

Научный рецензируемый журнал

# ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.

Периодичность 8 выпусков в год

DOI 10.18699/VJGB-22-01

## Учредители

Сибирское отделение Российской академии наук

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

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

## Главный редактор

А.В. Кочетов – чл.-кор. 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. Rogozin – 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 – чл.-кор. 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-22-01

**Founders**

Siberian Branch of the Russian Academy of Sciences

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

The Vavilov Society of Geneticists and Breeders

**Editor-in-Chief**

A.V. Kochetov, Corr. Member of the RAS, 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), Russia

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

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

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

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

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

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

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

S.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology), Russia

V.A. Stepanov, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

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

A.V. Vasiliev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

M.I. Voevoda, Full Member of the RAS, Dr. Sci. (Medicine), Russia

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

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

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

## Молекулярная и клеточная биология

- 5 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Фосфорилгуанидиновые олигонуклеотиды как праймеры для РНК-зависимого ДНК синтеза с помощью обратной транскриптазы вируса лейкемии мышей.  
Е.С. Дюдеева, И.А. Пышина

- 14 **ОБЗОР**  
Опухоль-ассоциированные фибробласты и их роль в опухолевой прогрессии.  
М.С. Ермаков, А.А. Нуштаева, В.А. Рихтер, О.А. Коваль

## Генетика и селекция растений

- 22 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Особенности побегообразования в популяциях *Miscanthus sacchariflorus* (Poaceae) под влиянием экологических факторов и паспортизация с помощью ISSR-маркеров. О.В. Дорогина, Н.С. Нуждина, Г.А. Зуева, Ю.А. Гисматулина, О.Ю. Васильева

- 30 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Динамические изменения содержания бетанина в столовой свекле в течение вегетационного периода: их взаимодействие с абиотическими факторами. Д.В. Соколова

## Генетика человека

- 40 **ОБЗОР**  
Взаимосвязь ламинов с эпигенетическими факторами при старении. Р.Н. Мустафин, Э.К. Хуснутдинова

- 50 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Транскрипционная активность генов репарации, апоптоза и клеточного цикла (*TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PAD14*, *MAPK8*, *NF-KB1*, *STAT3*, *GATA3*) у хронически облученных людей с различной интенсивностью апоптоза лимфоцитов периферической крови.  
В.С. Никифоров, Е.А. Блинова, А.И. Котикова, А.В. Аклев

## Медицинская генетика

- 59 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Ассоциация полиморфизма генов *TP53*, *MDM2* и *NQO1* с вирусной нагрузкой среди женщин с вирусом папилломы человека. А.Х. Албасал, Е.В. Машкина (на англ. языке)

- 65 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Функциональный подход к изучению генетики сердечно-сосудистых заболеваний: post-GWAS исследование. Л.О. Брызгалов, Е.Е. Корболина, И.С. Дамаров, Т.И. Меркулова (на англ. языке)

## Популяционная генетика

- 74 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Разграничение видов и анализ криптического разнообразия микроводорослей рода *Micractinium* (*Chlorophyta*). Е.С. Кривина, А.Д. Темралеева, Ю.С. Букин

- 86 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
ДНК-метабаркодинг бентосных водорослей и ассоциированных с ними эукариот оз. Байкал в условиях быстрых экологических изменений. Ю.С. Букин, Л.С. Кравцова, Т.Е. Перетолчина, А.П. Федотов, А.Е. Тупикин, М.Р. Кабилов, Д.Ю. Щербаков, Е.В. Минчева

## Актуальные технологии

- 96 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
Биоинформатический поиск соответствия дифференциально экспрессируемых генов домашних и диких животных с ортологичными генами, изменяющими репродуктивный потенциал человека. М.П. Пономаренко, И.В. Чадаева, П.М. Пономаренко, А.Г. Богомолов, Д.Ю. Ощепков, Е.Б. Шарыпова, В.В. Суслов, А.В. Осадчук, Л.В. Осадчук, Ю.Г. Матушкин

- 109 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**  
*De novo* сборка и анализ транскриптома сибирской лягушки *Rana amurensis*. Д.Н. Смирнов, С.В. Шеховцов, А.А. Шипова, Г.Р. Газизова, Е.И. Шагмарданова, Н.А. Булахова, Е.Н. Мещерякова, Т.В. Полубоярова, Е.Е. Храмеева, С.Е. Пельтек, Д.И. Берман (на англ. языке)

## Molecular and cell biology

- 5 **ORIGINAL ARTICLE**  
Phosphoryl guanidine oligonucleotides as primers for RNA-dependent DNA synthesis using murine leukemia virus reverse transcriptase. *E.S. Dyudeeva, I.A. Pyshnaya*

- 14 **REVIEW**  
Cancer-associated fibroblasts and their role in tumor progression. *M.S. Ermakov, A.A. Nushtaeva, V.A. Richter, O.A. Koval*

## Plant genetics and breeding

- 22 **ORIGINAL ARTICLE**  
Specific shoot formation in *Miscanthus sacchariflorus* (Poaceae) under different environmental factors and DNA passportization using ISSR markers. *O.V. Dorogina, N. S. Nuzhdina, G.A. Zueva, Yu.A. Gismatulina, O.Yu. Vasilyeva*

- 30 **ORIGINAL ARTICLE**  
Dynamic changes in betanin content during the growing season of table beet: their interplay with abiotic factors. *D.V. Sokolova*

## Human genetics

- 40 **REVIEW**  
The relationship of lamins with epigenetic factors during aging. *R.N. Mustafin, E.K. Khusnutdinova*

- 50 **ORIGINAL ARTICLE**  
Transcriptional activity of repair, apoptosis and cell cycle genes (*TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PADI4*, *MAPK8*, *NF-KB1*, *STAT3*, *GATA3*) in chronically exposed persons with different intensity of apoptosis of peripheral blood lymphocytes. *V.S. Nikiforov, E.A. Blinova, A.I. Kotikova, A.V. Akleyev*

## Medical genetics

- 59 **ORIGINAL ARTICLE**  
Association between *TP53*, *MDM2* and *NQO1* gene polymorphisms and viral load among women with human papillomavirus. *A.H. AlBosale, E.V. Mashkina*

- 65 **ORIGINAL ARTICLE**  
The functional insight into the genetics of cardiovascular disease: results from the post-GWAS study. *L.O. Bryzgalov, E.E. Korbolina, I.S. Damarov, T.I. Merkulova*

## Population genetics

- 74 **ORIGINAL ARTICLE**  
Species delimitation and microalgal cryptic diversity analysis of the genus *Micractinium* (*Chlorophyta*). *E.S. Krivina, A.D. Temraleeva, Yu.S. Bukin*

- 86 **ORIGINAL ARTICLE**  
DNA metabarcoding of benthic algae and associated eukaryotes from Lake Baikal in the face of rapid environmental changes. *Yu.S. Bukin, L.S. Kravtsova, T.E. Peretolchina, A.P. Fedotov, A.E. Tupikin, M.R. Kabilov, D.Yu. Sherbakov, E.V. Mincheva*

## Mainstream technologies


- 96 **ORIGINAL ARTICLE**  
A bioinformatic search for correspondence between differentially expressed genes of domestic versus wild animals and orthologous human genes altering reproductive potential. *M.P. Ponomarenko, I.V. Chadaeva, P.M. Ponomarenko, A.G. Bogomolov, D.Yu. Oshchepkov, E.B. Sharypova, V.V. Suslov, A.V. Osadchuk, L.V. Osadchuk, Yu.G. Matushkin*

