

Научный рецензируемый журнал

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.

Периодичность 8 выпусков в год

DOI 10.18699/VJGB-22-40

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-22-40

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

Молекулярная биология

- 341 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Характеристика деметилирующей ДНК-гликозилазы ROS1 из *Nicotiana tabacum* L. Д.В. Петрова, Н.В. Пермякова, И.Р. Грин, Д.О. Жарков

Генетика и селекция растений

- 349 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Изучение генетического полиморфизма российских сортов рапса и сурепицы с использованием SSR- и SRAP-маркеров. И.А. Клименко, В.Т. Воловик, А.А. Антонов, В.А. Душкин, А.О. Шамустакимова, Ю.М. Мавлютов

- 359 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Идентификация гена, кодирующего альбумин семян SCA, в геноме гороха (*Pisum* L.). А.В. Мглинец, В.С. Богданова, О.Э. Костерин (на англ. языке)

Генетика животных

- 365 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Плотность нейронов в коре головного мозга и гиппокампе мышей линии C1snt2-KO – модели расстройств аутистического спектра. И.Н. Рожкова, С.В. Окотруб, Е.Ю. Брусенцев, Е.Е. Ульданова, Э.А. Чуйко, Т.В. Липина, Т.Г. Амстиславская, С.Я. Амстиславский

- 371 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Наследственная предрасположенность водяных полевок (*Arvicola amphibius* L.) к судорожным припадкам в ответ на хэндлинг и ее связь с продолжительностью жизни. Г.Г. Назарова, Л.П. Проскурняк

378

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Платформа GWAS-MAPlovis для хранения и анализа результатов полногеномных ассоциативных исследований овец. А.В. Кириченко, А.С. Злобин, Т.И. Шашкова, Н.А. Волкова, Б.С. Иолчиев, В.А. Багиров, П.М. Бородин, Л.С. Карссен, Я.А. Цепилов, Ю.С. Аульченко

Биотехнология микроорганизмов

- 385 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Таксономическая структура бактериальных сообществ в хлебных заквасках спонтанного брожения. В.К. Хлесткин, М.Н. Локачук, О.А. Савкина, Л.И. Кузнецова, Е.Н. Павловская, О.И. Парахина

Тест-системы и вакцинопрофилактика

- 394 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Стабильность генома вакцинного штамма VACΔ6. Р.А. Максюттов, С.Н. Якубицкий, И.В. Колосова, Т.В. Трегубчак, А.Н. Швалов, Е.В. Гаврилова, С.Н. Щелкунов

402

ОБЗОР

Создание трансгенных мышей, восприимчивых к коронавирусам: платформа изучения вирусного патогенеза и тестирования вакцин. Н.Р. Баттулин, О.Л. Серов

Molecular biology

- 341 ORIGINAL ARTICLE
Characterization of demethylating DNA glycosylase ROS1 from *Nicotiana tabacum* L. D.V. Petrova, N.V. Permyakova, I.R. Grin, D.O. Zharkov

Plant genetics and breeding

- 349 ORIGINAL ARTICLE
Investigation of genetic polymorphism of Russian rape and turnip rape varieties using SSR and SRAP markers. I.A. Klimenko, V.T. Volovik, A.A. Antonov, V.A. Dushkin, A.O. Shamustakimova, Yu.M. Mavlyutov
- 359 ORIGINAL ARTICLE
Identification of the gene coding for seed cotyledon albumin SCA in the pea (*Pisum* L.) genome. A.V. Mglinets, V.S. Bogdanova, O.E. Kosterin

Animal genetics

- 365 ORIGINAL ARTICLE
Neuronal density in the brain cortex and hippocampus in *Clstn2*-KO mouse strain modeling autistic spectrum disorder. I.N. Rozhkova, S.V. Okotrub, E.Yu. Brusentsev, E.E. Uldanova, E.A. Chuyko, T.V. Lipina, T.G. Amstislavskaya, S.Ya. Amstislavsky
- 371 ORIGINAL ARTICLE
Hereditary predisposition of water voles (*Arvicola amphibius* L.) to seizures in response to handling. G.G. Nazarova, L.P. Proskurnyak

- 378 ORIGINAL ARTICLE
The GWAS-MAP|ovis platform for aggregation and analysis of genome-wide association study results in sheep. A.V. Kirichenko, A.S. Zlobin, T.I. Shashkova, N.A. Volkova, B.S. Iolchiev, V.A. Bagirov, P.M. Borodin, L.C. Karssen, Y.A. Tsepilov, Y.S. Aulchenko

Microbial biotechnology

- 385 ORIGINAL ARTICLE
Taxonomic structure of bacterial communities in sourdoughs of spontaneous fermentation. V.K. Khlestkin, M.N. Lockachuk, O.A. Savkina, L.I. Kuznetsova, E.N. Pavlovskaya, O.I. Parakhina

Test kits and preventive vaccination

- 394 ORIGINAL ARTICLE
Genome stability of the vaccine strain VACΔ6. R.A. Maksyutov, S.N. Yakubitskiy, I.V. Kolosova, T.V. Tregubchak, A.N. Shvalov, E.V. Gavrilova, S.N. Shchelkunov
- 402 REVIEW
Creation of transgenic mice susceptible to coronaviruses: a platform for studying viral pathogenesis and testing vaccines. N.R. Battulin, O.L. Serov

Original Russian text www.bionet.nsc.ru/vogis/

Characterization of demethylating DNA glycosylase ROS1 from *Nicotiana tabacum* L.

D.V. Petrova¹, N.V. Permyakova², I.R. Grin¹, D.O. Zharkov^{1, 3} 

¹ Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

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

³ Novosibirsk State University, Novosibirsk, Russia

 dzharkov@niboch.nsc.ru

Abstract. One of the main mechanisms of epigenetic regulation in higher eukaryotes is based on the methylation of cytosine at the C5 position with the formation of 5-methylcytosine (mC), which is further recognized by regulatory proteins. In mammals, methylation mainly occurs in CG dinucleotides, while in plants it targets CG, CHG, and CHH sequences (H is any base but G). Correct maintenance of the DNA methylation status is based on the balance of methylation, passive demethylation, and active demethylation. While in mammals active demethylation is based on targeted regulated damage to mC in DNA followed by the action of repair enzymes, demethylation in plants is performed by specialized DNA glycosylases that hydrolyze the *N*-glycosidic bond of mC nucleotides. The genome of the model plant *Arabidopsis thaliana* encodes four paralogous proteins, two of which, DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1), possess 5-methylcytosine-DNA glycosylase activity and are necessary for the regulation of development, response to infections and abiotic stress and silencing of transgenes and mobile elements. Homologues of DME and ROS1 are present in all plant groups; however, outside *A. thaliana*, they are poorly studied. Here we report the properties of a recombinant fragment of the ROS1 protein from *Nicotiana tabacum* (NtROS1), which contains all main structural domains required for catalytic activity. Using homologous modeling, we have constructed a structural model of NtROS1, which revealed folding characteristic of DNA glycosylases of the helix-hairpin-helix structural superfamily. The recombinant NtROS1 protein was able to remove mC bases from DNA, and the enzyme activity was barely affected by the methylation status of CG dinucleotides in the opposite strand. The enzyme removed 5-hydroxymethylcytosine (hmC) from DNA with a lower efficiency, showing minimal activity in the presence of mC in the opposite strand. Expression of the NtROS1 gene in cultured human cells resulted in a global decrease in the level of genomic DNA methylation. In general, it can be said that the NtROS1 protein and other homologues of DME and ROS1 represent a promising scaffold for engineering enzymes to analyze the status of epigenetic methylation and to control gene activity.

Key words: epigenetic demethylation; 5-methylcytosine; 5-hydroxymethylcytosine; DNA glycosylases; REPRESSOR OF SILENCING 1; *Nicotiana tabacum*.

For citation: Petrova D.V., Permyakova N.V., Grin I.R., Zharkov D.O. Characterization of demethylating DNA glycosylase ROS1 from *Nicotiana tabacum* L. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(4):341-348. DOI 10.18699/VJGB-22-41

Характеристика деметилирующей ДНК-гликозилазы ROS1 из *Nicotiana tabacum* L.

Д.В. Петрова¹, Н.В. Пермякова², И.Р. Грин¹, Д.О. Жарков^{1, 3} 

¹ Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия

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

³ Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 dzharkov@niboch.nsc.ru

Аннотация. Один из главных механизмов эпигенетической регуляции у высших эукариот основан на метилировании цитозина по положению C5 с образованием 5-метилцитозина (mC), который далее узнается регуляторными белками. У млекопитающих метилирование преимущественно протекает в динуклеотидах CG, тогда как у растений его мишенью служат последовательности CG, CHG и CHH (H – любое основание, кроме G). Корректное поддержание статуса метилирования ДНК требует баланса процессов метилирования, пассивного и активного деметилирования. В то время как у млекопитающих активное деметилирование происходит за счет направленного регулируемого повреждения mC в ДНК с последующим действием ферментов репарации, у растений функции деметилирования выполняют специализированные ДНК-гликозилазы, гидролизующие *N*-гликозидную связь mC-нуклеотидов. Геном модельного растения *Arabidopsis thaliana* кодирует че-

тыре паралогичных белка, два из которых – DEMETER (DME) и REPRESSOR OF SILENCING 1 (ROS1) – обладают 5-метилцитозин-ДНК-гликозилазной активностью и необходимы для регуляции развития, ответа на инфекции и абиотический стресс и сайленсинга трансгенов и мобильных элементов. Гомологи DME и ROS1 присутствуют во всех группах растений, однако за пределами *A. thaliana* исследованы крайне слабо. В статье приведены результаты изучения свойств рекомбинантного фрагмента белка ROS1 из *Nicotiana tabacum* (NtROS1), содержащего основные структурные домены, необходимые для каталитической активности. Методами гомологичного моделирования была построена структурная модель NtROS1, в которой выявлена укладка, характерная для ДНК-гликозилаз структурного суперсемейства «спираль–шпилька–спираль». Рекомбинантный белок NtROS1 был способен удалять из ДНК основания мС, причем активность фермента слабо зависела от статуса метилирования СG-динуклеотидов в противоположной цепи. С меньшей эффективностью фермент удалял из ДНК 5-гидроксиметилцитозин (hmC), проявляя минимальную активность при наличии мС в противоположной цепи. При экспрессии гена *NtROS1* в клетках человека в культуре происходило глобальное снижение уровня метилирования геномной ДНК. В целом можно сказать, что белок NtROS1 и другие гомологи DME и ROS1 представляют собой многообещающую основу для инженерии ферментов с целью анализа статуса эпигенетического метилирования и управления активностью генов.

Ключевые слова: эпигенетическое деметилирование; 5-метилцитозин; 5-гидроксиметилцитозин; ДНК-гликозилазы; REPRESSOR OF SILENCING 1; *Nicotiana tabacum*.

Introduction

DNA methylation is a dedicated mechanism of gene regulation, especially developed in higher eukaryotes. The 5-methylcytosine (mC) nucleobase, formed by cytosine methylation at the C5 position, serves as a reversible epigenetic mark that plays an important role in the control of gene expression and the protection of the genome from mobile elements. DNA methylation occurs in a wide range of multicellular eukaryotes; however, its significance and functions in these organisms differ greatly (Lee et al., 2010; Zemach, Zilberman, 2010). For example, in mammals, methylation most often occurs in CpG dinucleotides, while in plants, a significant proportion of mC occurs in trinucleotides CHG and CHH (where H is “not G”). The consequences of mC for gene activity are mainly mediated by proteins containing methyl-binding domains that form complexes with histone deacetylases or themselves have the activity of histone-specific methyltransferases, chromatin remodeling factors, etc., which leads to chromatin condensation and transcription suppression (Ballestar, Wolffe, 2001; Baubec et al., 2013). Recent studies have shown that an oxidized derivative of 5-methylcytosine, 5-hydroxymethylcytosine (hmC), also plays an epigenetic role in the mammalian genome (Branco et al., 2011). Unlike mC, hmC is enriched in promoters and bodies of actively expressed genes and is considered an activating epigenetic marker (Pastor et al., 2011; Yu et al., 2012).

Correct methylation of various sites in the genome is extremely important, since the transcriptional activity of genes depends on it. Errors in DNA methylation can have grave consequences. In particular, in humans, global DNA demethylation or hypermethylation of tumor suppressor genes serve as cancer markers. Maintaining the status of DNA methylation in cells requires a balance of methylation and active and passive demethylation. The mechanisms of active demethylation of the genomic DNA in higher eukaryotes have only been discovered in the last decade. In mammals, active demethylation is initiated by regulated damage to mC, which can occur in two ways: either deamination of mC to T by AID/APOBEC enzymes, or oxidation of mC to hmC and further derivatives (5-formylcytosine and 5-carboxycytosine) by TET family dioxygenases (Pastor et al., 2013; Bochtler et al., 2017; Wu,

Zhang, 2017). The modified bases are further perceived by cellular repair systems as damaged and are removed by the DNA base excision repair pathway.

Unlike mammals, plants have unique enzymes that directly hydrolyze *N*-glycosidic bonds of mC nucleotides. These DNA glycosylases – DEMETER (DME) and REPRESSOR OF SILENCING 1 (ROS1, also known as DEMETER-LIKE 1 or DML1) (Choi Y. et al., 2002; Gong et al., 2002; Agius et al., 2006; Morales-Ruiz et al., 2006) – are involved in the regulation of the methylation status of the sites in the plant genome that determine gene imprinting during paternal or maternal inheritance and silence or activate specific promoters during plant development and stress response (Li Y. et al., 2018; Parrilla-Doblas et al., 2019; Roldán-Arjona et al., 2019). After removal of mC, the resulting apurinic-apyrimidinic site (AP site) is cleaved either by the enzyme’s own AP lyase activity or by the AP endonucleases APE1L or ARP, then a normal nucleotide is incorporated by one of the DNA polymerases, and the nick is ligated by the LIG1 DNA ligase. Interestingly, DME and ROS1 can also excise hmC, which is not regarded as an epigenetic base in plants (Jang et al., 2014). The localization of demethylation by DME/ROS1 is regulated by small RNAs that bind to the enzyme itself or to the protein complex that contains it (Penterman et al., 2007; Li X. et al., 2012).

In addition to DME and ROS1, the genome of the model plant *Arabidopsis thaliana* contains three more genes that are homologous to DME and ROS1: DEMETER-LIKE 2 (DML2), DEMETER-LIKE 3 (DML3), and AT3G47830, which has not been characterized thus far except for participation of DML2 and DML3 in maintaining correct DNA methylation (Ortega-Galisteo et al., 2008; Le et al., 2014). All these proteins belong to the DNA helix–hairpin–helix (HhH) structural superfamily of DNA glycosylases. Other members of this superfamily are involved in the removal of oxidized, alkylated and deaminated nucleobases from the genome (Zharkov, 2008; Fedorova et al., 2010). DME/ROS1 enzymes attract attention as potential tools for targeted regulation of gene activity: for example, the possibility of targeted DNA demethylation in human cells by *A. thaliana* ROS1 (AtROS1) fused to the RNA-guided Cas9 protein has been shown (Devesa-Guerra et al., 2020), and

A. thaliana DME (AtDME) was used to analyze the level of mC in genomic DNA (Choi W.L. et al., 2021).

In plants other than *A. thaliana*, there were few studies on the role of DME-like proteins in active epigenetic demethylation; some data exist for rice, wheat, barley, and tomato (Ono et al., 2012; Wen et al., 2012; Kapazoglou et al., 2013; Liu et al., 2015). In 2007, a ROS1 homolog from *Nicotiana tabacum* (NtROS1) was cloned, and the recombinant protein produced in insect cell culture was shown to cleave methylated tobacco genomic DNA (Choi C.-S., Sano, 2007). None of these studies included a detailed biochemical characterization of the protein. We have previously shown that the NtROS1 fragment corresponding to the minimal catalytically active AtROS1 fragment has the activity of 5-methylcytosine-DNA glycosylase (Gruber et al., 2018).

Here, in view of the potential value of plant demethylation enzymes as tools for genetic technologies, we characterize the substrate specificity of the recombinant NtROS1 catalytic fragment on mC and hmC in different contexts of methylated CpG dinucleotides and show that the expression of NtROS1 in human cells causes a global decrease in DNA methylation.

Materials and methods

ProtoScript II reverse transcriptase, Q5 Hot Start High-Fidelity DNA polymerase, *Escherichia coli* uracil DNA glycosylase, *Cla*I and *Sac*I restriction endonucleases were purchased from New England Biolabs (USA), and bacteriophage T4 polynucleotide kinase, from Biosan (Novosibirsk, Russia). Oligonucleotides listed in the Table were synthesized at the SB RAS ICBFM Laboratory of Biomedical Chemistry using commercially available phosphoramidites (Glen Research, USA). If necessary, the oligonucleotides were ³²P-labeled at the 5'-end using γ [³²P]ATP (SB RAS ICBFM Laboratory of Biotechnology) and T4 polynucleotide kinase.

To build a model of the NtROS1 catalytic domain in the Swiss-Model program (Waterhouse et al., 2018), the AtDME (AF-Q8LK56-F1-model_v1) and AtROS1 (AF-Q9SJQ6-F1-

model_v1) templates from the AlphaFold database (Jumper et al., 2021) were used.

To obtain a catalytically inactive NtROS1 with the Asp1359Asn substitution, a pLATE31 plasmid with an insert encoding the catalytically active NtROS1 fragment (amino acid residues 754–1796) (Gruber et al., 2018) was mutagenized using primers D1359Nfwd and D1359Nrev (see the Table) and the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). The mutation was confirmed by Sanger sequencing. Wild-type NtROS1 and NtROS1 D1359N were overproduced and purified as described previously (Gruber et al., 2018).

To study the activity of NtROS1, double-stranded substrates were obtained by annealing oligonucleotides C1, C2, M1, M2, H1, and H2 (see the Table). The reaction mixture contained 50 nM substrate, 50 mM Tris–HCl (pH 8.0), 1 mM EDTA, 1 mM DTT, 0.1 % bovine serum albumin, and 100 nM NtROS1. The mixture was incubated at 37 °C; aliquots were withdrawn at various times (2–300 min) and mixed with an equal volume of the stop solution (80 % formamide, 20 mM EDTA, 0.1 % xylene cyanol, 0.1 % bromophenol blue). If necessary, the aliquots were preheated for 2 min at 95 °C in the presence of 0.1 M NaOH and neutralized with an equimolar amount of HCl. The reaction products were resolved by electrophoresis in a 20 % polyacrylamide gel containing 7.2 M urea, visualized by phosphorimaging using the Typhoon FLA 9500 system (GE Healthcare, USA), and quantified using the Quantity One v4.6.3 software (Bio-Rad Laboratories, USA). The apparent rate constants were determined using the SigmaPlot v11.0 software (Systat Software, USA) by non-linear regression to the equation $[P] = [P]_{\max}(1 - e^{-kt})$, where $[P]$ is product concentration, $[P]_{\max}$ is the maximum product concentration, k is the reaction rate constant, and t is time.

To assess the status of global DNA methylation upon NtROS1 expression in human cells, wild-type and D1359N NtROS1 coding sequences were cloned into the pIRES-eGFP-puro plasmid (Clontech, USA) at the *Sac*I and *Cla*I restriction

Oligonucleotides used in this work

ID	Sequence (5'→3')
Reverse transcription PCR	
RTPCRfwd	AGAAGGAGATATAACTATGTTCATTCATTAGAAGACGGAAACCG
RTPCRrev	GTGGTGGTGATGGTGATGGCCGTTTTTCATCTGGCTTTCCTTTAGTCC
Site-directed mutagenesis	
D1359Nfwd	CCTGTCAACACAACGTTGGC
D1359Nrev	GAAAGCAAGGTGGTGAAGTGT
Enzyme activity and specificity studies	
C1	GCTTGTACTTTAGCGCATTGATTCTCACCACG
C2	CGTGGTGAGAATCAATGCGCTAAAGTACAAGC
M1	GCTTGTACTTTAGMGCATTGATTCTCACCACG (M = mC)
M2	CGTGGTGAGAATCAATGMGCTAAAGTACAAGC (M = mC)
H1	GCTTGTACTTTAGHGCATTGATTCTCACCACG (H = hmC)
H2	CGTGGTGAGAATCAATGHGCTAAAGTACAAGC (H = hmC)

sites. HEK293 Phoenix cells ($1.2 \cdot 10^6$) were transfected with 5 μg of the plasmid by the calcium phosphate method and grown in a monolayer in DMEM with 10 % fetal calf serum (HyClone, USA). After 24 and 48 h, the medium was changed with the addition of 3 $\mu\text{g}/\text{ml}$ puromycin. Transfection efficiency was determined by flow cytometry (NovoCyte 3000, ACEA Biosciences, USA) by detection of the fluorescence of eGFP encoded by the same plasmid. NtROS1 expression in the transfected cells was confirmed by reverse transcription PCR using β -actin mRNA as a control. Genomic DNA was isolated from the cells ($5 \cdot 10^6$) using a QIAamp DNA Mini Kit (Qiagen, the Netherlands), and the relative content of mC was determined using anti-mC antibodies (MethylFlash Methylated DNA Quantification Kit, EpiGenetek, USA). The results were compared using Student's *t*-test.

Results and discussion

Plant mC-specific DNA glycosylases are proteins of considerable size: for example, AtDME and AtROS1, as well as their DML2 and DML3 paralogs, are over 1000 amino acid residues long (Fig. 1). The extended N-terminal regions of these polypeptides are unstructured, although some of their parts are necessary for enzyme activity. The C-terminal regions contain a conserved HhH catalytic domain and an iron-sulfur cluster (FeS cluster), characteristic of many DNA glycosylases that recognize oxidative DNA damage, as well as an RNA-binding motif (RNA Recognition Motif, RRM) and a CXXC type permuted zinc finger unique to the DME/ROS1 family (see Fig. 1, *a*). Unlike in all other HhH superfamily DNA glycosylases, the catalytic domain in DME/ROS1 proteins is disrupted by a long non-conserved insert (Ponferrada-Marrín et al., 2011). Based on the literature data on the AtROS1 protein, we previously cloned the NtROS1 cDNA fragment encoding amino acid residues 754–1796 (Gruber et al., 2018). This region contains all elements necessary for the catalytic activity in AtROS1 (Hong et al., 2014).

Since the structures of the DME/ROS1 family proteins are currently unknown, for a more detailed understanding of the

organization of the NtROS1 catalytic fragment, we carried out homology modeling based on the AtDME and AtROS1 models from the AlphaFold template collection (Jumper et al., 2021). The two resulting models were almost identical except for the structure of the long non-homologous regions. The disrupted HhH domain folded into the α -helical structure characteristic of DNA glycosylases of this superfamily, in which a DNA binding groove with the catalytic residues Lys1341 and Asp1359 is evident (see Fig. 1, *b*). In addition, several more peripheral α -helices buttress the HhH domain and the FeS cluster and are obviously important for maintaining their structure. The FeS cluster, zinc finger, and RRM motif form separate structural elements that leave free access to the DNA binding groove (see Fig. 1, *b*). All disordered regions of the structure are also located on the side of the protein globule opposite to the DNA binding groove.

To analyze the catalytic activity and substrate specificity of NtROS1, we performed a cleavage reaction of double-stranded oligonucleotides containing a CpG dinucleotide in which the cytosine was unmethylated, methylated, or hydroxymethylated in one or both chains (Fig. 2). The strand to be cleaved was ^{32}P -labeled at the 5'-end. NtROS1 showed virtually no activity on the substrate containing an unmethylated CpG site, which is consistent with the literature data claiming that the C5 position of the cytosine must carry a substitution to be cleaved by AtROS1 (Morales-Ruiz et al., 2006). Activity towards mC and hmC was observed; however, its level differed markedly for both types of substrates. Comparing the efficiency of cleavage of M1/C2 and M1/M2 substrates (see Fig. 2, lanes 5 and 8) with H1/C2 and H1/H2 substrates (lanes 11 and 14), one can see that the enzyme prefers to excise mC over hmC both from CpG sites modified at only one strand and from fully modified sites. Small differences in the cleavage of M1/C2, M1/M2 and M1/H2 substrates (lanes 5, 8, and 17) indicate that modifications in the complementary strand have little effect on the removal of mC, and hmC in the complementary strand might even increase the cleavage. The enzyme showed no activity against DNA substrates containing uracil

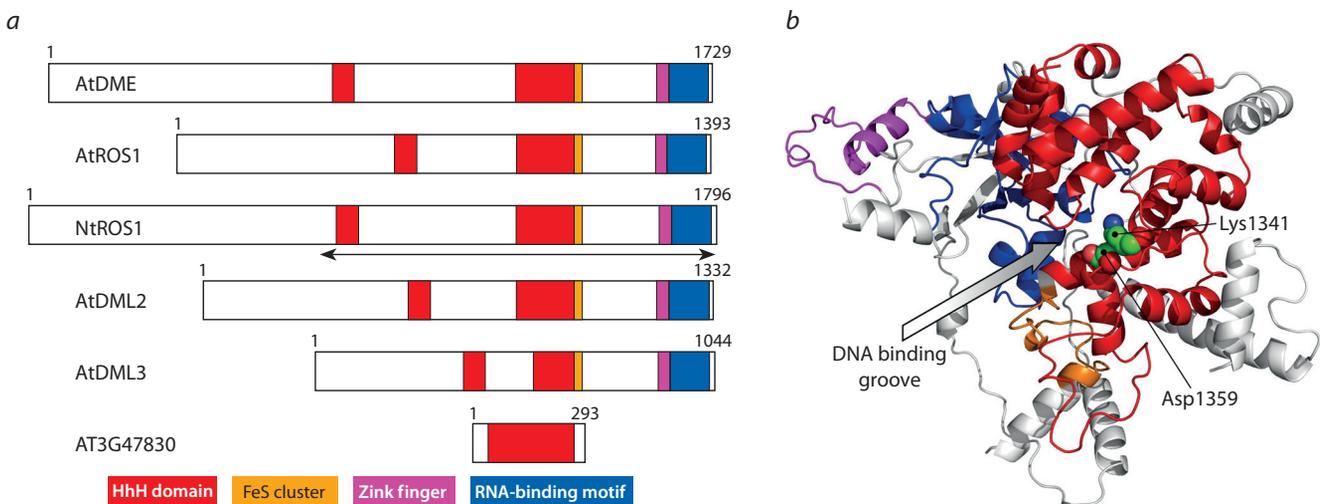


Fig. 1. Structure of NtROS1. *a*, Organization of DME/ROS1 family proteins. Arrows mark the catalytic fragment of NtROS1. *b*, Model of the NtROS1 catalytic fragment (long disordered regions are not shown).

The HhH domain is shown in red in both figures, the FeS cluster is in orange, the CXXC type zinc finger is in purple, and the RRM RNA-binding motif is in blue.

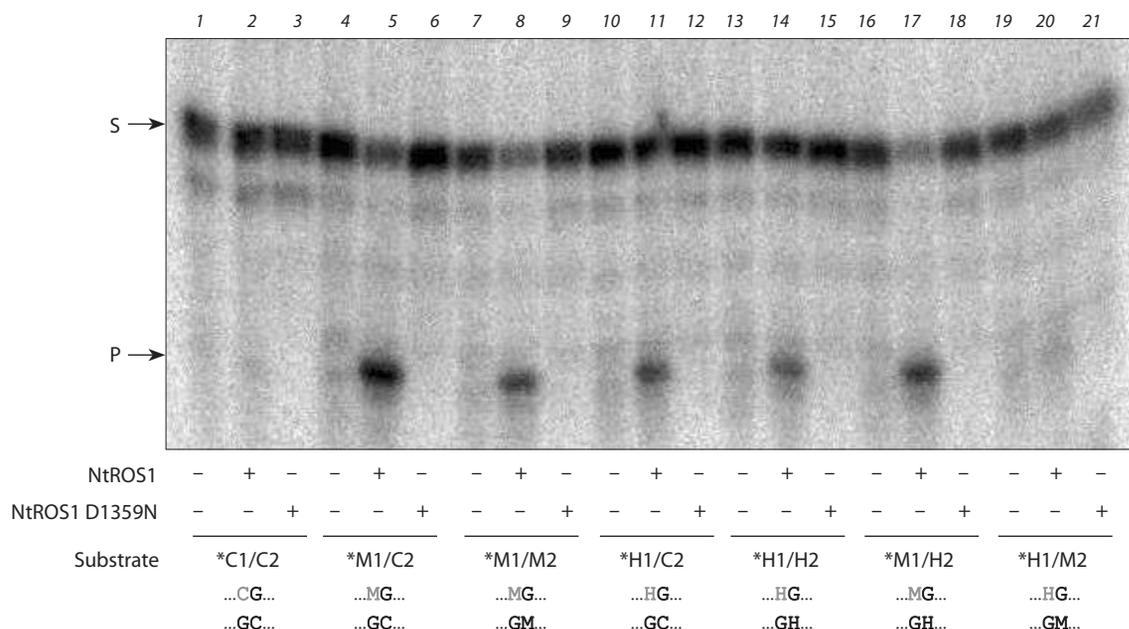


Fig. 2. Cleavage of substrates (50 nM) with wild-type NtROS1 and NtROS1 D1359N (100 nM) for 60 min. The methylation context of the CpG site in each substrate is indicated below the gel image. Lanes 1, 4, 7, 10, 13, 16, and 19, substrates without the enzyme; lanes 2, 5, 8, 11, 14, 17, and 20, wild-type NtROS1; lanes 3, 6, 9, 12, 15, 18, and 21, NtROS1 D1359N. Arrows mark the substrate (S) and the cleavage product (P). In the schematic representations of substrates below the gel, the excised base is shown in gray.

or 8-oxoguanine. The NtROS1 D1359N mutant, as expected, exhibited no glycosylase activity against any substrate with various combinations of mC and hmC, which confirms the importance of the Asp residue in the ROS1 catalytic center for the removal of modified bases from DNA.

AtROS1 was reported to be an enzyme with a very low turnover number (Ponferrada-Marín et al., 2009), so it was not feasible to use steady-state kinetics to characterize the activity of NtROS1. To determine the apparent reaction rate constant, we used the conditions close to the single-turnover kinetic regime ($[E]_0 > [S]_0$). Under these conditions, all DNA substrate rapidly binds to the enzyme, and the reaction rate is limited not by substrate binding or product release, but by the chemical step of the pseudo-first order reaction, the hydrolysis of the *N*-glycosidic bond of the modified nucleotide (Porello et al., 1998). Thus, following the accumulation of the product over time (Fig. 3, a), one can estimate the reaction rate constant. The time courses for all substrates are shown in Fig. 3, b, and the rate constants calculated from these data are summarized below:

Substrate	$k \cdot 10^4, s^{-1}$
*M1/C2	6.5 ± 1.1
*M1/M2	2.4 ± 0.5
*M1/H2	3.0 ± 0.7
*H1/C2	1.9 ± 0.3
*H1/M2	1.6 ± 0.6
*H1/H2	2.0 ± 0.3
*C1/C2	Not cleaved

* 32 P-labeled strand.

In all cases, hmC was a worse substrate for NtROS1 compared to mC. Based on these results, we conclude that NtROS1 cleaves hemimethylated CpG sites most efficiently, while hmC

in the context of HG/GM, on the contrary, is the worst substrate for this enzyme. The kinetic constants are consistent with the qualitative data on the relative cleavage efficiency for different substrates (see Fig. 2).

Many slow-turnover DNA glycosylases have lower AP lyase activity in comparison with their DNA glycosylase activity. In this case, after the removal of the modified base, a part of the reaction product exists as an AP site for a long time, and the true amount of the product can be revealed only upon treatment with alkali or nucleophilic amines (Porello et al., 1998). However, in the case of NtROS1, additional treatment with NaOH did not lead to a noticeable increase in the accumulation of the reaction product (see Fig. 3, c). Apparently, the rate of the reaction catalyzed by NtROS1 is limited by the hydrolysis of the *N*-glycosidic bond, which was also suggested for the enzyme from *A. thaliana* (Hong et al., 2014).

To assess the ability of the NtROS1 protein to act as a demethylase when expressed in mammalian cells, we constructed plasmids based on the pIRES-eGFP-puro vector encoding wild-type NtROS1 and its catalytically inactive mutant NtROS1 D1359N. Using anti-mC antibodies, we have estimated the level of this epigenetic base in HEK293 cells after transfection with these plasmids. When cells were observed post-transfection, the proliferation of cells with the pIRES-eGFP-puro-NtROS1 plasmid was reduced by about 30 % compared to the control cells transfected with the plasmid with no insert and the cells transfected with the plasmid encoding the catalytic mutant. The fraction of live cells was the same in all cases, which indicates that the cell cycle may be slower in the presence of active NtROS1 due to the need to repair a large number of breaks introduced into DNA at mC residues. An analysis of the mC level revealed a ~2-fold decrease relative to control samples when wild-type NtROS1

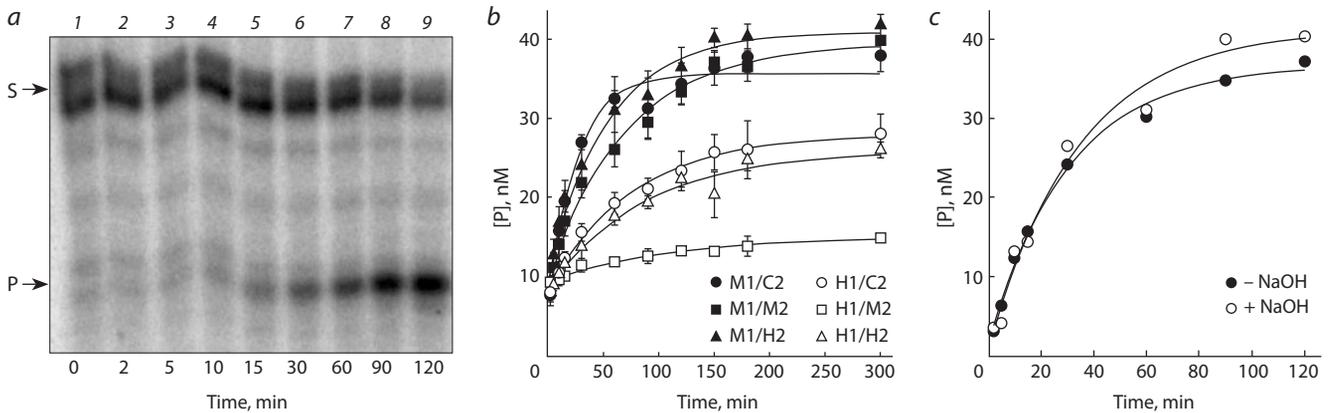


Fig. 3. Time course of the hydrolysis of various substrates by NtROS1.

a, Representative gel showing the accumulation of the reaction products after cleavage of the M1/M2 substrate for 0–120 min; b, cleavage of the M1/C2, M1/M2, M1/H2, H1/C2, H1/M2 and H1/H2 substrates (mean \pm S.D. for 3–4 experiments are shown); c, cleavage of the M1/M2 substrate without and with NaOH treatment for the complete elimination of AP sites.

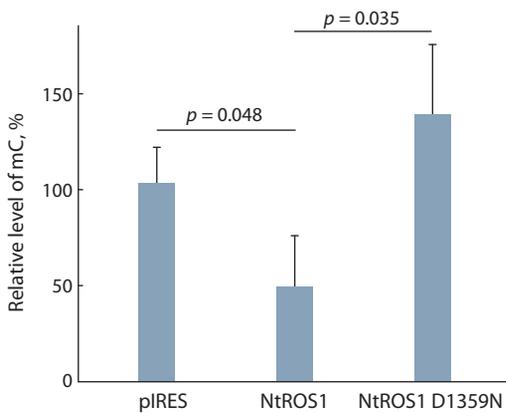


Fig. 4. Relative amounts of mC in the DNA of HEK293 Phoenix cells transfected with the pIRES-eGFP-puro plasmid carrying no insert or carrying an insert encoding the NtROS1 catalytic fragment or its inactive variant NtROS1 D1359N.

Mean \pm S.D. of 3 experiments are shown.

was expressed and the absence of statistically significant changes with the expression of NtROS1 D1359N (Fig. 4). In general, it can be considered that the transient expression of the catalytic domain of 5-methylcytosine–DNA glycosylase NtROS1 in human cells indeed leads to the global erasure of mC epigenetic marks. The amount of hmC in cells could not be measured using anti-hmC antibodies, probably because this modified base is about an order of magnitude less abundant than mC (Yu et al., 2012; Zahid et al., 2016).

Thus, it can be concluded that NtROS1 is a DNA glycosylase specific for 5-methylcytosine and, to a lesser extent, for 5-hydroxymethylcytosine. Its biological functions, as in the case of AtROS1, most likely consist in the regulation of methylation status and gene expression during embryonic development (Yamamuro et al., 2014) or in the response to infections and abiotic stress (Gong et al., 2002; Le et al., 2014), and the regulation of silencing of transgenes and transposable elements (Gong et al., 2002; Kapoor et al., 2005; Zhu et al., 2007). The mechanism of RNA-dependent addressing of

DME/ROS1 family proteins to specific demethylation sites, which has not yet been elucidated, is of great interest. The RRM motif in these polypeptides is homologous to the motifs responsible for nonspecific interactions with RNA in many other proteins (Cléry et al., 2008) and, by its position in the structure (see Fig. 1, b), could bind small RNAs complementary to the DNA stretch 5' of the targeted mC. Zinc fingers of the CXXC type are used by many mC-recognizing proteins; however, they predominantly bind unmethylated DNA and are presumably required for accurate positioning of the enzyme in the presence of several methylation sites located at a short distance (Iyer et al., 2011).

The prospects for using NtROS1 and other proteins of the DME/ROS1 family as tools for genetic technologies largely depend on the possibility of reducing the size of the catalytic fragment. A deletion of the long insert between the two parts of the HhH domain in AtROS1 fully preserves its activity, but a deletion of the C-terminal tail after the FeS cluster results in an inactive enzyme (Hong et al., 2014). Judging from the structural models of AtROS1 and NtROS1, the HhH and RRM domains interact with each other, and shortening the protein here may only be done by trimming the insert between them. In any case, DME/ROS1 proteins, including NtROS1, represent a promising scaffold for enzyme engineering to analyze epigenetic methylation status and control gene activity.

Conclusion

The study of the demethylating DNA glycosylase ROS1 from *Nicotiana tabacum* reported in this work presents the only biochemical investigation of ROS1 beyond its homologue from *Arabidopsis thaliana*. In addition to 5-methylcytosine, NtROS1 showed the ability to remove 5-hydroxymethylcytosine from DNA, but the efficiency of this reaction was lower than for 5-methylcytosine, apparently because plants rarely if at all use 5-hydroxymethylcytosine as an epigenetic marker. The observed decrease in global methylation upon expression of NtROS1 in human cells suggests that this protein or its optimized variants can be used as a tool for epigenetic regulation, either on its own or as an active module in constructs targeted to certain genome regions.

References

- Agius F., Kapoor A., Zhu J.-K. Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. USA.* 2006;103(31):11796-11801. DOI 10.1073/pnas.0603563103.
- Ballestar E., Wolffe A.P. Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur. J. Biochem.* 2001;268(1):1-6. DOI 10.1046/j.1432-1327.2001.01869.x.
- Baubec T., Ivánek R., Lienert F., Schübeler D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell.* 2013;153(2):480-492. DOI 10.1016/j.cell.2013.03.011.
- Bochtler M., Kolano A., Xu G.-L. DNA demethylation pathways: additional players and regulators. *Bioessays.* 2017;39(1):1-13. DOI 10.1002/bies.201600178.
- Branco M.R., Ficz G., Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* 2011;13(1):7-13. DOI 10.1038/nrg3080.
- Choi C.-S., Sano H. Identification of tobacco genes encoding proteins possessing removal activity of 5-methylcytosines from intact tobacco DNA. *Plant Biotechnol.* 2007;24(3):339-344. DOI 10.5511/plantbiotechnology.24.339.
- Choi W.L., Mok Y.G., Huh J.H. Application of 5-methylcytosine DNA glycosylase to the quantitative analysis of DNA methylation. *Int. J. Mol. Sci.* 2021;22(3):1072. DOI 10.3390/ijms22031072.
- Choi Y., Gehring M., Johnson L., Hannon M., Harada J.J., Goldberg R.B., Jacobsen S.E., Fischer R.L. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell.* 2002;110(1):33-42. DOI 10.1016/s0092-8674(02)00807-3.
- Cléry A., Blatter M., Allain F.H.-T. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* 2008;18(3):290-298. DOI 10.1016/j.sbi.2008.04.002.
- Devesa-Guerra I., Morales-Ruiz T., Pérez-Roldán J., Parrilla-Doblas J.T., Dorado-León M., García-Ortiz M.V., Ariza R.R., Roldán-Arjona T. DNA methylation editing by CRISPR-guided excision of 5-methylcytosine. *J. Mol. Biol.* 2020;432(7):2204-2216. DOI 10.1016/j.jmb.2020.02.007.
- Fedorova O.S., Kuznetsov N.A., Koval V.V., Knorre D.G. Conformational dynamics and pre-steady-state kinetics of DNA glycosylases. *Biochemistry (Moscow).* 2010;75(10):1225-1239. DOI 10.1134/S0006297910100044.
- Gong Z., Morales-Ruiz T., Ariza R.R., Roldán-Arjona T., David L., Zhu J.-K. ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell.* 2002;111(6):803-814. DOI 10.1016/s0092-8674(02)01133-9.
- Gruber D.R., Toner J.J., Miers H.L., Shernyukov A.V., Kiryutin A.S., Lomzov A.A., Endutkin A.V., Grin I.R., Petrova D.V., Kupryushkin M.S., Yurkovskaya A.V., Johnson E., Okon M., Bagryan-skaya E.G., Zharkov D.O., Smirnov S.L. Oxidative damage to epigenetically methylated sites affects DNA stability, dynamics, and enzymatic demethylation. *Nucleic Acids Res.* 2018;46(20):10827-10839. DOI 10.1093/nar/gky893.
- Hong S., Hashimoto H., Kow Y.W., Zhang X., Cheng X. The carboxy-terminal domain of ROS1 is essential for 5-methylcytosine DNA glycosylase activity. *J. Mol. Biol.* 2014;426(22):3703-3712. DOI 10.1016/j.jmb.2014.09.010.
- Iyer L.M., Abhiman S., Aravind L. Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* 2011;101:25-104. DOI 10.1016/B978-0-12-387685-0.00002-0.
- Jang H., Shin H., Eichman B.F., Huh J.H. Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. *Biochem. Biophys. Res. Commun.* 2014;446(4):1067-1072. DOI 10.1016/j.bbrc.2014.03.060.
- Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., Tunyasuvunakool K., Bates R., Židek A., Potapenko A., Bridgland A., Meyer C., Kohl S.A.A., Ballard A.J., Cowie A., Romera-Paredes B., Nikolov S., Jain R., Adler J., Back T., Petersen S., Reiman D., Clancy E., Zielinski M., Steinegger M., Pacholska M., Berghammer T., Bodenstern S., Silver D., Vinyals O., Senior A.W., Kavukcuoglu K., Kohli P., Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583-589. DOI 10.1038/s41586-021-03819-2.
- Kapazoglou A., Drosou V., Argiriou A., Tsaftaris A.S. The study of a barley epigenetic regulator, *HvDME*, in seed development and under drought. *BMC Plant Biol.* 2013;13:172. DOI 10.1186/1471-2229-13-172.
- Kapoor A., Agarwal M., Andreucci A., Zheng X., Gong Z., Hasegawa P.M., Bressan R.A., Zhu J.-K. Mutations in a conserved replication protein suppress transcriptional gene silencing in a DNA-methylation-independent manner in *Arabidopsis*. *Curr. Biol.* 2005;15(21):1912-1918. DOI 10.1016/j.cub.2005.09.013.
- Le T.-N., Schumann U., Smith N.A., Tiwari S., Au P.C.K., Zhu Q.-H., Taylor J.M., Kazan K., Llewellyn D.J., Zhang R., Dennis E.S., Wang M.-B. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. *Genome Biol.* 2014;15(9):458. DOI 10.1186/s13059-014-0458-3.
- Lee T.-F., Zhai J., Meyers B.C. Conservation and divergence in eukaryotic DNA methylation. *Proc. Natl. Acad. Sci. USA.* 2010;107(20):9027-9028. DOI 10.1073/pnas.1005440107.
- Li X., Qian W., Zhao Y., Wang C., Shen J., Zhu J.-K., Gong Z. Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA.* 2012;109(28):11425-11430. DOI 10.1073/pnas.1208557109.
- Li Y., Kumar S., Qian W. Active DNA demethylation: mechanism and role in plant development. *Plant Cell Rep.* 2018;37(1):77-85. DOI 10.1007/s00299-017-2215-z.
- Liu R., How-Kit A., Stammitt L., Teyssier E., Rolin D., Mortain-Bertrand A., Halle S., Liu M., Kong J., Wu C., Degraeve-Guibault C., Chapman N.H., Maucourt M., Hodgman T.C., Tost J., Bouzayen M., Hong Y., Seymour G.B., Giovannoni J.J., Gallusci P. A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc. Natl. Acad. Sci. USA.* 2015;112(34):10804-10809. DOI 10.1073/pnas.1503362112.
- Morales-Ruiz T., Ortega-Galisteo A.P., Ponferrada-Marín M.I., Martínez-Macías M.I., Ariza R.R., Roldán-Arjona T. DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci. USA.* 2006;103(18):6853-6858. DOI 10.1073/pnas.0601109103.
- Ono A., Yamaguchi K., Fukada-Tanaka S., Terada R., Mitsui T., Iida S. A null mutation of *ROS1a* for DNA demethylation in rice is not transmittable to progeny. *Plant J.* 2012;71(4):564-574. DOI 10.1111/j.1365-3113X.2012.05009.x.
- Ortega-Galisteo A.P., Morales-Ruiz T., Ariza R.R., Roldán-Arjona T. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* 2008;67(6):671-681. DOI 10.1007/s11103-008-9346-0.
- Parrilla-Doblas J.T., Roldán-Arjona T., Ariza R.R., Córdoba-Cañero D. Active DNA demethylation in plants. *Int. J. Mol. Sci.* 2019;20(19):4683. DOI 10.3390/ijms20194683.
- Pastor W.A., Aravind L., Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 2013;14(6):341-356. DOI 10.1038/nrm3589.
- Pastor W.A., Pape U.J., Huang Y., Henderson H.R., Lister R., Ko M., McLoughlin E.M., Brudno Y., Mahapatra S., Kapranov P., Tahiliani M., Daley G.Q., Liu X.S., Ecker J.R., Milos P.M., Agarwal S., Rao A. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 2011;473(7347):394-397. DOI 10.1038/nature10102.
- Penterman J., Uzawa R., Fischer R.L. Genetic interactions between DNA demethylation and methylation in *Arabidopsis*. *Plant Physiol.* 2007;145(4):1549-1557. DOI 10.1104/pp.107.107730.
- Ponferrada-Marín M.I., Parrilla-Doblas J.T., Roldán-Arjona T., Ariza R.R. A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. *Nucleic Acids Res.* 2011;39(4):1473-1484. DOI 10.1093/nar/gkq982.

