNA source / Reference	Candidate genes
<i>Bos taurus</i> (female Holstein–Friesian cattle)	GWAS / whole blood of female cattle / (Ilska-Warner et al., 2019)	<u>At birth:</u> <i>NUP93</i> , <i>CCSER1</i> , <i>MMRN1</i> , <i>SNCA</i> , <i>GPRIN3</i> , <i>HDGFL1</i> , <i>RF00026</i> , <i>DOK6</i> , <i>RF00001</i> , <i>CCDC102B</i> , <i>TMX3</i> , <i>DSEL</i> , <i>bta-mir-138-2</i> , <i>bta-mir-2284c</i> <u>In first lactation:</u> <i>PTPRD</i> , <i>CYTL1</i> , <i>MSX1</i> , <i>STX18</i> , <i>NSG1</i> , <i>ACOX3</i> , <i>TRMT44</i> , <i>CPZ</i> , <i>HMX1</i>
<i>Passer domesticus</i> (house sparrow)	GWAS / whole blood of nestlings aged 5–14 days / (Pepke et al., 2021)	<i>FRMD4B</i> , <i>LMOD3</i> , <i>ARL6IP5</i> , <i>UBA3</i> , <i>TMF1</i> , <i>EOGT</i> , <i>AQP1</i> , <i>GHRHR</i> , <i>OXR1</i> , <i>ORAI1</i> , <i>MORN3</i> , <i>KDM2B</i> , <i>RNF34</i> , <i>TMEM120B</i> , <i>RHOF</i> , <i>ANAPC5</i> , <i>SHCBP1</i> , <i>CDCA4</i> , <i>SCN4A</i> , <i>GH</i> , <i>GOSR2</i> , <i>WNT9B</i>
<i>Caenorhabditis elegans</i> (soil-inhabiting nematode)	GWAS using genome-wide sequencing data / cells of whole-body nematodes / (Cook et al., 2016)	<i>pot-2</i> (<i>POT1</i>)*, <i>mms-19</i> (<i>MMS19</i>), <i>ZK1127.4</i> (<i>BCCIP</i>), <i>ZC487.2</i> , <i>srd-35</i> , <i>T06D8.3</i> (<i>PLPPR1</i> , <i>PLPPR5</i>), <i>ZK783.5</i> , <i>F58F6.3</i> , <i>C12D5.10</i>

* *C. elegans* genes are cited with human orthologs in parentheses.

Conclusions

In the present paper, a compilation of telomere length-associated genes identified based on GWA studies and including the data on 270 human genes (see Suppl. Material 3), as well as 23, 22, and nine genes identified in cattle, house sparrow, and nematode (see the Table) is presented. The analysis of functions of 52 human genes with functional interpretation available (see Fig. 4, Suppl. Material 4) showed that telomere length may be affected by variants of genes encoding: (1) the structural components of telomerase; (2) the protein components of telomeric chromosome regions (shelterin complex and CST complex); (3) the proteins involved in telomerase biogenesis and regulating its activity; (4) the proteins regulating functional activity of shelterin subunits; (5) the proteins involved in telomere replication and/or capping; (6) the proteins controlling the alternative telomere lengthening pathway; (7) DNA damage response and repair proteins; (8) RNA exosome components.

Candidate human genes identified by several research groups in population samples of different ethnic origin are determined: genes encoding telomerase components (*TERC* and *TERT*) and *STN1* encoding a subunit of CST complex (see Fig. 5). It seems that polymorphic loci that affect the functions of these genes can potentially be the most reliable predisposition markers for telomere length-associated diseases.

Comparison of the data obtained from GWA studies in humans (see Suppl. Material 3) with the results of similar experiments obtained for other animal species (see the Table and Suppl. Materials 9–11) confirmed and expanded the understanding of the complex polygenic nature of telomere length regulation. In addition, a pair of orthologous genes encoding a shelterin protein (*POT1* in humans and *pot-2* in *C. elegans*) was identified; this finding demonstrates the high biological significance of this gene in various species.