- 109 **ORIGINAL ARTICLE**  
*De novo* assembly and analysis of the transcriptome of the Siberian wood frog *Rana amurensis*. *D.N. Smirnov, S.V. Shekhovtsov, A.A. Shipova, G.R. Gazizova, E.I. Shagimardanova, N.A. Bulakhova, E.N. Meshcheryakova, T.V. Poluboyarova, E.E. Khrameeva, S.E. Peltek, D.I. Berman*

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

## Phosphoryl guanidine oligonucleotides as primers for RNA-dependent DNA synthesis using murine leukemia virus reverse transcriptase

E.S. Dyudeeva, I.A. Pyshnaya 

Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia  
 pyshnaya@niboch.nsc.ru

**Abstract.** Modern approaches to the detection and analysis of low-copy-number RNAs are often based on the use of RNA-dependent DNA polymerases, for example, in reverse-transcription PCR. The accuracy and efficiency of cDNA synthesis in the reverse-transcription reaction catalyzed by reverse transcriptase (RNA-dependent DNA polymerase) significantly affect the correctness of the results of PCR diagnostic assays and/or RNA sequencing. In this regard, many studies are focused on the optimization of the reverse-transcription reaction, including the search for more perfect primers necessary to obtain a full-length DNA copy of RNA under study. The best-known completely uncharged analogs of oligonucleotides – morpholine oligonucleotides and peptide nucleic acids – cannot be substrates for enzymes that process nucleic acids. The aim of this work was to conduct a pilot study of uncharged phosphoryl guanidine oligodeoxyribonucleotides (PGOs) as primers for mouse leukemia virus reverse transcriptase (MMLV H<sup>-</sup>). Specific features of elongation of partially and completely uncharged PGO primers were investigated. It was demonstrated that PGOs can be elongated efficiently, e.g., in the presence of a fragment of human ribosomal RNA having complex spatial structure. It was shown that the proportion (%) of abortive elongation products of a PGO primer depends on buffer ionic strength, nucleotide sequence of the primer, and the presence and location of phosphoryl guanidine groups in the primer. The results indicate the suitability of PGOs, including completely electroneutral ones, as primers for reverse-transcription PCR, thereby opening up new prospects for the creation of experimental models for the analysis of highly structured RNA. Key words: uncharged analogs of oligonucleotides; phosphoryl guanidine oligonucleotides; reverse transcription; reverse transcriptase; mouse leukemia virus reverse transcriptase; MMLV H<sup>-</sup>.

**For citation:** Dyudeeva E.S., Pyshnaya I.A. Phosphoryl guanidine oligonucleotides as primers for RNA-dependent DNA synthesis using murine leukemia virus reverse transcriptase. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(1):5-13. DOI 10.18699/VJGB-22-02

## Фосфорилгуанидиновые олигонуклеотиды как праймеры для РНК-зависимого ДНК синтеза с помощью обратной транскриптазы вируса лейкемии мышей

Е.С. Дюдеева, И.А. Пышная 

Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия  
 pyshnaya@niboch.nsc.ru

**Аннотация.** Современные подходы к обнаружению и анализу низкокопийных РНК часто основываются на использовании РНК-зависимых ДНК-полимераз, например, метод ОТ-ПЦР. Точность и эффективность синтеза кДНК в реакции обратной транскрипции (ОТ), катализируемой ревертазой (РНК-зависимой ДНК-полимеразой), значительно влияют на корректность результатов ПЦР-диагностики и/или секвенирования РНК. В связи с этим множество исследований посвящено оптимизации реакции ОТ, в том числе поиску более совершенных праймеров, необходимых для наработки полноразмерной ДНК-копии исследуемой РНК. Наиболее известные полностью незаряженные аналоги олигонуклеотидов – морфолиновые олигонуклеотиды и пептидонуклеиновые кислоты – не могут являться субстратами для ферментов, процессирующих НК. Целью нашей работы было пилотное исследование незаряженных фосфорилгуанидиновых олигодезоксирибонуклеотидов (ФГО) в качестве праймеров для обратной транскриптазы вируса лейкемии мышей (MMLV H<sup>-</sup>). Изучены особенности превращения частично и полностью незаряженных ФГО-праймеров. Продемонстрировано, что ФГО способны эффективно удлиняться, в том числе в присутствии фрагмента рРНК человека, обладающего сложной пространственной структурой. Доля abortивных про-



дуктов удлинения ФГО-праймера зависит от ионной силы использованного буферного раствора, нуклеотидной последовательности праймера и наличия и расположения в его составе ФГ-групп. Полученные данные показывают возможность использования ФГО, включая полностью электронейтральные, как праймеров в реакциях ОТ-ПЦР, что открывает новые перспективы для создания систем анализа высокоструктурированных РНК.

Ключевые слова: незаряженные аналоги олигонуклеотидов; фосфорилгуанидиновые олигонуклеотиды; обратная транскрипция; ревертаза; обратная транскриптаза вируса лейкемии мышей; MMLV H-.

## Introduction

The detection of certain RNAs – some of the most important natural biopolymers – is an extremely relevant task in molecular diagnostics because such biomolecules are indicators of the state of the cell or an organism as a whole, of its metabolic status, or the presence of infection. Rapid and highly sensitive RNA detection can facilitate early diagnosis and hence more effective therapy for many diseases. Among RNA detection methods, the most popular is reverse-transcription polymerase chain reaction (RT-PCR), which consists of two sequential steps: (1) enzymatic synthesis of cDNA from an RNA template using reverse transcriptase and (2) obtaining the corresponding fragment of double-stranded DNA by standard PCR. With fluorophore-containing primers or probes, it is possible to carry out PCR in real time via monitoring an increase in the fluorescence signal, from which it is possible to accurately calculate the amount of analyzed RNA per cell/sample. RT-PCR is widely employed in the research on mRNA expression patterns (Bustin, 2000), and this method is used at the stage of creation of libraries for high-throughput RNA sequencing (RNA-Seq): a powerful tool for investigation of transcriptomes and other purposes (Mortazavi et al., 2008; Haas, Zody, 2010).

A popular enzyme for the reverse-transcription reaction is the Moloney murine leukemia virus (MMLV) reverse transcriptase. This reverse transcriptase is capable of (a) catalyzing DNA synthesis from a DNA or RNA template (Palikša et al., 2018; Li R. et al., 2020); (b) template-free synthesis of short DNA fragments (Ohtsubo et al., 2017); (c) DNA synthesis with strand displacement (Kelleher, Champoux, 1998; Malik et al., 2017); (d) switching a template strand (Wulf et al., 2019); and (e) cleavage of RNA as part of a hybrid RNA–DNA complex (Schultz, Champoux, 2008; Li R. et al., 2020). In laboratory studies, a recombinant variant devoid of RNase H activity (MMLV H-) is used. An extensive body of research deals with optimizing the reverse-transcription reaction, including a search for more efficient primers based on modified oligonucleotides with enhanced affinity for complementary nucleic acids (Heuverswyn et al., 2016; Menéndez-Arias et al., 2016; Palikša et al., 2018; Li R. et al., 2020). The currently known completely uncharged analogs of oligonucleotides – morpholine oligonucleotides and peptide nucleic acids – cannot serve as a substrate of enzymatic reactions, probably owing to the unusual/foreign backbone, which is substantially different in structure from the natural sugar-phosphate backbone. Partially modified P-alkyl phosphonate oligonucleotides and phosphoryl guanidine oligodeoxyribonucleotides (PGOs), in which the ribose moiety remains unchanged, are utilized as primers to distinguish the “wrong” complex during PCR (Li T.-L. et al., 2019; Chubarov et al., 2020). For PGOs, it has been demonstrated that they form complexes with complementary DNA or RNA

in solutions with low ionic strength, and even in deionized water; thermal stability of these complexes is comparable to that of the native complex under conditions close to physiological (Kupryushkin et al., 2014; Dyudeeva et al., 2019). Moreover, it is reported that spatial structure of duplexes containing PGOs is virtually the same as the structure of the double helix of two native nucleic acids (Lomzov et al., 2019). The combination of these factors makes phosphoryl guanidine (PG)-containing oligonucleotides a promising platform for the development of highly specific probes and, possibly, enables their applications in some enzymatic transformations of nucleic acids catalyzed by either reverse transcriptase or DNA-dependent DNA polymerases (Kupryushkin et al., 2017; Chubarov et al., 2020).