- Ponferrada-Marín M.I., Roldán-Arjona T., Ariza R.R. ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. *Nucleic Acids Res.* 2009;37(13):4264-4274. DOI 10.1093/nar/gkp390.
- Porello S.L., Leyes A.E., David S.S. Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates. *Biochemistry.* 1998;37(42):14756-14764. DOI 10.1021/bi981594+.
- Roldán-Arjona T., Ariza R.R., Córdoba-Cañero D. DNA base excision repair in plants: an unfolding story with familiar and novel characters. *Front. Plant Sci.* 2019;10:1055. DOI 10.3389/fpls.2019.101055.
- Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., de Beer T.A.P., Rempfer C., Bordoli L., Lepore R., Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296-W303. DOI 10.1093/nar/gky427.
- Wen S., Wen N., Pang J., Langen G., Brew-Appiah R.A.T., Mejias J.H., Osorio C., Yang M., Gemini R., Moehs C.P., Zemetra R.S., Kogel K.-H., Liu B., Wang X., von Wettstein D., Rustgi S. Structural genes of wheat and barley 5-methylcytosine DNA glycosylases and their potential applications for human health. *Proc. Natl. Acad. Sci. USA.* 2012;109(50):20543-20548. DOI 10.1073/pnas.1217927109.
- Wu X., Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* 2017;18(9):517-534. DOI 10.1038/nrg.2017.33.
- Yamamuro C., Miki D., Zheng Z., Ma J., Wang J., Yang Z., Dong J., Zhu J.-K. Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation. *Nat. Commun.* 2014;5:4062. DOI 10.1038/ncomms5062.
- Yu M., Hon G.C., Szulwach K.E., Song C.-X., Zhang L., Kim A., Li X., Dai Q., Shen Y., Park B., Min J.-H., Jin P., Ren B., He C. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012;149(6):1368-1380. DOI 10.1016/j.cell.2012.04.027.
- Zahid O.K., Zhao B.S., He C., Hall A.R. Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores. *Sci. Rep.* 2016;6:29565. DOI 10.1038/srep29565.
- Zemach A., Zilberman D. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Biol.* 2010;20(17):R780-R785. DOI 10.1016/j.cub.2010.07.007.
- Zharkov D.O. Base excision DNA repair. *Cell. Mol. Life Sci.* 2008; 65(10):1544-1565. DOI 10.1007/s00018-008-7543-2.
- Zhu J., Kapoor A., Sridhar V.V., Agius F., Zhu J.-K. The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. *Curr. Biol.* 2007;17(1):54-59. DOI 10.1016/j.cub.2006.10.059.

ORCID ID

N.V. Permyakova orcid.org/0000-0001-9291-1343
I.R. Grin orcid.org/0000-0002-5685-1248
D.O. Zharkov orcid.org/0000-0001-5013-0194

Acknowledgements. The study was supported by the Russian Foundation for Basic Research and the Government of the Novosibirsk Region (project No. 20-44-540007). The modeling of the structure of NtROS1 was supported by the Russian Ministry of Science and Education (State budget project No. 121031300056-8). DNA sequencing was performed at the SB RAS Genomics Core Facility.

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

Received November 23, 2021. Revised February 24, 2022. Accepted April 6, 2022.

Original Russian text www.bionet.nsc.ru/vogis/

Investigation of genetic polymorphism of Russian rape and turnip rape varieties using SSR and SRAP markers

I.A. Klimenko , V.T. Volovik, A.A. Antonov, V.A. Dushkin, A.O. Shamustakimova, Yu.M. Mavlyutov

Federal Williams Research Center of Forage Production and Agroecology, Lobnya, Moscow region, Russia

 iaklimenko@mail.ru

Abstract. Rapeseed (*Brassica napus* L.) and turnip rape (*B. rapa* L. subsp. *campestris* (L.)) are important agricultural plants widely used for food, fodder and technical purposes and as green manure. Over the past decades, a large number of perspective varieties that are being currently cultivated in every region of Russia have been developed. To increase the breeding efficiency and facilitate the seed production, modern molecular-genetic techniques should be introduced as means to estimate species and varietal diversity. The objective of the presented research study was to investigate DNA polymorphism of the rapeseed and turnip rape varieties developed at Federal Williams Research Center of Forage Production and Agroecology and detect informative markers for varietal identification and genetic certification. To genotype 18 gDNA samples, 42 and 25 combinations of respective SSR and SRAP primers were used. The results obtained demonstrate that SRAP markers were more effective for polymorphism analysis: 36 % of the tested markers revealed genetic polymorphism compared with only 16.7 % of microsatellite loci. Molecular markers to detect differences at interspecific and intervarietal levels have also been found. For the investigated set, such microsatellite loci as Na12A02, Ni2C12, Ni02-D08a, Ra02-E01, Ni03H07a and SRAP-marker combinations as F13-R9, Me4-R7, F11-Em2, F10-R7, F9-Em2 and F9-R8 proved to be informative. Application of the two marker techniques made it possible to detect a higher level of DNA polymorphism in plants of different types (spring and winter varieties) if compared against the intervarietal differences within a species or a group. According to Nei's genetic diversity index, in the cluster of winter rapeseed, VIK 2 and Gorizont varieties had the longest genetic distance, and in the spring cluster, these were Novosel and Veles. A high level of similarity was found between Vikros and Bizon winter rapeseed varieties. The results obtained have a high practical value for varietal specification of seed material and genetic certification of rapeseed and turnip rape varieties.

Key words: forage crops; *Brassica napus* L.; *B. rapa* L. *campestris*; bulk samples; genetic polymorphism; SSR markers; SRAP markers.

For citation: Klimenko I.A., Volovik V.T., Antonov A.A., Dushkin V.A., Shamustakimova A.O., Mavlyutov Yu.M. Investigation of genetic polymorphism of Russian rape and turnip rape varieties using SSR and SRAP markers. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(4):349-358. DOI 10.18699/VJGB-22-42

Изучение генетического полиморфизма российских сортов рапса и сурепицы с использованием SSR- и SRAP-маркеров

И.А. Клименко , В.Т. Воловик, А.А. Антонов, В.А. Душкин, А.О. Шамустакимова, Ю.М. Мавлютов

Федеральный научный центр кормопроизводства и агроэкологии им. В.Р. Вильямса, г. Лобня, Московская область, Россия

 iaklimenko@mail.ru

Аннотация. Рапс (*Brassica napus* L.) и сурепица (*B. rapa* L. subsp. *campestris* (L.)) – важные сельскохозяйственные культуры, широко используются для продовольственных, кормовых и технических целей, а также в качестве сидератов. За последние десятилетия создано большое количество перспективных сортов, культивируемых практически во всех регионах России. Для повышения эффективности селекционного процесса и успешного развития семеноводства необходимо внедрять современные молекулярно-генетические методы оценки видового и сортового разнообразия. Цель настоящей работы заключалась в изучении ДНК-полиморфизма сортов рапса и сурепицы селекции Федерального научного центра кормопроизводства и агроэкологии им. В.Р. Вильямса и выявлении информативных маркеров для сортовой идентификации и генетической паспортизации. Для генотипирования 18 образцов геномной ДНК использовали 42 и 25 комбинаций SSR- и SRAP-праймеров соответственно. Результаты показали, что маркеры SRAP более эффективны для анализа полиморфизма изучаемого материала: 36 % от общего числа испытанных маркеров демонстрировали генетический полиморфизм, тогда как для микросателлитных локусов этот показатель равнялся 16.7 %. Определены молекулярные маркеры для выявления различий на межвидовом и межсортовом уровнях. Информативными для исследуемой выборки сортов оказались микросателлитные локусы Na12A02, Ni2C12, Ni02-D08a, Ra02-E01, Ni03H07a и комбинации SRAP-маркеров F13-R9, Me4-R7, F11-Em2, F10-R7, F9-Em2 и F9-R8. Анализ сортового материала по двум системам

маркирования показал более высокий уровень ДНК-полиморфизма у образцов растений разного типа развития (яровой/озимый) в сравнении с различиями между сортами в пределах вида или группы. Согласно индексам генетического разнообразия Нея, в кластере сортов озимого рапса наибольшей генетической удаленностью выделялись ВИК 2 и Горизонт, среди яровых – Новосёл и Велес. Высокий уровень сходства обнаружен между яровыми сортами рапса Викрос и Бизон. Полученная информация имеет практическое значение для контроля сортовой принадлежности и генетической паспортизации семенного материала сортов рапса и сурепицы.

Ключевые слова: кормовые культуры; *Brassica napus* L.; *B. rapa* L. *campestris*; «балк-образцы», генетический полиморфизм; SSR-маркеры; SRAP-маркеры.

Introduction

Cabbage oilseed crops such as rapeseed (*Brassica napus* L.) and turnip rape (*B. rapa* L. subsp. *campestris* (L.)) are cultivated in almost every region of Russia, and, for the foreseeable future, are regarded as the main reserve for increasing the production of vegetable oil and fodder protein. These plants are widely used in food, fodder, technical purposes and as green manure that increases soil fertility thanks to the plants' root remains containing up to 6 tons of organic matters, 80 kg of nitrogen, 60 kg of phosphorus and 90 kg of potassium per hectare. As for their food and fodder properties, rapeseed and turnip rape exceed many other cultivated crops since their seeds are 40–48 % fat and 21–33 % protein and contain a high amount of essential amino acids (Volovik, 2015). Rapeseed can provide livestock with green forage from early spring to late fall thanks to their cold hardiness and fast regrowth after mowing. They are also an excellent silage material, and their seeds and seed by-pass products are processed to produce seed cake and coarse meal. In the recent years the varieties of rapeseed and turnip rape with low or no erucic-acid content became available and seed production has increased more than 7 times to reach the world's third place after soybeans and cotton. Russia's short-term plans are to increase rapeseed planting acreage to 2.5 mln ha.

As for Russian research institutions working intensely to select cabbage oilseed crops, the leading ones are All-Russian Research Institute of Rapeseed, All-Russian Research Institute of Oilseed Crops and All-Russian Williams Fodder Research Institute. For the two last decades, they have produced the perspective varieties of rapeseed, turnip rape, white mustard and oil radish that have been recommended for oil production, livestock and poultry green forage, combination fodder, seed cake and coarse meal production. In 2021, "State Register" of the Russian Federation included 13 varieties of rapeseed and 3 varieties of turnip rape selected by Federal Williams Research Center of Forage Production and Agroecology (Kosolapov et al., 2019; State Register..., 2021).

For preservation and rational use of newly available varieties, intensification of the selection process and protection of intellectual property, modern and effective methods to estimate species and varietal diversity at a genetic level are to be introduced. One of such techniques that has been successfully applied in the recent years is molecular DNA markers, which, if compared against the traditional morphological indicators, possess a number of advantages. These include a high level of polymorphism; even genome distribution; reliability; a possibility to automate the assay procedure that does not depend on environmental conditions or a plant development phase

(Agarwal et al., 2008; Khlestkina, 2011; Chesnokov, 2018). If the most informative and convenient DNA markers are selected, their capabilities to estimate the genetic variability of selection material are regarded as unlimited.

Laboratory for Molecular and Genetic Studies in Federal Williams Research Center of Forage Production and Agroecology has been developing a system for DNA identification and genetic certification of Russian fodder crops. For the time being, the varietal identification techniques have been adapted for perennial legume grasses such as red clover and different species of alfalfa (Klimenko et al., 2020a, b). The assay uses samples of the summary total DNA obtained through a modified method from an arbitrary selected sample of every variety's germinants. Two types of molecular markers were used: SSR (simple sequence repeats), which detect the variability of microsatellite genome sequences, and SRAP (sequence related amplified polymorphism), which is based on PCR with a pair of primers for amplification of intron/exon regions (open reading frames). The techniques have been tested on different species of fodder crops to optimize the amplification conditions, detection and analysis of results.

A problem of reliable varietal identification is particularly topical for rapeseed due to its limited genetic variability conditioned by the intensive selection aimed at higher content and quality of oil. Currently, a significant number of published studies have been devoted to using different DNA markers for estimation of the genetic diversity of rapeseed varieties and hybrids (Plieske, Struss, 2001; Snowdon, Friedt, 2004; Klyachenko et al., 2018; Mozgova et al., 2019); to genetic mapping (Piquemal et al., 2005; Gao et al., 2007; Geng, 2012) and marking the genes of economically valuable traits (Chen et al., 2010; Ananga et al., 2012). However, only a few such studies have investigated Russian varieties. Four varieties of winter and spring rapeseed (Podmoskovniy, Vikros, VIK 2 and Severyanin) were studied by Byelorussian researchers to identify the gene alleles determining the concentration of oleic and linolic acids in rapeseed oil (Lemesh et al., 2015). The same varieties were investigated to detect the DNA markers of the genes responsible for erucic-acid synthesis (Amosova et al., 2014). Microsatellite markers were used to study the genetic polymorphism of Russian varieties Ratnik and SNK-198 (Satina, 2010) as well as the genetic homogeneity of spring rapeseed varieties Bulat and Forward (Rogozhina et al., 2015). Such winter varieties as Stolychniy, Laureat, Gorizont, Nord and Severyanin were investigated to detect the quantitative trait loci (QTLs) associated with high winter hardiness (Mozgova et al., 2019).

The objective of the presented study was to investigate DNA polymorphism of rapeseed and turnip rape varieties developed by breeders of Federal Williams Research Center of Forage Production and Agroecology and to identify the informative markers for varietal differentiation and genetic certification.

Materials and methods

Plant material. The study investigated 15 varieties of winter (Severyanin, Stolychniy, VIK 2, Nord, Laureat, Gorizont, Garant) and spring (Vikros, Novik, Novosel, Veles, Grant, Podmoskovniy, Lugovskoy, Bizon) rapeseed and 3 varieties of winter (Zarya) and spring (Nadezhda, Svetlana) turnip rape.

DNA extraction and PCR analysis. The gDNA was extracted from 30 germinants of each abovementioned variety (bulk samples) using the basic SDS method (Kirby, Cook, 1967; Dellaporta et al., 1983) with some modifications (Klimenko et al., 2020b). The quality and concentration of the obtained DNA fractions were verified with agarose gel (1.5 %) electrophoresis and using a Nabi spectrophotometer (MicroDigital, South Korea).

To carry out SSR analysis, 42 markers from the database Brassica info (<https://www.brassica.info>) and available publications were applied. The efficiency of the primers devised for these markers had been demonstrated in the studies devoted to development of the technology of rapeseed genotyping (Satina, 2010) and selection of the samples with low erucic-acid and glucosinolate content (Hasan et al., 2008). A part of the markers included in the analysis was used for hybridization control and detection of *Alternaria blight* resistant genotypes in Indian mustard (*B. juncea* L.) (Chandra et al., 2013; Sharma et al., 2018).

The PCR-mixture of 20 µl contained 3 µl 10×PCR buffer (Taq Turbo Buffer), 0.5 µl 50× dNTPs mix, 0.4 µl Taq polymerase (5U), forward and reverse primers (0.1 µl each, 100 µM) and 0.1 µl of DNA sample (20 ng/µl). The amplification was performed in a T-1000 thermal cycler (Bio-Rad, USA) at two different temperature regimes. The first amplification program was an initial 3-min denaturation at 95 °C followed by 30 cycles of 30 s at 94 °C, 30 s at 55–57 °C, 30 s at 72 °C and a final 5-min elongation at 72 °C (Satina, 2010). The second program included an initial 5-min denaturation at 95 °C followed by 39 cycles of 1 min at 94 °C, 2 min at 46–51 °C (depending on the primer pair in use), 2 min at 72 °C and

a final 10-min elongation at 72 °C (Chandra et al., 2013). The reproducibility of obtained results was attested in three-fold replication.

SRAP analysis was carried out using 25 primer combinations comprised from 10 single oligonucleotides: F9, F13, Me4, F10, F11, R9, R7, Em2, R14, R8 (Li, Quiros, 2001; Rhouma et al., 2017). The amplification program was an initial 4-min denaturation at 94 °C followed by 10 cycles with changing temperature and duration parameters (1 min at 94 °C, 1 min at 35 °C, 1 min at 72 °C); followed by 30 cycles (1 min at 94 °C, 1 min at 50 °C, 1 min at 72 °C) and a final 5-min elongation step run at 72 °C. The PCR-mixture composition was similar to that used for the microsatellite analysis.

PCR-products were separated using 90-min 50-V agarose-gel electrophoresis (4 % MetaPhorR Agarose, Rockland or 1.6 % LE, Lonza, USA). As the reference markers, 20 bp DNA Ruler (Bio-Rad), 100 kb DNA Ladder (Thermo Fisher Scientific, USA) and 100 bp + 1.5 kb (SibEnzyme, Russia) were applied.

Analysis of the obtained results. PCR-product detection and size measurement was performed using a GelDoc XR+ imaging system (Bio-Rad) and the ImageLab software (Bio-Rad Lab., Inc.) for molecular-mass markers. The obtained results were transformed into a binary matrix, and PopGene v. 1.32 (Yeh et al., 2000) was applied to determine such genetic diversity indices as the effective number of alleles per locus; Shannon's index; expected heterozygosity; Nei's genetic distance (Nei, Li, 1979). Polymorphism information content (PIC) for every pair of primers was calculated by the formula presented in the study (Chesnokov, Artemyeva, 2015). To build the genetic similarity dendrogram, the unweighted pair group method with arithmetic averages was applied in NTSYSpc v 2.10 (Rohlf, 2000).

Results

To obtain gDNA from the rapeseed and turnip rape germinants, a modified SDS method was used. The applied protocol proved more effective and less costly compared to other known protocols and commercial reagents kits. The results of electrophoresis and spectrophotometry attested to the DNA's high concentration and purification degree from protein compounds and polysaccharides for all experimental samples (Fig. 1, 2).

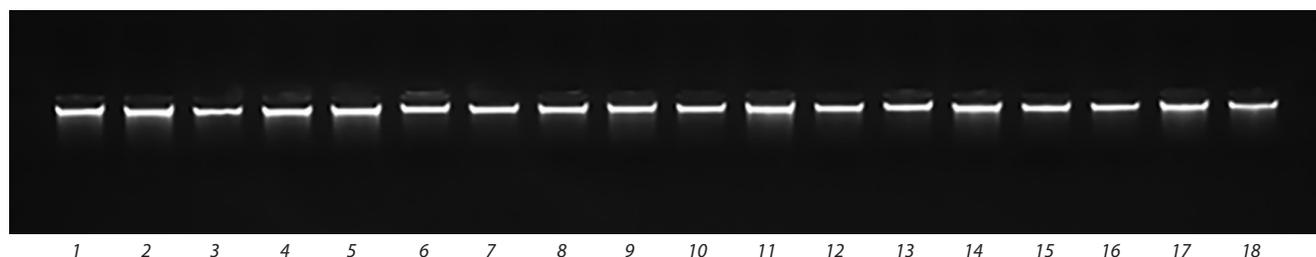


Fig. 1. Electrophoregram of the gDNA extracted from the rapeseed and turnip rape germinants.

Lanes 1–15 (rape varieties): Severyanin, Stolychniy, VIK 2, Nord, Laureat, Gorizont, Garant, Vikros, Novik, Novosel, Veles, Grant, Podmoskovniy, Lugovskoy, Bizon; 16–18 (turnip rape varieties): Zarya, Nadezhda, Svetlana.

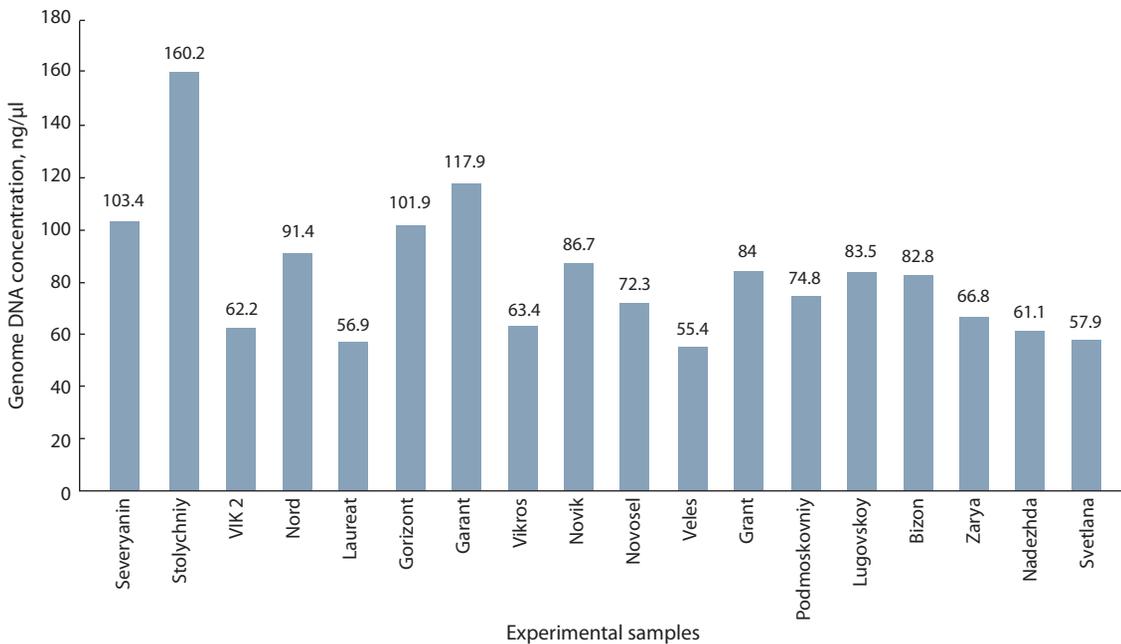


Fig. 2. gDNA concentrations.

Table 1. SSR primers used for rapeseed and turnip rape DNA polymorphism analysis

Primer	Nucleotide sequence (5'→3')	PCR fragment size according to literature data, bp	References
Ni-F02a	TGCAACGAAAAGGATCAGC/ TGCTAATTGAGCAATAGTGATTCC	250	Chandra et al., 2013
Ni03H07a	GCTGTGATTTTAGTGACCG/ AGCCGTTGATGGAATTTTTG	250–280	
Ni02-D08a	ACAACAACCCATGTCTTCCG/ ACAACAACCCATGTCTTCCG	< 100–300	
Ra02-E01a	GCACACACACTCAAACCC/ TCTATATTAACGCGGACGG	< 100- > 1000	
Ni2C12 (Bna.M.002)	AAAGGTCAAGTCCTTCTTCG/ ACATTCTTGGATCTTGATTCC	112–150	Satina, 2010
Na12A02 (Bna.M.001)	AGTGAATCGATGATCTGCC/ AGCCTTGTGCTTTTCAACG	160–218	
Bna.M.010	CATTGTTGTCTTGGGAGAGC/ AGGACACCAGGCACCATATA	100–150	

SSR-analysis

For genotyping the full variety collection, out of 42 SSR primers, 7 primers providing stable and reproducible amplification were selected (Table 1).

Analysis of the amplification fragments obtained using the listed primers detected 42 alleles. Their number per locus was 6 on average, varying from 3 (Ni2C12 and Bna.M.010) to 10 (Ra02-E01a). The fragment size varied from 110 bps (Ni2C12) to 1200 bps (Ni02-D08a). The maximum allele frequency was registered for Bna.M.010 (0.83), and the minimum – for Ni03H07a (0.27); the mean value was 0.42. The

primers developed for Ni03H07a, Ni02-D08a and Ra02-E01a markers made it possible to detect 8–10 alleles per locus and had the highest PIC (0.82).

SRAP-analysis

Based on the results of preliminary testing, the initial 25 combinations of SRAP primers were reduced to 10 pairs, amplifying stable polymorphic DNA fragments (Table 2). In total, 53 PCR fragments of 132–1674 nucleotide pairs in size were obtained. One combination contained from 4 (F9-R9) to 7 (F10-R8, F11-Em2, F10-R7) amplicons. A part of the

markers proved to be informative to detect the amplification fragments for differentiating the type of plants (winter/spring). Using 6 combinations made it possible to obtain the amplicons specific for varieties identification (marked with a star in the Table 2).

Fig. 3 demonstrates the electrophoregram of PCR results with the F9-R8 primer combination. Significant DNA profile differences were found between winter (I) and spring (II) rapeseed varieties (joined in curly brackets). The arrows mark the variety-specific PCR products characteristic for Stolychniy winter rapeseed (508 bps) and Nadezhda spring turnip rape (700 bps) as well as the absence of an amplicon in size of 460 bps in spring rapeseed Podmoskovniy though it was a specific characteristic for other varieties in this group.

The performed analysis demonstrated that it is possible to identify rapeseed varieties Grant and Novosel with 3 marker combinations (F11-Em2, F10-R7 and Me4-R7), and Gorizont and Lugovskoy – with 2 (F13-R9 and Me4-R7). Variety VIK 2 was identified with SRAP primers F9-Em2, and spring ones Veles – with F10-R7. Specific DNA spectra for rapeseed varieties Stolychniy, Podmoskovniy and turnip rape Nadezhda were obtained with F9-R8 combination.

The obtained data were transformed into a binary matrix to calculate Nei's genetic distances (Table 3). The lowest genetic similarity coefficient (0.7069) was found between rapeseed varieties Gorizont, Novosel and Grant, the highest – between spring varieties Vikros and Bizon (1.0) as well as Veles and Bizon (0.9655). A similarly high genetic distance (0.3228) indicated significant differences between pairs: Grant and VIK 2, and Lugovskoy and Stolychniy. Low distance values and high genetic similarity were demonstrated by spring varieties Bizon and Vikros (zero distance) and winter varieties Garant, Severyanin, Stolychniy, Nord, Laureat (0.0174).

Table 2. SRAP primers used for rapeseed and turnip rape DNA polymorphism analysis

Primer combination	Nucleotide sequence	PCR fragment size, bps
F13-R7	CGAATCTTAGCCGGCAC/ GACTGCGTACGAATTGAG	266–624
F10-R8	GTAGCACAAGCCGGAAG/ GACACCGTACGAATTGAC	248–1140
*F13-R9	CGAATCTTAGCCGGCAC/ GACTGCGTACGAATTCA	353–1674
F9-R9	GTAGCACAAGCCGGACC/ GACTGCGTACGAATTCA	145–849
*Me4-R7	CGAATCTTAGCCGGAAT/ GACTGCGTACGAATTGAG	300–806
*F11-Em2	CGAATCTTAGCCGGATA/ GACTGCGTACGAATTCGG	175–712
*F10-R7	GTAGCACAAGCCGGAAG/ GACTGCGTACGAATTGAG	132–628
*F9-Em2	GTAGCACAAGCCGGACC/ GACTGCGTACGAATTCGG	209–1025
F13-R8	CGAATCTTAGCCGGCAC/ GACACCGTACGAATTGAC	107–688
*F9-R8	GTAGCACAAGCCGGACC/ GACACCGTACGAATTGAC	205–700

Note. The data on nucleotide sequences, published by (Li, Quiros, 2001), were used for primers synthesis.

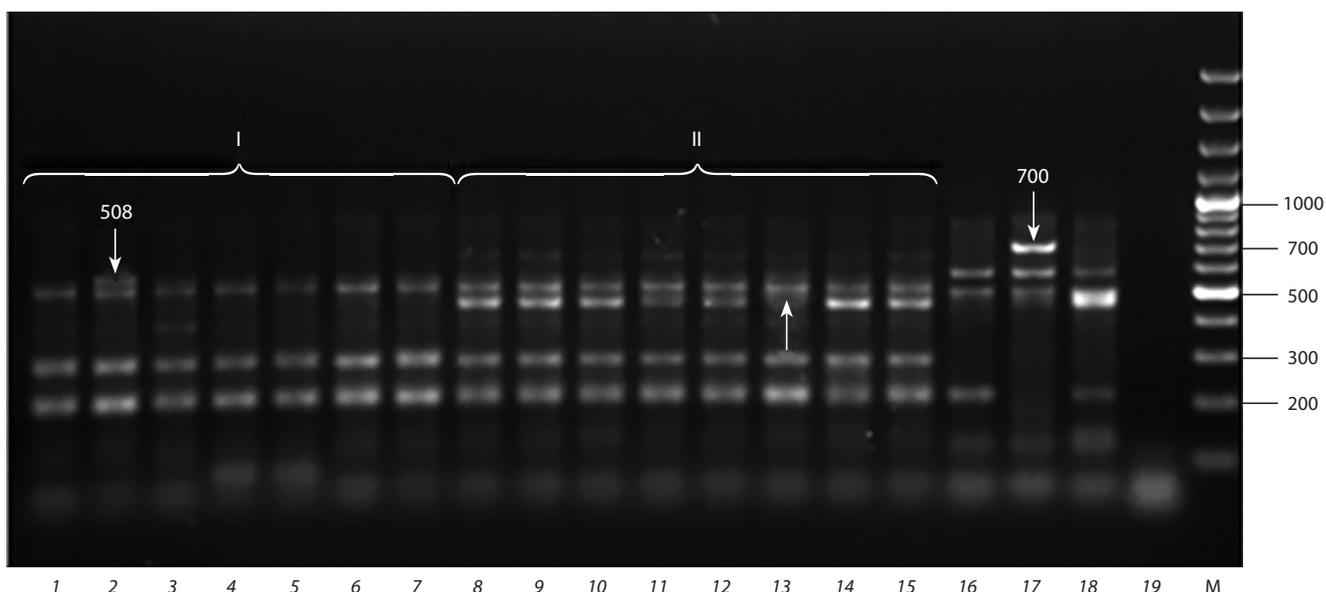


Fig. 3. Electrophoregram of the PCR products obtained during amplification of rapeseed and turnip rape varieties with SRAP primers F9-R8. Winter rapeseed varieties: Severyanin (1), Stolychniy (2), VIK 2 (3), Nord (4), Laureat (5), Gorizont (6), Garant (7); spring rapeseed varieties: Vikros (8), Novik (9), Novosel (10), Veles (11), Grant (12), Podmoskovniy (13), Lugovskoy (14), Bizon (15). Winter turnip rape: Zarya (16); spring turnip rape: Nadezhda (17), Svetlana (18). H₂O control (19). M – molecular weight marker (100 kb DNA Ladder).

Table 3. Genetic similarity indices (above the diagonal) and Nei's distances (below the diagonal) calculated from SRAP analysis results

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	****	0.9655	0.8966	0.9655	0.9655	0.9138	0.9828	0.8103	0.7931	0.7586	0.8448	0.7586	0.7931	0.7586	0.8103
2	0.0351	****	0.8966	0.9655	0.9655	0.9138	0.9828	0.7759	0.7586	0.7241	0.8103	0.7241	0.7586	0.7241	0.7759
3	0.1092	0.1092	****	0.931	0.8966	0.9138	0.9138	0.7414	0.7586	0.7586	0.7759	0.7241	0.7586	0.7586	0.7414
4	0.0351	0.0351	0.0715	****	0.9655	0.9483	0.9828	0.8103	0.7931	0.7586	0.8448	0.7586	0.7931	0.7586	0.8103
5	0.0351	0.0351	0.1092	0.0351	****	0.9138	0.9828	0.8103	0.7931	0.7586	0.8448	0.7586	0.7931	0.7586	0.8103
6	0.0902	0.0902	0.0902	0.0531	0.0902	****	0.931	0.7586	0.7414	0.7069	0.7931	0.7069	0.7759	0.7414	0.7586
7	0.0174	0.0174	0.0902	0.0174	0.0174	0.0715	****	0.7931	0.7759	0.7414	0.8276	0.7414	0.7759	0.7414	0.7931
8	0.2103	0.2538	0.2992	0.2103	0.2103	0.2763	0.2318	****	0.9483	0.9138	0.9655	0.9138	0.9483	0.8793	1
9	0.2318	0.2763	0.2763	0.2318	0.2318	0.2992	0.2538	0.0531	****	0.8966	0.9483	0.8966	0.9655	0.8621	0.9483
10	0.2763	0.3228	0.2763	0.2763	0.2763	0.3469	0.2992	0.0902	0.1092	****	0.9138	0.8966	0.8966	0.8621	0.9138
11	0.1686	0.2103	0.2538	0.1686	0.1686	0.2318	0.1892	0.0351	0.0531	0.0902	****	0.9138	0.9483	0.8793	0.9655
12	0.2763	0.3228	0.3228	0.2763	0.2763	0.3469	0.2992	0.0902	0.1092	0.1092	0.0902	****	0.8966	0.931	0.9138
13	0.2318	0.2763	0.2763	0.2318	0.2318	0.2538	0.2538	0.0531	0.0351	0.1092	0.0531	0.1092	****	0.8966	0.9483
14	0.2763	0.3228	0.2763	0.2763	0.2763	0.2992	0.2992	0.1286	0.1484	0.1484	0.1286	0.0715	0.1092	****	0.8793
15	0.2103	0.2538	0.2992	0.2103	0.2103	0.2763	0.2318	0	0.0531	0.0902	0.0351	0.0902	0.0531	0.1286	****

Note. No. 1–15 – rapeseed varieties Severyanin, Stolychniy, VIK 2, Nord, Laureat, Gorizont, Garant, Vikros, Novik, Novosel, Veles, Grant, Podmoskovniy, Lugovskoy, Bizon.

The results of PCR analysis for SSR and SRAP markers were used to determine the genetic variability indices and build an UPGMA dendrogram depicting the varieties' phylogenetic relationships. The variety material had a low degree of genetic heterogeneity, while higher values of expected heterozygosity (H_e) and the number of effective alleles (n_e) were determined with SSR markers: 0.25 on average against 0.14 and 1.47 per locus if compared to 1.24, respectively. However, the SRAP method has enabled obtaining more PCR products applicable for varietal differentiation (Table 4).

Analysis of the UPGMA dendrogram demonstrated that the winter/spring rapeseed varieties were divided into two distinguishable clusters (Fig. 4). The first one united such winter cultivars as Severyanin, Garant, Stolychniy, Nord, Laureat, Gorizont, VIK 2; the second – all the spring ones. In the winter cluster VIK 2 and Gorizont were the most distant from the other varieties. The distances between Stolychniy, Nord, Laureat as well as between Garant and Severyanin were much shorter, which was confirmed by their high genetic similarity indices being 0.9655 and 0.9828, respectively (see Table 3). The most distant among spring rapeseed were two-zero varieties Novosel, Grant and Lugovskoy, which had the longest genetic distances in the cluster (0.3469 and 0.3228). Bizon and Vikros belonged to one subgroup, sharing a common branch of the dendrogram.

Discussion

The bulk strategy of DNA sampling from 30 germinants per variety has significantly reduced the labor efforts and cost of the research if compared to the traditional method of individual sample genotyping. The method has proved its efficiency for different cultures especially in large-scale studies of vast populations (Liu et al., 2018). However, this approach is only justified if the analyzed set of samples is representative. For cross-pollinating species with a high level of intrapopulation variations, it should include at least 30–50 plants per variety, which significantly increases the likelihood of registering a rare alleles, the occurrence of which in the population does not exceed 10 % (Crossa, 1989; Semerikov et al., 2002). The plants of winter rapeseed are known for their high self-pollination capacity (up to 70 % of flowers) (Shpaar, 2012), many varieties are linear; while in spring rapeseed this capacity reaches 40 % (Osipova, 1998). That's why in our study we used budk samples that combined 30 seedlings from each variety.

A significant part of SSR primers tested in our study generated monomorphic amplification fragments. They did not allow us to properly estimate the genetic variability and had low reproducibility in replicated experiments. A proportion of the markers proven effective for intervarietal DNA polymorphism detection comprised 16.7 %, being much lower than in

Table 4. The indexes of genetic variability of rapeseed varieties according to the results of SSR and SRAP analysis

Marker	PCR fragment size, bp	Total number of PCR fragments	Number of effective alleles (ne)	Heterozygosity (He)	Shannon index (I)
SRAP markers					
¹ F13-R7	266–624	6	1.20	0.12	0.18
¹ F10-R8	248–1140	7	1.33	0.18	0.25
¹ F13-R9	353–1674	6	1.18	0.12	0.20
¹ F9-R9	145–849	4	1.25	0.12	0.17
¹ Me4-R7	300–806	6	1.02	0.02	0.04
¹ F11-Em2	175–712	7	1.23	0.14	0.22
¹ F10-R7	132–628	7	1.51	0.27	0.38
¹ F9-Em2	209–1025	10	1.27	0.17	0.28
¹ F13-R8	107–688	5	1.20	0.10	0.14
Mean	–	6.44	1.24	0.14	0.21
SSR markers					
² Na12A02	159–209	5	1.54	0.32	0.49
² Ni2C12	110–140	3	1.67	0.13	0.22
² Bna.M.010	145–173	3	1.25	0.19	0.34
³ Ni03H07a	132–1000	8	1.88	0.47	0.66
³ NiF02a	175–736	5	1.20	0.15	0.27
³ Ni02-D08a	135–1200	8	1.41	0.23	0.35
³ Ra02E01a	191–928	10	1.37	0.23	0.36
Mean	–	6.00	1.47	0.25	0.38

Note. According to the data of ¹ (Rhouma et al., 2017); ² (Сатина, 2010); ³ (Chandra et al., 2013).

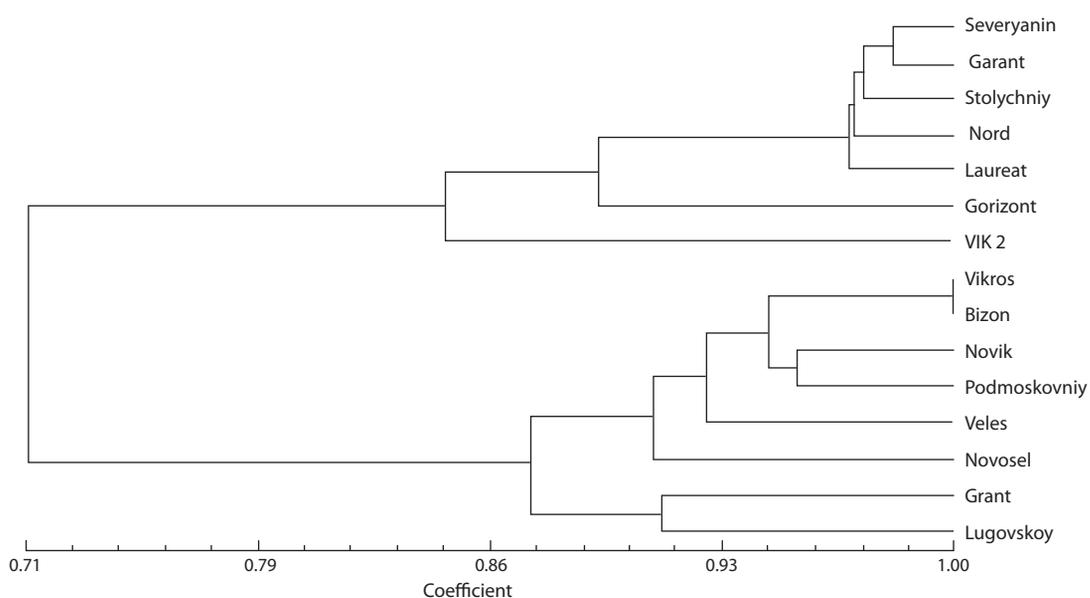


Fig. 4. Genetic similarity dendrogram for rapeseed varieties of Federal Williams Research Center of Forage Production and Agroecology.

other studies (Plieske, Struss, 2001; Hasan et al., 2008; Tian et al., 2017). It was probably due to the composition of the tested collection that had a narrow genetic basis considering the varieties' pedigree. At the same time, such parameters of genetic variability as the number of allelic variants, single-allele frequency, PIC and H_e were comparable to those found in published data (Satina, 2010; Klyachenko et al., 2018).

In general, the used markers made it possible to detect DNA polymorphism between rapeseed and turnip rape as well as between the winter and spring varieties within each species. However, Na12A02 marker turned out to be variety-specific for Bizon winter rapeseed and Zarya spring turnip rape, and Ra02-E01a – for VIK 2 winter rapeseed and Svetlana spring turnip rape. The unique alleles of Podmoskovniy and Lugovskoy rapeseed were detected using Ni02-D08a loci. The indicated markers can be used for varietal DNA identification and genetic certification.

SSR primers for the markers of Indian mustard's *Alternaria blight* resistance genes (Chandra et al., 2013), such as Ni02-D08a, Ni03H07a and RA02-E01a, proved to be the most effective. Their application enabled us to detect the specific amplification fragments for linear winter rapeseed variety VIK 2. They also proved effective for Gorizont, which had been obtained on the base of VIK 2 by seed freezing followed by their selection at low-temperature stress. These two varieties share high winter hardiness and are resistant to *Alternaria blight*. Thereby the results of our study can be useful for further selection of perspective breeding material and QTL analysis on disease resistance.

Among the spring rapeseed, Veles variety turned out to be substantially different while Lugovskoy and Garant had many similarities in the studied microsatellite parts of regions of the genome. Veles is a new perspective variety that has been approved for use since 2021 and was selected based on Vikros using the method of chemical mutagenesis, producing a high frequency of nucleotide changes. This is possibly the reason for Veles having unique alleles in three loci: Ni2C12, Ra02-E01a, Na12A02. For Vikros variety, a specific DNA profile was also obtained with Ni2C12 marker.

Rapeseed Grant was selected using the method of interspecies and intervarietal hybridization of early-maturing foreign breeding samples and the high-yielding varieties Lugovskoy and Vikros, developed at Federal Williams Research Center of Forage Production and Agroecology. Their common origin is probably the reason for the genetic similarity found between Grant and Lugovskoy varieties.

In general, SSR analysis failed to achieve optimum effect in identification of the investigated varieties: from the total set, including 42 primers for microsatellite genome loci, only four were attested as variety-specific for rapeseed, and only one (Ni03H07a) – for Nadezhda spring turnip rape.

For further investigation of DNA polymorphism, SRAP analysis was applied. SRAP is the third generation of molecular markers that were initially designed for the genes of *B. oleracea* L. (Li, Quiros, 2001) and are successfully used these days for genetic variability estimation and genetic mapping in different plants (Aneja et al., 2012; Rhouma et al.,

2017; Liu et al., 2018). This is a cheap, effective and highly reproducible technique.

In our study six combinations of SRAP primers were determined as applicable for identification of eight rapeseed and one turnip rape varieties. Two combinations of these informative primers (F10-R7 and F9-Em2) can be effectively used for both interspecies and intervarietal differentiation within a tested set.

The final dendrogram of phylogenetic relations made it possible to visually estimate the degrees of genetic similarities and differences of the studied material. For instance, close placing of such rapeseed varieties as Stolychniy, Nord and Laureat was probably determined by the features of their origin: they were selected for winter hardiness from a combination, in which one of the parental forms was Promin', a well-known winter rapeseed variety.

Garant, selected for winter hardiness, and Severyanin, which was obtained by seed freezing in a climatic chamber and the following individual-family selection, turned out to be in the common subgroup and at a short genetic distance (0.0174) from each other. In addition to high winter hardiness, these varieties are resistant to lodging and to damage by pathogenic fungi.

A two-zero spring variety Novosel takes a special position in his group (Nei's distance is 0.3469). Novosel was developed based on the foreign breeding samples and Russian varieties Lugovskoy and Vikros, characterized by early maturing and high yield. Specific properties of the new breeding achievement are shorter maturation period in comparison to standard varieties and high resistance to *Alternaria blight*.

Spring rapeseed Bizon and Vikros take the common branch of the dendrogram. The varieties were developed using the method of interspecies hybridization but from different parental forms; characterized by high yield productivity, early maturation and low glucosinolate content.

Conclusion

The presented study has proved the efficiency of SSR and SRAP markers for estimation of DNA polymorphism in rapeseed and turnip rape varieties developed in Federal Williams Research Center of Forage Production and Agroecology. During the study, SRAP technique has demonstrated a higher level of informativity: 36 % of the tested markers were polymorphic, while for the microsatellite loci this rate did not exceed 16.7 %.