Systematized data on genes and their functions may lay the foundation for development of prognostic criteria for human pathologies explicitly associated with telomere length. In addition, the data on biological processes affecting telomere length and genes regulating these processes may be used for marker-assisted and genomic selection of the farm animals aimed at increasing the duration of their productive lifetime.

References

Aviv A., Shay J.W. Reflections on telomere dynamics and ageing-related diseases in humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2018;373(1741):20160436. DOI 10.1098/rstb.2016.0436.

Bichet C., Bouwhuis S., Bauch C., Verhulst S., Becker P.H., Vedder O. Telomere length is repeatable, shortens with age and reproductive success, and predicts remaining lifespan in a long-lived seabird. *Mol. Ecol.* 2020;29(2):429-441. DOI 10.1111/mec.15331.

Bizarro J., Bhardwaj A., Smith S., Meier U.T. Nopp140-mediated concentration of telomerase in Cajal bodies regulates telomere length. *Mol. Biol. Cell.* 2019;30(26):3136-3150. DOI 10.1091/mbc.E19-08-0429.

Chen L., Roake C.M., Galati A., Bavasso F., Micheli E., Saggio I., Schoefner S., Cacchione S., Gatti M., Artandi S.E., Raffa G.D. Loss of human TGS1 hypermethylase promotes increased telomerase

RNA and telomere elongation. *Cell Rep.* 2020;30(5):1358-1372.e5. DOI 10.1016/j.celrep.2020.01.004.

Chen L.Y., Redon S., Lingner J. The human CST complex is a terminator of telomerase activity. *Nature.* 2012;488(7412):540-544. DOI 10.1038/nature11269.

Chik H.Y.J., Sparks A.M., Schroeder J., Dugdale H.L. A meta-analysis on the heritability of vertebrate telomere length. *J. Evol. Biol.* 2022;35(10):1283-1295. DOI 10.1111/jeb.14071.

Claussnitzer M., Dankel S.N., Kim K.-H., Quon G., Meuleman W., Haugen C., Glunk V., Sousa I.S., Beaudry J.L., Puvion-Randall V., Abdennur N.A., Liu J., Svensson P.-A., Hsu Y.-H., Drucker D.J., Mellgren G., Hui C.-Ch., Hauner H., Kellis M. FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 2015;373:895-907. DOI 10.1056/NEJMoa1502214.

Codd V., Nelson C.P., Albrecht E., Mangino M., Deelen J., Buxton J.L., Hottenga J.J., Fischer K., Esko T., Surakka I., ... Boomsma D.I., Jarvelin M.R., Slagboom P.E., Thompson J.R., Spector T.D., van der Harst P., Samani N.J. Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* 2013;45(4):422-427. DOI 10.1038/ng.2528.

Codd V., Wang Q., Allara E., Musicha C., Kaptoge S., Stoma S., Jiang T., Hamby S.E., Braund P.S., Bountziouka V., Budgeon C.A., Denniff M., Swinfield C., Papakonstantinou M., Sheth S., Natus D.E., Warner S.C., Wang M., Khera A.V., Eales J., Ouwehand W.H., Thompson J.R., Di Angelantonio E., Wood A.M., Butterworth A.S., Danesh J.N., Nelson C.P., Samani N.J. Polygenic basis and biomedical consequences of telomere length variation. *Nat. Genet.* 2021;53(10):1425-1433. DOI 10.1038/s41588-021-00944-6.

Cook D.E., Zdravljec S., Tanny R.E., Seo B., Riccardi D.D., Noble L.M., Rockman M.V., Alkema M.J., Braendle C., Kammenga J.E., Wang J., Kruglyak L., Félix M.A., Lee J., Andersen E.C. The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics.* 2016;204(1):371-383. DOI 10.1534/genetics.116.191148.

Crocco P., De Rango F., Dato S., Rose G., Passarino G. Telomere length as a function of age at population level parallels human survival curves. *Ageing (Albany NY).* 2021;13(1):204-218. DOI 10.18632/aging.202498.

de Lange T. Shelterin-mediated telomere protection. *Annu. Rev. Genet.* 2018;52:223-247. DOI 10.1146/annurev-genet-032918-021921.