The aim of this work was to study the effects of a number of PG groups and the profile of modification of a DNA primer with PG groups on the possibility of RNA-dependent elongation of such a primer by reverse transcriptase MMLV H- and to analyze the prospects of PGOs as primers for RT-PCR in the detection of relatively long RNAs.

## Materials and methods

**Synthesis of nucleic acids.** The synthesis of native oligodeoxyribonucleotides and their PG-containing analogs was carried out by the phosphoramidite method on an ASM-800 automated DNA/RNA synthesizer (Biosset, Russia) with commercial phosphoramidite monomers and porous polymer carriers (Glen Research, USA). For the introduction of PG groups, the protocol described by Kupryushkin et al. (2014) was used. A fragment of human 18S ribosomal RNA (rRNA) was kindly provided by M.R. Kabilov (Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences).

**The isolation** of oligonucleotides and their PGO analogs from a reaction mixture was implemented by reversed-phase high-performance liquid chromatography on an Agilent 1200 series chromatograph (Agilent, USA) with a column (4.6 × 150 mm) carrying the Eclipse XDB-C18 sorbent (5 μm) (Agilent). Elution of oligonucleotides was conducted in a linear concentration gradient of acetonitrile (0–90 %) in a 0.02 M solution of triethylammonium acetate for 30 min at a flow rate of 1.5 ml/min.

**Electrophoretic analysis** of oligonucleotides and their PGO analogs under denaturing conditions was performed in a 15 % polyacrylamide gel (PAAG; an acrylamide:N,N'-methylene bisacrylamide ratio of 29:1, with 8 M urea) in TBE buffer (89 mM Tris-borate, pH 8.3, 2 mM Na<sub>2</sub>EDTA) at a voltage of 50 V/cm. The results of electrophoretic separation were visualized by staining the gel with the StainsAll reagent. When a fluorophore was present in oligonucleotides, fluorescence scanning was performed by means of a VersaDoc

**Table 1.** Composition of reaction mixtures for reverse transcription in the case of a short synthetic RNA template and a relatively long fragment of rRNA

17-mer RNA template		Fragment of 18S rRNA	
Component	Amount	Component	Amount
DNA	10 <sup>-5</sup> M	DNA	10 <sup>-6</sup> M
RNA	10 <sup>-5</sup> M	RNA	10 <sup>-6</sup> M
MMLV H-	200 units of activity	MMLV H-	200 units of activity
dATP, dCTP, dGTP	300 μM each	dATP, dCTP, dGTP, dTTP	300 μM each
dUTP <sup>FAM</sup>	50 μM	dUTP <sup>FAM</sup>	–
Buffer	1 × St	Buffer	1 × St or L

MP 4000 Molecular Imager System (Bio-Rad, USA), which is a gel documentation system.

To separate relatively long products of primer elongation obtained by a reverse-transcription reaction, we used a 1.5 % agarose gel in TBE buffer in a horizontal electrophoretic unit (Bio-Rad). Double-stranded-DNA molecular-weight markers called the GeneRuler 100 bp DNA Ladder (ThermoScientific, Lithuania) were employed to assess relative electrophoretic mobility of the samples. The results of electrophoretic separation were visualized by staining the gel with ethidium bromide (0.0001 % aqueous solution) and by scanning the fluorescence signal as described above. The proportion (%) of major primer elongation products was calculated by analyzing the scanned fluorescence signal of the respective spots by means of the GelPro Analyzer 4.0 software (Media Cybernetics, USA).

**Reverse-transcription reactions.** Enzymatic elongation of oligonucleotides and their PGO analogs in a reverse-transcription reaction was performed using a recombinant enzyme: Murine leukemia virus reverse transcriptase MMLV H- (Sib-Enzyme, Russia). The following solutions served as buffers: standard (St): 30 mM Tris-HCl, 5 mM MgCl<sub>2</sub>, 50 mM KCl, 5 mM dithiothreitol, and low-salt (L): 30 mM Tris-HCl, 2 mM MgCl<sub>2</sub>, 5 mM dithiothreitol. Amounts of components of the reverse-transcription reaction mixtures are presented and compared in Table 1 (unless indicated otherwise); the total volume of the mixture was 10 μl.

The reverse-transcription reaction was conducted at 37 °C. The reaction time was 1 h in the case of the 17-mer RNA template and 3 h for a high-molecular-weight RNA template: a fragment of human 18S rRNA. At the end of the reaction, the enzyme was thermally inactivated by keeping the reaction mixture at 95 °C for 5 min, or the RNA template was hydrolyzed by adding an excess of concentrated aqueous ammonia to the reaction mixture and incubating at 56 °C for 2 h. Ammonia was removed by evaporating the solutions in a vacuum concentrator until the odor disappeared.

## Results and discussion

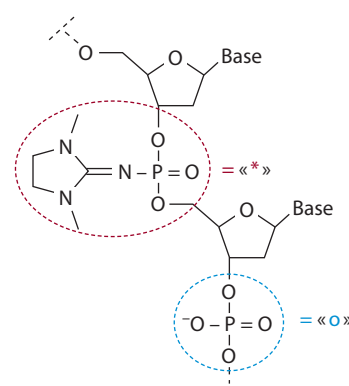
At the first stage of the work, we examined a series of decameric deoxyribonucleotides identical in nucleotide sequence but differing in the number of uncharged groups (i. e., PG groups) that were introduced (Fig. 1) and in their mutual arrangement (Table 2).