Both techniques of molecular analysis enabled detecting the DNA markers for identification of 10 out of 15 rapeseed varieties tested and for 2 turnip rape samples. Microsatellite loci Na12A02, Ni2C12, Ra02-E01 and Ni02-D08a allowed obtaining unique PCR products for Bizon, Veles, Vikros, VIK 2, Podmoskovniy and Lugovskoy rapeseed varieties. Marker Ni03H07a proved effective for identifying Nadezhda turnip rape. In the used SRAP test kit, such primers as F13-R9, Me4-R7, F11-Em2, F10-R7, F9-Em2 and F9-R8 proved effective for detecting variety-specific amplicons or obtaining unique DNA profiles for different types of plants (winter/spring) in rapeseed varieties Grant, Novosel, Gorizont,

Stolychniy, Lugovskoy, Podmoskovniy and in spring turnip rape Svetlana.

The results of the study can be used for development of the perspective breeding samples and hybrids, for genetic certification and seed material purity control.

References

- Agarwal M., Shrivastava N., Padh H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* 2008;27(4):617-631. DOI 10.1007/s00299-008-0507-z.
- Amosova A.V., Zemtsova L.V., Grushetskaya Z.E., Samatadze T.E., Mozgova G.V., Pilyuk Y.E., Volovik V.T., Melnikova N.V., Zeleinin A.V., Lemesh V.A., Muravenko O.V. Intraspecific chromosomal and genetic polymorphism in *Brassica napus* L. detected by cytogenetic and molecular markers. *J. Genet.* 2014;93:133-143. DOI 10.1007/s12041-014-0351-6.
- Ananga A.O., Ceberet E., Ochieng J.W., Kumar S., Kambiranda D., Vasanthaiiah H., Tsolova V., Senwo Z., Konan K., Anike F.N. Prospects for transgenic and molecular breeding for cold tolerance in canola (*Brassica napus* L.). In: Akpan U.G. (Ed.). Oilseeds. London: IntechOpen, 2012;108-129. DOI 10.5772/32721.
- Aneja B., Yadav N.R., Chawla V., Yadav R.C. Sequence related amplified polymorphism (SRAP) molecular marker system and its applications in crop improvement. *Mol. Breeding.* 2012;30:1635-1648. DOI 10.1007/s11032-012-9747-2.
- Chandra V., Pant U., Bhalan R., Singh A.K. Studies on genetic diversity among *Alternaria* blight tolerant Indian mustard genotypes using SSR markers. *Bioscan.* 2013;8(4):1431-1435.
- Chen G., Geng J., Rahman M., Liu X., Tu J., Fu T., Li G., McVetty P.B.E., Tahir M. Identification of QTL for oil content, seed yield, and flowering time in oilseed rape (*Brassica napus*). *Euphytica.* 2010;175:161-174. DOI 10.1007/s10681-010-0144-9.
- Chesnokov Yu.V. Genetic markers: comparative classification of molecular markers. *Ovoshchi Rossii = Vegetable Crops of Russia.* 2018;3:11-15. DOI 10.18619/2072-9146-2018-3-11-15. (in Russian)
- Chesnokov Y.V., Artemyeva A.M. Evaluation of the measure of polymorphism information of genetic diversity. *Agric. Biol.* 2015;50(5):571-578. DOI 10.15389/agrobiol.2015.5.571eng.
- Crossa J. Methodologies for estimating the sample-size required for genetic conservation of outbreeding crops. *Theor. Appl. Genet.* 1989;77:153-161.
- Dellaporta S.L., Wood J., Hicks J.B. A plant DNA mini preparation: Version II. *Plant Mol. Biol. Rep.* 1983;1(4):19-21. DOI 10.1007/BF02712670.
- Gao M., Li G., Yang B., Qiu D., Farnham M., Quiros C.F. High-density *Brassica oleracea* map: Identification of useful new linkages. *Theor. Appl. Genet.* 2007;115(2):277-287. DOI 10.1007/s00122-007-0568-3.
- Geng J., Javed N., McVetty P.B.E., Li G., Tahir M. An integrated genetic map for *Brassica napus* derived from double haploid and recombinant inbred populations. *Hered. Genet.* 2012;1(1):103. DOI 10.4172/2161-1041.1000103.
- Hasan M., Friedt W., Pons-Kühnemann J., Freitag N.M., Link K., Snowdon R.J. Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). *Theor. Appl. Genet.* 2008;116(8):1035-1049. DOI 10.1007/s00122-008-0733-3.
- Khlestkina E.K. Molecular methods of analysis of structural and functional organization of genes and genomes in higher plants. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2011;15(4):757-768. (in Russian)
- Kirby K.S., Cook E.A. Isolation of deoxyribonucleic acid from mammalian tissues. *Biochem. J.* 1967;104(1):254-257. DOI 10.1042/bj1040254.
- Klimenko I.A., Kostenko S.I., Mavlyutov Yu.M., Shamustakimova A.O. The efficiency of SSR-and PawS markers for genetic polymorphism evaluation of red clover (*Trifolium pratense* L.) varieties. *Trudy po Prikladnoy Botanike, Genetike i Selekcii = Proceedings on Applied Botany, Genetics, and Breeding.* 2020a;181(3):100-109. DOI 10.30901/2227-8834-2020-3-100-109. (in Russian)
- Klimenko I.A., Kozlov N.N., Kostenko S.I., Shamustakimova A.O., Mavlyutov Yu.M. Identification and Certification of Forage Grass Varieties (Red Clover, Sand Alfalfa, Common Alfalfa, and Black Medick Alfalfa) on the Base of DNA Markers: Methodological Recommendations. Moscow: Ugresha T Publ., 2020b. DOI 10.33814/978-5-6043194-9-9. (in Russian)
- Klyachenko O.L., Prysiazniuk L.M., Shofolova N.V., Piskova O.V. Polymorphism in spring and winter rapeseed varieties (*Brassica napus* L.) identified by SSR markers. *Plant Var. Stud. Prot.* 2018;14(4):366-374. DOI 10.21498/2518-1017.14.4.2018.151898.
- Kosolapov V.M., Shamsutdinov Z.Sh., Kostenko S.I., Pilipko S.V., Tyurin Yu.S., ..., Ivanov I.S., Saprykina N.V., Truzina L.A., Chuiikov V.A., Georgiadi N.I. Varieties of Forage Crops Bred at Williams Federal Scientific Center for Fodder Production and Agroecology. Moscow: Ugreshskaya Tipografiya Publ., 2019. (in Russian)
- Lemesh V.A., Mozgova G.V., Grushetskaya Z.E., Sidorenko E.V., Pilyuk Ya.E., Bakanovskaya A.V. The use of specific DNA markers for the identification of alleles of the *FAD3* genes in rape (*Brassica napus* L.). *Russ. J. Genet.* 2015;51(8):765-773. DOI 10.1134/S1022795415080049.
- Li G., Quiros C.F. Sequence-related amplified polymorphism (SRAP) a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theor. Appl. Genet.* 2001;103(2):455-461. DOI 10.1007/s001220100570.
- Liu S., Feuerstein U., Luesink W., Schulze S., Asp T., Studer B., Becker H.C., Dehmer K.J. DaRT, SNP, and SSR analyses of genetic diversity in *Lolium perenne* L. using bulk sampling. *BMC Genetics.* 2018;19(1):10. DOI 10.1186/s12863-017-0589-0.
- Mozgova G.V., Khoruzhy N.E., Amosova A.V., Pilyuk Ya.E., Belyavskiy V.M., Khramchenko S.Yu., Muravenko O.V., Lemesh V.A. Genetic polymorphism of *Brassica napus* related to cold tolerance. *Molekulyarnaya i Prikladnaya Genetika = Molecular and Applied Genetics.* 2019;26:34-44. (in Russian)
- Nei M., Li W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA.* 1979;76(10):5269-5273. DOI 10.1073/pnas.76.10.5269.
- Osipova G.M. Rapeseed in Siberia: Morphological, Genetic, and Breeding Aspects. Novosibirsk: Siberian Branch of the Russian Academy of Agricultural Sciences, 1998. (in Russian)
- Piquemal J., Cinquin E., Couton F., Rondeau C., Seignoret E., Doucet I., Perret D., Villeger M.-J., Vincourt P., Blanchard P. Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. *Theor. Appl. Genet.* 2005;111(8):1514-1523. DOI 10.1007/s00122-005-0080-6.
- Plieske J., Struss D. Microsatellite markers for genome analysis in *Brassica*. I. Development in *Brassica napus* and abundance in *Brassicaceae* species. *Theor. Appl. Genet.* 2001;102(5):689-694. DOI 10.1007/s001220051698.
- Rhouma H.B., Taski-Ajdukovic K., Zitouna N., Sdouga D., Milic D., Trifi-Farah N. Assessment of the genetic variation in alfalfa genotypes using SRAP markers for breeding purposes. *Chil. J. Agric. Res.* 2017;77(4):332-339. DOI 10.4067/S0718-58392017000400332.
- Rogozhina T.G., Aniskina Yu.V., Karpachev V.V., Shilov I.A. An application of microsatellite analysis for detection of biotypes from spring rapeseed cultivars (*Brassica napus* L.). *Maslichnye Kultury. Nauchno-Tekhnicheskiy Byulleten Vserossiyskogo Nauchno-Issledovatel'skogo Instituta Maslichnykh Kultur = Oil Crops. Scientific and Technical Bulletin of the All-Russian Research Institute of Oil Crops.* 2015;2(162):27-33. (in Russian).

- Rohlf F.J. NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System. Version 2.10 manual. Applied Biostatistics. Inc. New York: Exeter Software, 2000.
- Satina T.G. A protocol for genotyping based on microsatellite analysis in rapeseed (*Brassica napus* L.) breeding: Cand. Sci. (Biol.) Dissertation. Moscow, 2010. (in Russian)
- Semerikov V.L., Belyaev A.Y., Lascoux M. The origin of Russian cultivars of red clover (*Trifolium pratense* L.) and their genetic relationships to wild populations in the Urals. *Theor. Appl. Genet.* 2002; 106:127-132.
- Sharma M., Dolkar D., Salgotra R., Sharma D., Singh P.A., Gupta S.K. Molecular marker assisted confirmation of hybridity in Indian mustard (*Brassica juncea* L.). *Int. J. Curr. Microbiol. Appl. Sci.* 2018; 7(9):894-900. DOI 10.20546/ijcmas.2018.709.107.
- Snowdon R.J., Friedt W. Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. *Plant Breeding.* 2004; 123(1):1-8. DOI 10.1111/j.1439-0523.2003.00968.x.
- Shpaar D. Rapeseed and Wild Cabbage: Cultivation, Harvesting, Storage, and Utilization. Kiev: Zerno Publ., 2012. (in Russian)
- State Register of Selection Achievements Admitted for Use for Production Purposes. Vol. 1. Plant Varieties (official publication). Moscow: Rosinformagrotekh Publ., 2021. (in Russian)
- Tian H.Yu., Channa S.A., Hu Sh. Relationships between genetic distance, combining ability and heterosis in rapeseed (*Brassica napus* L.). *Euphytica.* 2017;213(1):1-11. DOI 10.1007/s10681-016-1788-x.
- Volovik V.T. Brassicas: the Economic Importance. In: The Basic Species and Varieties of Fodder Crops: Results of the Research Activity of the Central Breeding Center. Moscow: Nauka Publ., 2015;249-253. (in Russian)
- Yeh F.C., Yang R., Boyle T.J., Ye Z., Xiyan J.M. PopGene 32, Microsoft window-based freeware for population genetic analysis, version 1.32. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Alberta, Canada, 2000.

ORCID ID

I.A. Klimenko orcid.org/0000-0002-1850-3859
A.A. Antonov orcid.org/0000-0002-7684-0503
A.O. Shamustakimova orcid.org/0000-0003-3535-3108
Y.M. Mavlyutov orcid.org/0000-0002-5695-6242

Acknowledgements. The presented investigation was supported by the means of the federal budget, directed for performing the government assignment (project No. 0442-2019-0001AAAA-A19-119122590053-0).

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

Received August 2, 2021. Revised March 9, 2022. Accepted March 9, 2022.

Identification of the gene coding for seed cotyledon albumin SCA in the pea (*Pisum L.*) genome

A.V. Mglinets, V.S. Bogdanova, O.E. Kosterin

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
kosterin@bionet.nsc.ru

Abstract. Albumins SCA and SAA are short, highly hydrophilic proteins accumulated in large quantities in the cotyledons and seed axes, respectively, of a dry pea (*Pisum sativum L.*) seed. SCA was earlier shown to have two allelic variants differing in mobility in polyacrylamide gel electrophoresis in acid medium. Using them, the corresponding gene SCA was mapped on Linkage Group V. This protein was used as a useful genetic and phylogeographical marker, which still required electrophoretic analysis of the protein while the DNA sequence of the corresponding SCA gene remained unknown. Based on the length, the positive charge under acidic conditions and the number of lysine residues of SCA and SAA albumins, estimated earlier electrophoretically, the data available in public databases were searched for candidates for the SCA gene among coding sequences residing in the region of the pea genome which, taking into account the synteny of the pea and *Medicago truncatula* genomes, corresponds to the map position of SCA. Then we sequenced them in a number of pea accessions. Concordance of the earlier electrophoretic data and sequence variation indicated the sequence *Psat0s797g0160* of the reference pea genome to be the SCA gene. The sequence *Psat0s797g0240* could encode a minor related albumin SA-a2, while a candidate gene for albumin SAA is still missing (as well as electrophoretic variation of both latter albumins). DNA amplification using original primers SCA1_3f and SCA1_3r from genomic DNA and restriction by endonuclease *HindIII* made it possible to distinguish the SCA alleles coding for protein products with different charges without sequencing the gene. Thus, the gene encoding the highly hydrophilic albumin SCA accumulated in pea seeds, the alleles of which are useful for classification of pea wild relatives, has now been identified in the pea genome and a convenient CAPS marker has been developed on its basis.

Key words: late embryogenesis proteins; seed cotyledon albumin; peas; *Pisum sativum L.*; CAPS marker.

For citation: Mglinets A.V., Bogdanova V.S., Kosterin O.E. Identification of the gene coding for seed cotyledon albumin SCA in the pea (*Pisum L.*) genome. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(4):359-364. DOI 10.18699/VJGB-22-43

Идентификация гена, кодирующего альбумин семядолей SCA, в геноме гороха (*Pisum L.*)

А.В. Мглинец, В.С. Богданова, О.Э. Костерин

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
kosterin@bionet.nsc.ru

Аннотация. Альбумины SCA и SAA – короткие гидрофильные белки, содержащиеся в высокой концентрации в семядолях и оси семени сухих семян гороха (*Pisum sativum L.*). Ранее показано, что альбумин SCA имеет два аллельных варианта, различающихся подвижностью в электрофорезе в полиакриламидном геле в кислой среде. С их помощью соответствующий ген SCA картирован в группе сцепления V. Белок SCA был использован как генетический и филогеографический маркер, что до сих пор предполагало проведение электрофореза белков, тогда как последовательность кодирующего гена SCA оставалась неизвестной. На основе данных, доступных в публичных репозиториях, в районе генома гороха, соответствующего позиции гена SCA на генетической карте с учетом синтении геномов гороха и люцерны, осуществлен поиск кандидатов на роль этого гена в зависимости от длины его белкового продукта, положительного заряда в кислых условиях и количества остатков лизина, ранее оцененного электрофоретическими методами. Выявленные гены просеквенированы у ряда образцов гороха. Соответствие полученных электрофоретических данных и нуклеотидной изменчивости позволило идентифицировать последовательность *Psat0s797g0160* из референсного генома гороха в качестве гена SCA. Последовательность *Psat0s797g0240*, возможно, кодирует родственный минорный альбумин SA-a2, тогда как ген-кандидат альбумина SAA остается неидентифицированным (как и электрофоретическая изменчивость двух белков, упомянутых последними). Амплификация ДНК с использованием оригинальных праймеров SCA1_3f и SCA1_3r и ге-

номной ДНК в качестве матрицы, а также расщепление ее эндонуклеазой рестрикции *Hind*III позволяющая различать аллели гена *SCA*, белковые продукты которых имеют разный заряд, без секвенирования. Таким образом, ген, кодирующий высокогидрофильный альбумин *SCA*, накапливающийся в семядолях гороха, аллели которого полезны для классификации диких родственников культурного гороха, идентифицирован в геноме гороха, и на его основе разработан удобный CAPS-маркер.

Ключевые слова: белки позднего эмбриогенеза; альбумины семядолей; горох; *Pisum sativum L.*; CAPS-маркер.

Introduction

Mature pea (*Pisum sativum L.*) seeds contain a large amount of protein families, which are generally classified as globulins (soluble in salt solutions) and albumins (soluble in water), the former being mostly storage proteins, the latter having a number of functions, including substitution of water in dry tissues (Smirnova et al., 1990). A number of albumins are extractable from seed flour with 5 % perchloric acid, which were characterised in detail by O.G. Smirnova et al. (1990, 1992). Special attention was drawn to the two most abundant of these albumins, which are biochemically and immunologically related and quite short (about 100 amino acid residues) highly hydrophilic peptides. One of them predominates in the cotyledons and the other in the seed axis, both accumulating during seed formation and depleting during germination. They were respectively named *SCA* (seed cotyledon albumin) and *SAA* (seed axis albumin). Their amino acid content and the accumulation pattern in seeds left no doubt as to their participation in substitution of water in dry seed cells, that is in a dehydrin-like function, although they differ in many respects from the known dehydrins and are much smaller proteins (Smirnova et al., 1992). While the *SAA* protein was electrophoretically monomorphic in peas, two allelic variants of *SCA* were revealed to differ in electrophoretic mobility in 15 % polyacrylamide gels containing acetic acid and urea according to S. Panyim and R. Chalkley (1969), which allowed to genetically map the relevant gene *SCA* (Smirnova et al., 1992; Rozov et al., 1993; Gorel et al., 1998) on linkage group V (corresponding to chromosome 3; Smýkal et al., 2012; Kreplak et al., 2019).

O.G. Smirnova et al. (1992) found out that the fast electromorph *SCA^f* was frequent in the wild subspecies of the common pea (*P. sativum* subsp. *elatius* Aschers. et Graebn.) but was extremely rare in the cultivated subspecies *P. sativum L.* subsp. *sativum*. (Here, the inclusive taxonomic system of peas according to N. Maxted and M. Ambrose (2001) is followed.) O.E. Kosterin and V.S. Bogdanova (2008) and O.E. Kosterin et al. (2010) noticed a strong concordance of the occurrence of *SCA^f* with that of the plastid *rbcL* allele containing a recognition site for the *Hsp* AI restriction endonuclease, and a less strong concordance with that of the mitochondrial *cox1* allele containing the recognition site for the *Psi* I restriction endonuclease. This concordance was interpreted in terms of the common phyletic origin and as evidence of the existence of two different wild pea lineages. Based on this, different combinations (A, B and C) of alleles of the three mentioned dimorphic marker genes *SCA*, *rbcL* and *cox1* from different cellular genomes, respectively nuclear, plastid and mitochondrial, were proposed for a simple classification of evolutionary lineages of the wild pea subspecies *P. sativum* subsp. *elatius* (Kosterin, Bogdanova, 2008; Kosterin et al., 2010), which

was then used repeatedly (Zaytseva et al., 2012, 2015, 2017; Kosterin, Bogdanova, 2021; Bogdanova et al., 2021). So, the electromorphs of *SCA* appeared to be useful in the studies of genetic diversity of the pea crop wild relatives, which are important for the involvement of their potentially useful genetic resources into breeding (Ali et al., 1994; Maxted, Kell, 2009; Coyne et al., 2011; Ford-Lloyd et al., 2011; Maxted et al., 2012). However, while the plastidic and mitochondrial markers were scored by the CAPS approach involving DNA amplification and restriction, the *SCA* gene sequence remained unknown and analysis of this marker required more laborious protein electrophoresis. Molecular identification of this gene would be desirable to facilitate the analysis. This became possible when the pea nuclear genome was published by K. Kreplak et al. (2019). This communication is devoted to identification of the *SCA* gene in the pea genome, its brief characterisation and working out a convenient CAPS marker based on this gene.

Materials and methods

The gene *SCA* was sequenced from samples of DNA extracted in the course of our previous work (Kosterin, Bogdanova, 2008), from the following pea germplasm accessions: 721 (Israel), CE1 (Crimea), CE11 (= JI3557, Portugal), JI1794 (Golan Heights), L100 (Israel), PI344538, Pse001, Pe013, P015, P017 (Turkey), VIR320' (Palestine) (*P. sativum* subsp. *elatius*), VIR3429 (Egypt), VIR4911 (Tibet), VIR5414 (Ethiopia), VIR7335 (Tajikistan), WL1238 (a testerline), cultivar Cameor (*P. sativum* subsp. *sativum*), VIR4871 (*Pisum sativum* subsp. *transcaucasicum* Govorov), VIR2759 (Ethiopia) (*Pisum abyssinicum* A. Br.), and WL2140 (Israel) (*P. fulvum* Sibth. et Smith). VIR accessions were received from N.I. Vavilov All-Russian Institute of Plant Genetic Resources, Saint-Petersburg, accessions 721, Pse001, Pe013, P15, P17, JI1794, L100 and Cameor were kindly provided by Dr. Norman Weeden, Cornell University, New York, accession PI344538 was kindly provided by Dr. Petr Smýkal, Olomouc University, the progenitors of accessions CE1 and CE11 were collected in nature by the third author. Polymerase chain reaction to amplify the fragment corresponding to *Psat0s797g0160* was carried out using BIS 208 cycler in 20 µl of the PCR reaction mixture under a mineral oil layer with the following cycling parameters: 95 °C for 3 min; 45 cycles consisting of denaturation at 94 °C for 30 s, annealing at 56 °C for 25 s and elongation at 72 °C for 40 s; final elongation at 72 °C for 5 min; for amplification of the fragment corresponding to *Psat0s797g0240* the same parameters were used but annealing was at 59 °C. For restriction analysis, 5 µl of the resulting reaction mixture were digested with 1 unit of *Hind*III endonuclease according to manufacturer's instructions and the products analysed on 1.5 % agarose gel in TAE buffer. The 100-bp ladder (SibEn-

zyme, Novosibirsk) was used as a molecular mass marker. For sequencing, PCR products were purified by 20 % polyethylene glycol 800 in 2.5 M NaCl. Sanger reaction was carried out using BrightDye Terminator version 3-100 (Nimagen, Netherlands) with the conditions recommended by the manufacturer for 50 cycles. The Sanger reaction products were purified using Sephadex G-75. Sequencing was carried out in Genomic Core Facility SB RAS, Novosibirsk. The sequences obtained in this study were submitted to European Nucleotide Archive with the following entry numbers OU953856- OU953865, OU953869- OU953881 for SCA and OU953866- OU953868, OU953882- OU953894 for alleles of *Psat0s797g0240*.

Results and discussion

Candidate gene for SCA seed albumin in public databases

The SCA gene is mapped on linkage group V (LGV) between the loci *His1* (coding for histone H1 subtype 1) and *coch* (*cochleata*) (Gorel et al., 1998) (its earlier published position behind *coch* (Rozov et al., 1993) was tentative as based on non-additive data with respect to these three loci). The potential candidates were searched in the annotated genome of *Medicago truncatula* Gaertn. making use of its synteny with the genome of *Pisum* (Kalo et al., 2004). The bordering markers *coch* and *His1* of pea correspond to the loci with Gene IDs 11417633 and 25499208, respectively, therefore suitable candidates found in *Medicago truncatula* should map to chromosome 7, syntenic to LGV of *P. sativum* (Kalo et al., 2004; Kreplak et al., 2019) at physical position between 42,203,622 (GeneID: 11417633) and 45,488,994 (GeneID: 25499208) on NC_053048.1 (*M. truncatula* strain A17 chromosome 7). This region contained 421 coding sequences, of which three neighbouring loci were annotated as “18 kDa seed maturation protein”. Two of them, *LOC11421661* and *LOC11437338*, encoded polypeptides of 105 and 101 amino acids, respectively. The third locus, *LOC11437936* encoded polypeptide of 177 amino acids. These polypeptides were used as a query to search the *P. sativum* genome assembly at <https://urgi.versailles.inra.fr/blast>. All three searches retrieved the same hits, *Psat0s797g0240* and *Psat0s797g0160*, separated by about 25 Kb on the scaffold 00797 not attributed to any chromosome and *Psat3g068920* on chr3LG5 with physical position between the loci *coch* and *His1*. *Psat3g068920* encoded polypeptide of 190 amino acids and probably corresponded to the *LOC11437936* of *M. truncatula* while *Psat0s797g0240* and *Psat0s797g0160* probably corresponded to *LOC11421661* and *LOC11437338*. Both encoded polypeptides of 101 amino acids and were concluded to be ideal candidates to represent the SCA locus.

Psat0s797g0160 with position 57,488–57,939 on scaffold00797, annotated as “Late embryogenesis abundant (LEA) group 1”, encoded a polypeptide of 101 amino acid residues including 18 positively charged residues, of which 10 were lysines. *Psat0s797g0240* with position 83,885–84,190 on scaffold00797, also annotated as “Late embryogenesis abundant (LEA) group 1”, encoded a polypeptide of 101 residues including 18 positively charged residues, of which 11 were lysines. The earlier obtained data on amino acid composition of the SCA protein were as follows. The slow electromorph SCA^s, common in cultivated peas, was estimated by the incomplete

succinylation method to possess 17 positively charged residues, including 9 lysines, and, together with data on the amino acid content, as being ca 107 residues long (Smirnova et al., 1992). This was a rather good correspondence of data from protein chemistry and sequencing. The estimate of 9 rather than 10 lysine residues in SCA^s with the incomplete succinylation method (Smirnova et al., 1992) may be explained by a tandem of two lysine residues in positions 47–48 of the deduced protein product, which could not both bind to succinic acid residues for steric reasons.

Hence the size and amino acid content of the protein products and genome location, as concluded from the synteny with *M. truncatula*, of both *Psat0s797g0160* and *Psat0s797g0240* corresponded to SCA (Smirnova et al., 1992; Gorel et al., 1998). To choose between them for a candidate for SCA we made use of the genetic data. SCA was shown to be dimorphic with allelic variants differing with respect to the number of positively charged amino acid residues while SAA was monomorphic in this respect (Smirnova et al., 1992). Sequences available in sequence read archives (SRA) at NCBI containing data of high throughput resequencing of pea accessions were used to assemble alleles of *Psat0s797g0160* and *Psat0s797g0240* from pea accessions W6_2107 (BioProject PRJNA431567), JI1794, WL2140 (*Pisum fulvum* Sibth. et Smith) (BioProject PRJNA431567), JI2202 (*Pisum abyssinicum* A.Br.) (BioProject PRJNA285605), 711 (BioProject PRJEB30482), 721 (BioProject PRJNA431567, PRJEB30482). Five of these accessions were involved in our previous electrophoretic studies of SCA (Kosterin, Bogdanova, 2008); Cameor and JI1794 were shown to have the slow electromorph SCA^s while WL2140, 711 and 721 had the fast electromorph SCA^f. There was some sequence variability among alleles of *Psat0s797g0160*. Two nucleotide substitutions differed in the alleles of Cameor, W6_2107 and JI1794 from those of WL2140, JI2202, 711, 721, namely T/G in the position 215 (from start codon) and A/C in the position 238. The T/G substitution changed valine for glycine, and the A/C substitution changed asparagine to histidine, which is positively charged under conditions of acetic acid-urea PAAGE used for SCA analysis. Thus, the latter amino acid replacement affects electrophoretic mobility of the encoded protein and is associated with the fast electromorph. Allelic variants of *Psat0s797g0240* did not carry amino acid substitutions associated with the change of electrophoretic mobility. This allowed us to nominate *Psat0s797g0160* as a candidate for the SCA gene.

Concordance of the sequence variation of the candidate gene for SCA with SCA electrophoretic pattern

To confirm *Psat0s797g0160* to be the SCA gene we resequenced it in 20 pea accessions in which SCA was previously studied electrophoretically (Smirnova et al., 1992; Kosterin, Bogdanova, 2008; Kosterin et al., 2010). To design primers matching 3' and 5' non-coding regions we searched public databases for pea sequences coding for the same protein as *Psat0s797g0160*. The search retrieved the sequence *PUCA013656022.1* (*Pisum sativum*, cultivar Gradus No 2 flattened_line_3009, whole genome shotgun sequence) containing the coding sequence as well as 5' and 3' non-coding regions. The primers Ps_SCA1_3f (5' GCAT

Variants of the deduced amino acid sequences encoded by SCA alleles sequenced from pea accessions
(polymorphic positions boldfaced and highlighted)

Accession	Amino acid sequence
Cameor ¹	MEKTKETAANVGAAAKSGMEKTKANVQEKTERLTTRDPLEKELATQKKEERVAQAE ^{LDKQAARNHNAAATAVNTLGGQNHHTTGTGGNPNATGYGTGGTHR}
VIR4871 ²	MEKTKETAANVGASAKSGMEKTKANVQEKTERLTTRDPLEKELATQKKEERVAQAE ^{LDKQAARNHNAAATAVNTLGGQNHHTTGTGGNPNATGYGTGGTHR}
L100 ³	MEKTKETAANVGASAKSGMEKTKANVQEKTERLTTRDPLEKELATQKKEERVAQAE ^{LDKQAARNHNAAATAGNTLVGGQNHHTTGTGGNPNATGYGTGGTHR}
721 ⁴	MEKTKETAANVGAAAKSGMEKTKANVQEKTERLTTRDPLEKELATQKKEERVAQAE ^{LDKQAARNHNAAATAGNTLVGGQNHHTTGTGGNPNATGYGTGGTHR}
Pse001	MEKTKETAANVGASAKSGMEKTKANVQEKTERLTTRDPLEKELATQKKEERVAQAE ^{LDKQAARNHNAAATAGNTLVGGQNHHTTGTGGNPNATGYGTGGTHR}
WL2140	MEKTKETAANVGAAAKSGMEKTKANVQEKTERLTTRDPLEKELATQKKEERVAQAE ^{LDKQAARNHNAAATAGNTLVGGQNHHTTGTGGNPNATGYGTGGTHR}
Amino acid type	hnp ^h pn ^a aaahahaha ^a ph ^h hnp ^a pahah ⁿ ph ⁿ ph ⁿ ra ⁿ pn ^a ah ^h pp ⁿ np ^a ah ^a nan ⁿ pha ^a ph ^h ph ^a aaah ^a ah ^h hh ^h h ^h ph ^h hh ^h hh ^h rh ^h ah ^h yh ^h hh ^h pp ^p

Note. The lowest line shows amino acid types encoded as follows: a – aliphatic, h – hydrophilic, n – negatively charged in neutral conditions, p – positively charged in neutral conditions, r – proline, y – aromatic (tyrosine) (charged types boldfaced). ¹ the same in WL1238, VIR4911, VIR_5414, JI1794, CE1, P017, W6_10925; ² the same in P015; ³ the same in VIR320, VIR2759, VIR3429, VIR7335; ⁴ the same in W6_26109, Pe013, CE11, PI344009, PI344538.

CATACTCTTCAACACAT) and Ps_SCA1_3r (5' GTAG GAACATTCACAACATCA) were designed to match those non-coding regions. We sequenced the coding region of the SCA gene in two groups of 10 pea accessions each, including: (i) 711, 721, VIR320³, CE11, PI344538, Pe013 (*P. sativum* subsp. *elatius*), VIR2759 (*Pisum abyssinicum* A.B.), VIR3429, VIR7335 (*P. sativum* subsp. *sativum*) and WL2140 (*P. fulvum*) which were shown (Kosterin and Bogdanova, 2008) to have the fast SCA electromorph (SCA^f), frequent in wild peas, and (ii) accessions CE1, JI1794, Pse001, P015, P017 (*P. sativum* subsp. *elatius*), VIR4871 (*Pisum sativum* subsp. *transcaucasicum*), VIR4911, VIR5414, WL1238 and Cameor (*P. sativum* subsp. *sativum*) which were shown to have the slow SCA electromorph (SCA^s), predominating overwhelmingly in the cultivated pea but occurring in wild peas as well. The derived amino acid sequences had 19 and 18 positively charged amino acid residues in the first and second group, respectively. Full correspondence of the electrophoretic and sequence data, together with the genomic position and the size and content of the protein product, indicate that *Psat0s797g0160* is the SCA gene (the latter designation will be used in the text below).

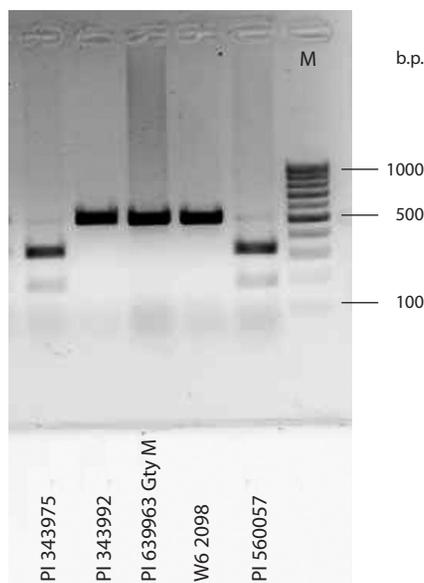
The related gene *Psat0s797g0240*

The gene *Psat0s797g0240* resides in the pea genome (cultivar Cameor) in about 25 Kb from *Psat0s797g0160*, the two loci have very similar sequence, and the difference in their coding sequences is 25 nucleotides (8.2 %). Obviously, these genes are paralogs originated by tandem duplication of a genome region. The inferred amino acid sequence of the *Psat0s797g0240* polypeptide product has the same length of 101 amino acid sequences and differs (in cv. Cameor) by eight amino acid substitutions from that of *Psat0s797g0160* (their positions are given in parenthesis): ala→ser (14), arg→lys (32), tre→ala (35), arg→his (36), tre→ser (70), val→gln (72), gln→arg (79), asn→his (80). Due to the two latter substitutions this polypeptide has two positively charged residues more than SCA. The mobility of polypeptides of equal length in the involved electrophoretic procedure in acid denaturing conditions is proportional to the number of positively charged residues, so we can expect the *Psat0s797g0240* product mobi-

lity to be 11 % greater than that of SCA^s. However, the SAA mobility is only 7 % greater than that of SCA^s (Smirnova et al., 1992). This is close with the mobility of SCA^f, which is 5 % greater than that of SCA^s (Smirnova et al., 1992), as expected from their difference of one positively charged residue. Most probably, *Psat0s797g0240* is not the gene coding for SAA but may encode an immunologically related minor protein SA-a2 with electrophoretic mobility 9 % greater than in SCA^s. Electrophoretic monomorphism of both SAA and SA-a2 so far observed (Smirnova et al., 1992) does not allow us to check these options genetically. We attempted resequencing the *Psat0s797g0240* alleles in the same set of accessions as for SCA (see above) using primers Ps_SCA2_1F (5' CACGTGTTCAATAATCTAACGC) and Ps_SCA2_1R (5' AAGAAAAAGAAACGAGCCATCA) matching the 5'- and 3' non-coding regions of PUCA011001169.1 (*Pisum sativum* cultivar Gradus No 2 flattened_line_64181, the whole genome shotgun sequence). Possibly, due to the abundance of poly-A and poly-T in the 5'- and 3' non-coding regions, respectively, amplification was not successful for 7 of 20 accessions involved (JI1794, 721, Pe013, P014, P017, VIR5414, WL1238, Cameor), so only 13 accessions were sequenced. The variation of protein products inferred from the obtained sequences was confined to 5 variable amino acid positions and did not affect electrophoretic mobility.

Protein product of SCA and its variation

The SCA gene has no introns. Its SCA protein product is remarkable for its extreme hydrophily and high content of charged (at neutral pH conditions) residues (Table). Among its 101 amino acid residues in Cameor, 70 are hydrophilic, of which 30 are charged, including 18 positively charged (lysine – 10, arginine – 5, histidine – 3), and 12 negatively charged (glutamate – 10, aspartate – 2) residues (the numbers of residues almost coinciding with their percentages) (see Table). Interestingly, 11 of 12 negatively charged residues have positively charged nearest neighbour(s) and 4 of 18 positively charged residues have negatively charged neighbour(s). There are tracts of three (glu-lys-glu) and five (lys-lys-glu-glu-arg) charged residues in a row. There are only two proline residues



An example of agarose gel electrophoresis of the PCR products obtained from genomic DNA of the indicated pea accessions with primers Ps_SCA1_3 and Ps_SCA1_3r, matching the SCA gene adjacent non-coding regions, digested with *Hind*III restriction endonuclease.

M stands for the molecular mass marker.

and only one aromatic residue (tyrosine). Such a content and structure, with alternating residues of opposite charge, suggest that in water solution, the SCA molecule has a rigid expanded (linear) structure.

Five variable amino acid positions were revealed in the SCA product: 14 (ala/ser), 72 (gly/val), 75 (leu/val), 80 (his/asn) and 95 (gly/val) (see Table), the fourth changing the molecule positive charge as discussed above. The SCA gene sequences obtained had 11 (3.6 %) variable nucleotide positions.

SCA gene and a CAPS marker based on it

The A→C substitution in position 238 of the SCA gene creates the recognition site GTCAAC for *Hind*III restriction endonuclease, missing in the rest of the gene. PCR amplification of genomic DNA with Ps_SCA1_3f and Ps_SCA1_3r primers resulted in a product of 512 bp, which was subsequently digested with *Hind*III endonuclease. The allele coding for SCA^f was cleaved into two fragments, 326 and 186 bp in size, while amplicon from the SCA^s allele remained 512 bp, the difference clearly seen on agarose gel (Figure). This makes the SCA gene the source of a convenient CAPS (cleaved amplified polymorphic sequence) marker, which permits scoring SCA allelic state without invoking protein electrophoresis. It should be noted that in spite of great similarity of the coding sequences of SCA and Psat0s797g0240, their adjacent non-coding regions appeared diverged enough to avoid cross-amplification.

References

Ali S.M., Sharma B., Ambrose M.J. Current status and future strategy in breeding pea to improve resistance to biotic and abiotic stresses. *Euphytica*. 1994;73:115-126. DOI 10.1007/BF00027188.
Bogdanova V.S., Shatskaya N.V., Mglinets A.V., Kosterin O.E., Vasiliev G.V. Discordant evolution of organellar genomes in peas (*Pisum L.*). *Mol. Phylogenet. Evol.* 2021;160:107136. DOI 10.1016/j.ympcv.2021.10713.6.

Coyne C.J., McGee R.J., Redden R.J., Ambrose M.J., Furman B.J., Miles C.A. Genetic adjustment to changing climates: pea. In: Yadav S.S., Redden R.J., Hatfield J.L., Lotze-Campen H., Hall A.E. (Eds.). *Crop Adaptation to Climate Change*. Oxford: John Wiley & Sons, Inc., 2011;238-250. DOI 10.1002/9780470960929.ch17.
Ford-Lloyd B.V., Schmidt M., Armstrong S.J., Barazani O., Engels J., Hadas R., Hammer K., Kell S.P., Kang D., Khoshbakht K., Li Y., Long C., Lu B.-R., Ma K., Nguyen V.T., Qiu L., Ge S., Wei W., Zhang Z., Maxted N. Crop wild relatives – undervalued, underutilized and under threat? *BioScience*. 2011;61(7):559-565. DOI 10.1525/bio.2011.61.7.10.
Gorel F.L., Rozov S.M., Berdnikov V.A. Mapping the locus *coch*. *Pisum Genet.* 1998;30:9-11.
Kalo P., Seres A., Taylor S.A., Jakab J., Kevei Z., Kereszt A., Endre G., Ellis T.H.N., Kiss G.B. Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Mol. Genet. Genomics*. 2004;272(3):235-246. DOI 10.1007/s00438-004-1055-z.
Kosterin O.E., Bogdanova V.S. Relationship of wild and cultivated forms of *Pisum L.* as inferred from an analysis of three markers, of the plastid, mitochondrial and nuclear genomes. *Genet. Resour. Crop Evol.* 2008;55(5):735-755. DOI 10.1007/s10722-007-9281-y.
Kosterin O.E., Bogdanova V.S. Reciprocal compatibility within the genus *Pisum L.* as studied in F₁ hybrids: 1. Crosses involving *P. sativum L.* subsp. *sativum*. *Genet. Resour. Crop Evol.* 2015;62(5):691-709. DOI 10.1007/s10722-014-0189-z.
Kosterin O.E., Bogdanova V.S. Reciprocal compatibility within the genus *Pisum L.* as studied in F₁ hybrids: 3. Crosses involving *P. sativum L.* subs. *elatius* (Bieb.) Aschers. et Graebn. *Br. Genet. Res. Crop Evol.* 2021;68(6):2565-2590. DOI 10.1007/s10722-021-01151-2.
Kosterin O.E., Zaytseva O.O., Bogdanova V.S., Ambrose M. New data on three molecular markers from different cellular genomes in Mediterranean accessions reveal new insights into phylogeography of *Pisum sativum L.* subsp. *elatius* (Beib.) Schmalh. *Genet. Resour. Crop Evol.* 2010;57(5):733-739. DOI 10.1007/s10722-009-9511-6.
Kreplak K., Madoui M.A., Cápál P., Novák P., Labadie K., Aubert G., Bayer P.E., Gali K.K., Syme R.A., Main D., Klein A., Bérard A., Vrbová I., Fournier C., d'Agata L., Belser C., Berrabah W., Toegelová H., Milec Z., Vrána J., Lee H., Kougbeadjo A., Térézold M., Huneau C., Turo C.J., Mohellibi N., Neumann P., Falque M., Gallardo K., McGee R., Tar'an B., Bendahmane A., Aury J.M., Batley J., Le Paslier M.C., Ellis N., Warkentin T.D., Coyne C.J., Salse J., Edwards D., Lichtenzweig J., Macas J., Doležel J., Wincker P., Burstin J. A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* 2019;51(9):1411-1422. DOI 10.1038/s41588-019-0480-1.
Maxted N., Ambrose M. Peas (*Pisum L.*). In: Maxted N., Bennett S.J. (Eds.). *Plant genetic resources of legumes in the Mediterranean*. Current Plant Science and Biotechnology in Agriculture, vol. 39. Dordrecht: Springer, 2001;181-190. DOI 10.1007/978-94-015-9823-1_10.
Maxted N., Kell S.P. Establishment of a global network for the *in situ* conservation of crop wild relatives: status and needs. *FAO Commission on Genetic Resources for Food and Agriculture*. Rome, 2009.
Maxted N., Kell S., Ford-Lloyd B., Dulloo E., Toledo Á. Toward the systematic conservation of global crop wild relative diversity. *Crop Sci.* 2012;52(2):774-785. DOI 10.2135/cropsci2011.08.0415.
Panyim S., Chalkley R. High resolution in acrylamide gel electrophoresis of histones. *Arch. Biochem. Biophys.* 1969;130(1):337-346. DOI 10.1016/0003-9861(69)90042-3.
Rozov S.M., Temnykh S.V., Gorel' F.L., Berdnikov V.A. A new version of pea linkage group 5. *Pisum Genet.* 1993;25:46-51.
Smirnova O.G., Rozov S.M., Kosterin O.E., Berdnikov V.A. Perchloric acid extractable low-M_r albumins SCA and SAA from cotyledons and seed axes of pea (*Pisum sativum L.*). *Plant Sci.* 1992;82(1):1-13. DOI 10.1016/0168-9452(92)90002-4.

- Smirnova O.G., Rozov S.M., Kosterin O.E. Perchloric acid extraction of pea seed proteins: characterisation and inheritance of electrophoretic variants. In: Shchapova A.I., Kolosova L.D. (Eds.). *Genetics of Economically Valuable Characteristics of Higher Plants*. Novosibirsk, 1990;158-179. (in Russian)
- Smýkal P., Aubert G., Burstin J., Coyne C.J., Ellis N.T., Flavell A.J., Ford R., Hýbl M., Macas I., Neumann P., McPhee K.E., Redden R.J., Rubiales D., Weller J.L., Warkentin T.D. Pea (*Pisum sativum L.*) in the genomic era. *Agronomy*. 2012;2(2):74-115. DOI 10.3390/agronomy2020074.
- Zaytseva O.O., Bogdanova V.S., Kosterin O.E. Phylogenetic reconstruction at the species and intraspecies levels in the genus *Pisum L.* (peas) using a histone H1 gene. *Gene*. 2012;504(2):192-202. DOI 10.1016/j.gene.2012.05.026.
- Zaytseva O.O., Bogdanova V.S., Mglinets A.V., Kosterin O.E. Refinement of the collection of wild peas (*Pisum L.*) and search for the area of pea domestication with a deletion in the plastidic *psbA-trnH* spacer. *Genet. Res. Crop Evol.* 2017;64:1417-1430. DOI 10.1007/s10722-016-0446-4.
- Zaytseva O.O., Gunbin K.V., Mglinets A.V., Kosterin O.E. Divergence and population traits in evolution of the genus *Pisum L.* as reconstructed using genes of two histone H1 subtypes showing different phylogenetic resolution. *Gene*. 2015;556(2):235-244. DOI 10.1016/j.gene.2014.

ORCID ID

O.E. Kosterin orcid.org/0000-0001-5955-4057

Acknowledgements. The work was supported by the Russian State Scientific Program No. FWNR-2022-0017. Sequencing was carried out in Genomic Core Facility SB RAS, Novosibirsk. Assembly of coding sequences from SRA archives was performed at Siberian Supercomputer Center Core Facility and Computational Facility of Novosibirsk State University. Plants were grown in the greenhouse at the SB RAS Artificial Plant Growing Facility.