Diotti R., Loayza D. Shelterin complex and associated factors at human telomeres. *Nucleus.* 2011;2(2):119-135. DOI 10.4161/nucl.2.2.15135.

Dorajoo R., Chang X., Gurung R.L., Li Z., Wang L., Wang R., Beckman K.B., Adams-Haduch J., M Y., Liu S., Meah W.Y., Sim K.S., Lim S.C., Friedlander Y., Liu J., van Dam R.M., Yuan J.M., Koh W.P., Khor C.C., Heng C.K. Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat. Commun.* 2019;10(1):2491. DOI 10.1038/s41467-019-10443-2.

Egan E.D., Collins K. Biogenesis of telomerase ribonucleoproteins. *RNA.* 2012;18(10):1747-1759. DOI 10.1261/rna.034629.112.

Fan H.C., Chang F.W., Tsai J.D., Lin K.M., Chen C.M., Lin S.Z., Liu C.A., Harn H.J. Telomeres and cancer. *Life (Basel).* 2021;11(12):1405. DOI 10.3390/life11121405.

Ilska-Warner J.J., Psifidi A., Seeker L.A., Wilbourn R.V., Underwood S.L., Fairlie J., Whitelaw B., Nussey D.H., Coffey M.P., Banos G. The genetic architecture of bovine telomere length in early life and association with animal fitness. *Front. Genet.* 2019;10:1048. DOI 10.3389/fgene.2019.01048.

Jafri M.A., Ansari S.A., Alqahtani M.H., Shay J.W. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med.* 2016;8(1):69. DOI 10.1186/s13073-016-0324-x.