The analyzed decanucleotides were complementary to the 3'-terminal region of a model 18-mer RNA template called rM (Fig. 2). The nucleotide sequence of the template as part of

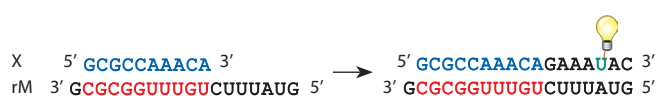
**Table 2.** Structures of decanucleotides

Name	Sequence, 5'→3'	<i>n</i>
X	G <sup>o</sup> C <sup>o</sup> G <sup>o</sup> C <sup>o</sup> C <sup>o</sup> A <sup>o</sup> A <sup>o</sup> A <sup>o</sup> C <sup>o</sup> A	0
X <sub>1,4;7</sub>	G <sup>*</sup> C <sup>o</sup> G <sup>o</sup> C <sup>*</sup> C <sup>o</sup> A <sup>o</sup> A <sup>*</sup> A <sup>o</sup> C <sup>o</sup> A	3
X <sub>1,5;9</sub>	G <sup>*</sup> C <sup>o</sup> G <sup>o</sup> C <sup>o</sup> C <sup>*</sup> A <sup>o</sup> A <sup>o</sup> A <sup>o</sup> C <sup>*</sup> A	3
X <sub>2,4;6,8</sub>	G <sup>o</sup> C <sup>*</sup> G <sup>o</sup> C <sup>*</sup> C <sup>o</sup> A <sup>*</sup> A <sup>o</sup> A <sup>*</sup> C <sup>o</sup> A	4
X <sub>1-2;5-6;9</sub>	G <sup>*</sup> C <sup>*</sup> G <sup>o</sup> C <sup>o</sup> C <sup>*</sup> A <sup>*</sup> A <sup>o</sup> A <sup>o</sup> C <sup>*</sup> A	5
X <sub>1-2;5;8-9</sub>	G <sup>*</sup> C <sup>*</sup> G <sup>o</sup> C <sup>o</sup> C <sup>*</sup> A <sup>o</sup> A <sup>o</sup> A <sup>*</sup> C <sup>*</sup> A	5
X <sub>1,3;5;7;9</sub>	G <sup>*</sup> C <sup>o</sup> G <sup>*</sup> C <sup>o</sup> C <sup>*</sup> A <sup>o</sup> A <sup>*</sup> A <sup>o</sup> C <sup>*</sup> A	5
X <sub>1-3;7-9</sub>	G <sup>*</sup> C <sup>*</sup> G <sup>*</sup> C <sup>o</sup> C <sup>o</sup> A <sup>o</sup> A <sup>*</sup> A <sup>*</sup> C <sup>*</sup> A	6
X <sub>2-3;5-6;8-9</sub>	G <sup>o</sup> C <sup>*</sup> G <sup>*</sup> C <sup>o</sup> C <sup>*</sup> A <sup>*</sup> A <sup>o</sup> A <sup>*</sup> C <sup>*</sup> A	6
X <sub>1-9</sub>	G <sup>*</sup> C <sup>*</sup> G <sup>*</sup> C <sup>*</sup> C <sup>*</sup> A <sup>*</sup> A <sup>*</sup> A <sup>*</sup> C <sup>*</sup> A	9

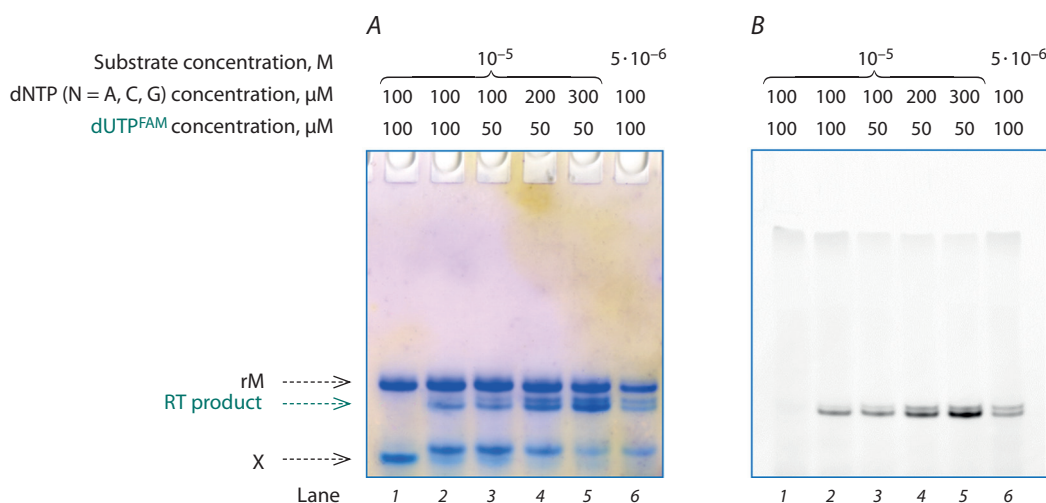
Note. "o": a native internucleotide phosphate group, *n*: the number of PG groups, which are denoted by "\*". In the proposed names, "X" is the general name of the series of 10-mer oligonucleotides, and the subscript designates position numbers (counted from the 5' end) of the PG groups. All decanucleotides were isolated and purified in a uniform manner (see the Materials and methods section).



**Fig. 1.** Structures of the modified phosphate group (i. e., the PG group, indicated by the asterisk) and native phosphate (indicated as "o") in a DNA strand. Base: a nitrogenous base.



**Fig. 2.** The scheme of elongation of an X series decanucleotide along the rM RNA template with the formation of a fluorescent product.



**Fig. 3.** Electrophoretic separation of the products of reverse-transcription reactions involving native substrate X-rM at 37 °C and 200 U of the enzyme at various concentrations of triphosphates and the substrate (i.e., a DNA–RNA complex, as shown in the figure) in a denaturing 15 % PAAG after visualization with the StainsAll reagent (A) and scanning of the fluorescence signal (B).

Arrows indicate the template (rM), a reverse-transcription reaction product, and the native decanucleotide (X).

the formed complex contained only one rA position in the single-stranded part, opposite to which it was possible to insert only one fluorescent dUMP<sup>FAM</sup> residue into the relatively long DNA chain using the DNA-dependent RNA polymerase under study: Moloney murine leukemia virus reverse transcriptase (MMLV H-).

Initially, optimal conditions of the reverse-transcription reaction were found using a native substrate (the X–rM complex) as an example. A series of experiments was conducted in which we varied (1) the reaction temperature, (2) the amount of the enzyme, or the concentration of (3) triphosphates or (4) the substrate DNA–RNA complex. The products of the reverse-transcription reaction were separated in a 15 % PAAG and visualized through scanning the fluorescence signal in the FAM (6-carboxyfluorescein) channel and then staining the same gel with the StainsAll reagent. It should be noted that staining the gel with StainsAll allows to examine the material balance of each reaction because this dye “reveals” all elongation products as well as the original template and primer. The fluorescent signal in turn only reflects the accumulation of the target products of the reverse-transcription reaction, with a length of at least 15 nucleotides, i.e., products containing the incorporated dUMP<sup>FAM</sup> residue. Figure 3 illustrates a typical result of a reverse-transcription reaction analysis.

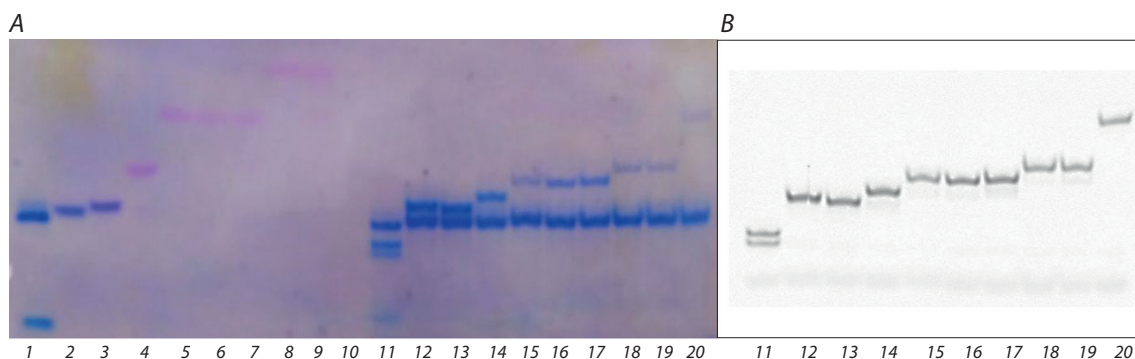
It is obvious that an increase in the concentration of fluorescent dUTP<sup>FAM</sup>, with all other things being equal (see Fig. 3, lanes 2 and 3), does not affect accumulation efficiency of elongation products of the native primer in the X–rM complex. On the contrary, when the concentration of unlabeled dNTP was increased at a constant concentration of dUTP<sup>FAM</sup> (see Fig. 3, A, lanes 3–5), a decrease in the proportion of abortive products of primer elongation was observed and, as a consequence, a higher yield of the fluorescent elongation product. The highest yield of the full-size reaction product (according to the fluorescence signal in Fig. 3, B) with the lowest amount of abortive elongation products is seen in lane 5.