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

Received February 3, 2022. Revised April 14, 2022. Accepted April 18, 2022.

Original Russian text www.bionet.nsc.ru/vogis/

Neuronal density in the brain cortex and hippocampus in *Clstn2*-KO mouse strain modeling autistic spectrum disorder

I.N. Rozhkova¹, S.V. Okotrub^{1, 2}, E.Yu. Brusentsev¹, E.E. Uldanova¹, E.A. Chuyko^{1, 2}, T.V. Lipina³, T.G. Amstislavskaya⁴, S.Ya. Amstislavsky¹ 

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

² Novosibirsk State University, Novosibirsk, Russia

³ University of Toronto, Toronto, Canada

⁴ Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

 amstis@yandex.ru

Abstract. Autistic spectrum disorders (ASD) represent conditions starting in childhood, which are characterized by difficulties with social interaction and communication, as well as non-typical and stereotyping models of behavior. The mechanisms and the origin of these disorders are not yet understood and thus far there is a lack of prophylactic measures for these disorders. The current study aims to estimate neuronal density in the prefrontal cortex and four hippocampal subfields, i.e. CA1, CA2, CA3, and DG in *Clstn2*-KO mice as a genetic model of ASD. In addition, the level of neurogenesis was measured in the DG area of the hippocampus. This mouse strain was obtained by a knockout of the calyntenin-2 gene (*Clstn2*) in C57BL/6J mice; the latter (wild type) was used as controls. To estimate neuronal density, serial sections were prepared on a cryotome for the above-mentioned brain structures with the subsequent immunohistochemical labeling and confocal microscopy; the neuronal marker (anti-NeuN) was used as the primary antibody. In addition, neurogenesis was estimated in the DG region of the hippocampus; for this purpose, a primary antibody against doublecortin (anti-DCX) was used. In all cases Goat anti-rabbit IgG was used as the secondary antibody. The density of neurons in the CA1 region of the hippocampus was lower in *Clstn2*-KO mice of both sexes as compared with controls. Moreover, in males of both strains, neuronal density in this region was lower as compared to females. Besides, the differences between males and females were revealed in two other hippocampal regions. In the CA2 region, a lower density of neurons was observed in males of both strains, and in the CA3 region, a lower density of neurons was also observed in males as compared to females but only in C57BL/6J mice. No difference between the studied groups was revealed in neurogenesis, nor was it in neuronal density in the prefrontal cortex or DG hippocampal region. Our new findings indicate that calyntenin-2 regulates neuronal hippocampal density in subfield-specific manner, suggesting that the CA1 neuronal subpopulation may represent a cellular target for early-life preventive therapy of ASD.

Key words: mice; calyntenin-2; brain; neuronal density; prefrontal cortex; hippocampus; autism spectrum disorder.

For citation: Rozhkova I.N., Okotrub S.V., Brusentsev E.Yu., Uldanova E.E., Chuyko E.A., Lipina T.V., Amstislavskaya T.G., Amstislavsky S.Ya. Neuronal density in the brain cortex and hippocampus in *Clstn2*-KO mouse strain modeling autistic spectrum disorder. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(4):365-370. DOI 10.18699/VJGB-22-44

Плотность нейронов в коре головного мозга и гиппокампе мышей линии *Clstn2*-KO – модели расстройств аутистического спектра

И.Н. Рожкова¹, С.В. Окотруб^{1, 2}, Е.Ю. Брусенцев¹, Е.Е. Ульданова¹, Э.А. Чуйко^{1, 2}, Т.В. Липина³, Т.Г. Амстиславская⁴, С.Я. Амстиславский¹ 

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

³ Университет Торонто, Торонто, Канада

⁴ Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия

 amstis@yandex.ru

Аннотация. Расстройства аутистического спектра (РАС) – это группа состояний, возникающих в детском возрасте, для которых характерны трудности с социальным взаимодействием и общением, а также нетипичные модели поведения и склонность к стереотипии. Механизмы возникновения этой группы расстройств до сих пор не вполне понятны, и, следовательно, отсутствуют соответствующие методы профилактики. Целью исследования была оценка плотности нейронов в медиальной префронтальной коре и четырех областях гиппо-

кампа, а именно CA1, CA2, CA3 и зубчатой извилины (DG) у мышей линии *Clstn2*-KO, которая может выступать в качестве генетической модели РАС. Кроме того, охарактеризовали уровень нейрогенеза в области DG гиппокампа. Данная линия получена путем нокаута гена кальсинтенина-2 (*Clstn2*) на основе мышей линии C57BL/6J; последняя была использована в настоящем исследовании в качестве контроля. Для определения плотности нейронов изготавливали серийные срезы соответствующих областей мозга на криотоме с последующим иммуногистохимическим окрашиванием и конфокальной микроскопией, для чего использовали нейрональный маркер (anti-NeuN) в качестве первичного антитела. Наряду с этим в области DG гиппокампа оценивали нейрогенез, для чего проводили иммуногистохимическое окрашивание с применением антитела против даблкортина (anti-DCX). В обоих случаях в качестве вторичного антитела был Goat anti-rabbit IgG. Плотность нейронов в области гиппокампа CA1 была снижена как у самцов, так и самок мышей *Clstn2*-KO по сравнению с контролем; у самцов обеих линий плотность нейронов была ниже в этой области по сравнению с самками. Помимо этого, были обнаружены различия между самцами и самками в двух других областях гиппокампа: в области CA2 – у мышей обеих исследованных линий, а в области CA3 лишь у мышей C57BL/6J плотность нейронов была меньше у самцов по сравнению с самками. Различий между исследованными группами в уровне нейрогенеза, а также в плотности нейронов в префронтальной коре и области DG гиппокампа не обнаружено. Полученные результаты показывают, что нокаут по гену *Clstn2* приводит к избирательному снижению плотности нейронов в области CA1 гиппокампа, что может представлять собой клеточную мишень для ранней профилактики и возможной терапии РАС.

Ключевые слова: мыши; кальсинтенин-2; мозг; плотность нейронов; префронтальная кора; гиппокамп; расстройства аутистического спектра.

Introduction

Diagnosis and prevention of autism spectrum disorders (ASD) at an early age is very important and requires identification of a specific molecular cellular target. Despite some progress in this area, for example, the discovery of the *Fragile X*, *SHANK3*, *CASPR2* genes as risk factors for ASD, the mechanisms of this group of disorders are still not fully understood and, therefore, there are no appropriate methods of their prevention. The main reason for this is that both genetic and environmental factors are involved in the pathogenesis of autism, including, for example, epigenetic modifications of the genome, chromosome remodeling, oxidative stress, and many others (Waye, Cheng, 2017).

The hippocampal regions (CA1, CA2, CA3 and dentate gyrus – DG) are involved in memory-related processes: CA1 is important for working memory (Newmark et al., 2013), while CA3, CA4 and DG are in the circuit of declarative memory (Coras et al., 2014), and CA2 is associated with episodic (Navratilova, Battaglia, 2015) and social (Hitti, Siegelbaum, 2014) memory.

Structural abnormalities of the hippocampus have been reported for many complex psychiatric disorders, including vascular dementia (Kim et al., 2015), Alzheimer's disease (Thomson et al., 2004), and ASD (Bauman, Kemper, 2005; Varghese et al., 2017). There are indications that people with ASD also have alterations in the prefrontal cortex, in particular, some neurons in this brain area are changed (Courchesne et al., 2011; Varghese et al., 2017). It has also been discussed that people with ASD demonstrate impaired neurogenesis (Gilbert, Man, 2017).

Several studies in ASD patients identified mutations in genes encoding synaptic proteins, including those involved in the regulation of cell adhesion (Bakkaloglu et al., 2008; Morrow et al., 2008; Bourgeron, 2015). Calcintenins (*Clstns*) are transmembrane synaptic proteins that belong to the superfamily of cadherin cell adhesion molecules. There are three types of *Clstns* (*Clstn*-1, -2 and -3) which are expressed postsynaptically (Hintsch et al., 2002) and contribute differently to

the balanced activity of excitatory and inhibitory neurons; an imbalance in these processes is characteristic of some patients with ASD (Yip et al., 2009).

The absence of *Clstn2* specifically reduces the density of inhibitory parvalbumin interneurons in some brain areas, which is manifested as insufficient inhibitory, but not excitatory, synaptic transmission in the pyramidal neurons of the hippocampal CA1 region (Lipina et al., 2016). In addition, a change in synapse architectonics was found in *Clstn2*-KO mice in the medial prefrontal cortex and hippocampus (Ranneva et al., 2020). Moreover, *Clstn2*-KO mice demonstrate signs characteristic for ASD, such as hyperactivity, stereotypy, insufficient spatial learning and memory, altered social behavior with impaired ultrasonic vocalization (Lipina et al., 2016; Ranneva et al., 2017; Klenova et al., 2021).

The density of neurons in the prefrontal cortex and hippocampus, as well as the level of neurogenesis in the brain, have not yet been studied on such a genetic model of ASD as *Clstn2*-KO mice. Thus, the aim of this study was to characterize neuronal density in the medial prefrontal cortex, as well as in CA1, CA2, CA3, and DG regions of the hippocampus; and to evaluate the level of neurogenesis in *Clstn2*-KO mice.

Materials and methods

Experimental animals. The study used 12 homozygous males and 14 females of the *Clstn2*-KO mouse strain knockout for the *Clstn2* gene, as well as 15 males and 15 females of C57BL/6J mice (control) at the age of three months. Animals were kept in 36 × 25 × 14 cm (length × width × height) cages with wood bedding. Males and females were kept individually in a conventional vivarium of the Scientific Research Institute of Neurosciences and Medicine (Novosibirsk, Russia), at 20–22 °C, 12 dark : 12 light cycle, with free access to dry granulated chew ("Chara", Assortiment-Agro, Russia) and purified water. All studies were performed in accordance with the European Convention for the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes.

Intracardiac perfusion. At the age of three months, the animals were perfused with phosphate buffered saline (PBS) and 10 % paraformaldehyde solution; thereafter, the brain was removed and fixed in phosphate buffer containing 30 % sucrose and 5 % formalin at +4 °C. Subsequently, the brain was immersed in Tissue-Tek O.C.T. compound (Sakura Finetek, USA), frozen and stored at -70 °C.

Brain sectioning. Cryosections were made for the following areas of the brain: 1) medial prefrontal cortex (MPC) at a distance of 2.46–2.22 mm from the bregma; 2) hippocampus (regions CA1, CA2, CA3 and DG) at a distance of -1.46...-1.82 mm from the bregma. Sections of 10 µm thick were done on an HM550 OP cryotome (Thermo Fisher Scientific, USA) at -25 °C and placed on Superfrost Plus, Menzel-Glaser glass slides (Thermo Fisher Scientific).

Immunohistochemical (IHC) staining. The slices were stained according to kit manufacturer's protocols with minor modifications. Brain sections were dehydrated before staining, followed by 5 min of rehydration in PBS. Then, after rehydration in 10 mM citrate buffer (pH = 9) at 95 °C in a TW-2.02 water bath (Elmi, Latvia), heat-induced epitope unmasking was performed. The sections were removed from the buffer and cooled to room temperature. Then, the samples were washed three times in PBS-Tween buffer: PBS with the addition of 0.1 % Tween-20 P9416-100ML (Merck, Germany). Protein Block ab64226 (Abcam, UK) was added to each section for 5 min as recommended by the manufacturer; excess liquid was removed.

After the procedures of washing and incubating with the Protein Block, the primary antibody was added to the slices and left overnight at +4 °C in a humid dark chamber. The antibody concentrations used were 1:750 and 1:750 for anti-NeuN ab177487 (Abcam) and anti-DCX ab18723 (Abcam), respectively. Then the sections were washed with PBS-Tween buffer, excess liquid was removed; thereafter, 50 µL of secondary antibody Goat anti-rabbit IgG H&L AF488 ab150077 (Abcam) at a concentration of 1:600 was added to the slices and left in a humid dark chamber for two hours at +4 °C. Then the samples were washed with PBS-Tween buffer, excess liquid was removed, and the samples were placed in

ProLong, Glass AntifadeMountant, Thermo P36982 (Thermo Fisher Scientific).

Neuron density analysis. Using a confocal laser scanning microscope LSM 780 (Carl Zeiss, Germany), EC Plan-Neofluar 20x/0.50 (Carl Zeiss), digital photographs were obtained, which were used to evaluate the density of neurons labeled with antibodies (<https://ckp.icgen.ru/ckpmabo>). The number of labeled neurons was counted using ImageJ. The number of antibody-labeled neurons was counted in at least three sections per animal, then the average per these sections was counted and the average volume density (mm³) was calculated.

Statistical analysis. The analysis of the obtained results was carried out using the software Statistica v. 10.0 (StatSoft, Inc., USA). All the data were tested for normality using the Shapiro-Wilk *W*-test. Data are presented as mean ± standard error of the mean (M ± SEM) and analyzed by two-way analysis of variance (ANOVA) followed by Fisher's multiple comparison. Differences at *p* < 0.05 were considered as statistically significant.

Results

Data on the density of neurons in the prefrontal cortex and hippocampus (CA1, CA2, CA3 and DG areas) are presented in the Table. A statistically significant effect on the density of neurons in the CA1 region of the hippocampus was found for the "sex" factor ($F_{1,35} = 29.53, p < 0.001$) and the "group" factor ($F_{2,35} = 16.68, p < 0.001$), while there were no interactions of these factors ($F_{1,35} < 1$). Post-hoc comparison confirmed that both male (*p* < 0.001) and female (*p* < 0.05) Clstn2-KO mice have fewer pyramidal neurons in the CA1 region of the hippocampus compared to control C57BL/6J (Fig. 1). Moreover, there were sex differences demonstrated: females of both strains had more pyramidal neurons (*p* < 0.001) in the CA1 region of the hippocampus compared to males (see Fig. 1).

The influence of the "sex" factor on the density of neurons in the CA2 region of the hippocampus was found ($F_{1,31} = 12.03, p < 0.05$). At the same time, the influence of the "group" factor ($F_{1,31} < 1$) and the interaction of these factors ($F_{1,31} < 1$) were not found for the density of neurons in the CA2 region. Post-hoc comparison showed that females of both strains had

Neuronal density in hippocampus and prefrontal cortex; neurogenesis in dentate gyrus of the hippocampus

Brain area	Males		Females		
	C57BL/6J ($N^*/\text{mm}^3 \pm \text{SEM}$)*10 ³ (n)	Clstn2-KO ($N^*/\text{mm}^3 \pm \text{SEM}$)*10 ³ (n)	C57BL/6J ($N^*/\text{mm}^3 \pm \text{SEM}$)*10 ³ (n)	Clstn2-KO ($N^*/\text{mm}^3 \pm \text{SEM}$)*10 ³ (n)	
PFC	150.84 ± 3.09 (5)	154.23 ± 5.21 (4)	143.41 ± 8.82 (4)	151.90 ± 9.47 (5)	
Hippocampus	CA1	109.54 ± 3.93 (5)	85.94 ± 4.84 (5)***	133.71 ± 4.71 (5)+++	116.31 ± 4.41 (5)**,+
	CA2	102.89 ± 17.22 (5)	111.68 ± 9.68 (4)	162.61 ± 9.13 (3) ⁺	147.85 ± 7.09 (5) ⁺
	CA3	138.11 ± 20.67 (5)	192.65 ± 28.39 (4)	217.95 ± 41.63 (3) ⁺	173.76 ± 9.72 (5)
	DG	400.72 ± 121.08 (5)	383.32 ± 87.66 (4)	504.41 ± 75.18 (3)	502.15 ± 30.73 (5)
Neurogenesis	19.96 ± 1.32 (6)	22.93 ± 1.98 (6)	20.29 ± 3.08 (5)	20.16 ± 2.83 (5)	

Note. N^* – number of neurons in the region of interest; SEM – standard error of the mean; *n* – number of animals; ***p* < 0.05 as compared with C57BL/6J of the same sex; ****p* < 0.001 as compared with C57BL/6J of the same sex; ⁺*p* < 0.05 as compared with males of the same strain; ⁺⁺⁺*p* < 0.001 as compared with males of the same strain.

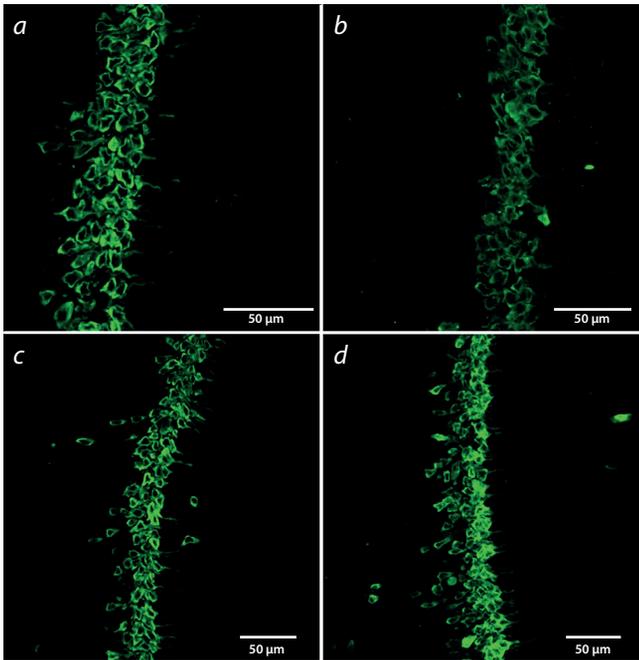


Fig. 1. Micrographs of hippocampal CA1 region, neurons labeled with antibodies against the NeuN.

a, b – females; *c, d* – males of the C57BL/6J (*a, c*) and *Clstn2*-KO (*b, d*) strains.

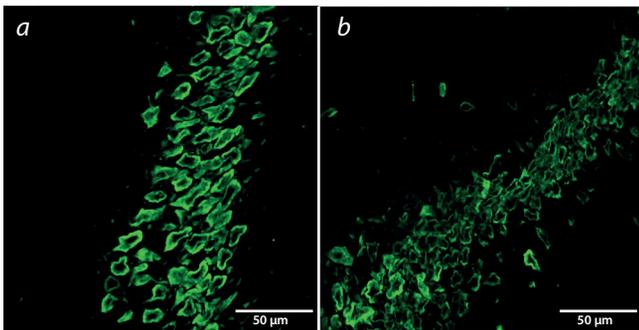


Fig. 2. Micrographs of the hippocampal CA2 region.

Neurons labeled with antibodies against NeuN in females (*a*) and males (*b*) of the *Clstn2*-KO strain.

more pyramidal neurons ($p < 0.05$) in the CA2 region of the hippocampus compared to males (Fig. 2). However, there were no interstrain differences in the density of neurons in this region.

No influence of the factors “sex” ($F_{1,35} = 1.66, p > 0.05$) and “group” ($F_{1,35} < 1$) on the density of neurons in the CA3 region of the hippocampus was found, but the interaction of these factors was revealed ($F_{1,35} = 4.36, p < 0.05$). Post-hoc comparison confirmed that C57BL/6J females had more pyramidal neurons than males of this strain ($p < 0.05$).

To test the possibility that one of the reasons for the decrease in the density of neurons in the CA1 region in *Clstn2*-KO mice may be caused by the altered level of neurogenesis, its evaluation was carried out in the DG region of the hippocampus. Hippocampal neurogenesis was not influenced by the factors “sex” ($F_{1,35} = 1.27, p > 0.05$), and “group” ($F_{1,35} < 1$), and

there was no interaction of these factors ($F_{1,35} < 1$). Post-hoc comparison confirmed that there are no interstrain and sex differences in the level of neurogenesis in the DG region (see the Table).

Discussion

In this study, no changes in the density of neurons in the prefrontal cortex were found in both male and female *Clstn2*-KO mice, although there is an alteration in neuronal density in this area of the brain in people with ASD (Courchesne et al., 2011). It should be noted, however, that human data are obtained by *post mortem* brain biopsy and should be discussed with caution, since only a few brain samples have been studied. In BTBR mouse strain, a model of idiopathic autism, no changes were found in the number of neurons in the prefrontal cortex (Stephenson et al., 2011), which is consistent with the results of our study in *Clstn2*-KO mice. However, other studies in BTBR mice have shown lower levels of extracellular acetylcholine and more kynurenic acid, but not serotonin, in this area of the brain (McTighe et al., 2013; Guo, Commons, 2017). It can be assumed that the development of ASD is associated with an imbalance of neurotransmitter systems in the prefrontal cortex, and the density of neurons in this structure is not a universal marker of these disorders.

In our current study, both male and female *Clstn2*-KO mice were found to have a reduced neuronal density in the CA1 region of the hippocampus. An earlier study had demonstrated a deficiency of inhibitory GABAergic neurons in the CA1 and CA3 regions of the hippocampus in mice of this strain (Lipina et al., 2016). We suppose that the decrease in the density of neurons in the CA1 region of the hippocampus revealed in *Clstn2*-KO mice in the current study is associated, among other things, with the detected decrease in GABAergic neurons mentioned above.

Structural studies using MRI have revealed a decrease in the relative volume of the hippocampus in patients with ASD aged 4 to 18 years (Sussman et al., 2015). Moreover, a change in the size of the hippocampus has also been found in adult patients with ASD (Braden et al., 2017). According to the results of the brain post mortem biopsy of people that suffered from ASD symptoms, changes in the neuronal density in certain areas of the hippocampus were found, the most pronounced changes were observed in the CA1 region (Lawrence et al., 2010; Greco et al., 2011).

Studies on various laboratory models also indicate the abnormalities in the CA1 region of the hippocampus are associated with the development of ASD. In particular, heterozygous mice deficient in the *Tef4* transcription factor, which demonstrate features of autism, showed an increased synaptic transmission in the CA1 region of the hippocampus (Kennedy et al., 2016). BTBR T+tf/J mice exhibit behavior characteristic for ASD, accompanied by loss of neurons in the CA1 region of the hippocampus (Zhang et al., 2019). Heterozygous SHANK-3 mice, which are a well-known model of ASD, exhibited perforated synapses in the radial layer of the CA1 region of the hippocampus (Uppal et al., 2015), which confirms impaired synaptic plasticity in this region of the brain (Moessner et al., 2007). Fragile X chromosome syndrome, one of the forms of ASD associated with a disruption of the *Fmr1* gene, also possessed specific changes in the pyramidal

neurons of the CA1 region of the hippocampus (Sawicka et al., 2019). Thus, both the results of the current study and the results of the above mentioned studies in other mouse strains indicate that disorders in the CA1 region of the hippocampus can be considered as a specific marker of ASD.

In our study, the density of neurons in Clstn2-KO females in the CA1 and CA2 regions of the hippocampus was higher than in males. Moreover, in our work, in female mice of the control strain C57BL/6J, the number of neurons in these areas, as well as in the CA3 region of the hippocampus, was also higher than in males, which may be a physiological feature non-related to ASD, despite the fact that, in humans, ASD is more common in boys than in girls in a ratio of 4.3:1 (Fombonne, 2003).

Disruption of the formation of new neurons in adulthood plays a significant role in the development of mental disorders (Schoenfeld, Cameron, 2015). Moreover, BTBR mice have been shown to have a significant impairment in adult neurogenesis (Stephenson et al., 2011). However, according to our data, neurogenesis in adult Clstn2-KO mice is not impaired. It is possible that the decrease in the density of neurons in the CA1 region found in Clstn2-KO mice is due to an increase in neurodegenerative processes, which, in particular, can lead to an imbalance in inhibitory and excitatory neurons.

Conclusion

A decrease in the density of neurons in the CA1 region of the hippocampus was found in both male and female Clstn2-KO mice compared to control C57BL/6J mice. Meanwhile, in Clstn2-KO mice, no changes in neuronal density were found in other areas of the hippocampus, and in the prefrontal cortex, the level of hippocampal neurogenesis was unaltered as well. The decrease in the density of neurons in the CA1 region of the hippocampus in Clstn2-KO mice can be considered as a specific characteristic for this model of autism spectrum disorders.

References

- Bakkaloglu B., O'Roak B.J., Louvi A., Gupta A.R., Abelson J.F., Morgan T.M., Chawarska K., Klin A., Ercan-Sencicek A.G., Stillman A.A., Tanriverog G., Abrahams B.S., Duvall J.A., Robbins E.M., Geschwind D.H., Biederer T., Gunel M., Lifton R.P., State M.W. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am. J. Hum. Genet.* 2008;82:165-173. DOI 10.1016/j.ajhg.2007.09.017.
- Bauman M.L., Kemper T.L. Neuroanatomic observations of the brain in autism: a review and future directions. *Int. J. Dev. Neurosci.* 2005; 23:183-187. DOI 10.1016/j.ijdevneu.2004.09.006.
- Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* 2015;16:551-563. DOI 10.1038/nrn3992.
- Braden B.B., Smith C.J., Thompson A., Glaspy T.K., Wood E., Vatsa D., Abbott A.E., McGee S.C., Baxter L.C. Executive function and functional and structural brain differences in middle-age adults with autism spectrum disorder. *Autism Res.* 2017;10:1945-1959. DOI 10.1002/aur.1842.
- Coras R., Pauli E., Li J., Schwarz M., Rossler K., Buchfelder M., Hamer H., Stefan H., Blumcke I. Differential influence of hippocampal subfields to memory formation: insights from patients with temporal lobe epilepsy. *Brain.* 2014;137:1945-1957. DOI 10.1093/brain/awu100.
- Courchesne E., Mouton P.R., Calhoun M.E., Semendeferi K., Ahrens-Barbeau C., Hallet M.J., Barnes C.C., Pierce K. Neuron number and size in prefrontal cortex of children with autism. *JAMA.* 2011; 306(18):2001-2010. DOI 10.1001/jama.2011.1638.
- Fombonne E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J. Autism Dev. Disord.* 2003; 33(4):365-382. DOI 10.1023/a:1025054610557.
- Gilbert J., Man H.-Y. Fundamental elements in autism: from neurogenesis and neurite growth to synaptic plasticity. *Front. Cell. Neurosci.* 2017;11:359. DOI 10.3389/fncel.2017.00359.
- Greco C.M., Navarro C.S., Hunsaker M.R., Maezawa I., Shuler J.F., Tassone F., Delany M., Au J.W., Berman R.F., Jin L.W., Schumann C., Hagerman P.J., Hagerman R.J. Neuropathologic features in the hippocampus and cerebellum of three older men with fragile X syndrome. *Mol. Autism.* 2011;2(1):2. DOI 10.1186/2040-2392-2-2.
- Guo Y.P., Commons K.G. Serotonin neuron abnormalities in the BTBR mouse model of autism. *Autism Res.* 2017;10(1):66-77. DOI 10.1002/aur.1665.
- Hintsch G., Zurlinden A., Meskenaite V., Steuble M., Fink-Widmer K., Kinter J., Sonderegger P. The calyntenins – a family of postsynaptic membrane proteins with distinct neuronal expression patterns. *Mol. Cell. Neurosci.* 2002;21:393-409. DOI 10.1006/mcne.2002.1181.
- Hitti F.L., Siegelbaum S.A. The hippocampal CA2 region is essential for social memory. *Nature.* 2014;508:88-92. DOI 10.1038/nature13028.
- Kennedy A.J., Rahn E.J., Paulukaitis B.S., Savell K.S., Kordasiewicz H.B., Wang J., Lewis J.W., Posey J., Strange S.K., Guzman-Karlsson M.C., Phillips S.E., Decker K., Motley S.T., Swayze E.E., Ecker D.J., Michael T.P., Day J.J., Sweatt J.D. Tcf4 regulates synaptic plasticity, DNA methylation, and memory function. *Cell Rep.* 2016;16:2666-2685. DOI 10.1016/j.celrep.2016.08.004.
- Kim G.H., Lee J.H., Seo S.W., Kim J.H., Seong J.K., Ye B.S., Cho H., Noh Y., Kim H.J., Yoon C.W., Oh S.J., Kim J.S., Choe Y.S., Lee K.H., Kim S.T., Hwang J.W., Jeong J.H., Na D.L. Hippocampal volume and shape in pure subcortical vascular dementia. *Neurobiol. Aging.* 2015;36:485-491. DOI 10.1016/j.neurobiolaging.2014.08.009.
- Klenova A.V., Volodin I.A., Volodina E.V., Ranneva S.V., Amtislavskaya T.G., Lipina T.V. Vocal and physical phenotypes of calyntenin2 knockout mouse pups model early-life symptoms of the autism spectrum disorder. *Behav. Brain Res.* 2021;412:113430. DOI 10.1016/j.bbr.2021.113430.
- Lawrence Y.A., Kemper T.L., Bauman M.L., Blatt G.J. Parvalbumin-, calbindin-, and calretinin-immunoreactive hippocampal interneuron density in autism. *Acta Neurol. Scand.* 2010;121(2):99-108. DOI 10.1111/j.1600-0404.2009.01234.x.
- Lipina T.V., Prasad T., Yokomaku D., Luo L., Connor S.A., Kawabe H., Wang Y.T., Brose N., Roder J.C., Craig A.M. Cognitive deficits in calyntenin-2-deficient mice associated with reduced GABAergic transmission. *Neuropsychopharmacology.* 2016;41:802-810. DOI 10.1038/npp.2015.206.
- McTighe S.M., Neal S.J., Lin Q., Hughes Z.A., Smith D.G. The BTBR mouse model of autism spectrum disorders has learning and attentional impairments and alterations in acetylcholine and kynurenic acid in prefrontal cortex. *PLoS One.* 2013;8(4):e62189. DOI 10.1371/journal.pone.0062189.
- Moessner R., Marshall C.R., Sutcliffe J.S., Skaug J., Pinto D., Vincent J., Zwaigenbaum L., Fernandez B., Roberts W., Szatmari P., Scherer S.W. Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* 2007;81:1289-1297. DOI 10.1086/522590.
- Morrow E.M., Yoo S.Y., Flavell S.W., Kim T.K., Lin Y., Hill R.S., Mukkaddes N.M., Balkhy S., Gascon G., Hashmi A., Al-Saad S., Ware J., Joseph R.M., Greenblatt R., Gleason D., Ertelt J.A., Apse K.A., Bodell A., Partlow J.N., Barry B., Yao H., Markianos K., Ferland R.J., Greenberg M.E., Walsh C.A. Identifying autism loci and genes by tracing recent shared ancestry. *Science.* 2008;321:218-223. DOI 10.1126/science.1157657.
- Navratilova Z., Battaglia F.P. CA2: it's about time – and episodes. *Neuron.* 2015;85:8-10. DOI 10.1016/j.neuron.2014.12.044.

- Newmark R.E., Schon K., Ross R.S., Stern C.E. Contributions of the hippocampal subfields and entorhinal cortex to disambiguation during working memory. *Hippocampus*. 2013;23:467-475. DOI 10.1002/hipo.22106.
- Ranneva S.V., Maksimov V.F., Korostyshevskaja I.M., Lipina T.V. Lack of synaptic protein, calyntenin-2, impairs morphology of synaptic complexes in mice. *Synapse*. 2020;74:e22132. DOI 10.1002/syn.22132.
- Ranneva S.V., Pavlov K.S., Gromova A.V., Amstislavskaya T.G., Lipina T.V. Features of emotional and social behavioral phenotypes of calyntenin2 knockout mice. *Behav. Brain Res*. 2017;332:343-354. DOI 10.1016/j.bbr.2017.06.029.
- Sawicka K., Hale C.R., Park C.Y., Fak J.J., Gresack J.E. FMRP has a cell-type-specific role in CA1 pyramidal neurons to regulate autism-related transcripts and circadian memory. *Elife*. 2019;8:e46919. DOI 10.7554/eLife.46919.
- Schoenfeld T.J., Cameron H.A. Adult neurogenesis and mental illness. *Neuropsychopharmacology*. 2015;40:113-128. DOI 10.1038/npp.2014.230.
- Stephenson D.T., O'Neill S.M., Narayan S., Tiwari A., Arnold E., Samaroo H.D., Du F., Ring R.H., Campbell B., Pletcher M., Vaidya V.A., Morton D. Histopathologic characterization of the BTBR mouse model of autistic-like behavior reveals selective changes in neurodevelopmental proteins and adult hippocampal neurogenesis. *Mol. Autism*. 2011;2(1):7. DOI 10.1186/2040-2392-2-7.
- Sussman D., Leung R.C., Vogan V.M., Lee W., Trelle S., Lin S., Cassel D.B., Chakravarty M.M., Lerch J.P., Anagnostou E., Taylor M.J. The autism puzzle: Diffuse but not pervasive neuroanatomical abnormalities in children with ASD. *Neuroimage Clin*. 2015;8:170-179. DOI 10.1016/j.nicl.2015.04.008.
- Thompson P.M., Hayashi K.M., De Zubicaray G.I., Janke A.L., Rose S.E., Semple J., Hong M.S., Herman D.H., Gravano D., Doddrell D.M., Toga A.W. Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage*. 2004;22(4):1754-1766. DOI 10.1016/j.neuroimage.2004.03.040.
- Uppal N., Puri R., Yuk F., Janssen W.G.M., Bozdagi-Gunal O., Harony-Nicolas H., Dickstein D.L., Buxbaum J.D., Hof P.R. Ultrastructural analyses in the hippocampus CA1 field in Shank3-deficient mice. *Mol. Autism*. 2015;6:41. DOI 10.1186/s13229-015-0036-x.
- Varghese M., Keshav N., Jacot-Descombes S., Warda T., Wicinski B., Dickstein D.L., Harony-Nicolas H., De Rubeis S., Drapeau E., Buxbaum J.D., Hof P.R. Autism spectrum disorder: neuropathology and animal models. *Acta Neuropathol*. 2017;134(4):537-566. DOI 10.1007/s00401-017-1736-4.
- Waye M.M.Y., Cheng H.Y. Genetics and epigenetics of autism: a review. *Psych. Clin. Neurosci*. 2017;72:228-244. DOI 10.1111/pcn.12606.
- Yip J., Soghomonian J.J., Blatt G.J. Decreased GAD65 mRNA levels in select subpopulations of neurons in the cerebellar dentate nuclei in autism: an *in situ* hybridization study. *Autism Res*. 2009;2:50-59. DOI 10.1002/aur.62.
- Zhang Q., Wu H., Zou M., Li L., Li Q., Sun C., Xia W., Cao Y., Wu L. Folic acid improves abnormal behavior via mitigation of oxidative stress, inflammation, and ferroptosis in the BTBR T+ tf/J mouse model of autism. *J. Nutr. Biochem*. 2019;71:98-109. DOI 10.1016/j.jnutbio.2019.05.002.

Acknowledgements. This work was funded by the Russian Foundation for Basic Research (grant No. 20-015-00162), as well as budget project No. FWNR-2022-0023 of the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia). The studies were performed using the equipment of the Centre of Microscopic Analysis of Biological Objects, and the Center for Genetic Resources of Laboratory Animals of the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia), supported by the Ministry of Education and Science of the Russian Federation (unique project RFMEFI62119X0023).

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

Received November 1, 2021. Revised January 11, 2022. Accepted March 1, 2022.

Original Russian text www.bionet.nsc.ru/vogis/

Hereditary predisposition of water voles (*Arvicola amphibius* L.) to seizures in response to handling

G.G. Nazarova , L.P. Proskurnyak 

Institute of Systematics and Ecology of Animals of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

 galinanazarova@mail.ru; luda_proskurnjak@mail.ru

Abstract. Finding out the hereditary predisposition to seizures in response to specific external stimuli is important for understanding the causes of epileptiform conditions, developing new methods for their prevention and therapies. In the water vole, individuals with convulsive seizures are found both in natural and laboratory conditions. The data of long-lasting maintenance and breeding of water voles in vivarium conditions were analyzed in order to establish a hereditary predisposition to convulsive seizures, and the influence of sex and age on their development. In the vivarium, seizures are provoked by handling and are observed in 2.4 % of voles caught in the natural population with cyclic fluctuations in abundance. Seizures are observed more often in individuals caught in the phases of decline and depression of abundance than in individuals caught in the phases of rise or peak. Convulsive states are probably an element of adaptive behavior formed in the predator-prey system. In natural conditions, individuals predisposed to convulsive seizures may have a selective advantage when under increasing pressure from predators. Convulsive seizures in response to handling were noted in 29.8 % of descendants of captive-bred water voles. The proportion of such individuals increased significantly if one or both parents had convulsive states, which indicates the presence of a hereditary predisposition to seizures. In parent-offspring pairs, a significant correlation was found between the average age of onset of the first seizures in parents and their offspring, $r = 0.42$, $p < 0.01$. The minimum age of registration of seizures in the water vole is 39 days, the maximum is 1105 days, and the median is 274 days. Predisposition to seizures is not related to sex. Genes that control the occurrence of seizures have a pleiotropic effect on life span, since individuals with seizures live longer in vivarium conditions than individuals with a normal phenotype. The water vole can serve as a suitable model object for studying the nature of convulsive states and the evolution of longevity. Key words: water vole; seizures; age; hereditary predisposition; population cycle; life span.

For citation: Nazarova G.G., Proskurnyak L.P. Hereditary predisposition of water voles (*Arvicola amphibius* L.) to seizures in response to handling. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(4): 371-377. DOI 10.18699/VJGB-22-45

Наследственная предрасположенность водяных полевок (*Arvicola amphibius* L.) к судорожным припадкам в ответ на хэндлинг и ее связь с продолжительностью жизни

Г.Г. Назарова , Л.П. Проскурняк 

Институт систематики и экологии животных Сибирского отделения Российской академии наук, Новосибирск, Россия

 galinanazarova@mail.ru; luda_proskurnjak@mail.ru

Аннотация. Выяснение наследственной предрасположенности к судорогам в ответ на специфические внешние стимулы важно для понимания причин epileptiformных состояний, нахождения новых методов их предупреждения и лечения. Особи водяной полевки с судорожными припадками встречаются как в природных, так и лабораторных условиях. Проанализированы данные многолетнего содержания и разведения водяных полевок в условиях вивария с целью установления наследственной предрасположенности к судорожным припадкам и влияния пола и возраста на их развитие. В условиях вивария припадки были спровоцированы взятием в руки и отмечены у 2.4 % полевок, отловленных в природной популяции с циклическими колебаниями численности. Судорожные припадки чаще встречали у особей, отловленных в фазы спада и депрессии численности, чем у особей с фаз подъема или пика. Судорожные состояния, вероятно, служат элементом адаптивного поведения, сформировавшегося в системе «хищник-жертва». В природных условиях особи, предрасположенные к судорожным припадкам, могут иметь селективное преимущество при усилении пресса хищников. Судорожные припадки в ответ на хэндлинг отмечены у 29.8 % потомков водяных полевок виварно-го разведения. Доля таких особей существенно увеличивалась, если у одного или обоих родителей регистрировали судорожные состояния, что свидетельствует о наследственной предрасположенности к припадкам. В парах «родители-потомки» обнаружена достоверная корреляция среднего возраста наступления первых

припадков ($r = 0.42$, $p < 0.01$). Минимальный возраст регистрации припадков у водяной полевки составляет 39 дней, максимальный – 1105, медианный – 274 дня. Предрасположенность к припадкам не связана с полом. Гены, контролирующие возникновение судорожных состояний, оказывают плейотропное действие на продолжительность жизни: особи с судорожными припадками в условиях вивария живут дольше, чем особи с нормальным фенотипом. Водяная полевка может служить подходящим модельным объектом для изучения природы судорожных состояний и эволюции долголетия.

Ключевые слова: водяная полевка; судорожные припадки; возраст; наследственная предрасположенность; циклы численности; продолжительность жизни.

Introduction

Tonic-clonic seizures – uncontrolled muscle tension or contraction – can occur in humans and other mammals: gray rats *Rattus norvegicus* (Poletaeva et al., 2017), house mice *Mus musculus* (Skradski et al., 1998), Mongolian gerbils *Meriones unguiculatus* (Buchhalter, 1993), deer mice *Peromyscus maniculatus* (Jackson, 1997), Syrian hamsters *Mesocricetus auratus* (Muñoz et al., 2017), meadow voles *Microtus pennsylvanicus* (Bronson and De La Rosa, 1994), dogs *Canis familiaris* (Catala et al., 2018), cats *Felis catus* (Pakozdy et al., 2014), and rabbits (Gülersoy et al., 2021). The most common cause of seizures is an imbalance between excitatory and inhibitory neurotransmitter mechanisms in the brain (Poletaeva et al., 2017). Seizures are categorized into spontaneous (provoked by unknown factors) and reflex ones (caused by specific external stimuli, for example, electrical, auditory, visual, or tactile) (Okudan, Özkara, 2018).

In response to tactile stimulation during handling, prolonged tonic-clonic seizures take place in bank voles *Myodes glareolus* (Schönecker, 2009), meadow voles *M. pennsylvanicus* (Bronson, De La Rosa, 1994), and Mongolian gerbils *M. unguiculatus* (Ludvig et al., 1991; Buckmaster, 2006). Other stimuli (e. g., audiogenic or olfactory) are not effective at inducing seizures in these species. The gray rat is a popular experimental model for research on audiogenic epilepsy (Poletaeva et al., 2017). It should be noted that genetic factors predisposing to various types of reflex epilepsy are poorly understood, and genes that control somatosensory epilepsy are not known at all (Okudan, Özkara, 2018). Clarification of hereditary predisposition to handling-induced seizures is important for understanding the causes of seizures and for improving the prevention and treatment methods.

In recent decades, rodents have been widely used in neurophysiological investigation into epilepsy owing to their small size, peacefulness, and rapid reproduction in captivity (Jackson, 1997). Comparative studies on mammals are needed to better understand the evolutionary and genetic factors that explain seizures because they should help to find common components of the pathogenesis of this neurological disorder that are of clinical relevance to humans (Grone, Baraban, 2015). To elucidate the evolutionary mechanisms underlying the onset of convulsive states, species in which individuals with seizures occur under natural conditions are of great interest. L.G. Krotova (1962) – who has investigated the function of adrenal glands and carbohydrate metabolism in water voles (*Arvicola amphibius* L.) living in the floodplain of the Chusova River (Sverdlovsk Oblast, Russia) – noticed that some captured animals had seizures accompanied by a

coma. Owing to detailed complex research data on various characteristics of population ecology, genetics, and physiology of the water vole and to excellent advances in the methods for its breeding in captivity (The Water vole..., 2001), the water vole can become a promising experimental model for investigation into predisposing factors of seizures at populational, genetic, and neurological levels.

In this study, data are analyzed that were obtained during long-term breeding of water voles at an animal facility to achieve the following aims: 1) to describe the picture of the seizures; 2) to estimate the proportion of individuals with seizures among animals caught in the wild and among animals born in the vivarium; 3) to determine the minimal, maximal, and average age at onset of seizures; 4) to find out whether the manifestation of convulsive seizures is associated with the sex of these animals and/or the presence of this neuropathology in the parents; and 5) to identify a possible relation between the predisposition to seizures and the life span.

Materials and methods

The study was performed on two groups of water voles kept at the animal facility of the Institute of Systematics and Ecology of Animals of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia), under natural lighting conditions with free access to water and feed (carrots, well-steamed grain mixtures, and fresh greens):

- 1) the wild group ($n = 589$), i. e., animals that were caught during 1983–2017 in the Ubinsky District of Novosibirsk Oblast (Russia), where comprehensive ecological and physiological studies on a population with pronounced high-amplitude fluctuations of population size (6–7-year cycle) were being conducted (The Water vole ..., 2001; Evsikov et al., 2017). After capture, these water voles lived in the vivarium for at least 3 months (290 females and 299 males). Ninety-four voles from the peak phase, 32 from the decline phase, 155 from the depression phase, and 308 from the increase phase were analyzed;
- 2) the vivarium group ($n = 1776$), i. e., water voles (parents and offspring) that were born in the vivarium and survived at least 22 days after birth. This group was founded by individuals caught in the aforementioned natural population. To preserve genetic heterogeneity, the group was regularly replenished with wild individuals: at least once every 3 years.

This approach allows, on the one hand, to identify populational factors that affect the prevalence of convulsive states in the wild, and on the other hand, to identify genetic and ontogenetic factors predisposing parents and offspring to seizures under standard maintenance conditions in a vivarium.

The work on the animals of the two groups throughout the study period was done according to a standard scheme: 1) all water voles were individually marked, sexed, and dates of birth (capture) and death were recorded; 2) until the age of 70 days, weekly, the water voles born in the vivarium were weighed and their body length was measured; the weight and body length of the vivarium animals older than 70 days and those of individuals caught in the natural population were measured once a month; 3) during the reproduction period, before forming of mating pairs (closely related animals were not used for setting up mating pairs), vaginal smears were taken from females, and anogenital distance was measured in males; 4) throughout their life span, the water voles were subjected to experiments aimed at examining the relation between ethological–physiological characteristics of water voles and their reproductive ability. The following procedures were performed: tests of social interactions, of olfactory mate choice, and of parental behavior as well as blood sampling from the retroorbital sinus and tail, urine collection. During the experiments, water voles were subjected to handling (at least once a month), and some animals developed convulsive seizures. The main diagnostic criterion of a convulsive state was assumed to be pronounced tonic or clonic contractions of muscles of the trunk and/or limbs. Such cases were recorded. The animals with convulsive seizures were not culled and were allowed to reproduce. There were no spontaneous seizures when the animals were in home cages and were not subjected to the handling.