- Joyce B.T., Zheng Y., Nannini D., Zhang Z., Liu L., Gao T., Kocherginsky M., Murphy R., Yang H., Achenbach C.J., Roberts L.R., Hoxha M., Shen J., Vokonas P., Schwart J., Baccarelli A., Hou L. DNA methylation of telomere-related genes and cancer risk. *Cancer Prev. Res. (Phila)*. 2018;11(8):511-522. DOI 10.1158/1940-6207.CAPR-17-0413.
- Kong C.M., Lee X.W., Wang X. Telomere shortening in human diseases. *FEBS J.* 2013;280(14):3180-3193. DOI 10.1111/febs.12326.
- Kroustallaki P., Lirussi L., Carracedo S., You P., Esbensen Q.Y., Götz A., Jobert L., Alsøe L., Sætrom P., Gagos S., Nilsen H. SMUG1 promotes telomere maintenance through telomerase RNA processing. *Cell Rep.* 2019;28(7):1690-1702.e10. DOI 10.1016/j.celrep.2019.07.040.
- Li C., Stoma S., Lotta L.A., Warner S., Albrecht E., Allione A., Arp P.P., Broer L., Buxton J.L., Da Silva Couto Alves A., ... Butterworth A.S., Danesh J., Samani N.J., Wareham N.J., Nelson C.P., Langenberg C., Codd V. Genome-wide association analysis in humans links nucleotide metabolism to leukocyte telomere length. *Am. J. Hum. Genet.* 2020;106(3):389-404. DOI 10.1016/j.ajhg.2020.02.006.
- Mangino M., Christiansen L., Stone R., Hunt S.C., Horvath K., Eisenberg D.T., Kimura M., Petersen I., Kark J.D., Herbig U., Reiner A.P., Benetos A., Codd V., Nyholt D.R., Sinnreich R., Christensen K., Nasr H., Hwang S.J., Levy D., Bataille V., Fitzpatrick A.L., Chen W., Berenson G.S., Samani N.J., Martin N.G., Tishkoff S., Schork N.J., Kyvik K.O., Dalgård C., Spector T.D., Aviv A. *DCAF4*, a novel gene associated with leukocyte telomere length. *J. Med. Genet.* 2015; 52(3):157-162. DOI 10.1136/jmedgenet-2014-102681.
- Mangino M., Richards J.B., Soranzo N., Zhai G., Aviv A., Valdes A.M., Samani N.J., Deloukas P., Spector T.D. A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length. *J. Med. Genet.* 2009;46(7):451-454. DOI 10.1136/jmg.2008.064956.
- Monaghan P., Ozanne S.E. Somatic growth and telomere dynamics in vertebrates: relationships, mechanisms and consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2018;373(1741):20160446. DOI 10.1098/rstb.2016.0446.
- Nagpal N., Wang J., Zeng J., Lo E., Moon D.H., Luk K., Braun R.O., Burroughs L.M., Keel S.B., Reilly C., Lindsley R.C., Wolfe S.A., Tai A.K., Cahan P., Bauer D.E., Fong Y.W., Agarwal S. Small-molecule PAPD5 inhibitors restore telomerase activity in patient stem cells. *Cell Stem Cell.* 2020;26(6):896-909.e8. DOI 10.1016/j.stem.2020.03.016.
- Nersisyan L., Nikoghosyan M., Genome of the Netherlands consortium, Arakelyan A. WGS-based telomere length analysis in Dutch family trios implicates stronger maternal inheritance and a role for *RRM1* gene. *Sci. Rep.* 2019;9(1):18758. DOI 10.1038/s41598-019-55109-7.
- Oh W., Ghim J., Lee E.W., Yang M.R., Kim E.T., Ahn J.H., Song J. PML-IV functions as a negative regulator of telomerase by interacting with TERT. *J. Cell Sci.* 2009;122(Pt.15):2613-2622. DOI 10.1242/jcs.048066.
- Ohira T., Kojima H., Kuroda Y., Aoki S., Inaoka D., Osaki M., Wanibuchi H., Okada F., Oshimura M., Kugoh H. PITX1 protein interacts with ZCCHC10 to regulate hTERT mRNA transcription. *PLoS One.* 2019;14(8):e0217605. DOI 10.1371/journal.pone.0217605.
- Pepke M.L., Kvalnes T., Lundregan S., Boner W., Monaghan P., Saezther B.E., Jensen H., Ringsby T.H. Genetic architecture and heritability of early-life telomere length in a wild passerine. *Mol. Ecol.* 2021. DOI 10.1111/mec.16288.
- Podlevsky J.D., Bley C.J., Omana R.V., Qi X., Chen J.J. The telomerase database. *Nucleic Acids Res.* 2008;36(Database issue):D339-D343. DOI 10.1093/nar/gkm700.
- Potts P.R., Yu H. The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nat. Struct. Mol. Biol.* 2007;14(7):581-590. DOI 10.1038/nsmb1259.
- Saxena R., Bjonnes A., Prescott J., Dib P., Natt P., Lane J., Lerner M., Cooper J.A., Ye Y., Li K.W., Maubaret C.G., Codd V., Brackett D., Mirabello L., Kraft P., Dinney C.P., Stowell D., Peyton M., Ralhan S., Wander G.S., Mehra N.K., Salpea K.D., Gu J., Wu X., Mangino M., Hunter D.J., De Vivo I., Humphries S.E., Samani N.J., Spector T.D., Savage S.A., Sanghera D.K. Genome-wide association study identifies variants in casein kinase II (*CSNK2A2*) to be associated with leukocyte telomere length in a Punjabi Sikh diabetic cohort. *Circ. Cardiovasc. Genet.* 2014;7(3):287-295. DOI 10.1161/CIRCGENETICS.113.000412.
- Schrumpfová P.P., Fajkus J. Composition and function of telomerase-A polymerase associated with the origin of eukaryotes. *Biomolecules.* 2020;10(10):1425. DOI 10.3390/biom10101425.
- Seeker L.A., Underwood S.L., Wilbourn R.V., Dorrens J., Froy H., Holland R., Ilska J.J., Psifidi A., Bagnall A., Whitelaw B., Coffey M., Banos G., Nussey D.H. Telomere attrition rates are associated with weather conditions and predict productive lifespan in dairy cattle. *Sci. Rep.* 2021;11(1):5589. DOI 10.1038/s41598-021-84984-2.
- Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C., Imamichi T., Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 2022;50(W1):W216-221. DOI 10.1093/nar/gkac194.
- Snow B.E., Erdmann N., Cruickshank J., Goldman H., Gill R.M., Robinson M.O., Harrington L. Functional conservation of the telomerase protein Est1p in humans. *Curr. Biol.* 2003;13(8):698-704. DOI 10.1016/s0960-9822(03)00210-0.
- Sobinoff A.P., Pickett H.A. Alternative lengthening of telomeres: DNA repair pathways converge. *Trends Genet.* 2017;33(12):921-932. DOI 10.1016/j.tig.2017.09.003.
- Sobinoff A.P., Pickett H.A. Mechanisms that drive telomere maintenance and recombination in human cancers. *Curr. Opin. Genet. Dev.* 2020;60:25-30. DOI 10.1016/j.gde.2020.02.006.
- Stuparević I., Novačić A., Rahmouni A.R., Fernandez A., Lamb N., Primig M. Regulation of the conserved 3'-5' exoribonuclease EXOSC10/Rp6 during cell division, development and cancer. *Biol. Rev. Camb. Philos. Soc.* 2021;96(4):1092-1113. DOI 10.1111/brv.12693.
- Tennen R.I., Bua D.J., Wright W.E., Chua K.F. SIRT6 is required for maintenance of telomere position effect in human cells. *Nat. Commun.* 2011;2:433. DOI 10.1038/ncomms1443.
- Tseng C.K., Wang H.F., Burns A.M., Schroeder M.R., Gaspari M., Baumann P. Human telomerase RNA processing and quality control. *Cell Rep.* 2015;13(10):2232-2243. DOI 10.1016/j.celrep.2015.10.075.
- Tutton S., Azzam G.A., Stong N., Vladimirova O., Wiedmer A., Monteith J.A., Beishline K., Wang Z., Deng Z., Riethman H., McMahon S.B., Murphy M., Lieberman P.M. Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J.* 2016;35(2):193-207. DOI 10.15252/embj.201490880.
- Van de Vosse D.W., Wan Y., Lapetina D.L., Chen W.M., Chiang J.H., Aitchison J.D., Wozniak R.W. A role for the nucleoporin Nup170p in chromatin structure and gene silencing. *Cell.* 2013;152(5):969-983. DOI 10.1016/j.cell.2013.01.049.
- Vera E., Bernardes de Jesus B., Foronda M., Flores J.M., Blasco M.A. The rate of increase of short telomeres predicts longevity in mammals. *Cell Rep.* 2012;2(4):732-737. DOI 10.1016/j.celrep.2012.08.023.