It is these conditions that were chosen subsequently to compare the efficiency of the reverse-transcription reaction of native primers versus that of PG-containing primers, the backbone of which partially or completely lacked the negative charge. Figure 4 presents data from the electrophoretic analysis of the products of the reverse-transcription reaction involving complexes of rM with each decanucleotide as a substrate (the decanucleotides are described in Table 2).

According to both the results on the gel stained with StainsAll (see Fig. 4, A) and the fluorescence signal data (see Fig. 4, B), it can be argued that there was an enzymatic elongation of all the oligonucleotides, including completely electroneutral X<sub>1–9</sub>. It should be pointed out that the accumulation of a full-size fluorescent product slightly decreased with an increase in the number of PG groups in a primer (see Fig. 4, B, lanes 11–20). This result is probably due to disturbances of electrostatic interactions within the enzyme–substrate complex. Nevertheless, our data indicate that uncharged PGOs can serve as primers in a reverse-transcription reaction; to the best of our knowledge, this finding has never been reported regarding completely modified uncharged oligonucleotide analogs.

It was not possible to analyze in detail the degree of elongation of PGO primers in this model system (see Fig. 4, A, lanes 2–10 and 11–20) because a higher degree of modification of the oligos sharply reduces the efficiency of their staining with dyes specific to nucleic acids, for example, StainsAll (Dyudeeva et al., 2019). Taking this circumstance into account, native and PG-containing oligonucleotides carrying a fluorophore at the 5' end were designed for the next stage of the work (Table 3). These materials enabled subsequent detection of the entire spectrum of products of enzymatic elongation of primers on a relatively long template: a fragment of human 18S rRNA. The primer structures varied in the degree of modification of their sugar-phosphate backbone at the 3' end. Along with native primers (J1.o and J2.o), we used their partially uncharged analogs containing a backbone modified at the 5' end





**Fig. 4.** Electrophoretic separation of the products of reverse-transcription reactions in a denaturing 15 % PAAG after staining with the StainsAll reagent (A) and scanning the fluorescence signal (B).

Lanes 1–10 represent controls without the enzyme: mixtures of rM and primer X (lane 1) or primer X<sub>1,4;7</sub> (lane 2), X<sub>1,5;9</sub> (lane 3), X<sub>2,4;6;8</sub> (lane 4), X<sub>1–2,5;8–9</sub> (lane 5), X<sub>1,3;5;7;9</sub> (lane 6), X<sub>1–2,5–6;9</sub> (lane 7), X<sub>1–3;7–9</sub> (lane 8), X<sub>2–3,5–6;8–9</sub> (lane 9), X<sub>1–9</sub> (lane 10); lanes 11–20 represent reverse-transcription reactions involving primer X (lane 11), X<sub>1,4;7</sub> (lane 12), X<sub>1,5;9</sub> (lane 13), X<sub>2,4;6;8</sub> (lane 14), X<sub>1–2,5;8–9</sub> (lane 15), X<sub>1,3;5;7;9</sub> (lane 16), X<sub>1–2,5–6;9</sub> (lane 17), X<sub>1–3;7–9</sub> (lane 18), X<sub>2–3,5–6;8–9</sub> (lane 19), X<sub>1–9</sub> (lane 20).

and unmodified at the 3' end with six unmodified phosphate groups (J1.6 and J2.6) or three unmodified phosphate groups (J1.3 and J2.3) as well as completely electroneutral analogs (J1.x and J2.x, fully modified backbone).

The nucleotide sequences of these primers were complementary to some regions of the fragment of human 18S rRNA (the length of the fragment was 558 nucleotides) (Fig. 5).

The first of the two sites selected for hybridization – the J1 series in Fig. 5 – is located closer to the 3' end region of the fragment of rRNA. Through reverse transcription from this site, it is possible to obtain products, the theoretical length of which is 439 nucleotides. The second binding site of the J2 series primers is located closer to the middle of the rRNA fragment; theoretical size of the full-length elongation products of these primers is 276 nucleotides. When choosing binding sites for oligonucleotides in structured RNA templates, we employed the principles laid down by Petyuk et al. (1999).

PGOs, as previously demonstrated, can form a complementary complex in solutions with ultralow ionic strength (Kupryushkin et al., 2014; Dyudeeva et al., 2019); therefore, it was essential to compare the efficiency rates of the reverse-transcription reactions involving native versus PG-containing primers in buffers with different ionic strengths. For this purpose, potassium chloride was completely removed from the “standard” buffer, and the concentration of magnesium chloride was reduced from 5 to 2 mM. A decrease in ionic strength of the buffer may help to destabilize internal secondary structure of an RNA template, thereby “facilitating” the reverse-transcription reaction. Relatively long products of the reverse-transcription reaction were analyzed by electrophoresis in a 1.5 % agarose gel, with detection of the primer elongation products by means of the fluorescence signal and by staining with ethidium bromide. The results of electrophoretic separation of reverse-transcription reaction products obtained in buffers with “standard” and “low” ionic strength for primers of the J1 series are given in Fig. 6. Readers can see that the introduction of PG groups into the primer leads to an expected decrease in its electrophoretic mobility owing to a decrease in the total charge of the molecule (see Fig. 6, A, lanes 1, 4, 7, and 10). Enzymatic elongation of a primer