Statistical analysis was performed in SPSS Statistics 15.0 (IBM Corp., USA) and Statistica 6.1 (StatSoft Inc., USA) software. Comparisons of proportions (%) between groups were conducted by Pearson's χ^2 test. Effects of categorical variables and continuous variables on age at onset of the first seizure (hereafter: "age at seizure onset") or on the life span were assessed via the linear mixed model, with dependent variables subjected to the natural-logarithm (ln) transformation in order to make their distributions normal. Survival curves were constructed by the Kaplan–Meier method. Differences between the curves were evaluated by the Gehan–Wilcoxon test. The text and tables show means of parameters (X), standard error (\pm SE), and sample size (n). Statistical significance was assumed at $p < 0.05$.

Results

Description of the seizures

Each seizure was convulsive (tonic and/or clonic) and occurred in the first minute after a water vole was subjected to handling. The seizure began with a twitching of the vibrissae and tonic tension of anterior trunk muscles, followed by their rhythmic contractions; meanwhile, the animal was likely to arch its back and tilt its head back. At the first signs of a seizure, individuals were placed into an arena, and their state was examined. The convulsions spread throughout the whole body, and in some cases, the seizure was accompanied by gnashing of teeth and very rarely by loss of consciousness. After the seizures stopped, the animals experienced locomotor agitation: "wild running".

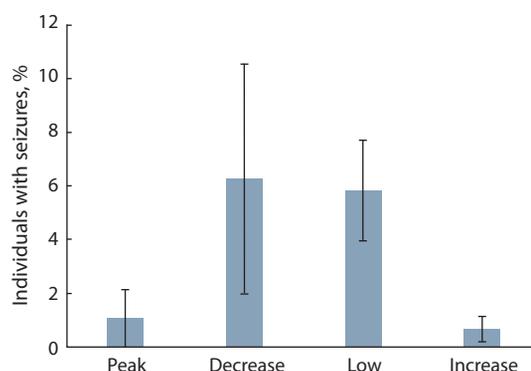


Fig. 1. The proportion of individuals (%) with seizures in different phases of population cycle.

The proportion of individuals with convulsive seizures among water voles taken from the natural population during different phases of cyclic population dynamics

Seizures were noted in 2.4 % of the 589 water voles in the wild group; the seizures were caused by handling. There were no sex differences in susceptibility to seizures ($\chi^2 = 0.004$, $p = 0.950$).

The proportion of individuals with seizures was affected by the population phase at which the animals were taken from the population ($\chi^2 = 14.58$, $df = 3$, $p = 0.002$), Fig. 1. Among the water voles from the decline or low phase, the proportion of those susceptible to seizures was significantly higher than that among the water voles from the increase or peak phase (5.88 ± 1.72 % and 0.97 ± 0.56 %, respectively; $\chi^2 = 10.25$, $df = 1$, $p = 0.014$).

Seizures among the water voles born in the vivarium

Seizures were observed in 29.8 % of the 1776 animals in the vivarium group. Susceptibility to seizures was not influenced by sex (males: 29.4 %, females: 30.1 %; $\chi^2 = 0.116$, $df = 1$, $p = 0.734$). There were no seizures during the suckling period, which ended at age 3 weeks.

Age at seizure onset

The date of onset of the first seizure was recorded only for 480 animals. The minimum age at seizure onset was 39 days, and the maximum age at seizure onset was 1105 days. In the seizure group, modal age was 189 days, median age 274 days, and the mean age 324.6 ± 7.8 days. A histogram of the distribution of age at seizure onset is presented in Fig. 2.

Analysis of the data using a mixed linear model, taking into account the year of birth and litter identity (ID) as random factors, showed no effect of sex as a fixed factor ($F_{1,416.5} = 0.312$, $p = 0.577$) on the age of onset of the first seizure (males: 351.5 ± 25.3 days, females: 344.4 ± 25.1 days). The influence of the year of birth and ID was significant ($Z = 2.877$, $p = 0.004$ and $Z = 3.841$, $p < 0.001$, respectively).

Hereditary predisposition to convulsive states

To clarify hereditary predisposition of water voles to handling-induced seizures, we used animals that lived in the vivarium for at least 39 days, which was the minimum age at seizure

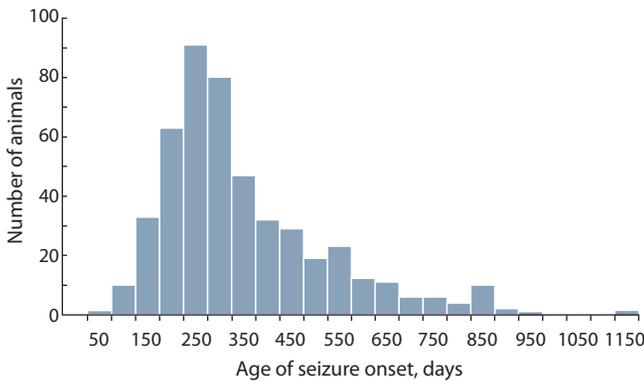


Fig. 2. The histogram of the distribution of age of seizure onset.

Table 1. The percentage of offspring with seizures depending on the convulsive status of parents

Seizures in parents		Seizures in offspring	
Mother	Father	N	%, mean ± SE
No	No	627	15.00 ± 1.42
	Yes	316	39.56 ± 2.75
Yes	No	441	37.87 ± 2.31
	Yes	272	52.57 ± 3.02

onset in this work. Thus, 1656 water voles from 445 litters were analyzed. The results indicated that the susceptibility of offspring to seizures correlates with seizure status of their parents (Table 1). Namely, the proportion of offsprings with seizures was significantly higher if one or both parents had seizures ($\chi^2 = 151.67$, $df = 3$, $p < 0.001$).

Within “parent–offspring” pairs with known age at seizure onset in both parents and offspring (39 parental pares, 90 offspring), a significant correlation between average ages at seizure onset was found ($r = 0.42$, $p < 0.01$; Fig. 3), indicating a significant additive contribution of genes to this trait and the possibility of selection for “age at seizure onset”.

The relation between the presence of seizures and the life span

To clear up the relation between the presence of seizures and the life span, the linear mixed model was used, which included categorical variables (sex and the presence/absence of seizures) and the birth year and ID as random factors. As a result, we detected significant effects of sex ($F_{1,1410.3} = 5.342$, $p = 0.021$) and of seizure status of an individual ($F_{1,1392.2} = 132.926$, $p < 0.001$) on the life span. Males lived longer than females ($\beta = 24.041 \pm 10.402$, $t = 2.311$, $p = 0.021$). The life span of animals without seizures was significantly shorter than that of animals with seizures ($\beta = -134.458 \pm 10.402$, $t = -11.529$, $p < 0.001$), Fig. 4. A significant impact of the birth year on the life span was noticed (Wald-Z = 3.295, $p = 0.001$),

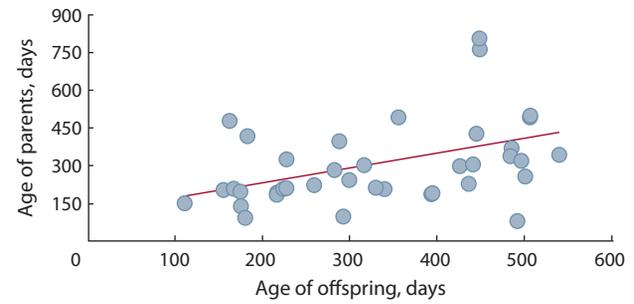


Fig. 3. The correlation between mean ages of seizure onsets within “parent–offspring” pairs.

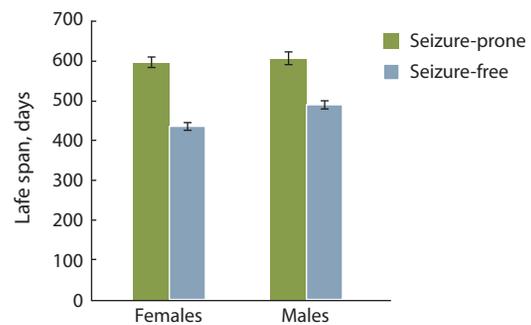


Fig. 4. The relation between the life span and the predisposition to seizures among male and female water voles.

whereas such an effect of ID was statistically insignificant (Wald-Z = 1.419, $p = 0.156$).

Figure 5 shows survival curves of two groups of water voles (with seizures and without seizures). Significant differences were found between them according to the Gehan–Wilcoxon test (GW = -11.452, $p < 0.001$).

The effect of seizure status of parents on the life span of the offspring

To clarify the dependence of life span on the convulsive status of parents, using a linear mixed model, we assessed the influence of fixed factors: sex, predisposition to convulsions of the mother, father and interaction of parents’ phenotypes on the life span of offspring. The birth year and ID were included in the model as random factors. It was found that the life span is affected by sex ($F_{1,1413.6} = 4.580$, $p = 0.033$), by maternal seizure status ($F_{1,366.2} = 7.280$, $p = 0.007$), and by an interaction of these maternal and paternal phenotypes ($F_{1,376.4} = 4.024$, $p = 0.046$). The offspring of parental pairs in which either the mother or both parents did not have seizures manifested a significantly shorter life span ($\beta = -39.442 \pm 12.664$, $t = -3.115$, $p = 0.002$ and $\beta = -50.504 \pm 25.176$, $t = -2.006$, $p = 0.046$, respectively). The effect of paternal seizure status on the offspring lifespan was not statistically significant ($F_{1,393.5} = 0.253$, $p = 0.615$), and neither was the effect of ID (Wald-Z = 1.760, $p = 0.078$). Data on the mean life span of offspring and seizure status of the parents is given in Table 2.

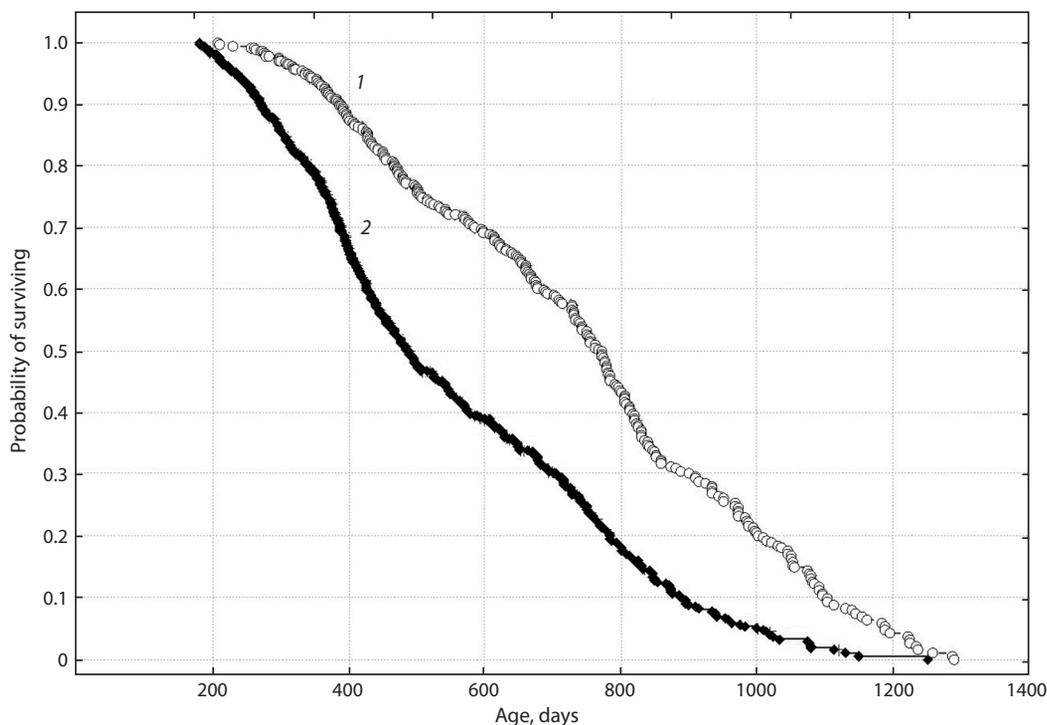


Fig. 5. Survival curves of two groups of water voles: seizure-prone (1) and seizure-free (2).

Table 2. Mean life span of the offspring in relation to seizure status of the parents

Seizures in parents		Total number of offspring	Life span (days), mean ± SE
Mother	Father		
No	No	540	482.0 ± 19.2
	Yes	282	513.8 ± 21.5
Yes	No	386	541.9 ± 20.4
	Yes	241	523.2 ± 22.6

Discussion

These studies establish the occurrence of genetically based, tonic-clonic convulsions in water voles in response to handling. The incidence of seizures did not depend on sex. Prolonged tonic-clonic seizures in response to handling have been documented for the bank vole (Schönecker, 2009), the meadow vole (Bronson, De La Rosa, 1994), and the Mongolian gerbil (Ludvig et al., 1991; Buckmaster, 2006). The pattern of seizures in the water vole turned out to be similar to that in bank voles, meadow voles, and Krushinsky–Molodkina audiogenic rat strain, except that in bank and meadow voles, there is no “wild running” phase observed in Krushinsky–Molodkina rats (Bronson, De La Rosa, 1994; Schönecker, 2009; Poletaeva et al., 2017) and in water voles.

According to the literature, reflex epileptiform seizures caused by tactile or acoustic stimuli are typical only for small rodents, which is associated with the structural and functional features of their central nervous system, due to which pathological, from a human point of view, behavioral responses to

stimuli that inform about danger are carried out (Poletaeva et al., 2017; Fedotova et al., 2021). Convulsive states are probably a component of adaptive behavior formed in a “predator–prey” system. Short-term clonic convulsions in response to tactile or acoustic stimulation can have a frightening effect on raptors that capture prey with their claws, and rapid running, as the next phase of a convulsive seizure, allows the prey to quickly take refuge in a safe place.

Individuals susceptible to seizures may have a selective advantage during high predator pressure. It is known that the number of specialized bird predators with great mobility changes synchronously with the dynamics of population size of water voles (Weber et al., 2002). The results of crossings of water voles, differing in their predisposition to seizures, indicate the hereditary transmission of this trait: the proportion of offspring with seizures significantly increases if one or, especially, both parents had seizures. Therefore, an increase in the prevalence of individuals with seizures during the phases of decline and depression of cyclic population dynamics may

be explained by positive selection for seizure susceptibility at the peak phase of population numbers accompanied by an increase in predator pressure, or greater inbreeding at the low phase, or random processes.

In our study, the minimum age at seizure onset among water voles was 39 days, the maximum age at seizure onset was 1105 days, and the median age at seizure onset, 274 days. Among bank voles, median age at seizure onset is 157 days (Schönecker, 2009). Seizures develop at a later age in voles and gerbils than in seizure-prone laboratory rats and mice.

According to our findings, in the water vole, age at seizure onset is an inherited trait that can be selected for. Identification of the main mutation that triggers seizures and elucidation of the mechanisms underlying the observed correlations between age and seizures may provide insights into the pathogenesis of seizures in other mammals. It is now known that 70–80 % of epilepsy cases are associated with one or more genetic factors, with the remaining cases attributable to acquired conditions such as a tumor, stroke, or head injury (Myers, Mefford, 2015).

Genes that control the predisposition to convulsive states in response to handling can be stored in populations due to their positive pleiotropic influence on life span. Here, we demonstrated that under vivarium conditions, individuals with seizures live longer than individuals with a normal phenotype. These results contradict mouse studies that have addressed the effects of individual genes on epileptogenesis and life span (Marshall et al., 2021); this discrepancy points to diversity of genetic mechanisms mediating the association of such a behavioral phenotype with viability.

The presence of common links in the effects of genes that control the onset of handling-induced seizures and life span confirms the positive effect of the mother's convulsive status on the life span of the offspring in water voles. The offspring of mothers with seizures were found to live longer than the offspring of mothers with the normal phenotype, whereas the offspring life span was not affected by seizure status of the father. It is possible that genetic predisposition of females to seizures correlates with physiological parameters during pregnancy, with lactation capacity, or with maternal behavior, which in turn determine offspring viability. We are planning on investigating this topic in the future.

Thus, the water vole can serve as a suitable experimental model for researching not only the nature of convulsive states but also the evolution of longevity. Further studies are needed regarding physiological and genetic mechanisms behind the handling-induced seizures in the water vole as well as regarding the relation of seizure susceptibility to life history traits.

Conclusion

It is shown for the first time that the water vole can have convulsive seizures in response to handling. When water voles are kept in a vivarium, the proportion of animals susceptible to convulsive seizures is higher among water voles caught in the decline or depression phase of a cycling natural population than among water voles caught in the increase or peak phase.

The susceptibility to seizures is an inherited trait correlating with the life span; this finding implies the possibility of natural and artificial selection for this trait. Our data open up opportunities for the creation of water vole strains that differ in their susceptibility to handling-induced epileptiform seizures, for research on the mechanisms of epileptogenesis and longevity.

References

- Bronson F.H., De La Rosa J. Tonic-clonic convulsions in meadow voles. *Physiol. Behav.* 1994;56(4):683-685. DOI 10.1016/0031-9384(94)90227-5.
- Buchhalter J.R. Animal models of inherited epilepsy. *Epilepsia*. 1993; 34(Suppl. 3): S31-S41. DOI 10.1111/j.1528-1157.1993.tb05921.
- Buckmaster P.S. Inherited epilepsy in Mongolian gerbils. In: Pitkänen A., Schwartzkroin P.A., Moshé S.L. (Eds.). *Models of seizures and epilepsy*. Ch. 21. Burlington: Academic Press, 2006;273-294. DOI 10.1016/b978-012088554-1/50023-2.
- Catala A., Cousillas H., Hausberger M., Grandgeorge M. Dog alerting and/or responding to epileptic seizures: A scoping review. *PLoS One*. 2018;13(12):e0208280. DOI 10.1371/journal.pone.0208280.
- Evsikov V.I., Muzyka V.Y., Nazarova G.G., Potapova O.F., Potapov M.A. Effect of hydrological conditions on reproduction rate and population structure of the European water vole, *Arvicola amphibious*. *Ekologia = Russian Journal of Ecology*. 2017;48(3):290-293. DOI 10.1134/S1067413617030055
- Fedotova I.B., Surina N.M., Nikolaev G.M., Revishchin A.V., Poletaeva I.I. Rodent brain pathology, audiogenic epilepsy. *Biomedicines*. 2021;9(11):1641. DOI 10.3390/biomedicines9111641.
- Grone B.P., Baraban S.C. Animal models in epilepsy research: legacies and new directions. *Nat. Neurosci.* 2015;18(3):339-343. DOI 10.1038/nn.3934.
- Gülseroy E., İyigün S.S., Erol B.B. An overview of seizures and epilepsy in rabbits: etiological differences and clinical management. *Nauk. Visn. Vet. Med.* 2021;1(165):159-164. DOI: 10.33245/2310-4902-2021-165-1-159-164.
- Jackson R.K. Unusual laboratory rodent species: research uses, care, and associated biohazards. *ILAR J.* 1997;38(1):13-21. DOI 10.1093/ilar.38.1.13.
- Krotova L.G. Changes in the adrenal glands and carbohydrate metabolism in the water vole (*Arvicola terrestris*) in the spring-summer period. In: *Issues of the intraspecific variability of mammals. Proceedings of the Institute of Biology. Sverdlovsk, 1962*;29:129-140. (in Russian)
- Ludvig N., Farias P.A., Ribak C.E. An analysis of various environmental and specific sensory stimuli on the seizure activity of the Mongolian gerbil. *Epilepsy Res.* 1991;8(1):30-35. DOI 10.1016/0920-1211(91)90033-c.
- Marshall G.F., Gonzalez-Sulser A., Abbott C.M. Modelling epilepsy in the mouse: challenges and solutions. *Dis. Model. Mech.* 2021;14(3): dmm047449. DOI 10.1242/dmm.047449.
- Muñoz L.J., Carballosa-Gautam M.M., Yanowsky K., García-Atarés N., López D.E. The genetic audiogenic seizure hamster from Salamanca: The GASH:Sal. *Epilepsy Behav.* 2017;71(Pt. B):181-192. DOI 10.1016/j.yebeh.2016.03.002.
- Myers C.T., Mefford H.C. Advancing epilepsy genetics in the genomic era. *Genome Med.* 2015;7(1):91. DOI 10.1186/s13073-015-0214-7.
- Okudan Z.V., Özkara Ç. Reflex epilepsy: triggers and management strategies. *Neuropsychiatr. Dis. Treat.* 2018;14:327-337. DOI 10.2147/NDT.S107669.
- Pakozdy A., Halasz P., Klang A. Epilepsy in cats: theory and practice. *J. Vet. Intern. Med.* 2014;28(2):255-263. DOI 10.1111/jvim.12297.

- Poletaeva I.I., Kostyna Z.A., Surina N.M., Fedotova I.B., Zorina Z.A. The Krushinsky–Molodkina genetic rat strain as a unique experimental model of seizure states. *Vavilovskii Zhurnal Genetiki i Selektzii = Vavilov Journal of Genetics and Breeding*. 2017;21(4):427-434. DOI 10.18699/VJ17.261. (in Russian)
- Schönecker B. Handling-induced tonic/clonic seizures in captive born bank voles (*Clethrionomys glareolus*). *Appl. Anim. Behav. Sci.* 2009; 118(1-2):84-90. DOI 10.1016/j.applanim.2009.02.02.
- Skradski S.L., White H.S., Ptáček L.J. Genetic mapping of a locus (mass1) causing audiogenic seizures in mice. *Genomics*. 1998;49(2): 188-192. DOI 10.1006/geno.1998.5229.
- The Water vole: an image of the species. Moscow: Nauka, 2001;527. (in Russian)
- Weber J.-M., Aubry S., Ferrari N., Fischer C., Lachat Feller N., Meia J.-S., Meyer S. Population changes of different predators during a water vole cycle in a central European mountainous habitat. *Ecography*. 2002;25(1):95-101. DOI 10.1034/j.1600-0587.2002.250111.

ORCID ID

G.G. Nazarova orcid.org/0000-0001-8067-7596
L.P. Proskurnyak orcid.org/0000-0001-6138-2696

Acknowledgements. The study was supported by Russian-government-funded project No. 122011800268-1.

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

Received February 8, 2022. Revised April 18, 2022. Accepted April 19, 2022.

Original Russian text www.bionet.nsc.ru/vogis/

The GWAS-MAP|ovis platform for aggregation and analysis of genome-wide association study results in sheep

A.V. Kirichenko¹✉, A.S. Zlobin¹✉, T.I. Shashkova², N.A. Volkova³, B.S. Iolchiev³, V.A. Bagirov³, P.M. Borodin², L.C. Karssen⁴, Y.A. Tsepilov^{1, 5}, Y.S. Aulchenko¹

¹ Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

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

³ Federal Research Center for Animal Husbandry named after Academy Member L.K. Ernst, Dubrovitsy, Moscow region, Russia

⁴ PolyKnomics BV, 's-Hertogenbosch, The Netherlands

⁵ Novosibirsk State University, Novosibirsk, Russia

✉ kianvl@bionet.nsc.ru; zlobin@bionet.nsc.ru

Abstract. In recent years, the number of genome-wide association studies (GWAS) carried out for various economically important animal traits has been increasing. GWAS discoveries provide summary statistics that can be used both for targeted marker-oriented selection and for studying the genetic control of economically important traits of farm animals. In contrast to research in human genetics, GWAS on farm animals often does not meet generally accepted standards (availability of information about effect and reference alleles, the size and direction of the effect, etc.). This greatly complicates the use of GWAS results for breeding needs. Within the framework of human genetics, there are several technological solutions for researching the harmonized results of GWAS, including one of the largest, the GWAS-MAP platform. For other types of living organisms, including economically important agricultural animals, there are no similar solutions. To our knowledge, no similar solution has been proposed to date for any of the species of economically important animals. As part of this work, we focused on creating a platform similar to GWAS-MAP for working with the results of GWAS of sheep, since sheep breeding is one of the most important branches of agriculture. By analogy with the GWAS-MAP platform for storing, unifying and analyzing human GWAS, we have created the GWAS-MAP|ovis platform. The platform currently contains information on more than 34 million associations between genomic sequence variants and traits of meat production in sheep. The platform can also be used to conduct colocalization analysis, a method that allows one to determine whether the association of a particular locus with two different traits is the result of pleiotropy or whether these traits are associated with different variants that are in linkage disequilibrium. This platform will be useful for breeders to select promising markers for breeding, as well as to obtain information for the introduction of genomic breeding and for scientists to replicate the results obtained.

Key words: genome-wide association study; marker-based breeding; sheep; database; web interface.

For citation: Kirichenko A.V., Zlobin A.S., Shashkova T.I., Volkova N.A., Iolchiev B.S., Bagirov V.A., Borodin P.M., Karssen L.C., Tsepilov Y.A., Aulchenko Y.S. The GWAS-MAP|ovis platform for aggregation and analysis of genome-wide association study results in sheep. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(4):378-384. DOI 10.18699/VJGB-22-46

Платформа GWAS-MAP|ovis для хранения и анализа результатов полногеномных ассоциативных исследований овец

А.В. Кириченко¹✉, А.С. Злобин¹✉, Т.И. Шашкова², Н.А. Волкова³, Б.С. Иолчиев³, В.А. Багиров³, П.М. Бородин², Л.С. Карссен⁴, Я.А. Цепилов^{1, 5}, Ю.С. Аулченко¹

¹ Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

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

³ Федеральный исследовательский центр животноводства – ВИЖ им. академика Л.К. Эрнста, пос. Дубровицы, Московская область, Россия

⁴ PolyKnomics BV, Хертогенбос, Нидерланды

⁵ Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ kianvl@bionet.nsc.ru; zlobin@bionet.nsc.ru

Аннотация. В последние годы увеличивается количество полногеномных исследований ассоциаций (ПГИА, GWAS), проведенных для различных экономически важных признаков животных. Результаты этих исследований представлены в виде суммарных статистик, которые можно использовать для изучения генетического контроля экономически важных признаков сельскохозяйственных животных, в том числе и при разра-

ботке методик маркер-ориентированной селекции. В большинстве случаев ПГИА сельскохозяйственных животных не соответствуют общепринятым в области исследований генетики человека стандартам формата публикаций результатов ПГИА в виде суммарных статистик (наличие информации об эффекторном и референсном аллелях, значение и направление эффекта и др.). Это существенно затрудняет использование суммарных статистик для нужд селекции. В области исследований генетики человека имеется несколько технологических решений для анализа результатов ПГИА, в том числе одно из самых крупных – платформа GWAS-MAP. Для других видов живых организмов, включающих и экономически важных сельскохозяйственных животных, подобных решений нет. В настоящей работе мы сфокусировались на создании схожей платформы для работы с суммарными статистиками ПГИА различных признаков овец, так как овцеводство в последнее время становится все более актуальной областью сельского хозяйства. По аналогии с платформой GWAS-MAP для хранения, унификации и анализа GWAS человека мы создали платформу GWAS-MAP|ovis. На сегодняшний день платформа содержит информацию о более чем 34 млн ассоциаций между вариантами геномной последовательности и признаками мясной продуктивности. Платформа может быть использована и для проведения анализа колокализации – метода, который позволяет установить, является ли ассоциация определенного локуса с двумя разными признаками результатом плейотропии или же данные признаки ассоциированы с разными вариантами, которые находятся в неравновесии по сцеплению. Эта платформа будет полезна как селекционерам для выбора перспективных маркеров для селекции (эффекты и аллели различных маркеров, влияющих на изучаемые признаки), так и для ученых, ведущих исследования в области генетики овец.

Ключевые слова: полногеномное исследование ассоциаций; маркер-ориентированная селекция; овцы; база данных; веб-интерфейс.

Introduction

A genome-wide association study (GWAS) is modern technique to detect genome loci associated with different traits both quantitative and binary ones. GWASs have been widely used in human and animal genetics (Visscher et al., 2017). Since 2007, an annual increase of GWAS-based investigations to search for trait-associated loci has been observed. A GWAS results in the so-called summary statistics, a text tables where each line is a result of association analysis between a studied trait and a single nucleotide polymorphism (SNP). In breeding, such summary statistics can be used to devise marker-assisted and genomic selection models and identify target candidates for genome editing. Thanks to the new-generation technologies of genome-wide genotyping and next-generation sequencing, modern GWASs enable one to study hundreds of thousands and oftentimes millions of SNP associations with dozens of traits in huge samples including some tens thousands of animals. In other words, GWAS summary statistics are huge datasets (Visscher et al., 2017), whose processing is far from being trivial and requires, apart from their collection, a platform to provide proper quality control, storage, access and processing tools.

In the field of human genetics, such a platform, known as GWAS-MAP (Shashkova et al., 2020), serves for unification, storing and analysis of millions of associations for thousands of human traits. However, to our knowledge, there is still no such a solution even for a single kind of livestock farming. There are only databases to process quantitative trait loci significantly associated with studied traits. The biggest of such databases is AnimalQTLdb that contains information on quantitative-trait loci for different kinds of livestock (Hu et al., 2013). There are similar solutions for particular kinds of animals such as the iSheep

database that provides access to the known quantitative-trait loci, genotyping and sequencing data of sheep (Wang et al., 2021). Since none of the mentioned solutions provides opportunities for storing, unification and analysis of GWAS summary statistics, the objective of the presented study was to create such a platform for sheep.

Sheep breeding is an important branch of animal farming with mutton being its most popular product. Since 2013, more than 20 GWASs of economically important sheep traits have been known. Here, S. Bolormaa et al. (2016) is of particular note being the biggest GWAS for a meat productivity traits in sheep. The authors investigated a cohort of more than 10,000 animals to study information on 56 different genetic traits related to meat productivity and found more than 70 significantly associated SNPs included in 23 loci.

Here we present the GWAS-MAP|ovis platform that is analogous to GWAS-MAP and can be used for aggregation, unification, storing and analysis of the GWAS summary statistics of the economically important traits affecting meat production in sheep. Currently, the platform contains more than 34 mln associations for 80 such traits. Breeders may find it useful for selection based on the effects of the alleles associated with studied traits and on perspective selection markers. It may also be useful for the scientists investigating sheep genetics. In particular, the platform enables for GWAS quality control; search for significantly associated SNPs; GWAS results visualization and colocalization analysis.

Materials and methods

Developing the GWAS-MAP|ovis platform. GWAS results are commonly summary statistics being big text tables in which a single line is the results of association analysis

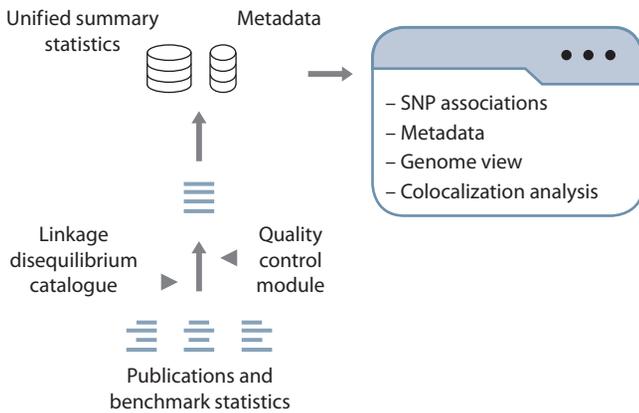


Fig. 1. GWAS-MAP_{ovis} flow chart.

to match a studied trait and SNP. Summary statistics can include different columns such as SNP identifier (if any); SNP position in a physical genome map (chromosome and position), SNP imputation or genotyping quality control metrics; effective and reference alleles; effective allele frequency; SNP effect estimation; standard error and *p*-value; sample size and other technical variables specific for particular research. The minimum set of the variables necessary for summary statistics processing includes an SNP identifier, effective and reference alleles, effect estimation and a standard error. Without loss of information, the standard error can be replaced by either the *p*-value or by *z*-statistic. Since the number of SNPs is huge, the size of the file containing a summary statistics table can reach several gigabytes in a case the number of characterized SNPs exceeds 20 million.

The GWAS-MAP platform (Shashkova, 2020; Shashkova, Aulchenko, 2020) to store GWAS results in the human was used as a basis to design the solution in question. For presentation and visualization of genetic data, the PheLiGe web interface was introduced (Shashkova et al., 2021). The flow chart of GWAS-MAP_{ovis} can be seen in Fig. 1.

To store and process data, GWAS-MAP uses two database management systems, ClickHouse and PostgreSQL. While ClickHouse serves to store matched summary statistics, PostgreSQL is used to store the metadata describing the sources summary statistics come from, a GWAS dataset and data analysis features. They may include e. g., a reference to the publication data come from; trait name; description of the model used for analysis; the number of studied animals, etc. Their full list is available in Supplementary material 1¹. The two systems are deployed in an LXC container and operate using an Ubuntu 18.04 operating system image.

One of the key steps in setting up the database was creating a linkage disequilibrium catalogue and the so-called reference table (see the description below). To provide con-

venient access to the database and enable data visualization as interactive graphs, the PheLiGe-ovis web interface was deployed in another LXC container and operated using the same operating system image. The web interface is available at <https://pheligeovis.icgbio.ru/>.

The designed platform having two databases to store summary statistics and metadata and a web interface was called GWAS-MAP_{ovis} and is available via public (web) and private (ssh) interfaces. The general principles of how to operate its web interface and main tools are described in the on-line assistant program, whose window pops up when pressing the red button under the Associations menu item in the top right corner of the platform's main page.

Linkage disequilibrium catalogue is an SNP list that contains the following information for each included SNP: identifier (*rs_id*); chromosome and position (base on the OAR_V3.1 genome assembly); effective and reference allele, and effective allele frequency in referent population. To create the catalog were used the genome data of a reference population (*N* = 96) including 18 Romanov breed sheep, 6 Katahdin sheep, 10 argali, 48 F₁ hybrids from crossing F₁ (Romanov and argali) and Romanov breeds, and 14 F₁ hybrids from crossing Romanov and Katahdin breeds. All the animal had been genotyped using Infinium[®] HD SNP Bead-Chip (606,060 OHII, Illumina Inc., San Diego, CA, USA) and used for replication in our recent paper (Zlobin et al., 2021). Preparing the reference table involved removing all duplicated and monomorphic SNPs; the alleles were placed in lexicographic order; sex chromosome data were removed. For the time being, the disequilibrium catalogue contains information on 523.578 SNPs. Their autosomal distribution can be found in Supplementary material 2 and their minor allele frequencies in Supplementary material 3.

The linkage disequilibrium catalogue is files containing the genotypes of the reference population animals for every SNP included in the reference table. The files are in binary format PLINK (.bed, .bim and .fam). The catalogue is used to calculate the linkage disequilibrium between SNPs for finding a proxy SNP in the PheLiGe-ovis web interface.

Module of unification and quality control GWAS results. This module is used to download summary statistics in the database. Since these files are obtained from different sources and often have different formats, they have to be unified first. At this stage, matches between *rs_id* and the position in the chromosome and between the effective and reference alleles are checked.

As the next stage, the summary statistics pass quality control that includes three stages: (1) comparison of GWAS allele frequencies against those in the linkage disequilibrium catalogue; (2) checking effect-size distribution, and (3) verifying if the *p*-value corresponds to *z*-statistic (effect estimation divided by a standard estimation error).

¹ Supplementary materials 1–4 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Kirichenko_engl.pdf

Data collections included in the GWAS-MAP|ovis database

Collection	Source	Number of animals	Number of traits	Number of SNPs for each trait	Included breeds
Bolormaa et al., 2016	Publication	10,613	56	4,880,774	Merino, Paul Dorset, Border, Lester, Suffolk, White Suffolk, Texel, Corriedale, Coopworth and different hybrids
MARH	Original data	50–108	24	453,024	Hybrid sheep (Romanov crossed with Katahdin, argali and moufflons)

The module generates an html report that can be used to detect unification and quality-control errors. If data pass the control, the summary statistics are stored in the platform's database.

Results and discussion

Content management

To fill the database, a search for published summary statistics was carried out in an early published database of the QTL and genes associated with meat productivity traits in sheep (Zlobin et al., 2019). Out of 46 publications analyzed, only S. Bolormaa et al. (2016) provided access to genome-wide summary statistics. Before being loaded in the GWAS-MAP|ovis database, all the data were unified and they passed the quality control. As a result, 488,074 associations for each of the 56 traits listed in the paper were downloaded in.

In addition to the data provided by S. Bolormaa et al. (2016), the GWAS-MAP|ovis database also contains the summary statistics obtained from the Russian sheep population ($N = 50–108$) early used for replication analysis (Zlobin et al., 2021). This summary statistics described 7 composite indices related to the meat productivity measured for three different time periods (6, 42 and 90 days after birth) and an animal's weight also measured for three time periods (for more details, see Supplementary material 4). As a result, the platform's database has accumulated information on more than 34 million SNP associations for 80 different traits related to meat productivity in sheep. The available dataset and its short description are presented in the Table.

Pleiotropy analysis

One of the advantages of the designed platform is that it can be used to perform colocalization analysis (SMR- θ) (Momozawa et al., 2018), which enables one to compare the association patterns of a particular genome locus for different traits and make conclusions whether this locus has a pleiotropic effect on a certain trait. In a nutshell, in presence of this effect, the ratio of estimated SNP effects on two traits in a studied locus is expected to be insignificant. However, if the association patterns are different i. e., the effect ratio changes significantly from one SNP to another,

it is most likely that the locus contains different functional polymorphisms for each of the traits and they are being in linkage disequilibrium.

Statistic θ is a weighted correlation, whose computation requires information on p -values and an effect direction. The high absolute value (e. g., $|\theta| > 0.7$) means a locus has a pleiotropic effect on investigated traits. If θ is positive, the SNPs of the locus have a similar effect direction towards the traits, and a divergent one if negative. An example of colocalization analysis is presented in Fig. 2.

In the framework of the presented study, colocalization analysis was carried for the most significantly associated SNP of the *LCORLI* gene in the data by (Bolormaa et al., 2016). For SNP rs401834107, 12 traits of genome-wide significance level (p -value $< 5E-07$) associated with this locus were indicated. The SMR- θ method was applied to compare association patterns for all the 12 traits, the results are demonstrated in Fig. 3. The heat map characterizing θ -values shows the traits form two clusters with different effect correlation directions, which means the locus had a pleiotropic effect inside each of the clusters. At the same time, the same strong effect was observed between traits from different clusters with divergent SNP effect on the traits (negative θ value). It is noteworthy that the slaughter weight trait was a dropout since some of its θ -values turned out to be low (possible pleiotropy absence). However, due to the pleiotropic effect the locus had on the other traits, the low θ -values in this particular case can be considered either as a statistic artifact or as a value produced due to insignificant analysis capacity.

Thus, it has been found that locus rs401834107 had a pleiotropic effect on at least 12 different traits associated with meat production in sheep. The presented data is an example of a first in the world colocalization analysis performed using sheep GWAS results.

Searching for candidate DNA markers for marker-assisted selection

In addition to the described pleiotropy analysis, another advantage of the GWAS-MAP|ovis platform is that it allows one to use previously loaded summary statistics to search for candidate markers to perform marker-assisted selection (MAS). In the presented study, such a search for

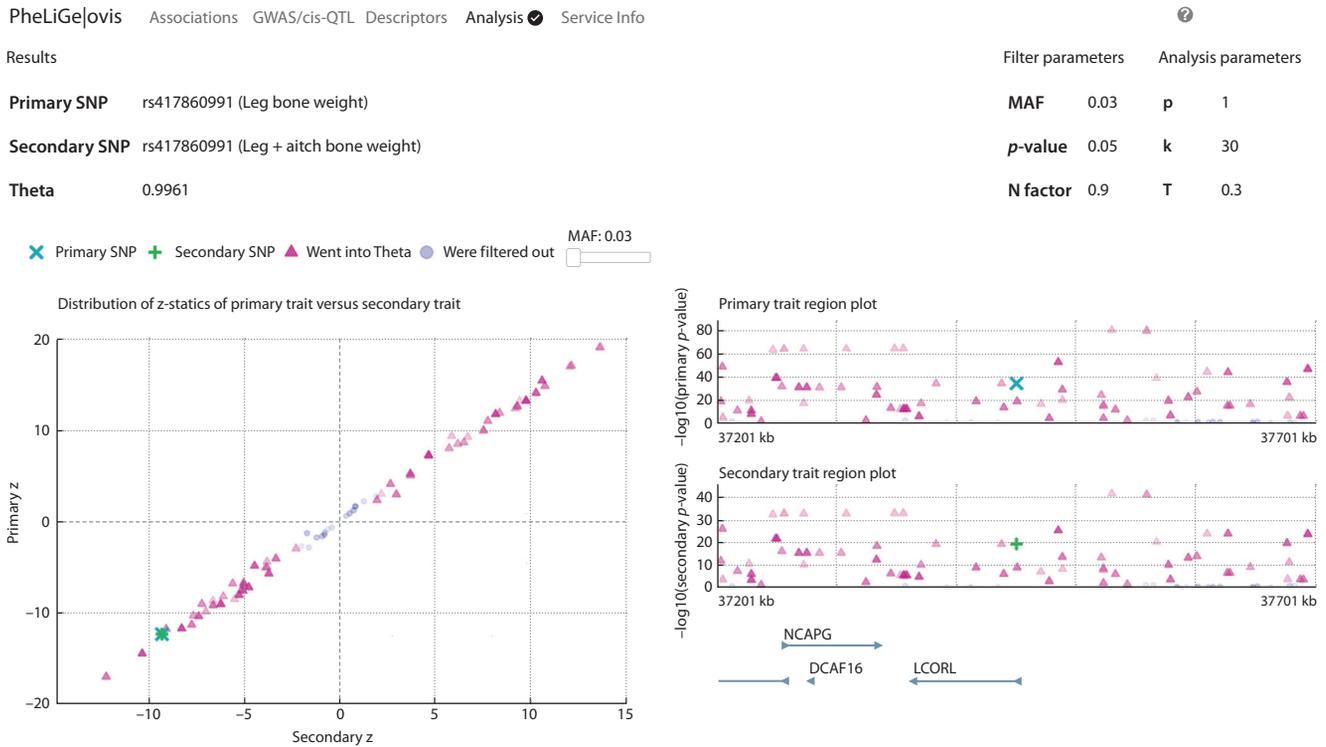


Fig. 2. Results of the colocalization analysis performed using the GWAS-MAP|ovis platform. The left part is a joint distribution of z-statistics for a main (leg bone weight, LEGBONE, X-axis) and secondary (leg + aitch bone weight, BONE, Y-axis) traits. The right part is regional association graphs for each of the traits (LEGBONE and BONE, respectively). The θ is indicated in the top left corner over the graph.

the hot-carcass-weight (HCWT) trait was performed to estimate the markers' predictive potential in a Russian sheep cohort. To begin with, a clumping procedure was carried out. To do this, the private (ssh) interface was used to select the summary statistics of interest and set up their significant threshold of p -value $< 1E-07$. Four independent loci were detected (rs406365427, rs423487570, rs161042491, rs427891980). A table containing information on SNP with the highest associations for each of the loci, their chromosomes, positions, effector and reference alleles and p -values was formed for further analysis.

The found loci were used to estimate the breeding value (BV) of 94 animals (Zlobin et al., 2021) to be the back-cross hybrids of Romanov and argali sheep. Their BV was estimated as:

$$BV = \sum(b \cdot g),$$

where b is the effect of the effective allele of the most associated SNP in every locus; g is the genotypes of 94 animals for that SNP encoded as 0, 1, 2 in relation to the effective allele's quantity.

Cases then underwent linear regression to find how significant the estimated BVs were for predicting a species' phenotype. As a trait, an animal's weight in six days after birth was considered. The used model was $p \sim BV + cov$, where p is a phenotype's value; cov is the covariates includ-

ing information about animal gender, yield number, and two first main components of their kinship matrix. As a result, the estimated BV was found to be significantly associated (p -value = 0.03) with the trait, which confirmed that the selected loci had a significant effect on an animal's weight in six days after birth.

In addition to the performed analysis, it was also estimated what body weight increase could be expected in the animals whose BV was included in the fourth (top) quartile ($> 75\%$). Student's t -criterion was used to compare the mean values of the phenotype corrected for the above-mentioned covariates for the whole cohort and the animals included in top BV quartile. It was found that the weight of the top-quartile animals was 1.8% (around 55 grams, p -value = 0.67) higher than that in the averaged population. Thus, the selected markers can be potentially used for development of MAS test systems in sheep.