Wilbourn R.V., Moatt J.P., Froy H., Walling C.A., Nussey D.H., Boonekamp J.J. The relationship between telomere length and mortality risk in non-model vertebrate systems: a meta-analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2018;373(1741):20160447. DOI 10.1098/rstb.2016.0447.

Wright W.E., Piatyszek M.A., Rainey W.E., Byrd W., Shay J.W. Telomerase activity in human germline and embryonic tissues and cells. *Dev. Genet.* 1996;18(2):173-179. DOI 10.1002/(SICI)1520-6408(1996)18:2<173::AID-DVG10>3.0.CO;2-3.

Zeiger A.M., White M.J., Eng C., Oh S.S., Witonsky J., Goddard P.C., Contreras M.G., Elhawary J.R., Hu D., Mak A.C.Y., Lee E.Y., Keys K.L., Samedy L.A., Risse-Adams O., Magaña J., Huntsman S., Salazar S., Davis A., Meade K., Brigino-Buenaventura E., LeNoir M.A., Farber H.J., Bibbins-Domingo K., Borrell L.N., Burchard E.G. Genetic determinants of telomere length in African American youth. *Sci. Rep.* 2018;8(1):13265. DOI 10.1038/s41598-018-31238-3.

ORCID ID

E.V. Ignatieva orcid.org/0000-0002-8588-6511

N.S. Yudin orcid.org/0000-0002-1947-5554

D.M. Larkin orcid.org/0000-0001-7859-6201

Acknowledgements. The work was supported by the Russian Scientific Foundation, Project No. 22-26-00143, <https://rscf.ru/project/22-26-00143/>.

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

Received November 17, 2022. Revised December 4, 2022. Accepted December 5, 2022.