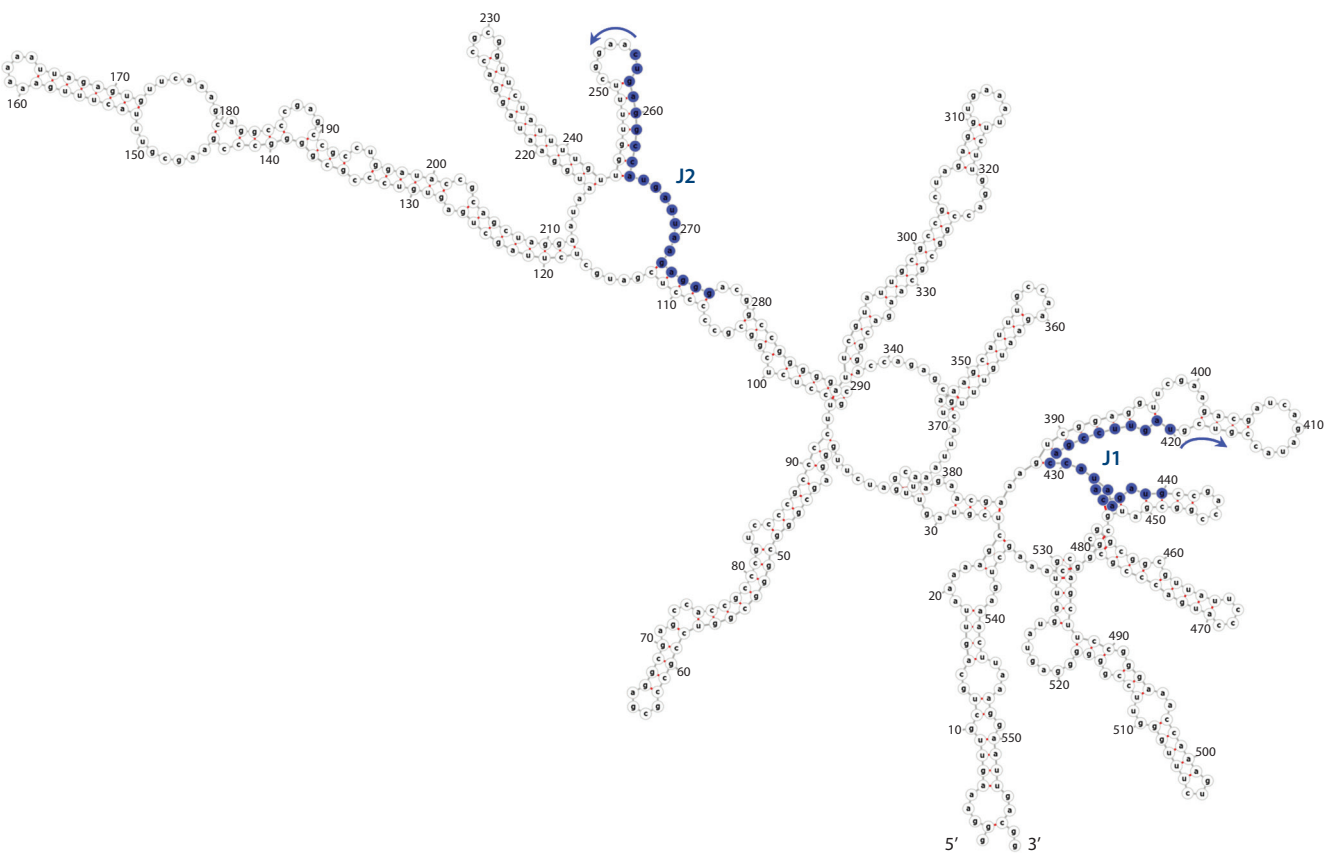
**Table 3.** Structures of primers for the reverse-transcription reaction involving a high-molecular-weight RNA template: a fragment of human 18S rRNA

Name	Sequence, 5'→3'
J1.o	[Flu] <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> G <sup>o</sup> A <sup>o</sup> A <sup>o</sup> C <sup>o</sup> T <sup>o</sup> A
J1.6	[Flu] <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> G <sup>o</sup> A <sup>o</sup> A <sup>o</sup> C <sup>o</sup> T <sup>o</sup> A
J1.3	[Flu] <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> G <sup>o</sup> A <sup>o</sup> A <sup>o</sup> C <sup>o</sup> T <sup>o</sup> A
J1.x	[Flu] <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> T <sup>o</sup> C <sup>o</sup> G <sup>o</sup> G <sup>o</sup> A <sup>o</sup> A <sup>o</sup> C <sup>o</sup> T <sup>o</sup> A
J2.o	[Flu] <sup>o</sup> C <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> G <sup>o</sup> T
J2.6	[Flu] <sup>o</sup> C <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> G <sup>o</sup> T
J2.3	[Flu] <sup>o</sup> C <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> G <sup>o</sup> T
J2.x	[Flu] <sup>o</sup> C <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> T <sup>o</sup> T <sup>o</sup> A <sup>o</sup> A <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> T <sup>o</sup> G <sup>o</sup> G <sup>o</sup> C <sup>o</sup> C <sup>o</sup> T <sup>o</sup> C <sup>o</sup> A <sup>o</sup> G <sup>o</sup> T

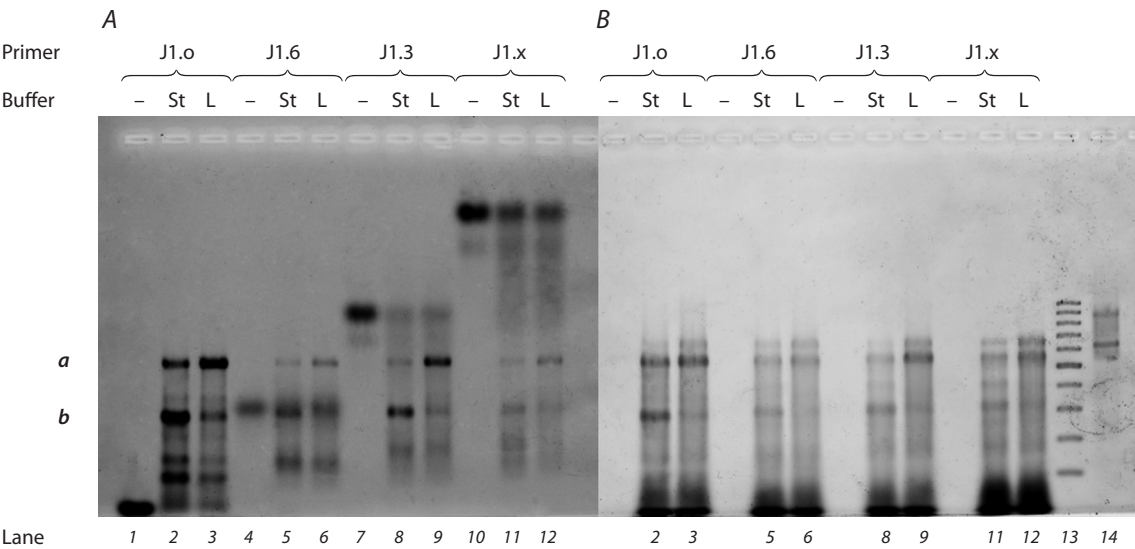
Note. “o”: a native internucleotide phosphate group, “x”: a PG group, [Flu]: a fluorescein residue.

was observed in all cases, even for primer J1.x, which has a completely electroneutral sugar-phosphate backbone (see Fig. 6, A, lanes 11 and 12).

Using native primer J1.o as an example, let us examine the reaction mixtures obtained via the reverse transcription overall (see Fig. 6, lanes 1–3). Original primer J1.o (lane 1) is virtually absent among the products of its reverse-transcription reaction (lanes 2 and 3), indicating an efficient course of the reaction. Nevertheless, the products of primer elongation vary in their electrophoretic mobility and hence in length: abortive elongation products are present. This finding is probably due to structural complexity of the RNA template because reverse transcriptase, lacking a helicase activity, is unable to overcome RNA hairpins or other secondary structures, which can lead to dissociation of the enzyme from the substrate (DNA–RNA complex) (Okano et al., 2017). As a consequence, expected processivity of the reverse-transcription reaction is lower. In the examined reaction mixtures (see Fig. 6, lanes 2 and 3), two prevailing products (a and b) can be distinguished.



**Fig. 5.** Hypothetical secondary structure of the fragment of human 18S rRNA, as built using an online service, RNAfold web server. J1 and J2: sites selected for primer hybridization. Arrows indicate the direction of primer elongation.