Conclusion

The GWAS-MAP|ovis platform has been designed for aggregation, unification, storage and analysis of GWAS summary statistics in sheep. Today, it contains more than 34 million SNP associations for 80 different traits related to meat productivity in sheep. The platform enables one to search for information on associations for a particular SNP, plot Manhattan graphs and carry out colocalization

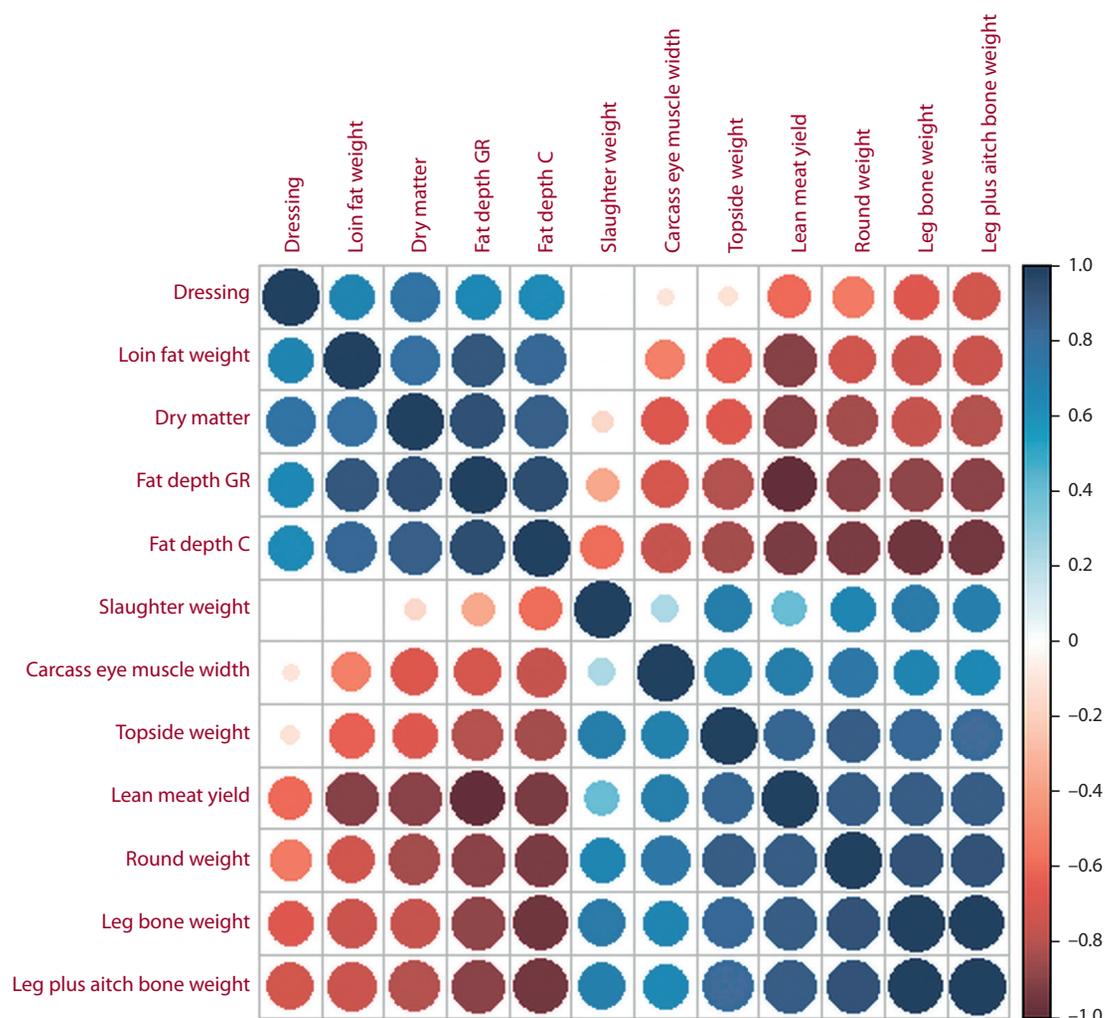


Fig. 3. Results of colocalization analysis for SNP rs401834107 for 12 original traits, from (Bolormaa et al., 2016).

The red color marks negative θ -values, the blue – positive ones. The brighter the color and bigger the circle, the closer is the absolute value to one.

analysis. The platform can be useful for scientists as an investigating tool to study sheep genetics and for breeders as a solution for searching for potential MAS loci.

References

Bolormaa S., Hayes B.J., van der Werf J.H.J., Pethick D., Goddard M.E., Daetwyler H.D. Detailed phenotyping identifies genes with pleiotropic effects on body composition. *BMC Genomics*. 2016;17:224. DOI 10.1186/s12864-016-2538-0.

Hu Z.-L., Park C.A., Wu X.-L., Reecy J.M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res*. 2013; 41(D1):D871-D879. DOI 10.1093/nar/gks1150.

Momozawa Y., Dmitrieva J., Théâtre E., Deffontaine V., Rahmouni S., Charlotiaux B., Crins F., Docampo E., Elansary M., Gori A.-S., Lecut C., Mariman R., Mni M., Oury C., Altukhov I., Alexeev D., Aulchenko Y., Amininejad L., Bouma G., Hoentjen F., Löwenberg M., Oldenburg B., Pierik M.J., van der Meulen-de Jong A.E., van der Woude C.J., Visschedijk M.C., International IBD Genetics

Consortium, Lathrop M., Hugot J.-P., Weersma R.K., De Vos M., Franchimont D., Vermeire S., Kubo M., Louis E., Georges M. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun*. 2018;9(1):2427. DOI 10.1038/s41467-018-04365-8.

Shashkova T.I., Aulchenko Y.S. Database for storing and quickly accessing the results of genome-wide and regional association studies. Patent RF No. 2020620869. 2020. (in Russian)

Shashkova T.I., Gorev D.D., Pakhomov E.D., Shadrina A.S., Sharapov S.Z., Tsepilov Y.A., Karssen L.C., Aulchenko Y.S. The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP[homo] database of 70 billion genetic associations of human traits. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):876-884. DOI 10.18699/VJ20.686.

Shashkova T.I., Pakhomov E.D., Gorev D.D., Karssen L.C., Joshi P.K., Aulchenko Y.S. PheLiGe: an interactive database of billions of human genotype-phenotype associations. *Nucl. Acids Res*. 2021; 49(D1):D1347-D1350.

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.

Wang Z.-H., Zhu Q.-H., Li X., Zhu J.-W., Tian D.-M., Zhang S.-S., Kang H.-L., Li C.-P., Dong L.-L., Zhao W.-M., Li M.-H. iSheep: an integrated resource for sheep genome, variant and phenotype. *Front. Genet.* 2021;12:714852. DOI 10.3389/fgene.2021.714852.

Zlobin A.S., Nikulin P.S., Volkova N.A., Zinovieva N.A., Iolchiev B.S., Bagirov V.A., Borodin P.M., Aksenovich T.I., Tsepilov Y.A. Multivariate analysis identifies eight novel loci associated with meat productivity traits in sheep. *Genes.* 2021; 12(3):367. DOI 10.3390/genes12030367.

Zlobin A.S., Volkova N.A., Borodin P.M., Aksenovich T.I., Tsepilov Y.A. Recent advances in understanding genetic variants associated with growth, carcass and meat productivity traits in sheep (*Ovis aries*): an update. *Arch. Anim. Breed.* 2019;62(2):579-583. DOI 10.5194/aab-62-579-2019.

ORCID ID

A.V. Kirichenko orcid.org/0000-0002-9972-9517

A.S. Zlobin orcid.org/0000-0001-9473-2410

N.A. Volkova orcid.org/0000-0002-7784-2201

P.M. Borodin orcid.org/0000-0002-6717-844X

Y.A. Tsepilov orcid.org/0000-0002-4931-6052

Acknowledgements. The study was supported by Kurchatov Genomic Center of Institute of Cytology and Genetics of SB RAS.

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

Received October 26, 2021. Revised April 4, 2022. Accepted April 11, 2022.

Original Russian text www.bionet.nsc.ru/vogis/

Taxonomic structure of bacterial communities in sourdoughs of spontaneous fermentation

V.K. Khlestkin¹✉, M.N. Lockachuk², O.A. Savkina², L.I. Kuznetsova², E.N. Pavlovskaya², O.I. Parakhina²

¹ All-Russian Research Institute of Genetics and Breeding of Farm Animals – Branch of L.K. Ernst Federal Research Center for Animal Husbandry, Pushkin, St. Petersburg, Russia

² Saint-Petersburg Branch of the Scientific Research Institute for the Baking Industry, Pushkin, St. Petersburg, Russia

✉ dir2645@yandex.ru

Abstract. The article is devoted to the study of the microbiome of spontaneously fermented sourdoughs. The aim of the work was to study the influence of the technological parameters of sourdough propagations on the taxonomic structure of the microbiome of spontaneously fermented sourdoughs. Two spontaneously fermented sourdoughs were studied: dense rye sourdough and liquid rye sourdough, both prepared using the same batch of peeled rye flour. To study the taxonomic structure of the sourdough microbiome in dynamics, the method of high-throughput sequencing of 16S rRNA gene fragments of microorganisms was used. It was shown that the technological parameters of sourdough (humidity, temperature) do not affect the taxonomic composition of the microbiome of dense rye or liquid rye sourdough at the phylum/class/genus level. It was found that during the first three days of propagations, bacteria from the phyla Proteobacteria and Firmicutes dominated in the microbial community. In the phylum Proteobacteria, microorganisms from the order Enterobacterales took a large share, which persisted for three days of backslipping. The phylum Firmicutes was represented by lactic acid bacteria of the genera *Weissella*, *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Lactococcus*. It was established by classical microbiological methods that after a day of fermentation, the number of lactic acid bacteria cells was significantly higher in liquid rye sourdough compared to dense one. However, with further propagation of sourdoughs, the number of cells was comparable, while significant changes occurred at the level of genera and species. It was shown that as the relative number of lactic acid bacteria of the genus *Lactobacillus* increased, a gradual displacement of the coccal forms of *Lactococcus*, *Leuconostoc*, *Weissella*, *Pediococcus* happened. With further propagation of sourdough after 10 days, the position of the dominant groups of bacteria was occupied by representatives of the phylum Firmicutes, lactic acid bacteria of the genus *Lactobacillus*. The influence of the mode and parameters of the sourdough on the species composition of lactobacilli, which demonstrated a low bacterial diversity, is shown. In the first three days of propagations, lactobacilli *L. curvatus*, *L. brevis*, and *Lactiplantibacillus* sp. dominated in both sourdoughs. After a month of backslipping, *Fructilactobacillus sanfranciscensis* and *Companilactobacillus* sp. dominated in dense rye sourdough, and *L. pontis* dominated in liquid rye sourdough. Key words: rye sourdough; microbiome; microbial community; lactobacillus; high-throughput sequencing; fermentation; bakery products.

For citation: Khlestkin V.K., Lockachuk M.N., Savkina O.A., Kuznetsova L.I., Pavlovskaya E.N., Parakhina O.I. Taxonomic structure of bacterial communities in sourdoughs of spontaneous fermentation. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(4):385-393. DOI 10.18699/VJGB-22-47

Таксономическая структура бактериальных сообществ в хлебных заквасках спонтанного брожения

В.К. Хлесткин¹✉, М.Н. Локачук², О.А. Савкина², Л.И. Кузнецова², Е.Н. Павловская², О.И. Парахина²

¹ Всероссийский научно-исследовательский институт генетики и разведения сельскохозяйственных животных –

филиал Федерального исследовательского центра животноводства – ВИЖ им. академика Л.К. Эрнста, Пушкин, Санкт-Петербург, Россия

² Санкт-Петербургский филиал Научно-исследовательского института хлебопекарной промышленности, Пушкин, Санкт-Петербург, Россия

✉ dir2645@yandex.ru

Аннотация. Статья посвящена исследованию микробиома хлебных заквасок спонтанного брожения. Цель работы – изучение влияния технологических параметров ведения заквасок на таксономическую структуру микробиома хлебных заквасок спонтанного брожения. Объектами исследования являлись две закваски спонтанного брожения – густая ржаная и жидкая ржаная без заварки, приготовленные с использованием одной партии муки ржаной обдирной. Для изучения таксономической структуры заквасочного микробиома в динамике применяли метод высокопроизводительного секвенирования фрагментов генов 16S рРНК микроорганизмов. Показано, что технологические параметры ведения заквасок (влажность, температура) не оказывают влияния на таксономический состав микробиома густой ржаной и жидкой ржаной заквасок на уровне филумов/классов/родов (но не видов). Установлено, что в течение первых трех суток ведения в микробном сообществе доминирова-

ли бактерии из филумов Proteobacteria и Firmicutes. В филуме Proteobacteria большую долю занимали микроорганизмы из порядка Enterobacteriales, которые сохранялись в течение трех суток ведения заквасок. Филум Firmicutes был представлен молочнокислыми бактериями родов *Weissella*, *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Lactococcus*. Классическими микробиологическими методами установлено, что через одни сутки брожения количество клеток молочнокислых бактерий было значительно выше в жидкой ржаной закваске по сравнению с густой, однако при дальнейшем ведении заквасок количество клеток было сопоставимым, при этом происходили существенные изменения на уровне родов и видов. Выявлено, что по мере увеличения относительной численности молочнокислых бактерий рода *Lactobacillus* происходило постепенное вытеснение кокковых форм *Lactococcus*, *Leuconostoc*, *Weissella*, *Pediococcus*. При дальнейшем ведении заквасок через 10 суток положение доминирующих групп бактерий занимали представители филума Firmicutes – молочнокислые бактерии рода *Lactobacillus*. Показано влияние режима и параметров ведения заквасок на видовой состав лактобацилл, который демонстрировал низкое бактериальное разнообразие. В первые трое суток ведения в обеих заквасках доминировали лактобациллы *L. curvatus*, *L. brevis* и *Lactiplantibacillus* sp. Через месяц ведения в густой ржаной закваске доминировали *Fructilactobacillus sanfranciscensis* и *Companilactobacillus* sp., а в жидкой ржаной – *L. pontis*.
Ключевые слова: ржаная закваска; микробиом; микробное сообщество; лактобактерии; высокопроизводительное секвенирование; ферментация; хлебобулочные изделия.

Introduction

In recent years, an increased interest in the development of sourdough bakery products, including in handicraft and home conditions, has been observed. This is due to the fact that bakery products on sourdough are characterized by improved taste, aroma, nutritional value and resistance to microbial spoilage. Sourdough is a semi-finished bakery product obtained by fermentation of a nutrient mixture of flour and water by lactic acid bacteria or lactic acid bacteria and baking yeast, which can enter the starter from the feedstock or from industrial starter microbial compositions (Auerman, 2009; De Vuyst et al., 2017).

Sourdough maintenance is a technological process that includes regular refreshing of the sourdough with a portion of flour and water, followed by fermentation until ready. After fermentation, part of the ripe sourdough goes to kneading the dough, and part is used for a new refreshment. This ensures a continuous process of fermentation. This makes it possible to maintain the starter microbiota in an active state and obtain a sourdough with the specified biotechnological and physicochemical parameters, which ensures the production of bakery products with the required consumer properties (Kosovan, 2008; Auerman, 2009).

We have previously shown for the first time (Lokachuk et al., 2020) that during the long-term management of national sourdoughs bred using starting microbial compositions, significant changes in the species diversity of lactobacilli occur, leading to predominance of species other than those introduced in the first phase of the breeding cycle, and, nevertheless, allowing to obtain bakery products corresponding to GOST 2077-84. Thus, we have established one of the reasons why consumers may notice a difference in the physicochemical and organoleptic characteristics of products of the same type, since changes in the microbiome during the long-term management of thick rye sourdough lead to a significant change in the content of lactic and acetic acids, in titrated acidity, lifting force, alcohol content in the starter and dough, and hence, the finished product. Samples of rye bread prepared with a long-term starter culture were characterized by higher acidity, lower alcohol content and a large amount of volatile acids (mainly acetic acid).

At the same time, much attention is paid worldwide to the study of the microbiome of spontaneous sourdoughs, in which the initial microflora of raw materials develops. At the same time, the microbiome of domestic starter cultures of spontaneous fermentation and its changes during technological processes remain unexplored, despite its crucial role in shaping the quality and safety of bakery products.

Currently, the use of high-performance sequencing of the 16S rRNA gene makes it possible to expand our knowledge about the taxonomic structure of the starter culture microbiome. This is of great importance, since a large number of factors influence the diversity and structure of starter microbial communities.

Despite the instability of quality and non-sterility of raw materials, starter cultures are stable ecosystems, which may be due to metabolic adaptations in the starter ecosystem (Müller et al., 2001; Minervini et al., 2012; Viiard et al., 2016).

In the fermentation process of sourdoughs, temperature acts as the main factor affecting the dynamics of the microbial community and the kinetics of metabolite production. The fermentation temperature affects the fermentation coefficient, which is the ratio of the concentrations of lactic and acetic acids. Higher temperatures cause a shift towards an increase in the lactic acid content, thereby increasing the acidity of the starter. Homofermentative and facultatively heterofermentative species of lactic acid bacteria (LAB), belonging, for example, to the group *Lactobacillus delbrueckii* and often prevailing in sourdoughs managed at elevated temperatures, cause a rapid decrease in the pH of the water-flour nutrient mixture mainly due to the formation of lactic acid. Heterofermentative types of LAB, as a rule, predominate in sourdoughs that are managed at lower temperatures and during long periods of fermentation, produce a mixture of lactic, acetic acids and/or ethanol (De Vuyst et al., 2017).

There is a positive correlation between the fermentation temperature (< 30 °C) and the frequency of detection of *L. sanfranciscensis* in sourdoughs. On the contrary, such temperatures negatively correlate with the presence of *L. fermentum* and *L. plantarum* species in the sourdough. For example, it has been shown that *L. sanfranciscensis* prevails in traditional renewable thick sourdoughs, which are managed at tempera-

tures below 30 °C and have pH of about 4 – type I sourdoughs according to European classification (Böcker et al., 1995; Hammes et al., 2005) – being optimally associated with yeast *C. humilis* at a temperature of 25–30 °C (Van Kerrebroeck et al., 2017). Since *C. humilis* has a temperature optimum of 27–28 °C and cannot grow at temperatures above 35 °C, elevated temperatures negatively affect this mutualistic symbiosis. *L. sanfranciscensis* is also uncompetitive at elevated temperatures (Gänzle et al., 1998; Vogelmann, Hertel, 2011) compared to other types of LAB.

Humidity and pH also have a significant effect on microbial diversity in sourdoughs. Low pH values stimulate the development of acid-resistant lactobacilli, while higher pH values are favorable for *Enterococcus*, *Lactococcus*, *Leuconostoc*, *Pediococcus* and *Weissella* species (De Vuyst et al., 2017). For example, relatively high pH values, usually exceeding 4.0, in type I French sourdoughs may explain the detection of acid-sensitive species *P. pentosaceus*, *Leuconostoc* and *Weissella* (Robert et al., 2009).

It is known that the acidity of the sourdough affects the content of the specie *L. sanfranciscensis*, the optimal pH value for the growth of which is 5.0. Nevertheless, this type of LAB demonstrates adaptation to acid stress, like many other strains of LAB isolated from sourdoughs.

The *L. sanfranciscensis* species is mainly found in type I sourdoughs with low humidity and is displaced in sourdoughs with too low a pH value, which persists for a significant period of time during the fermentation, since this type of lactobacilli does not grow at a pH below 3.8. On the contrary, it has been found that acid-resistant species of LAB, such as *L. fermentum*, *L. plantarum*, *L. reuteri*, *L. rossiae* and/or *L. pontis*, predominate in liquid non-renewable sourdoughs, which are derived using starting microbial compositions, and are characterized by a higher maintenance temperature (above 30 °C) and prolonged fermentation from 15 h to 3 days – type II sourdoughs (Meroth et al., 2003; Vogelmann, Hertel, 2011; De Vuyst et al., 2017).

Dried or stabilized liquid sourdoughs (type III) contain LAB, which are resistant to the drying process and are able to maintain viability in this form, for example, obligately heterofermentative *L. brevis*, facultatively heterofermentative *L. plantarum*, *P. pentosaceus* (De Vuyst, Neysens, 2005; Hammes et al., 2005; Settanni et al., 2013).

The above data suggest that significant differences in the microbiome and, in particular, in the species composition of lactobacilli can also be detected in domestic sourdoughs with different humidity and maintenance temperature.

For many decades, five types of rye sourdoughs have been widely used at baking plants in Russia for the production of rye and rye-wheat bread: thick sourdough, liquid sourdough without welding, liquid sourdough with welding, concentrated lactic acid – enriched sourdough and thermophilic sourdough. According to Scientific Research Institute for the Baking Industry (Kuznetsova et al., 2021), thick rye sourdough and liquid one without welding are most often used. Both sourdoughs are used in the production of bakery products from a mixture of rye and wheat flour with full or partial replacement of pressed yeast and allow to get high-quality bread.

The choice of a particular sourdough at the plant is determined by technological capabilities (equipment, operating mode). Liquid sourdoughs are used in plants designed to transport the sourdough through pipes from the starter shop to the place where the dough is kneaded, and having tanks with a jacket to ensure the desired temperature. To prepare such a sourdough, a nutritious mixture of flour and water with a moisture content of 70–75 % is used, the fermentation temperature is 30–32 °C. Thick rye sourdough differs significantly in technological parameters: it has a moisture content of 48–50 % and a fermentation temperature of 26–28 °C. It is easier to preserve it during breaks in work, and it does not require the use of tanks with a jacket for fermentation, since the temperature in the workshop conditions is sufficient (Kosovan, 2008). A higher fermentation temperature of the liquid sourdough allows to achieve the necessary technological parameters (acidity and lifting force) in a shorter time, which speeds up the technological process. Given the significant difference in the parameters of the sourdoughs, it is of great interest to study the differences in the formation of their microbiome.

The aim of the work was to study the influence of technological parameters of sourdoughs on the taxonomic structure of the microbiome of bread sourdoughs of spontaneous fermentation.

Materials and methods

Preparation and maintenance of sourdoughs. Two sourdoughs of spontaneous fermentation were studied: thick rye sourdough and liquid rye sourdough without welding, derived using one batch of rye flour (OAO Kirov LKHP, Russia). Flour quality indicators: the drop number is 193 s, the moisture content is 12.4 %. The moisture content of flour was determined according to GOST 9404-88, the drop number – according to GOST 27676-88. The experiment was carried out in two repetitions.

Lactic acid bacteria of the genera *Lactobacillus*, *Weissella*, *Pediococcus* and *Leuconostoc* were detected in rye flour, studied in the form of a water-flour nutrient mixture for sourdough. Genus *Lactobacillus* was the dominant. Lactobacilli of the species *F. sanfranciscensis*, *L. pontis*, *L. brevis*, *L. plantarum*, *Companilactobacillus* sp., *L. curvatus* were also found.

The sourdoughs were managed under laboratory conditions for a month. Nutritional mixtures with a moisture content of 50 and 70 % by weight of 1000 g each were prepared from the same batch of rye flour. To do this, flour and water were mixed to a homogeneous consistency in a ratio of 1:0.76 and 1:1.9, respectively.

The obtained water-flour nutrient mixtures with a moisture content of 50 and 70 % were left to ferment in thermostats for two days at temperatures of 26±1 °C and 32±1 °C, respectively.

The fermented sourdoughs were updated with appropriate nutrient mixtures in the ratio of sourdough: the nutrition 1:1 and left for another day of fermentation at temperatures of 26±1 °C and 32±1 °C, depending on the moisture content (50 or 70 %) of the starter.

During the following days, the sourdoughs were renewed in a ratio of 1:1 after 6.5–7 h and 16 h and fermented at appro-

priate temperatures. After 16-h fermentation, the sourdoughs were renewed in a ratio of 1:1, each sourdough was fermented for 3.5–4 h at a given temperature, tested for quality (lifting force, titrated acidity, volume increase in % to the original volume) and sent for storage in the refrigerator at a temperature of 5 ± 1 °C for 2.5 days. The acidity was determined by titration with 0.1 N sodium hydroxide solution in the presence of phenolphthalein and presented in degrees. The lifting force was determined by the “ball” method and expressed in minutes (Puchkova, 2004).

Over the next four weeks, from Monday to Thursday, the sourdoughs with a moisture content of 50 % were refreshed in the ratio sourdough: the nutrient mixture 1:3 and fermented at a temperature of 26 °C for 7 h, and then refreshed in a ratio of 1:5 at a temperature of 20 °C and left for 16 h at a temperature of 17–18 °C. During the same period, the sourdoughs with a moisture content of 70 % were refreshed in a ratio of 1:2 with a temperature of 32 °C and fermented for 7 h. Then they were refreshed in the same ratio, but with a temperature of about 20 °C and left to ferment for 16 h at a temperature of 17–18 °C. On Friday, the sourdoughs were updated according to the regime established for each sourdough, fermented for 3.5–4 h at a given temperature and after quality control were placed in the refrigerator and stored at a temperature of 4–5 °C for 2.5 days.

These two temperature and humidity regimes are actually used at the plants in Russia due to technical reasons. So, for a thick rye sourdough, it is technically impossible to provide a higher fermentation temperature, since it is impossible to unload from a container with a jacket, and transportation of the sourdough through a pipeline is impossible. It is possible to manage such a sourdough only under those regimes that are installed at plants.

The objective of this study was to identify the influence of these modes and parameters characteristic of our industry on the microbiome of starter cultures.

Microbiological analysis of sourdoughs. The number of viable cells of lactic acid bacteria was controlled during the management of sourdoughs. For that, 10 g of the sourdough sample was homogenized manually in a mortar in 90 ml of 0.9 % sodium chloride solution. A series of tenfold consecutive dilutions was prepared and seeded on Sanfrancisco agar (Picozzi et al., 2005). To create anaerobic cultivation conditions, gas packages (AnaeroGen) were used, providing a carbon dioxide level of 9 to 13 % and an oxygen content of less than 1 %. The crops were incubated at 30 °C.

Determination of the composition of microbial communities by high-throughput sequencing of a fragment of the 16S rRNA gene. In each sample, the taxonomic structure of the microbiome was determined by high-performance sequencing based on the Collective Use Center of “Genomic Technologies, Proteomics and Cell Biology” of the Federal State Budget Scientific Institution All-Russia Research Institute for Agricultural Microbiology.

To isolate DNA from the samples, a set of reagents (MACHEREY-NAGEL NucleoSpin Soil) from MACHEREY-NAGEL (Germany) was used according to the manufacturer’s instructions. The taxonomic composition of the bacterial com-

munity was determined in each sample based on the analysis of amplicon libraries of ribosomal operon fragments. The taxonomic analysis of the bacterial community was performed with universal primers F515/R806 for a variable region of the 16S rRNA-v4 gene (GTGCCAGCMGCCGCGGTAA/GGACTACVSGGGTATCTAAT) specific for a wide range of microorganisms, including bacteria and archaea (Bates et al., 2010). All primers included service sequences containing linkers and barcodes (necessary for sequencing using Illumina technology).

PCR was performed in 15 µl of a reaction mixture containing 0.5–1 unit of activity of Q5[®] High-Fidelity DNA Polymerase polymerase (NEB, USA), 5 pM of direct and reverse primers, 10 ng of DNA matrix and 2nM of each dNTP (LifeTechnologies, USA). The mixture was denatured at 94 °C for 1 min, followed by 35 cycles: 94 °C – 30 s, 50 °C – 30 s, 72 °C – 30 s. The final elongation was carried out at 72 °C for 3 min. PCR products were purified according to the Illumina recommended method using AMPureXP (BeckmanCoulter, USA). Further preparation of libraries was carried out in accordance with the instructions of the manufacturer MiSeq Reagent Kit Preparation Guide (Illumina, USA). The libraries were sequenced in accordance with the manufacturer’s instructions on the Illumina MiSeq device (Illumina, USA) using the MiSeq[®] ReagentKit v3 reagent kit (600 cycles) with two-way reading (2*300 n).

The data obtained as a result of sequencing samples were processed using software packages Trimomatic (Bolger et al., 2014) and QIIME (Caporaso et al., 2010). At the first stage, the primary analysis of the reading quality, the selection of sequences based on the reading quality of individual bases (base pair quality), the combination of paired-terminal sequences with an overlap area of at least 35 bases was performed, as well as the removal of sequences the length of which is less than 180 bp. At the second stage of processing, all service sections were removed from libraries (primers), as well as sequences containing extended homopolymer repeats. *De novo* OTE-picking was used in the analysis of bacterial communities. Taxonomic identification of OTE was carried out using the RDP database (<http://rdp.cme.msu.edu>).

Results

Lactic acid bacteria of the genera *Lactobacillus*, *Weissella*, *Pediococcus* and *Leuconostoc*, characteristic for the natural seeding of grain and flour, were detected in rye flour, studied in the form of a water-flour nutrient mixture for starter culture. Genus *Lactobacillus* was the dominant. Lactobacilli of the species *F. sanfranciscensis*, *L. pontis*, *L. brevis*, *L. plantarum*, *Companilactobacillus* sp., *L. curvatus* were also found.

According to the data obtained, during the first three days of management, the bacterial complex of the studied sourdough microbiomes was formed by representatives of two phyla: Proteobacteria (the Gammaproteobacteria class dominated) and Firmicutes (lactic acid bacteria of the genera *Weissella*, *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Lactococcus*). The phylum Protobacteria was represented by bacteria of the order Enterobacterales (about 40 %), which after 48 h of fermentation were also detected in significant quantities (Fig. 1).

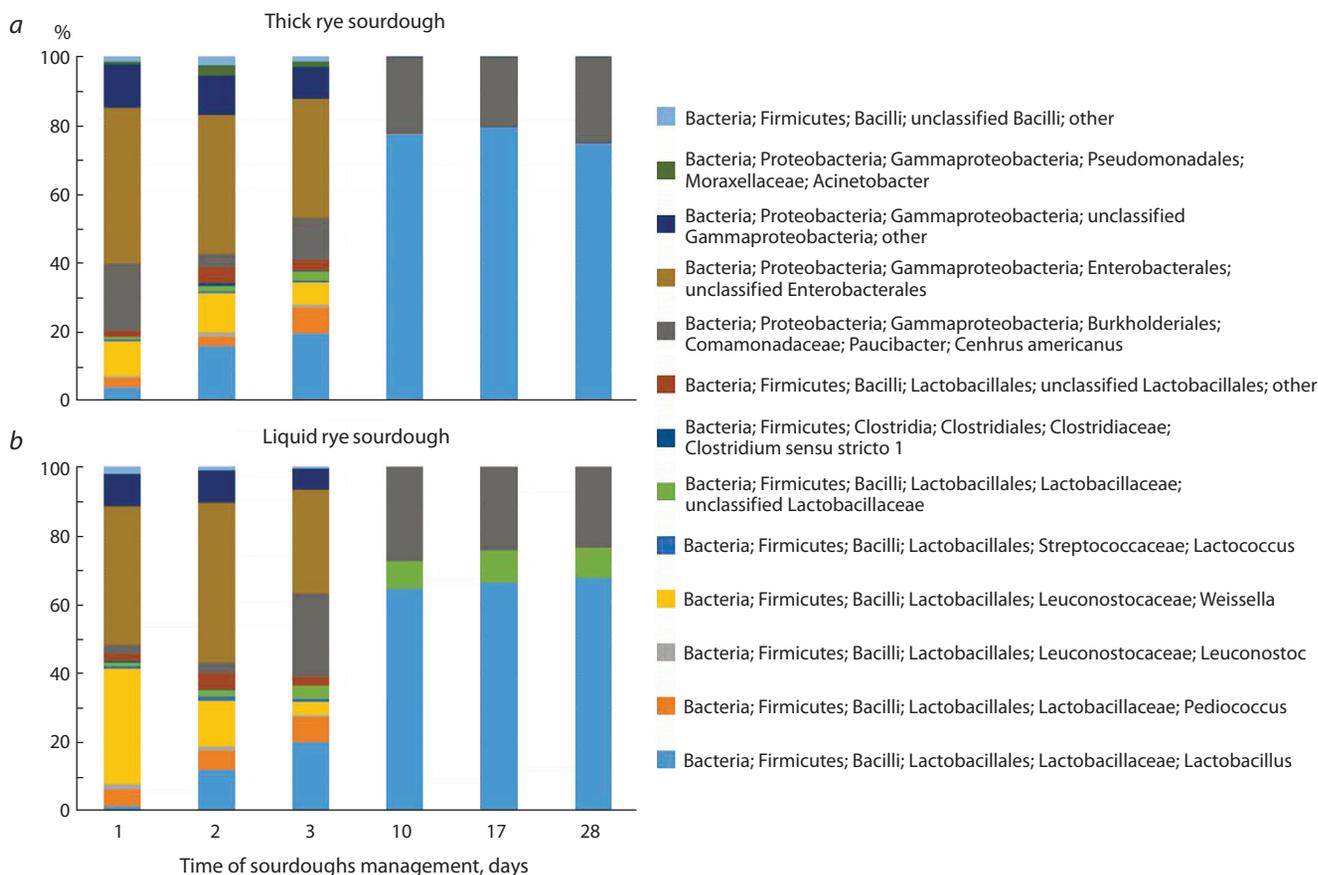


Fig. 1. Composition of microbial communities according to the analysis of sequences of 16S rRNA gene fragments in spontaneous sourdoughs: dense rye sourdough (a), liquid rye sourdough (b).

On the third day of sourdough management, the number of bacteria of the order Enterobacterales decreased slightly to 30%. It is worth noting that during this period the sourdoughs had an unpleasant putrid smell. With further management of sourdoughs after 10 days, representatives of the phylum Firmicutes dominated lactic acid bacteria of the genus *Lactobacillus*, extraneous bacteria of the order Enterobacterales were not detected. Both sourdoughs acquired a characteristic sourdough smell at that time.

Control of the number of viable LAB cells by classical microbiological methods showed that after a day of fermentation, the number of LAB cells was significantly higher in liquid rye sourdough. This may be explained by a higher temperature of 32 °C compared to a thick sourdough, which was conducted at a temperature of 26 °C. With further management of sourdoughs, the number of LAB cells was comparable, but significant changes occurred at the level of genera and species (Fig. 2).

Lactic acid bacteria of the genus *Weissella* dominated in both sourdoughs after a day of fermentation (53% of the total amount of LAB in thick and 74% in liquid sourdough), LAB of the genera *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Lactococcus* were also detected (Fig. 3). At the same time, an extremely low content of the most significant for sourdoughs genus *Lactobacillus* was noted in the liquid sourdough (3%).

With further management of sourdoughs, as the number of LAB of the genus *Lactobacillus* increased, there was a gradual displacement of coccoid forms of *Lactococcus*, *Leuconostoc*, *Weissella*, *Pediococcus*. After three days, the number of lactobacilli increased in both sourdoughs to 50%. On the 10th day of management, only LAB of the genus *Lactobacillus* were detected.

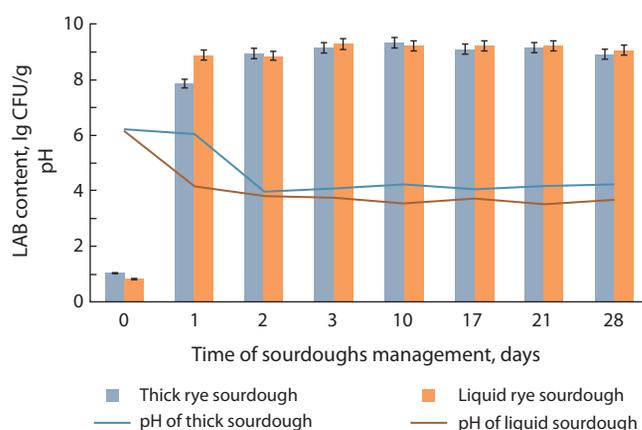


Fig. 2. Changes in number of lactic acid bacteria cells and pH in sourdoughs during 28 days.

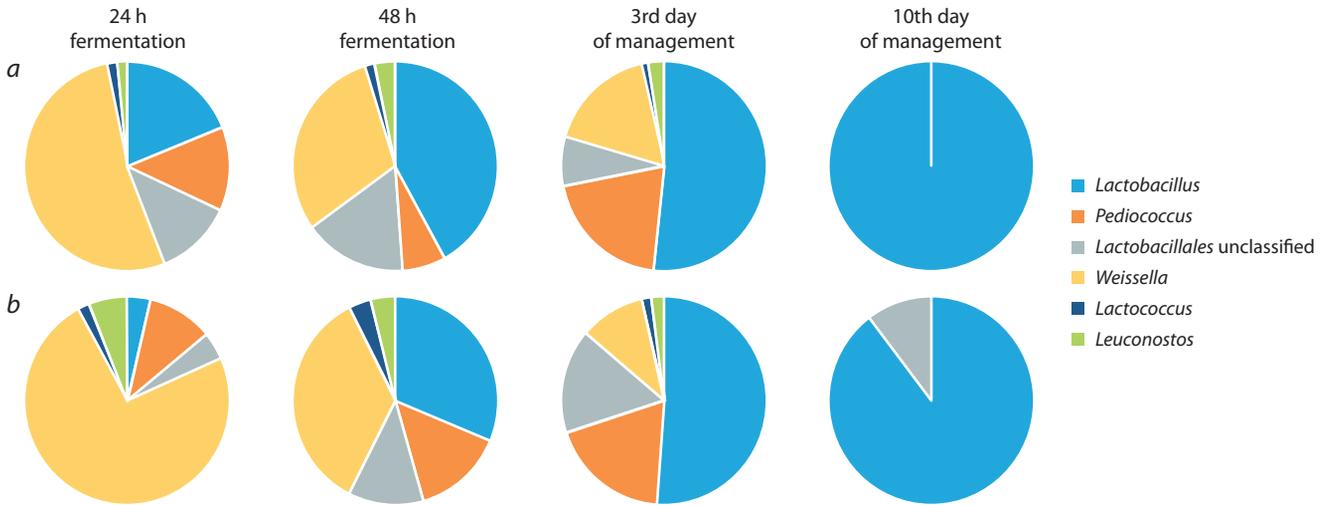


Fig. 3. Representation of LAB genera (%) according to the analysis of sequences of 16S rRNA gene fragments in: dense rye sourdough (a); liquid rye sourdough (b).

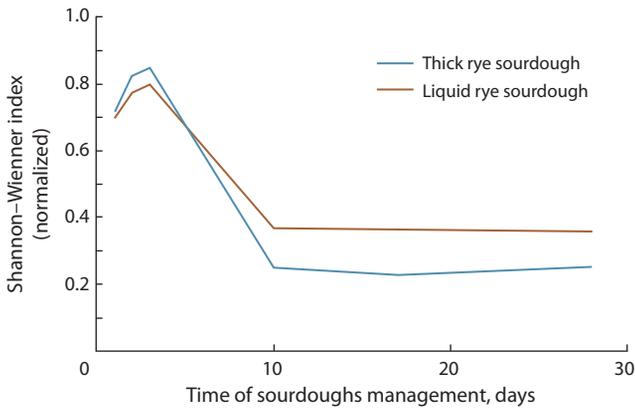


Fig. 4. Biodiversity change in the process of the rye sourdoughs.

It was found that during the management of sourdoughs, significant changes occurred not only at the level of LAB genera, but also at the level of lactobacillus species.

After 24 h of fermentation, *Lactobacillus curvatus* dominated in both sourdoughs. The microbiota after 48 h of fermentation in both sourdoughs was represented by lactobacilli *Latilactobacillus curvatus*, *Levilactobacillus brevis* and *Lactiplantibacillus plantarum/paraplantarum/pentosus/fabifermentans* (could not be accurately identified due to the low difference in the nucleotide sequences of the 16S rRNA gene in these species), which were found in the initial water-flour nutrient mixture before fermentation. With further management of sourdoughs, the content of these genera decreased.

After 10 days of management and throughout the rest of the period, obligately heterofermentative lactobacilli of the species *Fructilactobacillus sanfranciscensis*, also found in rye flour, dominated the thick rye sourdough. After a month of sourdough management, a species belonging to *Companilactobacillus* sp. was identified, which could not be uniquely identified using the RDP database. The source of this type of LAB was also rye flour. During the first two weeks of the

sourdough management this type was present in the sourdough in small quantities.

Obligately heterofermentative LAB of the species *Limosilactobacillus pontis* prevailed in the liquid rye sourdough without welding, and were also found in the flour sample.

The alpha diversity was evaluated (Fig. 4). The Shannon–Wiener index was calculated using the formula

$$H = -\sum p_i \cdot \lg(p_i),$$

where i is the number of lactobacillus species found in the starter culture, p_i is the proportion occupied by a particular species in the total population of lactobacillus species.

The obtained results fully confirm the fact that the parameters of sourdoughs (temperature, humidity) significantly affect the species diversity of lactobacilli, which was low in both sourdoughs.

Discussion

In this paper, for the first time, the influence of technological parameters of sourdoughs on the taxonomic structure of the microbiome of domestic bread sourdoughs of spontaneous fermentation was studied. Using a culture-independent method of high-performance sequencing of the 16S rRNA gene, it was shown that bacteria belonging to two phyla: Firmicutes and Proteobacteria were detected in the sourdoughs during the first three days. The relative number of proteobacteria decreased during fermentation and after 10 days of management they were not detected. This is completely consistent with previous results of other researchers showing that several phyla of bacteria in addition to Firmicutes (for example, Actinobacteria, Bacteroidetes, Cyanobacteria and Proteobacteria) may be present in the sourdoughs before fermentation begins. Most of them are inactive populations and are completely displaced by Firmicutes (Ercolini et al., 2013; Rizzello et al., 2015; Menezes et al., 2020).

At the same time, the source of extraneous microbiota is raw materials. Typically, the bacterial population of rye and wheat flour that does not belong to the phylum Firmicutes consists

of representatives of the phylum Proteobacteria (for example, the genus *Erwinia*, *Acinetobacter*, *Pantoea*, *Pseudomonas*, *Comamonas*, *Enterobacter* and *Sphingomonas*) and Bacteroidetes (for example, *Chryseobacterium*). This population is usually almost completely suppressed on the first day of sourdough management. The exception is representatives of the Enterobacteriaceae family, which are detected up to 5 days of sourdough, probably due to the ability to synthesize organic acids and some resistance to acid stress (Ercolini et al., 2013). Thus, during fermentation, microbial diversity changes with an increase in the proportion of the phylum Firmicutes and the displacement of representatives of the phylum Proteobacteria.

The phylum Firmicutes was represented by lactic acid bacteria. It was found that on the first day, representatives of the genus *Weissella* dominated in both sourdoughs after 24 h of fermentation. The bacteria of the genus *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Lactococcus* were contained in smaller quantities, and after three days the number of LAB of the genus *Lactobacillus* increased in both sourdoughs to 50 % of the total number of LAB, and coccoid forms of *Lactococcus*, *Leuconostoc*, *Weissella*, *Pediococcus* were displaced. On the 10th day of management, only LAB of the genus *Lactobacillus* were detected. In the study of Polish rye starter cultures, a similar dynamics was noted (Boreczek et al., 2020): after 24 h of fermentation in the sourdough, the content of bacteria of the genus *Weissella* was 36 %, and after 72 h – only 5 %, while the content of the genus *Lactobacillus* increased from 30 to 67 % by the third day of the sourdough management.

The data obtained differ somewhat from the data obtained for Italian sourdoughs, in which, after 10 days, the content of representatives of the genus *Weissella* was almost 2 times greater than that of the genus *Lactobacillus* (Ercolini et al., 2013). The authors suggest that this is due to the fact that the genus *Weissella* dominated in Italian rye flour. Indeed, *Lactococcus*, *Enterococcus*, *Leuconostoc* and *Weissella* are usually found in grain and flour, respectively, but are not able to withstand a long acidification process, since their development requires higher pH values compared to lactobacilli (Van Kerrebroeck et al., 2017). A number of studies have shown that the growth of representatives of the genus *Weissella* is inhibited at pH 4.5, but they are capable of growth at pH 6.5–6.8. The acidic environment of the sourdough stimulates the change of LAB communities in the sourdoughs and creates a niche favorable for the development of acid-resistant lactobacilli, such as *L. brevis* and *L. sanfranciscensis* (Oshiro et al., 2020).

Summarizing the known and obtained data, it can be noted that the stabilization of the microbiome with the predominance of the genus *Lactobacillus* occurs by 10 days (Van der Meulen et al., 2007; Weckx et al., 2010). This correlates with our data on the change in alpha diversity (see Fig. 4) of thick and liquid rye sourdoughs.

It is noted that during the same period there are significant changes in the species composition of lactobacilli. The dominant species after 24 h of fermentation, *Latilactobacillus curvatus*, was discovered after 48 h together with the species *Levilactobacillus brevis* and *Lactiplantibacillus plantarum/paraplantarum/pentosus/fabifermentans*, which were found in

small quantities and in the initial water-flour nutrient mixture before fermentation.

A decrease in the content of *Latilactobacillus curvatus* in the process of sourdoughs was also noted by foreign researchers. It was found that at the first stages of the management of Korean sourdoughs, the content of certain types of LAB was: *L. curvatus* (9.5 log CFU/g), *F. sanfranciscensis* (< 5 log CFU/g), *L. brevis* (6.5 log CFU/g), whereas after the 11th refreshment, the number of *F. sanfranciscensis* significantly increased (> 9.0 log CFU/g), while the content of *L. curvatus* and *L. brevis* decreased, which, according to the authors, is due to the negative effect of increased lactic acid content in the medium on the development of *L. curvatus* (Baek et al., 2021). Studies (Landis et al., 2021) have shown that *L. plantarum* and *L. brevis* were the most frequently detected pair of simultaneously occurring taxa (in 177 out of 500 sourdoughs), while the species *L. sanfranciscensis* dominated in most long-term sourdoughs, and its content negatively correlated with the content of *L. plantarum* and *L. brevis* species, which dominated in young sourdoughs.

With further management of sourdoughs, significant differences were noted in the formation of microbiota in liquid and thick rye sourdoughs. In thick rye sourdough, an increase in the content of bacteria of the species *Fructilactobacillus sanfranciscensis* was noted. According to the literature data, this species is considered the most adapted and is an autochthonous microorganism of the type I sourdough microbiota (Siragusa et al., 2009; Vogel et al., 2011; Rogalski et al., 2020). The dominance of this species in thick rye sourdough is explained by the creation of optimal conditions for its development (temperature less than 30 °C, pH within 4.1–4.3, moisture content 50 %). However, after a month of keeping in the sourdough in addition to *F. sanfranciscensis* lactic acid bacteria *Companilactobacillus* sp. were found, the source of which was flour.

Obligately heterofermentative LAB of the species *Limosilactobacillus pontis* found in liquid rye sourdough without welding developed at higher temperatures (32 °C) and moisture content (70 %) and lower pH (3.6–4.0). It is obvious that these conditions are favorable for the development of this species, which is confirmed by the literature data (De Vuyst et al., 2017).