**Fig. 6.** Data on the scanning of fluorescence signals of FAM (A) or ethidium bromide (B) after separation of the reverse-transcriptional reaction products in a 1.5 % agarose gel:  
individual DNA primers J1.o (lane 1), J1.6 (lane 4), J1.3 (lane 7), and J1.x (lane 10), which are visible only when FAM fluorescence is scanned; products of the reverse-transcription reaction involving primer J1.o (lane 2), J1.6 (lane 5), J1.3 (lane 8), or J1.x (lane 11) under “standard” buffering conditions (St); products of the reverse-transcription reaction involving primer J1.o (lane 3), J1.6 (lane 6), J1.3 (lane 9), or J1.x (lane 12) in the buffer with low ionic strength (L); DNA molecular-weight markers 100–1000 bp (lane 13), which are visible only when the fluorescence of ethidium bromide is scanned; the RNA template (lane 14), which is visible only when the fluorescence of ethidium bromide is scanned.  
The types of primer and buffer are indicated in the figure; *a* and *b* correspond to positions of the prevalent products of the reverse-transcription reaction.

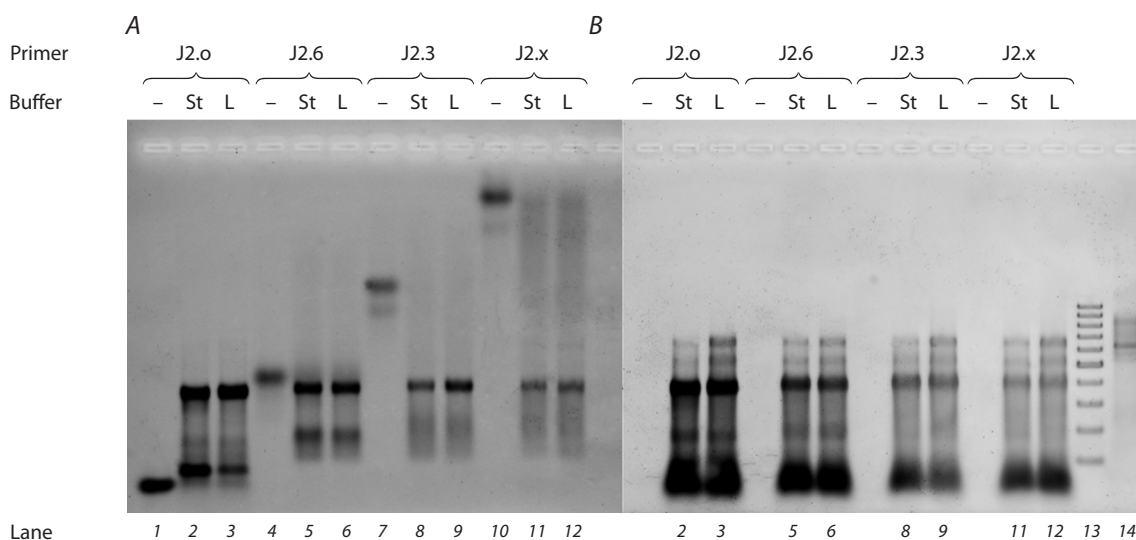
Electrophoretic mobility of product **a** is comparable to that of the initial RNA template (see Fig. 6, *B*, lane 14), which is 558 nucleotides long, and to the mobility of 500 and 600 bp bands among the DNA molecular-weight markers (lane 13). It can be assumed that this product is close to the theoretical full-length one. Product **b** has lower electrophoretic mobility approximately corresponding to 300 bp among the double-stranded-DNA molecular-weight markers. Even the introduction of an additional step of denaturation of rRNA at 95 °C to reduce the likelihood of structural complexity did not raise the yield of product **a** (data not shown).

Notably, electrophoretic mobility of initial primers J1.6, J1.3, and J1.x (see Fig. 6, *A*, lanes 4, 7, and 10) is lower than that of the products of their elongation by the reverse-transcription reaction (lanes 5–6, 8–9, and 11–12). Apparently, at initial stages of the reaction, the inclusion of dNTPs into the growing chain of the PGO primer causes an increase in the migration rate of the oligonucleotide in the gel owing to an appreciable change in the charge; the size of the molecule has not yet changed so much as to cause product retardation in agarose gel pores. Nevertheless, when a certain critical length is reached at later stages of the reaction, the total molecular size has a greater effect on the migration rate in the gel than does the increase in charge when each subsequent dNTP is incorporated. Thus, relatively long products of the reverse-transcription reaction involving PG-containing primers manifested “classic” behavior in the gel, by slowing down with increasing molecular size. In this context, the influence of the relatively small number of PG groups becomes negligible. For this reason, it is likely that electrophoretic mobility of the major products of the reverse-transcription reaction (**a** and **b**) is almost independent from the number of modifications in the initial PG-containing primer.

A decrease in ionic strength of reaction buffer (see Fig. 6, lanes 3, 6, 9, and 12) leads to the accumulation of longer products in the reverse-transcription reaction as compared to standard reaction conditions (lanes 2, 5, 8, and 11) regardless of the presence of modifications in the primer. Ionic strength caused the most pronounced change in the ratio of elongation products for native primer J1.o: according to fluorescence intensity levels, the proportion (%) of the longest product **a** increased threefold (from 9 to 27 %), while the proportion of abortive product **b** diminished (from 26 up to 10 %) under the influence of the decrease in buffer ionic strength. Perhaps the lower ionic strength of the buffer destabilized the internal structure of RNA but did not have a significant impact on the formation of the DNA–RNA–enzyme triple complex. For PG-containing primers, the redistribution of elongation products upon buffer change was observed too but at a level of only several percentage points.

It should be mentioned that for completely electroneutral PGO J1.x (see Fig. 6, *A*, lanes 10–12), we noticed less active elongation in the reverse-transcription reaction in comparison with the native primer (lanes 1–3). It is possible that the rate-limiting stage of the reverse-transcription reaction for the completely electroneutral PGO is the attachment of the first dNMPs to its 3' end (consistent with the findings of Skasko et al., 2005). After that, the enzyme focuses on the “preferred” substrate (which contains native unmodified nucleotides at the 3' end) and “ignores” even an excessive amount of the original primer.

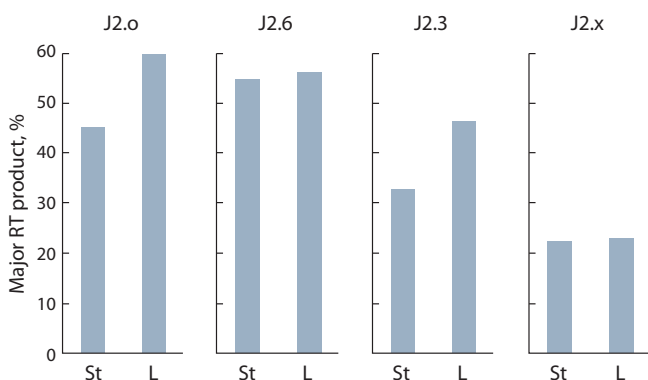
Similar results were obtained for primers of the J2 series (Fig. 7). In this case, fewer abortive products are observed among the primer elongation products. Most likely, the reason for this is that the theoretical maximal length of elongation products is shorter than that in the case of primer J1, and



**Fig. 7.** Data on the scanning of the fluorescence signal of FAM (*A*) or ethidium bromide (*B*) after separation of the reverse-transcription reaction products in a 1.5 % agarose gel:

for individual DNA primers J2.o (lane 1), J2.6 (lane 4), J2.3 (lane 7), and J2.x (lane 10), which are visible only when FAM fluorescence is scanned; for products of the reverse-transcription reaction of primer J2.o (lane 2), J2.6 (lane 5), J2.3 (lane 8), or J2.x (lane 11) under the “standard” buffering conditions (St); for products of the reverse-transcription reaction of primer J2.o (lane 3), J2.6 (lane 6), J2.3 (lane 9), or J2.x (lane 12) in the buffer with low ionic strength (L); DNA molecular-weight markers 100–1000 bp (lane 13), which are visible only when ethidium bromide fluorescence is scanned; the RNA template (lane 14), which is visible only when the fluorescence of ethidium bromide is scanned.