Conclusion

As a result of the research, the diversity of prokaryotes in domestic sourdoughs of spontaneous fermentation was studied for the first time by the method of high-performance sequencing. It was found that during the first three days of management, the bacterial complex of the studied sourdoughs was represented by the phyla Proteobacteria (class Gammaproteobacteria) and Firmicutes (lactic acid bacteria of the genera *Weissella*, *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Lactococcus*). In the further management of sourdoughs after 10 days, representatives of the phylum Firmicutes dominated lactic acid bacteria of the genus *Lactobacillus*.

A comparative analysis of the taxonomic composition of the microbiome of thick and liquid rye sourdough without welding did not demonstrate deep differences throughout the entire

period of sourdough management both at the phylum/class level and at the generic level. However, there was a difference at the level of lactobacilli species, which is due to the influence of exogenous factors, such as temperature and humidity of sourdoughs, on the formation of the starter microbiome.

Further studies of industrial and laboratory sourdoughs of long-term management will allow us to establish whether there is a stabilization of the microbiome with the dominance of one or two species, whether periodic fluctuations in the composition of the microbiome occur, or other scenarios are being implemented.

References

- Auerman L.Ya. Technology of Bakery Production. Saint-Petersburg, Professiya Publ., 2009. (in Russian)
- Baek H.W., Bae J.H., Lee Y.G., Kim S.A., Min W., Shim S., Han N.S., Seo J.H. Dynamic interactions of lactic acid bacteria in Korean sourdough during back-slopping process. *J. Appl. Microbiol.* 2021; 131(5):2325-2335. DOI 10.1111/jam.15097.
- Bates S.T., Berg-Lyons D., Caporaso J.G., Walters W.A., Knight R., Fierer N. Examining the global distribution of dominant archaeal populations in soil. *ISME J.* 2010;5:908-917. DOI 10.1038/ismej.2010.171.
- Böcker G., Stolz P., Hammes W.P. Neue Erkenntnisse zum Ökosystem Sauerteig und zur Physiologie der Sauerteigtypischen Stamme *Lactobacillus sanfrancisco* und *Lactobacillus pontis*. *Getreide Mehl. und Brot.* 1995;49:370-374.
- Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120. DOI 10.1093/bioinformatics/btu170.
- Boreczek J., Litwinek D., Żylińska-Urban J., Izak D., Buksa K., Gawor J., Gromadka R., Bardowski J.K., Kowalczyk M. Bacterial community dynamics in spontaneous sourdoughs made from wheat, spelt, and rye wholemeal flour. *MicrobiologyOpen.* 2020;9(4):1-13. DOI 10.1002/mbo3.1009.
- Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods.* 2010; 7(5):335-336. DOI 10.1038/nmeth.f.303.
- De Vuyst L., Neysens P. The sourdough microflora: biodiversity and metabolic interactions. *Trends Food Sci. Technol.* 2005;16(1-3): 43-56. DOI 10.1016/j.tifs.2004.02.012.
- De Vuyst L., Van Kerrebroeck S., Leroy F. Microbial ecology and process technology of sourdough fermentation. *Adv. Appl. Microbiol.* 2017;100:49-160. DOI 10.1016/bs.aams.2017.02.003.
- Ercolini D., Pontonio E., De Filippis F., Minervini F., La Stora A., Gobbetti M., Di Cagno R. Microbial ecology dynamics during rye and wheat sourdough preparation. *Appl. Environ. Microbiol.* 2013; 79(24):7827-7836. DOI 10.1128/AEM.02955-13.
- Gänzle M., Ehmann M.R., Hammes W. Modeling of growth of *Lactobacillus sanfranciscensis* and *Candida milleri* in response to process parameters of sourdough fermentation. *Appl. Environ. Microbiol.* 1998;64(7):2616-2623. DOI 10.1128/AEM.64.7.2616-2623.1998.
- Hammes W.P., Brandt M.J., Francis K.L., Rosenheim J. Microbial ecology of cereal fermentations. *Trends Food Sci. Technol.* 2005; 16(1-3):4-11. DOI 10.1016/j.tifs.2004.02.010.
- Kosovan A.P. (Ed.). Collection of Modern Bakery Technologies. Moscow, 2008. (in Russian)
- Kuznetsova L.I., Parakhina O.A., Savkina O.A., Burykina M.S. Starting compositions of microorganisms for sourdoughs. *Khlebopekar-nyy i Konditerskiy Forum = Bakery and Confectionery Forum.* 2021; 49:54-57. (in Russian)
- Landis E.A., Oliverio A.M., McKenney E.A., Nichols L.M., Kfoury N., Biango-Daniels M., Shell L.K., Madden A.A., Shapiro L., Sakunala S., Drake K., Robbat A., Booker M., Dunn R.R., Fierer N., Wolfe B.E. The diversity and function of sourdough starter microbiomes. *Elife.* 2021;10:e61644. DOI 10.7554/eLife.61644.
- Lokachuk M.N., Khlestkin V.K., Savkina O.A., Kuznetsova L.I., Pavlovskaya E.N. Changes in the microbiota of dense rye sourdough during long-term propagation. *Khleboprodukt = Bread Products.* 2020;11:33-37. DOI 10.32462/0235-2508-2020-29-11-33-37. (in Russian)
- Menezes L.A.A., Sardaro M.L.S., Duarte R.T.D., Mazzon R.R., Neviiani E., Gatti M., De Dea Lindner J. Sourdough bacterial dynamics revealed by metagenomic analysis in Brazil. *Food Microbiol.* 2020;85:103302. DOI 10.1016/j.fm.2019.103302.
- Meroth C.B., Hammes W.P., Hertel C. Monitoring the bacterial population dynamics in sourdough fermentation processes by using PCR-denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 2003;69(1):475-482. DOI 10.1128/AEM.69.1.475-482.2003.
- Minervini F., Lattanzi A., De Angelis M., Di Cagno R., Gobbetti M. Influence of artisan bakery- or laboratory-propagated sourdoughs on the diversity of lactic acid bacterium and yeast microbiotas. *Appl. Environ. Microbiol.* 2012;78(15):5328-5340. DOI 10.1128/AEM.00572-12.
- Müller M., Wolfrum G., Stolz P., Ehrmann M., Vogel R. Monitoring the growth of *Lactobacillus* species during rye flour fermentation. *Food Microbiol.* 2001;18:217-227. DOI 10.1006/fmic.2000.0394.
- Oshiro M., Tanaka M., Zendo T., Nakayama J. Impact of pH on succession of sourdough lactic acid bacteria communities and their fermentation properties. *Biosci. Microbiota Food Health.* 2020;39(3): 152-159. DOI 10.12938/bmfh.2019-038.
- Picozzi C., Gallina S., Dell Fera T., Foschino R. Comparison of cultural media for the enumeration of sourdough lactic acid bacteria. *Ann. Microbiol.* 2005;55(4):317-320.
- Puchkova L.I. Laboratory Workshop on the Technology of Bakery Production. Saint-Petersburg: GIRD Publ., 2004. (in Russian)
- Rizzello C.G., Cavoski I., Turk J., Ercolini D., Nionelli L., Pontonio E., De Angelis M., De Filippis F., Gobbetti M., Di Cagno R. Organic cultivation of *Triticum turgidum* subsp. *durum* is reflected in the flour-sourdough fermentation-bread axis. *Appl. Environ. Microbiol.* 2015;81(9):3192-3204. DOI 10.1128/AEM.04161-14.
- Robert H., Gabriel V., Fontagné-Faucher C. Biodiversity of lactic acid bacteria in French wheat sourdough as determined by molecular characterization using species-specific. *Int. J. Food Microbiol.* 2009; 135(1):53-59. DOI 10.1016/j.ijfoodmicro.2009.07.006.
- Rogalski E., Ehrmann M.A., Vogel R.F. Role of *Kazachstania humilis* and *Saccharomyces cerevisiae* in the strain-specific assertiveness of *Fructilactobacillus sanfranciscensis* strains in rye sourdough. *Eur. Food Res. Technol.* 2020;9:1817-1827. DOI 10.1007/s00217-020-03535-7.
- Settanni L., Ventimiglia G., Alfonzo A., Corona O. An integrated technological approach to the selection of lactic acid bacteria of flour origin for sourdough production. *Food Res. Int.* 2013;54(2):1569-1578. DOI 10.1016/j.foodres.2013.10.017.
- Siragusa S., Di Cagno R., Ercolini D., Minervini F., Gobbetti M., De Angelis M. Taxonomic structure and monitoring of the dominant population of lactic acid bacteria during wheat flour sourdough type I propagation using *Lactobacillus sanfranciscensis* starters. *Appl. Environ. Microbiol.* 2009;75(4):1099-1109. DOI 10.1128/AEM.01524-08.
- Van der Meulen R., Scheirlinck I., Van Schoor A., Huys G., Vancanneyt M., Vandamme P., De Vuyst L. Population dynamics and metabolite target analysis during laboratory fermentations of wheat and spelt sourdoughs. *Appl. Environ. Microbiol.* 2007;73(15):4741-4750. DOI 10.1128/AEM.00315-07.

- Van Kerrebroeck S., Maes D., De Vuyst L. Sourdoughs as a function of their species diversity and process conditions, a meta-analysis. *Trends Food Sci. Technol.* 2017;68:152-159. DOI 10.1016/j.tifs.2017.08.016.
- Viiard E., Bessmeltseva M., Simm J., Talve T., Aaspõllu A., Paalme T. Diversity and stability of lactic acid bacteria in rye sourdoughs of four bakeries with different propagation parameters. *PLoS One.* 2016;11(2):e0148325. DOI 10.1371/journal.pone.0148325.
- Vogel R.F., Pavlovic M., Ehrmann M.A., Wiezer A., Liesegang H., Offschanka S., Voget S., Angelov A., Böcker G., Lieb W. Genomic analysis reveals *Lactobacillus sanfranciscensis* as stable element in traditional sourdoughs. *Microb. Cell Factories.* 2011;10(Suppl. 1): 1-11. DOI 10.1186/1475-2859-10-S1-S6.
- Vogelmann S.A., Hertel C. Impact of ecological factors on the stability of microbial associations in sourdough fermentation. *Food Microbiol.* 2011;28(3):583-589. DOI 10.1016/j.fm.2010.11.010.
- Weckx S., Van der Meulen R., Maes D., Scheirlinck I., Huys G., Vandamme P. Lactic acid bacteria community dynamics and metabolite production of rye sourdough fermentations share characteristics of wheat and spelt sourdough fermentations. *Food Microbiol.* 2010; 27(8):1000-1008. DOI 10.1016/j.fm.2010.06.005.

ORCID ID

V.K. Khlestkin orcid.org/0000-0001-9605-8028
M.N. Lockachuk orcid.org/0000-0001-5074-2457
O.A. Savkina orcid.org/0000-0002-2372-4277
O.I. Parakhina orcid.org/0000-0002-0508-2813
L.I. Kuznetsova orcid.org/0000-0002-1149-6043

Acknowledgements. The study was carried out with the support of the Russian Foundation for Basic Research (project No. 19-016-00085 "Study of species diversity and symbiotic interactions in the microbiomes of starch-protein hydrocolloid systems (bread sourdoughs)"). The work was carried out using the equipment of the Collective Use Center of "Genomic Technologies, Proteomics and Cell Biology" of the Federal State Budget Scientific Institution All-Russia Research Institute for Agricultural Microbiology.

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

Received November 2, 2021. Revised March 27, 2022. Accepted March 28, 2022.

Original Russian text www.bionet.nsc.ru/vogis/

Genome stability of the vaccine strain VACΔ6

R.A. Maksyutov, S.N. Yakubitskiy, I.V. Kolosova, T.V. Tregubchak, A.N. Shvalov, E.V. Gavrilova, S.N. Shchelkunov 

State Research Center of Virology and Biotechnology "Vector", Rospotrebnadzor, Koltsovo, Novosibirsk region, Russia

 snshchel@rambler.ru

Abstract. Due to cessation of mass smallpox vaccination in 1980, the collective immunity of humans against orthopoxvirus infections has virtually been lost. Therefore, the risk of spreading zoonotic human orthopoxvirus infections caused by monkeypox and cowpox viruses has increased in the world. First-generation smallpox vaccines based on *Vaccinia virus* (VAC) are reactogenic and therefore not suitable for mass vaccination under current conditions. This necessitates the development of modern safe live vaccines based on VAC using genetic engineering. We created the VACΔ6 strain by transient dominant selection. In the VACΔ6 genome, five virulence genes were intentionally deleted, and one gene was inactivated by inserting a synthetic DNA fragment. The virus was passaged 71 times in CV-1 cells to obtain the VACΔ6 strain from the VAC L1VP clonal variant. Such a long passage history might have led to additional off-target mutations in VACΔ6 compared to the original L1VP variant. To prevent this, we performed a genome-wide sequencing of VAC L1VP, VACΔ6, and five intermediate viral strains to assess possible off-target mutations. A comparative analysis of complete viral genomes showed that, in addition to target mutations, only two nucleotide substitutions occurred spontaneously when obtaining VACΔ4 from the VACΔ3 strain; the mutations persisting in the VACΔ5 and VACΔ6 genomes. Both nucleotide substitutions are located in intergenic regions (positions 1431 and 189738 relative to L1VP), which indicates an extremely rare occurrence of off-target mutations when using transient dominant selection to obtain recombinant VAC variants with multiple insertions/deletions. To assess the genome stability of the resulting attenuated vaccine strain, 15 consecutive cycles of cultivation of the industrial VACΔ6 strain were performed in 4647 cells certified for vaccine production in accordance with the "Guidelines for Clinical Trials of Medicinal Products". PCR and sequencing analysis of six DNA fragments corresponding to the regions of disrupted genes in VACΔ6 showed that all viral DNA sequences remained unchanged after 15 passages in 4647 cells.

Key words: vaccinia virus; transient dominant selection; targeted gene inactivation; genome stability.

For citation: Maksyutov R.A., Yakubitskiy S.N., Kolosova I.V., Tregubchak T.V., Shvalov A.N., Gavrilova E.V., Shchelkunov S.N. Genome stability of the vaccine strain VACΔ6. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(4):394-401. DOI 10.18699/VJGB-22-48

Стабильность генома вакцинного штамма VACΔ6

Р.А. Максютлов, С.Н. Якубицкий, И.В. Колосова, Т.В. Трегубчак, А.Н. Швалов, Е.В. Гаврилова, С.Н. Щелкунов 

Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р. п. Кольцово, Новосибирская область, Россия

 snshchel@rambler.ru

Аннотация. В связи с прекращением после 1980 г. массовой противооспенной вакцинации в настоящее время практически полностью утрачен коллективный иммунитет человеческой популяции к ортопоксвирусным инфекциям. Вследствие этого увеличилась опасность распространения в мире зоонозных ортопоксвирусных инфекций человека, обусловленных вирусами оспы обезьян или оспы коров. Противооспенные вакцины первого поколения на основе вируса осповакцины (*Vaccinia virus*, VAC) являются реактогенными и поэтому в современных условиях не пригодны для массовой вакцинации. Это обуславливает необходимость разработки современных безопасных живых вакцин на основе VAC с применением методов генетической инженерии. С использованием метода временной доминантной селекции нами создан штамм VACΔ6, в геноме которого пять генов вирулентности направленно делетированы, а один ген инактивирован встройкой синтетического фрагмента ДНК. В процессе получения штамма VACΔ6 из клонового варианта VAC L1VP вирус прошел 71 пассаж в культуре клеток CV-1. Такая длинная пассажная история могла привести к дополнительным нецелевым изменениям в геноме штамма VACΔ6 относительно исходного L1VP. Поэтому для оценки возможных нецелевых изменений провели полногеномное секвенирование VAC L1VP, VACΔ6 и пяти промежуточных штаммов вируса. Сравнительный анализ полных вирусных геномов показал, что, помимо целевых нарушений, спонтанно произошли только две нуклеотидные замены при получении VACΔ4 из штамма VACΔ3 и сохранившиеся в геноме VACΔ5 и VACΔ6. При этом обе эти замены находятся в межгенных участках (позиции 1431 и 189738 относительно штамма L1VP), что указывает на крайне редкое возникновение нецелевых изменений при использовании методики временной доминантной селекции для получения рекомбинантных VAC со множественными встройками/делециями. Для выяснения стабильности генома полученного аттенуированного вакцинного штамма и

в соответствии с «Руководством по проведению клинических исследований лекарственных средств...» выполнено 15 последовательных циклов культивирования производственного штамма вируса VACΔ6 в культуре клеток 4647, аттестованной для производства вакцины. ПЦР-анализ и секвенирование шести фрагментов ДНК, соответствующих районам нарушаемых генов VACΔ6, показали, что после 15 пассажей в культуре клеток 4647 все последовательности вирусной ДНК остались неизменными.

Ключевые слова: вирус осповакцины; временная доминантная селекция; направленная инактивация генов; стабильность генома.

Introduction

In 1980, after the declaration of smallpox eradication by the World Health Organization (WHO), it was recommended to stop vaccinating people against this extremely dangerous disease. This decision was due to the fact that *Vaccinia virus* (VAC) vaccinations can cause severe post-vaccination complications and even death in some cases (Smallpox and its Eradication, 1988; Kretzschmar et al., 2006).

Due to cessation of vaccination against smallpox, there are fewer people with specific immunity against the disease every year. This makes humans susceptible not only to possible variola virus infection but also other closely related orthopoxviruses, whose natural reservoir is various animals, primarily rodents. These viruses include monkeypox and cowpox, which cause smallpox-like diseases in animals and humans. The spread of these viruses in the human population can potentially lead to their adaptation to the host antiviral defense and occurrence of viral variants that are epidemically dangerous for humans (Shchelkunov, 2013). In recent years, unusually large outbreaks of orthopoxvirus infections have been recorded among humans in various regions of the world (Singh et al., 2012; Nolen et al., 2016; Reynolds et al., 2019).

Vaccination is the only effective method to combat the growing threat of human orthopoxvirus infections (Moss, 2011; Shchelkunov, 2011). The accumulation of immunodeficiency states in the human population in recent decades has led to contraindication of the use of classical live VAC-based vaccines for mass vaccination, since it can cause a large number of adverse reactions and more severe manifestations than those observed during the smallpox eradication campaign (Albarnaz et al., 2018; Shchelkunov, Shchelkunova, 2020). Therefore, there is an urgent need to develop modern orthopoxvirus vaccines that should be both much safer than previous generations of smallpox vaccines and highly immunogenic, thus providing reliable protection against viral infection.

The first attenuated VAC strains were obtained by serial passage of the virus in heterologous host cells: the MVA strain was generated upon 572 passages of VAC Ankara in primary chicken embryo fibroblasts (Volz, Sutter, 2017), the LC16m8 strain was produced after 45 passages of VAC Lister in primary rabbit kidney cells (Kidokoro, Shida, 2014; Eto et al., 2015).

VAC MVA attenuation was due to spontaneous extended deletions and mutations in the viral genome that affect not only virulence genes but also genes responsible for viral replication and the range of virus-sensitive hosts (Blanchard et al., 1998; Drexler et al., 1998). MVA lost its ability to form infectious progeny in most mammalian cell cultures, including human cells. Many MVA genes are expressed in these cell cultures; however, only immature virions are produced. MVA has

retained its immunogenic properties as a smallpox vaccine; however, in order to achieve a sufficient immune response, the virus should be administered at higher doses and multiple times compared to the classical vaccine (Sanchez-Sampedro et al., 2015).

Attenuation of the VAC LC16m8 clonal variant is due to a single nucleotide deletion in the *B5R* gene encoding the envelope protein of extracellular virions. The mutation creates a translational frameshift. In mammalian cells, LC16m8 produces infectious viral particles with a reduced ability to spread in both cell cultures and infected/vaccinated organisms. LC16m8 is less attenuated than MVA and it is a replication-competent vaccine (Sanchez-Sampedro et al., 2015; Albarnaz et al., 2018).

Advances in genetic engineering techniques have made it possible to create modified VAC variants by either introducing the desired nucleotide sequences into the viral genome or deleting/disrupting viral genes. One of the most promising areas is the use of genetic engineering to develop highly attenuated VAC variants with the same levels of immunogenicity and protectiveness but lower pathogenicity compared to those of the classical smallpox vaccine (Shchelkunov, Shchelkunova, 2020).

Whole-genome sequencing of various strains and different orthopoxvirus species that are pathogenic to humans, accumulation of data on functions of numerous viral genes, and development of methods for introducing targeted changes into the viral genome have made it possible to formulate and implement a new approach to the development of highly attenuated VAC variants. The approach involves strictly localized sequential deletion/inactivation of individual virulence genes without affecting virus replication in a cell culture and the range of virus-susceptible hosts (Yakubitskiy et al., 2015).

We have previously obtained live attenuated vaccine strain VACΔ6 against smallpox and other human orthopoxvirus infections by targeted sequential inactivation of individual viral genes (Yakubitskiy et al., 2015, 2016). We used transient dominant selection to produce this strain (Falkner, Moss, 1990). Each stage of the method involved multiple cycles of viral reproduction (passages) in a cell culture. By that time, no information on how these procedures can affect preservation of the nucleotide sequence of the large VAC DNA genome had been available.

In this regard, the aim of the work was to study the degree of VAC genome preservation during generation of the VACΔ6 strain from the parental LIVP strain by whole-genome sequencing. Another equally important issue was the genomic stability of the industrial vaccine strain VACΔ6 after 15 passages in 4647 cells used specifically for production of the smallpox vaccine.

Materials and methods

Viruses and cell cultures. We used clone 14 of the VAC L1VP (L1VP) strain, which had been previously obtained by limiting dilution with threefold plaque purification using an agarose overlay (Yakubitskiy et al., 2015), and mutant L1VP-derived VAC variants with inactivation of the target genes (Yakubitskiy et al., 2015, 2016). Viruses were grown and titrated in African green monkey kidney cell lines CV-1 and 4647 from the cell culture collection of the State Research Center of Virology and Biotechnology “Vector” of Rospotrebnadzor. Cell line 4647 was certified by L.A. Tarasevich State Institute of Standardization and Control of Biomedical Preparations in accordance with the requirements of Guidance document 42-28-10-89 and recommended for production of preventive medical immunobiological preparations (MIBPs) (protocol No. 14 of the meeting of the Academic Council of L.A. Tarasevich State Institute of Standardization and Control of Biomedical Preparations dated October 28, 2003; protocol No. 9 of the MIBP Committee dated November 20, 2003).

Oligonucleotide primers. Oligonucleotides for PCR analysis of inactivated viral gene regions (Table 1) were synthesized at the Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences. The oligonucleotide primers were designed using Oligo software version 3.3 (Borland International, USA).

Virus cloning. Prior to cloning, viral suspension was sonicated, and the titer was estimated by plaque assay in CV-1 cells. The virus was cloned in 6-well culture plates by agarose overlay plaque assay. For this, monolayer CV-1 cells were infected with 10–20 PFU/well of a viral suspension. The virus was adsorbed for 60 min at 37 °C and 5 % CO₂. The medium with unadsorbed virus was replaced with 2 ml/well of DMEM medium (BioloT, Russia) supplemented with 2 % fetal bovine serum (FBS) (HyClone, USA) containing 1 % low melting agarose. After agarose solidified, the plates were incubated in a thermostat for 48 h at 37 °C and 5 % CO₂. Then, 1.5 ml/well of a 0.05 % neutral red solution in DMEM medium was added, and plates were incubated for 1 h at 37 °C and 5 % CO₂. The neutral red solution was then removed, and individual plaques were collected using an automatic pipette and then transferred to 100 µl of DMEM medium supplemented with 2 % FBS. The resulting virus sample after a single freeze–thaw cycle was inoculated onto CV-1 cell monolayers in individual wells of 6-well plates containing 1 ml/well of DMEM medium with 2 % FBS. Cells were incubated for 48–72 h at 37 °C and 5 % CO₂. The resulting virus suspension was frozen–thawed twice and used to further produce the clone at a working titer of 10⁶–10⁷ PFU/ml.

Generation of viral variants with targeted gene deletions. Recombinant VAC variants were obtained in CV-1 cells by cationic lipid-mediated transfection with Lipofectin Reagent (Invitrogen, USA) and a selective medium containing mycophenolic acid (MPA), xanthine, and hypoxanthine (Sigma, USA). For this, CV-1 cells at 80–90 % confluence in 6-well culture plates were infected with VAC at a multiplicity of infection of 1 PFU/cell and incubated for 1 h at 37 °C and 5 % CO₂. The cell monolayer was then washed with a selective serum-free medium and transfected with a recombinant integrative plasmid according to the following scheme: 3 µl

Table 1. Oligonucleotide primers for PCR analysis of modified regions in the VAC genome used in the VACΔ6 vaccine development

Gene	Oligonucleotide primers
A56R	5' TTTTGAGACTCATCAACGATGAAAACCTT 3'
	5' GTGGTATGGGACACCACAAATCCAA 3'
B8R	5' TCACAAATATGATGGTGATGAGCGA 3'
	5' CGTGATATACCCTAGCCATAGGCAT 3'
J2R	5' CGGACATATTCAGTTGATAATCG 3'
	5' AACACTTTCTACACACCGATTGA 3'
C3L	5' TCGCGCTTTACATTCTCGAATCT 3'
	5' TGTTCTGTGTTCTTGCGGTGA 3'
N1L	5' CTAGAATGGCAAATCTAAGT 3'
	5' ATGGCTGTTGAAGTGGGATCC 3'
A35R	5' ACGACGGATGCTGAAGCGTGTATA 3'
	5' AAACGATGTTACCAATCGTTGCTAGGT 3'

of 1 µg/µl plasmid was mixed with 15 µl of 1 mg/ml Lipofectin; 1 ml of DMEM medium containing 25 µg/ml MPA, 250 µg/ml xanthine, and 15 µg/ml hypoxanthine was added and incubated for 15 min at room temperature; the resulting mixture was added dropwise onto the cell monolayer, and the cells were incubated at 37 °C and 5 % CO₂. After 24 h, the medium was replaced with a selective maintenance medium (supplemented with 2 % FBS) and incubated under the same conditions for one day.

To enrich the viral progeny with the recombinant VAC variant, several passages (successive viral reproduction cycles in cells) were performed in 6-well plates with a selective medium. When a 90–100 % viral cytopathic effect (CPE) was achieved, the culture was frozen and thawed twice to destroy cells and release the virus in the culture medium. The resulting viral suspension was sonicated and further utilized for cloning using agarose overlay plaque assay under non-selective conditions (without addition of xanthine, hypoxanthine, and MPA to the DMEM medium). Several viral clones were selected and passaged under non-selective conditions in CV-1 cells until a 100 % CPE was achieved. The titers of resulting virus clone samples were determined in CV-1 cells as described previously (Yakubitskiy et al., 2015).

Viral DNA was isolated from individual clones using the QIAamp DNA Mini Kit (Qiagen, Germany) according to the manufacturer’s instructions. PCR analysis with the use of specific oligonucleotide primers (see Table 1) was performed to screen clones at viral genome regions containing either a target deletion or a nucleotide sequence insertion (Table 2).

Clones with the disrupted target gene were selected based on PCR results (cloning round No. 1) and re-cloned using the agarose overlay plaque assay to prevent possible contamination with the original virus variant. Having conducted addi-

Table 2. Lengths of PCR products for the original LIVP strain and VAC variants with inactivated genes

Gene	Original VAC LIVP, bp	VAC variants with inactivated genes, bp
A56R	2354	1537
B8R	1535	716
C3L	1543	743
N1L	1785	1549
J2R	512	616
A35R	1888	1372

tional passages of the resulting subclones under non-selective conditions and achieved 100 % CPE of the cell monolayer, we performed a PCR analysis and selected clones with inactivated target gene (cloning round No. 2). Infectious titers of viral preparations were determined. The preparations were aliquoted, frozen, and used for further experiments.

VAC whole-genome sequencing. Sequencing was performed on a MiSeq instrument (Illumina, USA). For this, viral DNA was isolated from cell cultures infected with different viruses. The QIAamp DNA Mini Kit (Qiagen) was used for DNA isolation according to the manufacturer’s instructions.

Whole-genome sequence analysis. Nucleotide sequences of VAC variants were analyzed using software packages MIRA (v. 4.9.6), BWA (v. 0.7.15) (Li, Durbin, 2009), IGV (v. 2.3.78) (Robinson et al., 2011), Samtools (v. 1.3.1) (Li et al., 2009), Tabix (v. 0.2.5) (Li et al., 2009), and Genome-AnalysisTK (v. 3.6) (McKenna et al., 2010). Whole-genome sequences were aligned in Ugene Alignment Editor v. 1.24.1 (Okonechnikov et al., 2012) using the MAFFT algorithm (Kato et al., 2002).

Results

Introduction of targeted deletions/insertions into the VAC genome

We used transient dominant selection to obtain VAC with a targeted genome modification (Falkner, Moss, 1990). The integration/deletion plasmid (Fig. 1, a) carries a dominant selective marker (the *gpt* gene of *Escherichia coli* encoding xanthine-guanine-phosphoribosyl transferase, under the control of the 7.5K VAC promoter) located outside the extended regions of homology with the viral DNA flanking the disrupted/deleted gene (R stands for right, L stands for left).

The bacterial enzyme xanthine-guanine-phosphoribosyl transferase, which is synthesized in mammalian cells, can restore purine nucleotide metabolism blocked by MPA. As a result of a single crossover between the integrative plasmid and viral DNA, a recombinant viral genome containing a fully integrated recombinant plasmid is formed (see Fig. 1, b). The genome contains both the *gpt* gene and long repeats R, R' and L, L'. A genetic construct containing long repeats is unstable

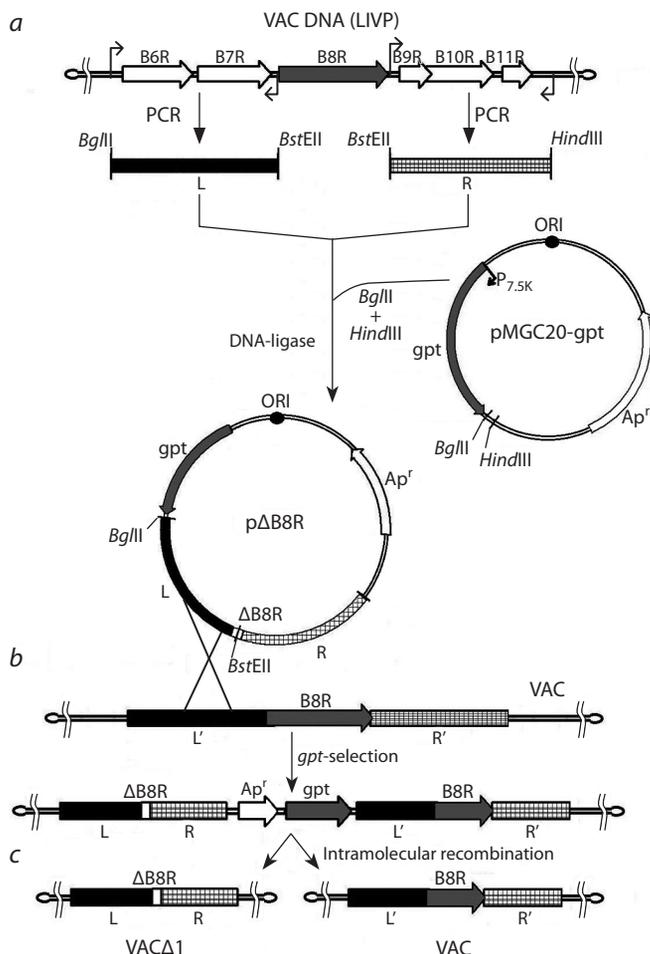


Fig. 1. General scheme for obtaining VACΔ1 virus with a targeted *B8R* gene deletion.

a – scheme of hybrid pΔB8R plasmid design; b – recombinant insertion of the hybrid plasmid into the viral genome; c – division of the recombinant virus progeny into two variants after removal of selective pressure on the *gpt* gene (see explanations in the text).

and can only exist under selective pressure on *gpt*. When selective conditions are removed (cultivation in a normal growth medium), intramolecular homologous recombination occurs in the viral genome (either R–R' or L–L'); it yields two virus variants: viruses with either a disrupted or an intact gene (see Fig. 1, c). It should be noted that intramolecular recombination deletes all foreign sequences from the viral genome, which is important when creating VAC-based immunobiological preparations. In addition, cleavage of the entire plasmid sequence from the viral genome makes it possible to further produce double-, triple-, etc. recombinant viruses with disrupted loci in different genomic regions using the same technique and selective marker (Fig. 2).

The scheme for generating recombinant VAC variants with six inactivated virulence genes is presented in Fig. 2. The thymidine kinase gene in integrative plasmid pΔJ2R is disrupted by insertion of a synthetic DNA fragment; in other recombinant plasmids, target genes are deleted (Yakubitskiy et al., 2015, 2016).

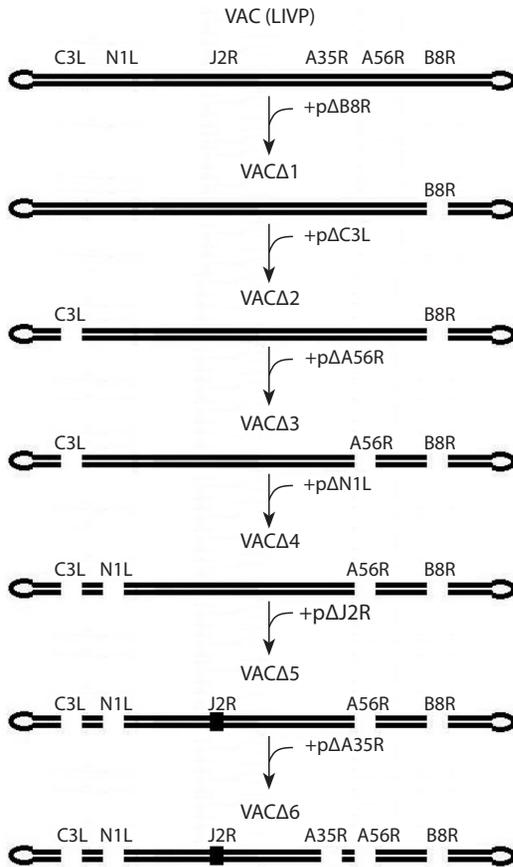


Fig. 2. Scheme of recombinant VACΔ6 production.

A total of 18 serial passages were carried out in CV-1 cells when producing recombinant VACΔ6 from the LIVP strain: 6 serial passages under selective pressure (3–7 passages), 6 serial passages without selective pressure after the first cloning (3–4 passages), and 6 serial passages without selective pressure after the second cloning (2–4 passages) (Table 3).

Whole-genome sequencing of VAC variants

VACΔ6 passage history includes 71 passages in total, of which 31 and 40 passages were conducted under selective pressure on *gpt* and without selective pressure, respectively. Such a long

passage history could cause additional off-target mutations in the VACΔ6 genome compared to the LIVP strain. Therefore, in order to assess possible off-target mutations occurring during transient dominant selection and the stability of both the original LIVP strain and its recombinant variants, we performed whole-genome sequencing of seven VAC strains (LIVP clone 14, VACΔ1, VACΔ2, VACΔ3, VACΔ4, VACΔ5, and VACΔ6).

As a result, seven whole-genome nucleotide sequences of the studied VACs were obtained; the genomic nucleotide sequence of the original clonal variant VAC LIVP was deposited in GenBank under the accession number KX781953.

A comparative analysis of complete viral genomes showed that, in addition to targeted mutations, only two nucleotide substitutions occurred spontaneously when generating VACΔ4 from VACΔ3; the mutations persisted in VACΔ5 and VACΔ6 genomes. Both nucleotide substitutions are located in intergenic regions (positions 1431 and 189738 relative to LIVP). Considering the total length of the viral genome (more than 190 kb) and a long passage history, this indicates the absence of off-target mutations in viral DNA when using transient dominant selection for obtaining recombinant VAC variants.

Analysis of genetic stability of the industrial strain VACΔ6

The discovered stability of the VAC genome after 71 passages in CV-1 cells indicated the importance of using a more reasonable approach to confirm strain identity for the future large-scale production of a VACΔ6-based vaccine. This method uses PCR with primers complementary to six regions of the viral genome at which the parental VAC LIVP strain was modified (see Table 1). Therefore, the genetic stability of the industrial strain VACΔ6 during multiple passages in 4647 cells (certified for cultivating the vaccine strain) was evaluated by PCR at preclinical stages.

A total of 15 serial cycles of 4647 cell monolayer infection with the industrial strain VACΔ6 and viral progeny production were carried out in accordance with the “Guidelines for Clinical Trials of Medicinal Products...” (2013). Viral DNA was isolated from both the original sample and 4647 cell cryolysate that had been infected with VACΔ6 obtained after the 14th passage. PCR analysis was performed using oligonucleotide primers shown in Table 1.

DNA isolated from intact 4647 cells and VAC LIVP DNA were used as a negative and positive PCR control, respectively.

Table 3. Passage history of producing the VACΔ6 strain from LIVP in CV-1 cells

Original VAC strain	Number of passages under selective pressure*	Number of passages without selective pressure		Resulting VAC strain
		Cloning No. 1	Cloning No. 2	
LIVP cl. 14	6	3	4	VACΔ1
VACΔ1	7	4	3	VACΔ2
VACΔ2	4	3	4	VACΔ3
VACΔ3	3	3	3	VACΔ4
VACΔ4	6	4	3	VACΔ5
VACΔ5	5	4	2	VACΔ6

* Without taking into account the transfection step under selective pressure.

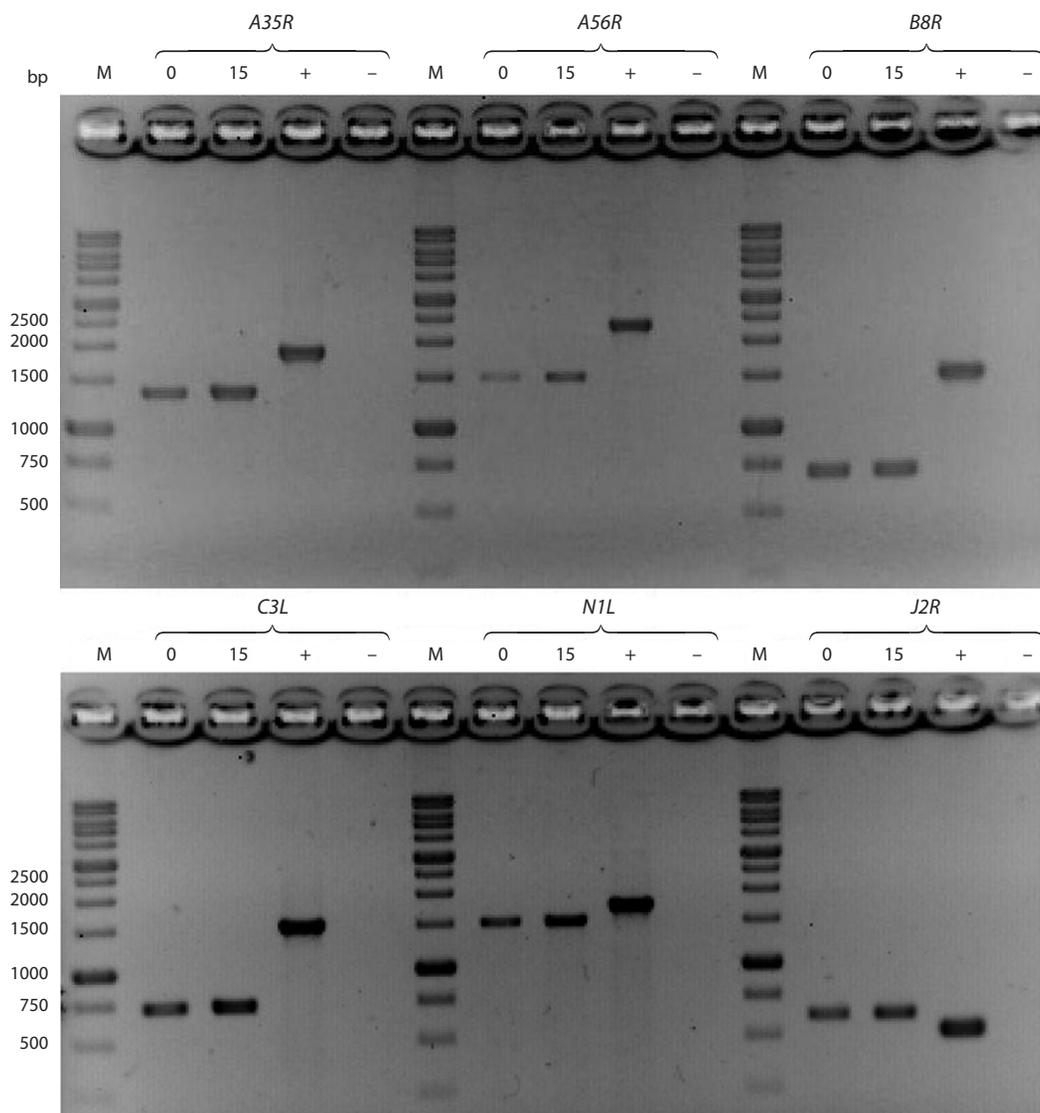


Fig. 3. Electrophoretic separation of the DNA products of PCR analysis of the regions of viral genes *A35R*, *A56R*, *B8R*, *C3L*, *N1L*, and *J2R*.

0 – PCR product obtained using genomic DNA of the original strain VACΔ6 as a template; 15 – PCR product obtained using genomic DNA of the VACΔ6 strain after 15 passages in 4647 cells as a template; M – DNA marker (length in bp is indicated on the left); “+” – PCR product obtained using VAC LIVP DNA; “-” – negative control.

The theoretically calculated lengths of the DNA products of PCR performed with specific primer pairs are presented in Table 2. The electrophoretic separation of the resulting amplicons indicates that, after 15 passages of the industrial strain VACΔ6 in 4647 cells, the PCR products correspond to theoretically calculated values that do not differ from those of the PCR products obtained using DNA of the industrial strain VACΔ6 before passing as the template (Fig. 3).

Sequencing of six DNA fragments corresponding to disrupted gene regions in VACΔ6 showed that all viral DNA sequences remained unchanged after 15 passages in 4647 cells.

Discussion

The *Vaccinia virus* (VAC) is a member of the genus *Orthopoxvirus* of the family Poxviridae. Representatives of this

genus are the largest mammalian viruses; their DNA genome contains about 200 genes. Poxviruses replicate in the cytoplasm of the host cells; the products of numerous viral genes control viral DNA replication, transcription, and translation of viral genes. In addition, many genes are involved in regulating the antiviral immune response, susceptible host range, pathogenicity, and other poxvirus properties. VAC is the most studied orthopoxvirus; it played a crucial role as a live vaccine in global eradication of smallpox (Shchelkunov, Shchelkunova, 2020).

Smallpox vaccines based on different VAC strains are moderately reactogenic. However, in mass vaccination, they can cause severe adverse reactions and even death in a small percentage of cases. Therefore, after confirmation of smallpox eradication in 1980, WHO strongly recommended to

cease smallpox vaccination (Smallpox and its Eradication, 1988).

The cessation of vaccination resulted in the loss of the immune protection not only against smallpox but also other zoonotic human orthopoxvirus infections such as monkeypox and cowpox in almost all humans. Of particular concern is monkeypox, whose clinical manifestations in humans resemble smallpox. All of this has raised the issue of the need to develop new safe vaccines against orthopoxvirus infections using modern techniques.

We have implemented an approach to introduce targeted deletions into (inactivate) individual VAC virulence genes by genetic engineering without affecting viral replication in mammalian cells (Yakubitskiy et al., 2015). The protocol of transient dominant selection used for inactivation of each of the selected viral genes is carried out through serial passages and cloning of the virus in cells. The selection of generated VAC variants is based on PCR data in the target region of the viral genome. Before the advent of modern methods of whole-genome sequencing, the complete nucleotide sequence of the long VAC DNA was not controlled. Therefore, there was no answer to the question of how multiple VAC passages in cells can affect viral genome stability in general.

For the first time, we performed whole-genome sequencing of the original clonal variant VAC LIVP, the VACΔ6 vaccine strain, and five intermediate virus variants with a series of disrupted target genes using the obtained VACΔ6 strain with inactivation of six genes in different viral genomic regions as an example (see Fig. 2). A comparative analysis of complete viral genomes showed that, in addition to targeted mutations, only two nucleotide substitutions occurred spontaneously when producing VACΔ4 from VACΔ3; the mutations persisted in VACΔ5 and VACΔ6 genomes. Both nucleotide substitutions are located in intergenic regions (positions 1431 and 189738 relative to LIVP). These results show that transient dominant selection does not introduce significant off-target mutations into the viral genome. Apparently, this also indicates that the original VAC LIVP variant is suitable for reproduction in CV-1 cells and therefore stably maintains the genome integrity under experimental conditions.

Cultivation of VAC, which is a mammalian virus, in heterologous primary avian cells (chicken embryo fibroblasts) seems to exert a high selective pressure on this virus and cause significant changes in the viral genome after multiple passages, which was observed in the MVA variant (Volz, Sutter, 2017).

In order to use VACΔ6 as a safe live vaccine for mass vaccination, one has to subject the resulting strain to multiple reproduction cycles in a cell culture certified for this purpose. Preservation of the attenuated phenotype/genotype of the vaccine strain is the most important criterion in large-scale virus production. For this reason, a total of 15 consecutive cycles of a 4647 cell monolayer infection with the industrial strain VACΔ6 were carried out in accordance with the “Guidelines for Clinical Trials of Medicinal Products...” (2012). The results of a PCR analysis and sequencing of six DNA fragments corresponding to the regions of disrupted VACΔ6 genes showed that viral DNA sequences in these regions remained unchanged after 15 passages in 4647 cells.

Conclusion

Thus, the obtained results demonstrate high genetic stability of the studied recombinant strains with a long passage history in CV-1 and 4647 cells, which is an important positive characteristic of recombinant VACΔ6 variant as a stable vaccine strain for obtaining a fourth-generation live vaccine against smallpox and other orthopoxvirus infections.

References

- Albarnaz J.D., Torres A.A., Smith G.L. Modulating vaccinia virus immunomodulators to improve immunological memory. *Viruses*. 2018; 10(3):101. DOI 10.3390/v10030101.
- Blanchard T.J., Alcamí A., Andrea P., Smith G.L. Modified vaccinia virus Ankara undergoes limited replication in human cells and lacks several immunomodulatory proteins: implications for use as a human vaccine. *J. Gen. Virol.* 1998;79(Pt. 5):1159-1167. DOI 10.1099/0022-1317-79-5-1159.
- Drexler I., Heller K., Wahren B., Erfle V., Sutter G. Highly attenuated modified vaccinia virus Ankara replicates in baby hamster kidney cells, a potential host for virus propagation, but not in various human transformed and primary cells. *J. Gen. Virol.* 1998;79(Pt. 2): 347-352. DOI 10.1099/0022-1317-79-2-347.
- Eto A., Saito T., Yokote H., Kurane I., Kanatani Y. Recent advances in the study of live attenuated cell-cultured smallpox vaccine LC16m8. *Vaccine*. 2015;33(45):6106-6111. DOI 10.1016/j.vaccine. 2015.07.111.
- Falkner F.G., Moss B. Transient dominant selection of recombinant vaccinia viruses. *J. Virol.* 1990;64(6):3108-3111. DOI 10.1128/JVI. 64.6.3108-3111.1990.
- Guidelines for Clinical Trials of Medicinal Products (Immunobiological Medicinal Products). Part 2. Moscow: Grif and K Publ., 2012. (in Russian)
- Katoh K., Misawa K., Kuma K., Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059-3066. DOI 10.1093/nar/gkf436.
- Kidokoro M., Shida H. Vaccinia virus LC16m8Δ as a vaccine vector for clinical applications. *Vaccines*. 2014;2(4):755-771. DOI 10.3390/vaccines2040755.
- Kretzschmar M., Wallinga J., Teunis P., Xing S., Mikolajczyk R. Frequency of adverse events after vaccination with different vaccinia strains. *PLoS Med.* 2006;3(8):e272. DOI 10.1371/journal.pmed. 0030272.
- Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. DOI 10.1093/bioinformatics/btp324.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. DOI 10.1093/bioinformatics/btp352.
- McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303. DOI 10.1101/gr.107524.110.
- Moss B. Smallpox vaccines: Targets of protective immunity. *Immunol. Rev.* 2011;239(1):8-26. DOI 10.1111/j.1600-065X.2010. 00975.x.
- Nolen L.D., Osadebe L., Katomba J., Likofata J., Mukadi D., Monroe B., Doty J., Hughes C.M., Kabamba J., Malekani J., Bomponda P.L., Lokota J.J., Balilo M.P., Likafi T., Lushima R.S., Ilunga B.K., Nkawa F., Pukuta E., Karhemere S., Tamfum J.J., Nguete B., Wema-koy E.O., McCollum A.M., Reynolds M.G. Extended human-to-human transmission during a monkeypox outbreak in the Democratic Republic of the Congo. *Emerg. Infect. Dis.* 2016;22(6):1014-1021. DOI 10.3201/eid2206.150579.

- Okonechnikov K., Golosova O., Fursov M., UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012; 28(8):1166-1167. DOI 10.1093/bioinformatics/bts091.
- Reynolds M.G., Doty J.B., McCollum A.M., Olson V.A., Nakazawa Y. Monkeypox re-emergence in Africa: a call to expand the concept and practice of One Health. *Expert Rev. Anti Infect. Ther.* 2019;17(2): 129-139. DOI 10.1080/14787210.2019.1567330.
- Robinson J.T., Thorvaldsdottir H., Winckler W., Guttman M., Lander E.S., Getz G., Mesirov J.P. Integrative genomics viewer. *Nat. Biotechnol.* 2011;29(1):24-26. DOI 10.1038/nbt.1754.
- Sanchez-Sampedro L., Perdiguero B., Mejias-Perez E., Garcia-Arriaza J., Di Pilato M., Esteban M. The evolution of poxvirus vaccines. *Viruses*. 2015;7(4):1726-1803. DOI 10.3390/v7041726.
- Shchelkunov S.N. Emergence and reemergence of smallpox: the need in development of a new generation smallpox vaccine. *Vaccine*. 2011;29(Suppl. 4):D49-D53. DOI 10.1016/j.vaccine.2011.05.037.
- Shchelkunov S.N. An increasing danger of zoonotic orthopoxvirus infections. *PLoS Pathog.* 2013;9(12):e1003756. DOI 10.1371/journal.ppat.1003756.
- Shchelkunov S.N., Shchelkunova G.A. Genes that control vaccinia virus immunogenicity. *Acta Naturae*. 2020;12(1):33-41. DOI 10.32607/actanaturae.10935.
- Singh R.K., Balamurugan V., Bhanuprakash V., Venkatesan G., Hosamani M. Emergence and reemergence of vaccinia-like viruses: global scenario and perspectives. *Indian J. Virol.* 2012;23(1):1-11. DOI 10.1007/s13337-012-0068-1.
- Smallpox and its Eradication. Geneva: World Health Organization, 1988.
- Volz A., Sutter G. Modified vaccinia virus Ankara. History, value in basic research, and current perspectives for vaccine development. *Adv. Virus Res.* 2017;97:187-243. DOI 10.1016/bs.aivir.2016.07.001.
- Yakubitskiy S.N., Kolosova I.V., Maksyutov R.A., Shchelkunov S.N. Attenuation of vaccinia virus. *Acta Naturae*. 2015;7(4):113-121. DOI 10.32607/20758251-2015-7-4-113-121.
- Yakubitskiy S.N., Kolosova I.V., Maksyutov R.A., Shchelkunov S.N. Highly immunogenic variant of attenuated vaccinia virus. *Dokl. Biochem. Biophys.* 2016;466:35-38. DOI 10.1134/S160767291601105.

ORCID ID

S.N. Yakubitskiy orcid.org/0000-0002-0496-390X
S.N. Shchelkunov orcid.org/0000-0002-6255-9745

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

Received Desember 24, 2021. Revised March 9, 2022. Accepted March 14, 2022.

Original Russian text www.bionet.nsc.ru/vogis/

Creation of transgenic mice susceptible to coronaviruses: a platform for studying viral pathogenesis and testing vaccines

N.R. Battulin^{1,2}, O.L. Serov^{1,2} 

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

² Novosibirsk State University, Novosibirsk, Russia

 serov@bionet.nsc.ru

Abstract. Over the past 20 years, coronaviruses have caused three epidemics: SARS-CoV, MERS-CoV, and SARS-CoV2, with the first two having a very high lethality of about 10 and 26 %, respectively. The last outbreak of coronavirus infection caused by SARS-CoV2 in 2019 in China has swept the entire planet and is still spreading. The source of these viruses in humans are animals: bats, Himalayan civets, and camels. The genomes of MERS-CoV, SARS-CoV and SARS-CoV2 are highly similar. It has been established that coronavirus infection (SARS-CoV and SARS-CoV2) occurs through the viral protein S interaction with the lung epithelium – angiotensin-converting enzyme receptor 2 (ACE2) – due to which the virus enters the cells. The most attractive model for studying the development of these diseases is a laboratory mouse, which, however, is resistant to coronavirus infection. The resistance is explained by the difference in the amino acid composition of mouse Ace2 and human ACE2 proteins. Therefore, to create mice susceptible to SARS-CoV and SARS-CoV2 coronaviruses, the human ACE2 gene is transferred into their genome. The exogenous DNA of the constructs is inserted into the recipient genome randomly and with a varying number of copies. Based on this technology, lines of transgenic mice susceptible to intranasal coronavirus infection have been created. In addition, the use of the technology of targeted genome modification using CRISPR/Cas9 made it possible to create lines of transgenic animals with the insertion of the human ACE2 gene under the control of the endogenous murine Ace2 gene promoter. This “humanization” of the Ace2 gene makes it possible to obtain animals susceptible to infection with coronaviruses. Thus, transgenic animals that simulate coronavirus infections and are potential platforms for testing vaccines have now been created.

Key words: coronaviruses CoVs; SARS-CoV; MERS-CoV; COVID-19; transgenesis; “humanization” of the mouse genome; CRISPR/Cas9 technology.

For citation: Battulin N.R., Serov O.L. Creation of transgenic mice susceptible to coronaviruses: a platform for studying viral pathogenesis and testing vaccines. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(4):402-408. DOI 10.18699/VJGB-22-49

Создание трансгенных мышей, восприимчивых к коронавирусам: платформа изучения вирусного патогенеза и тестирования вакцин

Н.Р. Баттулин^{1,2}, О.Л. Серов^{1,2} 

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 serov@bionet.nsc.ru

Аннотация. За последние 20 лет коронавирусы вызвали три эпидемии, SARS-CoV, MERS-CoV и SARS-CoV2, причем летальность первых двух была очень высокой: около 10 и 26 % соответственно. Последняя вспышка коронавирусной инфекции, вызванная SARS-CoV2 в 2019 г. в Китае, охватила всю планету, и она все еще продолжает распространяться. Источником этих вирусов у человека были животные: летучие мыши, гималайские циветы и верблюды. Геномы MERS-CoV, SARS-CoV и SARS-CoV2 имеют высокое сходство между собой. Установлено, что заражение коронавирусной инфекцией (SARS-CoV и SARS-CoV2) происходит посредством контакта вирусного белка S с рецептором легочного эпителия – ангиотензин-конвертирующим ферментом 2 (ACE2), благодаря чему вирус попадает в клетки. Наиболее привлекательной моделью для исследования особенностей развития этих заболеваний является лабораторная мышь, которая, однако, резистентна к коронавирусной инфекции. Резистентность объясняется различием аминокислотного состава белков Ace2 мыши и ACE2 человека. Поэтому при получении мышей, восприимчивых к коронавирусам SARS-CoV и SARS-CoV2, в их геном переносят ген ACE2 человека. Экзогенная ДНК конструкций встраивается в реципиентный геном случайным

образом и с варьирующим числом копий. На основе этой технологии были получены линии трансгенных мышей, восприимчивых к интраназальной коронавирусной инфекции. Применение технологии адресной модификации геномов с помощью CRISPR/Cas9 позволило получить линии трансгенных животных с inserцией гена *ACE2* человека под контроль промотора эндогенного гена *Ace2* мыши. Такая «гуманизация» гена *Ace2* дает возможность получить животных, наиболее близко имитирующих коронавирусную инфекцию человека. Таким образом, к настоящему времени создана серия линий трансгенных мышей – животных, моделирующих коронавирусные инфекции человека и потенциально способных служить платформами для тестирования вакцин.

Ключевые слова: коронавирусы CoVs; SARS-CoV; MERS-CoV; COVID-19; трансгенез; «гуманизация» генома мышей; технология CRISPR/Cas9.

Introduction

The viral infection that caused severe acute respiratory syndrome (SARS) was first recorded in December 2019 in Wuhan, China, and rapidly spread around the world on a pandemic scale (Li et al., 2020; Zhou P. et al., 2020; Zhu et al., 2020). Soon, at the end of January 2020, it was published that the causative agent of the disease is a new type of coronavirus isolated from the bronchoalveolar secretions of six patients and called 2019-nCov (Zhu et al., 2020). Later, on the recommendation of the WHO (World Health Organization), the disease caused by the new SARS-CoV-2 virus was called COVID-19. Almost simultaneously (in early February 2020), additional data on the novel coronavirus infection from seven patients, six of whom were seafood vendors in the Wuhan market, were published (Zhou F. et al., 2020). According to P. Zhou et al. (2020), the SARS-CoV-2 virus genome shares 85 % similarity with the bat coronavirus and 79.6 % similarity with the previously described human SARS-CoV.

Model animals are an important tool in the research of many human pathologies. However, the creation of adequate animal models of infectious diseases has specifics associated with the high rate of co-evolution of the host-parasite system, during which both participants gain many specific adaptations. An example of this is the tropism of infectious human viruses based on the specific interaction of viral proteins with cellular receptor proteins, which gives rise to infection. The absence of the specific binding of viral proteins and target cell proteins causes resistance to a specific viral infection in different species.

On the other hand, differences in the response of the immune system to a particular viral agent in humans and animals can also become an insurmountable obstacle to the use of animals as model objects. Despite these difficulties, many mouse models of human viral diseases (poliomyelitis, measles, hepatitis B and C) have been created, which makes it possible to study the fundamental aspects of the development of a particular disease, such as infection processes, the course of a viral disease, and the interaction of the virus and the immune system. Such models have proven to be highly demanded for preclinical trials of new vaccines and antivirals (Takaki et al., 2017).

This review is devoted to the creation of laboratory mice susceptible to the SARS-CoV-2 and SARS-CoV coronaviruses in order to create an experimental platform for studying both the coronavirus pathogenesis itself and testing pharmacological antiviral drugs and vaccines.

Human and animals coronavirus infections

The group of coronaviruses (CoVs) is represented by large enveloped viruses, the genome of which consists of single-stranded RNA (Lai et al., 2007). Coronaviruses are members of the subfamily Coronavirinae, family Coronaviridae, order Nidovirales. CoVs are composed of four genera: alpha-, beta-, gamma-, and deltacoronaviruses (Woo et al., 2009a, b), all of which cause zoonotic infections in animals. In the past two decades, of the entire cohort of coronaviruses, two have become pathogenic to humans, causing SARS and Middle East respiratory syndrome (MERS). In the first case, the outbreak originated in Guangzhou province in China in December 2002 and then spread to five continents (Peiris et al., 2003). According to the WHO, the SARS-CoV epidemic affected 8437 people, 813 of whom died. MERS-CoV coronavirus infection outbreak originated in the Arabian Peninsula in 2012 (Zaki et al., 2012) and spread to the Middle East, England and South Korea. According to the WHO, the epidemic affected 1728 people, with 624 cases of MERS-CoV infection resulting in death. It has been established that animals are the source of human infection with SARS-CoV and MERS-CoV (more details below).

The process of infection with coronaviruses in humans and animals

To enter human target cells, SARS-CoV and MERS-CoV use their “crown”, which is represented by many spike-shaped (S) proteins. It has been established that the SARS-CoV S protein interacts with angiotensin-converting enzyme 2 (ACE2; encoded by the *ACE2* gene) as an entry receptor (Li et al., 2003; Ge et al., 2013). During viral infection, the trimeric S protein is cleaved into S1 and S2 subunits, after which they are recognized by human cell receptors (Belouzard et al., 2009). Further, the S1 subunit containing the receptor-binding domain binds directly to the peptidase domain of the ACE2 protein, while S2 is responsible for membrane fusion (Li et al., 2005).

Based on this knowledge, several independent groups of researchers have proposed that the new SARS-CoV-2 coronavirus uses the same mode of entry into human cells as previously described SARS-like viruses. To confirm this hypothesis, comparisons were made between the protein sequences of SARS-like viruses and the new coronavirus SARS-CoV-2, and a high level of similarity was revealed. Further, to establish binding sites, a crystallographic analysis of the complex between the S1 subunit of the coronavirus and the human ACE2 protein was performed. As a result, it was found

that the ACE2 protein has five key amino acid sequences that are involved in the binding of the S1 subunit of the virus (Lan et al., 2020; Wan et al., 2020; Wang et al., 2020).

The *ACE2* gene in humans is expressed in the lungs, arteries, heart, brain, and small intestine and is an important component of the renin-angiotensin-aldosterone system (Bader, 2013). Expression of *ACE2* in the lungs is mainly limited to alveolar epithelial cells of the second type. During coronavirus infection, *ACE2* interacts with the receptor-binding domain of the virus spike protein, which leads to endocytosis of viral particles and their internalization (Kuba et al., 2010). These events result in severe acute respiratory syndrome, lung tissue damage, and extensive inflammation (Imai et al., 2005).

It is important to note that the first stages of coronavirus infections caused by SARS-CoV, MERS-CoV and SARS-CoV2 have significant similarities.

Technologies for the creation of transgenic mice for modeling coronavirus infections

As noted above, the source of human coronavirus diseases was animals. In 2003, Chinese researchers found that bats were carriers of the SARS-CoV coronavirus in humans through an intermediary – Himalayan civets, whose meat is considered a delicacy in Chinese cuisine (Guan et al., 2003; Peiris et al., 2003). The basis for this conclusion was 99.8 % similarity of the SARS-CoV genome with the virus isolated from bats and Himalayan civets. Bats are also natural carriers of the MERS-CoV coronavirus, and the camel is an intermediate carrier of the virus to humans (van Boheemen et al., 2012; Reusken et al., 2013).

It is worth noting that there are several animal species susceptible to SARS-CoV infection: ferrets, Syrian hamsters, cats, and several primates: macaques, African green monkeys, and marmosets (Glass et al., 2003; Martina et al., 2003; Roberts et al., 2005; Subbarao, Roberts, 2006). It is assumed that other animals may be susceptible to the novel coronavirus SARS-CoV-2 (Wan et al., 2020). However, these infected animals show minimal signs of impairment and generally lack the clinical symptoms associated with human coronavirus infection.

In connection with the above, the strategy for creating model animals is based on the technology of introducing the human *ACE2* gene, the main receptor for coronaviruses, into their genome. Indeed, in one of the first works on the creation of transgenic mice susceptible to coronavirus infection, a pK18-*hACE2* recombinant DNA construct was developed, including the 5'-promoter and the 1st intron (with a mutation in the 3'-splice of the acceptor) of the human *CK18* gene (encodes cytokeratin-18), as well as the translational enhancer alpha of the alfalfa mosaic virus (total size 2.5 kb), human *ACE2* cDNA and a 3' sequence including exon 6, intron 6, exon 7 and polyA signal element of the human *CK18* gene (McCray et al., 2007). According to the authors' intention, all elements were present in the construct to ensure a high level of its expression in epithelial cells. A purified DNA fragment of 6.8 kb excised from pK18-*hACE2* was injected into the pronuclei of hybrid (C57BL/6Jx SJL/J) zygotes to obtain transgenic animals.

In the experiment described above (McCray et al., 2007), three lines of transgenic mice were obtained from different

founders. It should be noted that the chosen technology provides random insertion of the transgene into the recipient genome, with a different number of copies. According to the authors, the number of transgene copies in the lines varied from 4 to 10. The transgene expression was observed in various tissues of the obtained mice: lungs, small intestine, liver and kidney, and at a low level was noted in the brain.

After intranasal SARS-CoV infection of transgenic CK18-*hACE2*, animals of all three lines died after the 7th day after virus inoculation. Moreover, mice carrying more transgene copies died already on the 4th day after infection. It is important to note that weight loss was observed in all transgenic mouse strains. The high titer of the virus was determined in the lungs compared with the control and reached the highest level on the 2nd day after infection. These data suggest increased viral replication as a key factor in the development of severe disease in transgenic animals. Interestingly, despite the expression of human *ACE2* in the small intestine, liver, and kidney, the presence of the virus was not found in them. Among three tested lines of transgenic mice, only in one virus was detected in the brain at a low level, although the level of transgene expression was at the background level.

Histological analysis of the lungs on the 2nd day of infection showed signs of vascularization and peribronchiolar inflammation, and then there was an expansion of the zone of the inflammatory process, cell infiltration and desquamation of the cell epithelium in two lines of transgenic mice. In general, the pattern of intranasal infection of transgenic CK18-*hACE2* lines showed similarities with the development of acute respiratory syndrome in humans caused by SARS-CoV infection, in other words, these animals can be used as model objects for studying the pathogenesis of coronavirus infection (McCray et al., 2007; Netland et al., 2008). More recently, CK18-*hACE2* infection of mice with SARS-CoV-2 has shown similarities with clinical manifestations of human COVID-19 (Yinda et al., 2020).

Almost simultaneously, another group of researchers created transgenic mice expressing human *ACE2* under the control of the constitutive CAG promoter (Tseng et al., 2007). The cDNA sequence of the human *ACE2* gene was inserted into the expression vector pCAGGS/MCS, which contained in the 5'-sequence of the enhancer of the early promoter of the cytomegalovirus fused with the promoter of the chicken actin gene, and in the 3'-region splicing sites of the rabbit globin gene. The total size of the pCAGGS-*ACE2* expression vector was 7750 bp. A DNA fragment of this cassette was injected into the pronuclei of C57BL/6J × C3H/HeJ hybrid zygotes. Among the F0 offspring born from experimental zygotes, five transgenic animals were identified, of which two founders, AC70 and AC63, gave rise to two lines. RT-PCR analysis showed the presence of human *ACE2* transgene transcripts in the stomach, heart, muscles, brain, kidneys, lungs, and small intestine.

Infection of transgenic mice with SARS-CoV showed the following symptoms: permanent weight loss, shortness of breath, and uncontrolled motor activity. The death of animals was observed after the 3rd day of infection and ended in total lethality by the 8th day. The reproduction of the virus occurred

mainly in the lung tissue, while in other samples: swabs from the oral cavity, blood, heart, spleen, kidney, urine or feces, the virus was not detected. Summing up, the authors (Tseng et al., 2007) concluded that the resulting transgenic mouse lines are susceptible to SARS-CoV infection and exhibit external signs similar to those of humans, including the lethal outcome of infected animals. According to the authors, such mice may be useful for studying the pathogenesis of SARS-CoV infection.

In 2007, a third article on the creation of transgenic mice susceptible to SARS-CoV infection appeared (Yang et al., 2007). This group of researchers used a construct that included the mouse *Ace2* gene promoter fused to the human *ACE2* gene. The DNA of this construct was injected into the zygotes of ICR mice and the birth of transgenic animals was observed. Human *ACE2* expression was detected in the lungs, heart, kidneys, and small intestine. On the 3rd and 7th days after infection with SARS-CoV, virus replication was observed in the lungs as well as signs of lung damage: interstitial hyperemia and hemorrhages, monocytic and lymphocytic infiltration, a proliferation of the alveolar epithelium and its desquamation. Interestingly, much later (Bao et al., 2020), after intranasal SARS-CoV-2 infection of these transgenic mice, a moderate weight loss was observed in the first five days, but no deaths were recorded in any case. The target and site of replication of COVID-19 was lung tissue, which resulted in the development of signs of pneumonia. Thus, the transgenic mouse line created in 2007 (Yang et al., 2007) has become a convenient platform for studying the pathogenesis of the two coronaviruses SARS-CoV and SARS-CoV-2.

Using the classical technology of transgenesis – microinjection of DNA expression vectors into the pronuclei of zygotes, which are randomly inserted into the recipient genome (Smirnov et al., 2020) and, which results in the expression of the transgene varying in different founders. One such study worthy of attention is the generation of transgenic C3B6 mice carrying human *ACE2* under the control of an HFH4 promoter specific for ciliated lung epithelial cells (Ostrowski et al., 2003; Menachery et al., 2016). Human *ACE2* expression was found in the lungs, brain, liver, and kidneys of transgenic HFH4-hACE2 mice. Intranasal infection with SARS-CoV or one of its WIV1-CoV strains caused weight loss in the first days of infection and death of animals after the 6th day from the moment of infection (Menachery et al., 2016). It is important that vaccines were tested on these mice and their positive effect was observed against both types of coronavirus. Later, these transgenic mice were successfully used to test antiviral therapy against COVID-19 (Jiang et al., 2020).

It makes sense to dwell on the study of the Russian group of A.V. Deikin, which for the first time provides protective measures to the researchers themselves against being infected with coronaviruses from transgenic mice (Bruter et al., 2021). The authors created a cassette consisting of two main elements: the pKB1 vector and the *hACE2* open reading frame. pKB1 ampicillin-resistant vector was designed for cloning of genes, the expression of which depends on Cre-recombination (Cre-recombinase is present in the prokaryotic genome, but

absent in eukaryotes), and contains insulators and terminators (“protecting” the transgene from the influence of nearby sequences), CAG promoter and STOP cassette. In addition, the vector contains the IRES element of the encephalomyocarditis virus, the *GFP* reporter gene, and the polyA signal of the SV40 virus. It is important that the expression of the human *ACE2* transgene is activated only after removal of the STOP cassette by Cre-recombinase.

The recombinant DNA cassette was microinjected into F₁ hybrid mouse zygotes (CBA×C57BL/6). The resulting transgenic animals did not express either human *ACE2* or the reporter gene. To activate the transgene, transgenic mice were crossed with B6 mice Cg-Ndor1Tg(UBC-cre/ERT2)1Ejb/1J (abbreviated as Ubi-Cre) carrying the Cre recombinase gene under the control of the UBC promoter. In heterozygous *ACE2*-GFP and Cre-transgene mice, the transgene was activated with tamoxifen, which activates Cre-UBC, which, in turn, cuts out the STOP cassette and activates the expression of *ACE2* and the reporter gene (Bruter et al., 2021). Thus, heterozygous mice become susceptible to coronavirus infection (Dolskiy et al., 2022).

Indeed, direct experiments demonstrated that the virus intranasally inoculated to transgenic mice SARS-CoV2 induced thickening of alveolar duct septa mediated by diffuse hyperplasia of alveolar type II epithelium. Also, the lung tissue underwent lymphocyte infiltration. It should be emphasized that erythrocyte aggregates were present in the lung tissue in abundance, which was indicative of clotting. In contrast to lung tissue samples, no aberrations were found in the histological examination of the brain except for abundant erythrocyte aggregates as a sign of clotting (Dolskiy et al., 2022). All experimental transgenic mice died on day 5 to 10 after the intranasal inoculation.

The COVID-19 pandemic has stimulated the search for new technologies for creating model animals – laboratory mice. The development of targeted modification of human and animal genomes using CRISPR/Cas9 technology has opened up the prospect of obtaining “humanized” animals, in the genome of which target endogenous genes can be replaced by homologous human genes. An example is the insertion of the human *ACE2* cDNA into the coding sequence of the endogenous *Ace2* gene in C57BL/6 mice using CRISPR/Cas9 technology (Sun et al., 2020). The cDNA of the human *ACE2* gene was inserted into exon 2 of the mouse *Ace2* gene in mouse zygotes. Such an insert inactivated the endogenous *Ace2* gene, and to visualize human *ACE2* expression, the fluorescent protein reporter gene tdTomato (red glow) was inserted at its 3'-end, together with the IRES site and the polyA sequence. Among mice born from experimental zygotes, transgenic animals with a target insertion of the human *ACE2* gene were identified.

Such mice were susceptible to intranasal SARS-CoV-2 infection at a young and adult age; the virus affected the lungs, trachea and brain. With intranasal infection, interstitial pneumonia developed, similar in manifestations to that of a person infected with SARS-CoV2, but without a lethal effect. It is clear that such genetically modified mice are seen as an attractive model of human coronavirus infection and could po-

Transgenic mice with cDNA of the human *ACE2* gene susceptible to SARS-CoV and SARS-CoV-2 coronavirus infection

Year of creation	Line name	Promoter	Used technology	References
2007	CK18-hACE2	Human CK18 gene	Random insertion of a transgene	McCray et al., 2007
2007	AC70 Tg ⁺ AC63 Tg ⁺	CAG – “merged” chicken actin gene promoter + enhancer megalovirus	Random insertion of a transgene	Tseng et al., 2007
2007	hACE2	Mouse <i>Ace2</i> gene promoter	Random insertion of a transgene	Yang et al., 2007
2003, 2016	HFH4-hACE2	Specific HFH 4/FOXJ1 promoter active in ciliated epithelial cells	Random insertion of a transgene	Ostrowski et al., 2003 Menachery et al., 2016
2021	hACE2(LoxP-Stop)	CAG promoter	Random insertion of a transgene	Bruter et al., 2021
2020	hACE2	Mouse <i>Ace2</i> gene promoter	Targeted transgene insertion with CRISPR/Cas 9	Sun et al., 2020
2021	C57BL/6N-Ace2 ^{em2} (hACE2-WPRE, pgk-puro)/CCLA BALB/c-Ace2 ^{em1} (hACE2-WPRE, pgk-puro)/CCLA	Mouse <i>Ace2</i> gene promoter	Targeted transgene insertion with CRISPR/Cas 9	Liu et al., 2021

tentially serve as a platform for vaccine trials and pharmacological drug testing.

An essentially similar approach to “humanization” of the mouse *Ace2* gene by insertion of the human *ACE2* cDNA, has been implemented in C57BL/6 and BALB/c mouse embryonic stem cell (ESC) lines using CRISPR/Cas9 technology (Liu et al., 2021). It is appropriate to recall that the used ESC lines are capable, after injection into the cavity of tetraploid blastocysts, to replace endogenous cells of the internal mass, as a result of which transgenic descendants developed from donor ESCs are born. Transgenic mice thus generated, named C57BL/6NAce2em2(hACE2-WPRE, pgk-puro)/CCLA and BALB/c-Ace2em1(hACE2-WPRE, pgk-puro)/CCLA, were susceptible to intranasal SARS-CoV2 infection, although they differed from those obtained by S.-H. Sun et al. (2020) by a number of traits from transgenic mice. Thus, to date, a number of lines of “humanized” mice that carry the human *ACE2* transgene and are susceptible to coronavirus infection and potentially capable of modeling human coronavirus pathology have been created.

Conclusion

To sum up, it should be noted that despite the variety of created transgenic mouse lines susceptible to coronavirus infection (see the Table), the most popular among researchers are CK18-hACE2 mice created by the group P.B. McCray et al. (2007). According to PubMed, from 2020 to 2022, 101 articles that used this line as a model animal for the study of pathogenesis and coronavirus infection were published. Nevertheless, the development of new models continues, since the source of supply of mice of the CK18-hACE2 line is the Jackson Laboratory (USA), which supplies them only for experiments, without the right to breed them in national animal facilities in other countries.

References

- Bader M. ACE2, angiotensin-(1-7), and Mas: the other side of the coin. *Pflugers Arch.* 2013;465(1):79-85. DOI 10.1007/s00424-012-1120-0.
- Bao L., Deng W., Huang B., Gao H., Liu J., Ren L., Wei Q., Yu P., Xu Y., Qi F., Qu Y., Li F., Lv Q., Wang W., Xue J., Gong S., Liu M., Wang G., Wang S., Song Z., Zhao L., Liu P., Zhao L., Ye F., Wang H., Zhou W., Zhu N., Zhen W., Yu H., Zhang X., Guo L., Chen L., Wang C., Wang Y., Wang X., Xiao Y., Sun Q., Liu H., Zhu F., Ma C., Yan L., Yang M., Han J., Xu W., Tan W., Peng X., Jin Q., Wu G., Qin C. The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature.* 2020;583(7818):830-833. DOI 10.1038/s41586-020-2312-y.
- Belouzard S., Chu V.C., Whittaker G.R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. USA.* 2009;106(14):5871-5876. DOI 10.1073/pnas.0809524106.
- Bruter A.V., Korshunova D.S., Kubekina M.V., Sergiev P.S., Kalina A.A., Ilchuk K.A., Yuliya Yu., Silaeva Y.Y., Korshunov E.K., Soldatov V.O., Deykin A.V. Novel transgenic mice with Cre-dependent co-expression of GFP and human ACE2: a safe tool for study of COVID-19 pathogenesis. *Transgenic Res.* 2021;30(3):289-301. DOI 10.1007/s11248-021-00249-8.
- Dolskiy A.A., Gudymo A.S., Taranov O.S., Grishchenko I.V., Shitik E.M., Prokopov D.Y., Soldatov V.O., Sobolevskaya E.V., Bodnev S.A., Danilchenko N.V., Moiseeva A.A., Torzhkova P.Y., Bulanovich Y.A., Onhonova G.S., Ivleva E.K., Kubekina M.V., Belykh A.E., Tregubchak T.V., Ryzhikov A.B., Gavrilova E.V., Maksyutov R.A., Deykin A.V., Yudkin D.V. The tissue distribution of SARS-CoV-2 in transgenic mice with inducible ubiquitous expression of hACE2. *Front. Mol. Biosci.* 2022;8:821506. DOI 10.3389/fmolb.2021.821506.
- Ge X.-Y., Li J.-L., Yang X.-L., Chmura A.A., Zhu G., Epstein J.H., Mazet J.K., Hu B., Zhang W.L., Peng C., Zhang Y.-J., Luo C.-M., Tan B., Wang N., Zhu Y., Crameri G., Zhang S.-Y., Wang L.-F., Daszak P., Shi Z.-L. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature.* 2013;503(7477):535-538. DOI 10.1038/nature12711.

- Glass W.G., Subbarao K., Murphy B., Murphy P.M. Mechanisms of host defense following severe acute respiratory syndrome-coronavirus (SARS-CoV) pulmonary infection of mice. *J. Immunol.* 2004; 173(6):4030-4039. DOI 10.4049/jimmunol.173.6.4030.
- Guan Y., Zheng B.J., He Y.Q., Liu X.L., Zhuang Z.X., Cheung C.L., Luo S.W., Li P.H., Zhang L.J., Guan Y.J., Butt K.M., Wong K.L., Chan K.W., Lim W., Shortridge K.F., Yuen K.Y., Peiris J.S., Poon L.L. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science.* 2003; 302(5643):276-278. DOI 10.1126/science.1087139.
- Imai Y., Kuba K., Rao S., Huan Y., Guo F., Guan B., Yang P., Sarao R., Wada T., Leong-Poi H., Crackower M.A., Fukamizu A., Hui C.-C., Hein L., Uhlig S., Slutsky A.S., Jiang C., Penninger J.M. Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature.* 2005;436(7047):112-116. DOI 10.1038/nature03712.
- Jiang R.-D., Liu M.-Q., Chen Y., Shan C., Zhou Y.-W., Shen X.-R., Li Q., Zhang L., Zhu Y., Si H.-R., Wang Q., Min J., Wang X., Zhang W., Li B., Zhang H.-J., Baric R.S., Zhou P., Yang X.-L., Shi Z.-L. Pathogenesis of SARS-CoV-2 in transgenic mice expressing human angiotensin-converting enzyme 2. *Cell.* 2020;182(1): 50-58.e8. DOI 10.1016/j.cell.2020.05.027.
- Kuba K., Imai Y., Ohto-Nakanishi T., Penninger J.M. Trilogity of ACE2: a peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. *Pharmacol. Ther.* 2010;128(1): 119-128. DOI 10.1016/j.pharmthera.2010.06.003.
- Lai M.M.C., Perlman S., Anderson J. Coronaviridae. In: Knipe D.M., Howley P.M. (Eds.). *Fields Virology*. Philadelphia: Lippincott Williams and Wilkins, 2007;1305-1335.
- Lan J., Ge J., Yu J., Shan S., Zhou H., Fan S., Zhang Q., Shi X., Wang Q., Zhang L., Wang X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020;581(7807):215-220. DOI 10.1038/s41586-020-2180-5.
- Li F., Li W., Farzan M., Harrison S.C. Structure of SARS coronavirus spike receptor binding domain complexed with receptor. *Science.* 2005;309(5742):1864-1868. DOI 10.1126/science.1116480.
- Li Q., Guan X., Wu P., Wang X., Zhou L., Tong Y., Ren R., Leung K.S.M., Lau E.H.Y., Wong J.Y., Xing X., Xiang N., Wu Y., Li C., Chen Q., Li D., Liu T., Zhao J., Liu M., Tu W., Chen C., Jin L., Yang R., Wang Q., Zhou S., Wang R., Liu H., Luo Y., Liu Y., Shao G., Li H., Tao Z., Yang Y., Deng Z., Liu B., Ma Z., Zhang Y., Shi G., Lam T.T.Y., Wu J.T., Gao G.F., Cowling B.J., Yang B., Leung G.M., Feng Z. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 2020; 382(13):1199-1207. DOI 10.1056/NEJMoa2001316.
- Li W., Moore M.J., Vasilieva N., Sui J., Wong S.K., Berne M.A., Somasundaran M., Sullivan J.L., Luzuriaga K., Greenough T.C., Choe H., Farzan M. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature.* 2003;426(6965):450-454. DOI 10.1038/nature02145.
- Liu F.-L., Wu K., Sun J., Duan Z., Quan X., Kuang J., Chu S., Pang W., Gao H., Xu L., Li Y.-C., Zhang H.-L., Wang X.-H., Luo R.-H., Feng X.-L., Schöler H.R., Chen X., Pei D., Wu G., Zheng Y.-T., Chen J. Rapid generation of ACE2 humanized inbred mouse model for COVID-19 with tetraploid complementation. *Natl. Sci. Rev.* 2021;8(2):nwaa285. DOI 10.1093/nsr/nwaa285.
- Martina B.E., Haagmans B.L., Kuiken T., Fouchier R.A., Rimmelzwaan G.F., Van Amerongen G., Peiris J.S.M., Lim W., Osterhaus A.D.M.E. Virology: SARS virus infection of cats and ferrets. *Nature.* 2003;425(6961):915. DOI 10.1038/425915a.
- McCray P.B. Jr., Pewe L., Wohlford-Lenane C., Hickey M., Manzel L., Shi L., Netland J., Jia H.P., Halabi C., Sigmund C.D., Meyerholz D.K., Kirby P., Look D.C., Perlman S. Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. *J. Virol.* 2007;81(2):813-821. DOI 10.1128/JVI.02012-06.
- Menachery V.D., Yount B.L. Jr., Sims A.C., Debbink K., Agnihothram S.S., Gralinski L.E., Graham R.L., Scobey T., Plante J.A., Royal S.R., Swanstrom J., Sheahan T.P., Pickles R.J., Corti D., Randell S.H., Lanzavecchia A., Marasco W.A., Baric R.S. ARS-like WIV1-CoV poised for human emergence. *Proc. Natl. Acad. Sci. USA.* 2016;113(11):3048-3053. DOI 10.1073/pnas.1517719113.
- Netland J., Meyerholz D.K., Moore S., Cassell M., Perlman S. Severe acute respiratory syndrome coronavirus infection causes neuronal death in the absence of encephalitis in mice transgenic for human ACE2. *J. Virol.* 2008;82(15):7264-7275. DOI 10.1128/JVI.00737-08.
- Ostrowski L.E., Hutchins J.R., Zakel K., O'Neal W.K. Targeting expression of a transgene to the airway surface epithelium using a ciliated cell-specific promoter. *Mol. Ther.* 2003;8(4):637-645. DOI 10.1016/S1525-0016(03)00221-1.
- Peiris J.S., Lai S.T., Poon L.L., Guan Y., Yam L.Y.C., Lim W., Nicholls J., Yee W.K.S., Yan W.W., Cheung M.T., Cheng V.C.C., Chan K.H., Tsang D.N.C., Yung R.W.H., Ng T.K., Yuen K.Y., SARS study group. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet.* 2003;361(9366):1319-1325. DOI 10.1016/S0140-6736(03)13077-2.
- Reusken C.B., Haagmans B.L., Müller M.A., Gutierrez C., Godeke G.J., Meyer B., Muth D., Raj V.S., Smits-De Vries L., Corman V.M., Drexler J.-F., Smits S.L., El Tahir Y.E., De Sousa R., van Beek J., Nowotny N., van Maanen K., Hidalgo-Hermoso E., Bosch B.-J., Rottier P., Osterhaus A., Gortázar-Schmidt C., Drosten C., Koopmans M.P. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet. Infect. Dis.* 2013;13(10):859-866. DOI 10.1016/S1473-3099(13)70164-6.
- Roberts A., Vogel L., Guarner J., Hayes N., Murphy B., Zaki S., Subbarao K. Severe acute respiratory syndrome coronavirus infection of golden Syrian hamsters. *J. Virol.* 2005;79(1):503-511. DOI 10.1128/JVI.79.1.503-511.2005.
- Smirnov A., Fishman V., Yunusova A., Korablev A., Serova I., Skryabin B.V., Rozhdvestvensky T.S., Battulin N. DNA barcoding reveals that injected transgenes are predominantly processed by homologous recombination in mouse zygote. *Nucleic Acids Res.* 2020;48(2): 719-735. DOI 10.1093/nar/gkz1085.
- Subbarao K., Roberts A. Is there an ideal animal model for SARS? *Trends Microbiol.* 2006;14(7):299-303. DOI 10.1016/j.tim.2006.05.007.
- Sun S.-H., Chen Q., Gu H.-J., Yang G., Wang Y.-X., Huang X.-Y., Liu S.-S., Zhang N.-N., Li X.-F., Xiong R., Guo Y., Deng Y.-Q., Huang W.-J., Liu Q., Liu Q.-M., Shen Y.-L., Zhou Y., Yang X., Zhao T.-Y., Fan C.-F., Zhou Y.-S., Qin C.-F., Wang Y.-C. A mouse model of SARS-CoV-2 infection and pathogenesis. *Cell Host Microbe.* 2020;28(1):124-133.e4. DOI 10.1016/j.chom.2020.05.020.
- Takaki H., Oshiumi H., Shingai M., Matsumoto M., Seya T. Development of mouse models for analysis of human virus infections. *Microbiol. Immunol.* 2017;61(3-4):107-113. DOI 10.1111/1348-0421.12477.
- Tseng C.T., Huang C., Newman P., Wang N., Narayanan K., Watts D.M., Makino S., Packard M.M., Zaki S.R., Chan T.-S., Peters C.J. Severe acute respiratory syndrome coronavirus infection of mice transgenic for the human angiotensin-converting enzyme 2 virus receptor. *J. Virol.* 2007;81(3):1162-1173. DOI 10.1128/JVI.01702-06.
- van Boheemen S., de Graaf M., Lauber C., Bestebroer T.M., Raj V.S., Zaki A.M., Osterhaus A.D., Haagmans B.L., Gorbalenya A.E., Snijder E.J., Fouchier R.A. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio.* 2012;3(6):e00473-12. DOI 10.1128/mBio.00473-12.
- Wan Y., Shang J., Graham R., Baric R.S., Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 2020;94(7): e00127-20. DOI 10.1128/jvi.00127-20.
- Wang Q., Zhang Y., Wu L., Niu S., Song C., Zhang Z., Lu G., Qiao C., Hu Y., Yuen K.-Y., Wang Q., Zhou H., Yan J., Qi J. Structural and

- functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*. 2020;181(4):894-904. DOI 10.1016/j.cell.2020.03.045.
- Woo P.C.Y., Lau S.K.P., Huang Y., Yuen K.-Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med. (Maywood)*. 2009a;234(10):1117-1127. DOI 10.3181/0903-MR-94.
- Woo P.C., Lau S.K., Lam C.S., Lai K.K., Huang Y., Lee P., Luk G.S., Dyrting K.C., Chan K.H., Yuen K.Y. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *J. Virol.* 2009b;83(2):908-917. DOI 10.1128/JVI.01977-08.
- Yang X.H., Deng W., Tong Z., Liu Y.X., Zhang L.F., Zhu H., Gao H., Huang L., Liu Y.L., Ma C.M., Xu Y.F., Ding M.X., Deng H.K., Qin C. Mice transgenic for human angiotensin-converting enzyme 2 provide a model for SARS coronavirus infection. *Comp. Med.* 2007; 57(5):450-459. PMID: 17974127.
- Yinda C.K., Port J.R., Bushmaker T., Owusu I.O., Avanzato V.A., Fischer R.J., Schulz J.E., Holbrook M.G., Hebner M.J., Rosenke R., Thomas T., Marzi A., Best S.M., de Wit E., Shaia C., van Doremalen N., Munster V.J. K18-hACE2 mice develop respiratory disease resembling severe COVID-19. *bioRxiv*. 2020;Aug 11; 2020.08.11.246314. DOI 10.1101/2020.08.11.246314.
- Zaki A.M., van Boheemen S., Bestebroer T.M., Osterhaus A.D.M.E., Fouchier R.A.M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 2012;367(19):1814-1820. DOI 10.1056/NEJMoa1211721.
- Zhou F., Yu T., Du R., Fan G., Liu Y., Liu Z., Xiang J., Wang Y., Song B., Gu X., Guan L., Wei Y., Li H., Wu X., Xu J., Tu S., Zhang Y., Chen H., Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062. DOI 10.1016/S0140-6736(20)30566-3.
- Zhou P., Yang X.-L., Wang X.-G., Hu B., Zhang L., Zhang W., Si H.-R., Zhu Y., Li B., Huang C.-L., Chen H.-D., Chen J., Luo Y., Guo H., Jiang R.-D., Liu M.-Q., Chen Y., Shen X.-R., Wang X., Zheng X.-S., Zhao K., Chen Q.-J., Deng F., Liu L.-L., Yan B., Zhan F.-X., Wang Y.-Y., Xiao G.-F., Shi Z.-L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. DOI 10.1038/s41586-020-2012-7.
- Zhu N., Zhang D., Wang W., Li X., Yang B., Song J., Zhao X., Huang B., Shi W., Lu R., Niu P., Zhan F., Ma X., Wang D., Xu W., Wu G., Gao G.F., Tan W. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 2020;382(8):727-733. DOI 10.1056/NEJMoa2001017.

ORCID ID

O.L. Serov orcid.org/0000-0003-1056-2966

Acknowledgements. This study was supported by the Russian Foundation for Basic Research, project No. 20-04-60094. The access to full-text publications was supported by state budgeted project No. FWNR-2022-0019.

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

Received February 16, 2022. Revised March 25, 2022. Accepted March 25, 2022.