The types of primer and buffer are indicated in the figure.



**Fig. 8.** The proportion of the main product of elongation for primers of the J2 series in the reverse-transcription reaction carried out in a solution with "standard" (St) and low (L) ionic strength.

therefore during elongation, fewer hairpins or other structural complexities of the RNA template have to be overcome, and this situation makes the enzyme more processive.

The main elongation product of the J2 series primers has an electrophoretic mobility comparable to that of a 400 bp double-stranded-DNA molecular-weight marker. Presumably, this product has a size similar to the theoretical full length (276 nucleotides for this primer). Its proportion (%) in the reaction mixtures is depicted in Fig. 8. As was the case for primers of the J1 series, we noticed greater accumulation of the target product in the buffer with lower ionic strength (L) for native primer J2.o (an increase in the proportion of the main product from 45 to 61 %) and for PGO primer J2.3 (an increase from 33 to 46 %, respectively).

Notably, other PGO primers turned out to be virtually indifferent to the changes in ionic strength. It is conceivable that the arrangement of PG groups in the active center of the enzyme affects the conformation of the enzyme–substrate complex and therefore its sensitivity to ionic composition of the medium and to solvation.

## Conclusion

Thus, we characterized the elongation of partially and completely uncharged PGOs in the reverse-transcription reaction carried out by MMLV H<sup>-</sup>. It was demonstrated that PG-containing primers can be elongated effectively, e.g., in the presence of a fragment of human rRNA with complicated spatial structure. A decrease in buffer ionic strength raises the efficiency of full-length cDNA synthesis when either a native or PG-containing primer is used. The findings require further research and point to the suitability of PGOs, including completely electroneutral ones, as primers for RT-PCR, thereby opening up new prospects for the creation of experimental models for the analysis of highly structured RNA.

## References

- Bustin S.A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 2000;25(2):169-193. DOI 10.1677/jme.0.0250169.
- Chubarov A.S., Oscorbin I.P., Filipenko M.L., Lomzov A.A., Pyshnyi D.V. Allele-specific PCR for *KRAS* mutation detection using

- phosphoryl guanidine modified primers. *Diagnostics (Basel)*. 2020; 10(11):872. DOI 10.3390/diagnostics10110872.
- Dyudeeva E.S., Kupryushkin M.S., Lomzov A.A., Pyshnaya I.A., Pyshnyi D.V. Physicochemical properties of the phosphoryl guanidine oligodeoxyribonucleotide analogs. *Russ. J. Bioorg. Chem.* 2019;45(6):709-718. DOI 10.1134/S1068162019060153.
- Haas B.J., Zody M.C. Advancing RNA-Seq analysis. *Nat. Biotechnol.* 2010;28(5):421-423. DOI 10.1038/nbt0510-421.
- Heuverswyn F.V., Karczmarczyk M., Schimmel H., Trapmann S., Emons H. Influence of primer & probe chemistry and amplification target on reverse transcription digital PCR quantification of viral RNA. *Biomol. Detect. Quantif.* 2016;9:20-28. DOI 10.1016/j.bdq.2016.08.003.
- Kelleher C.D., Champoux J.J. Characterization of RNA strand displacement synthesis by Moloney murine leukemia virus reverse transcriptase. *J. Biol. Chem.* 1998;273(16):9976-9986. DOI 10.1074/jbc.273.16.9976.
- Kupryushkin M.S., Pyshnaya I.A., Dmitrienko E.V., Stetsenko D.A., Filipenko M.L., Oscorbin I.P., Stepanov G.A., Richter V.A., Ivanov M.K., Pyshnyi D.V. Template-directed enzymatic DNA synthesis using phosphoryl guanidine oligonucleotides. 2017. Publication Number WO/2019/112485. Appl. Date 17.09.2018. Publ. Date 13.06.2019.
- Kupryushkin M.S., Pyshnyi D.V., Stetsenko D.A. Phosphoryl guanidines: a new type of nucleic acid analogues. *Acta Naturae*. 2014; 6(4):116-118.
- Li R., Wu R., Li H., Zhang Y., Peng D., Wang N., Shen X., Wang Z., Sun H. Characterizing stutter variants in forensic STRs with massively parallel sequencing. *Forensic Sci. Int. Genet.* 2020;45: 102225. DOI 10.1016/j.fsigen.2019.102225.
- Li T.-L., Wu M.-W., Lin W.-C., Lai C.-H., Chang Y.-H., Su L.-J., Chen W.-Y. Designed phosphate-methylated oligonucleotides as PCR primers for SNP discrimination. *Anal. Bioanal. Chem.* 2019; 411:3871-3880. DOI 10.1007/s00216-019-01865-4.
- Lomzov A.A., Kupryushkin M.S., Shernyukov A.V., Nekrasov M.D., Dovidenko I.S., Stetsenko D.A., Pyshnyi D.V. Diastereomers of a mono-substituted phosphoryl guanidine trideoxyribonucleotide: isolation and properties. *Biochem. Biophys. Res. Commun.* 2019; 513(4):807-811. DOI 10.1016/j.bbrc.2019.04.024.
- Malik O., Khamis H., Rudnizky S., Kaplan A. The mechano-chemistry of a monomeric reverse transcriptase. *Nucleic Acids Res.* 2017; 45(22):12954-12962. DOI 10.1093/nar/gkx1168.
- Menéndez-Arias L., Sebastián-Martín A., Álvarez M. Viral reverse transcriptases. *Virus Res.* 2016;234:153-176. DOI 10.1016/j.virusres. 2016.12.019.
- Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B. Mapping and quantifying mammalian transcriptome by RNA-Seq. *Nat. Methods*. 2008;5(7):621-628. DOI 10.1038/nmeth.1226.
- Ohtsubo Y., Nagata Y., Tsuda M. Efficient N-tailing of blunt DNA ends by Moloney murine leukemia virus reverse transcriptase. *Sci. Rep.* 2017;7:41759. DOI 10.1038/srep41769.
- Okano H., Katano Y., Baba M., Fujiwara A., Hidese R., Fujiwara S., Yanagihara I., Hayashi T., Kojima K., Takita T., Yasukawa K. Enhanced detection of RNA by MMLV reverse transcriptase coupled with thermostable DNA polymerase and DNA/RNA helicase. *Enzyme Microb. Technol.* 2017;96:111-120. DOI 10.1016/j.enzmictec. 2016.10.003.
- Palikša S., Alzbutas G., Skirgaila R. Decreased  $K_m$  to dNTPs in an essential M-MuLV reverse transcriptase adoption requires to perform efficient cDNA synthesis in One-Step RT-PCR assay. *Protein Eng. Des. Sel.* 2018;31(3):79-89. DOI 10.1093/protein/gzy003.
- Petyuk V.A., Zenkova M.A., Giege R., Vlassov V.V. Hybridization of antisense oligonucleotides with the 3' part of tRNA<sup>Phe</sup>. *FEBS Lett.* 1999;444(2-3):217-221. DOI 10.1016/S0014-5793(99)00063-0.



- Schultz S.J., Champoux J.J. RNase H activity: structure, specificity, and function in reverse transcription. *Virus Res.* 2008;134(1-2):86-103. DOI 10.1016/j.virusres.2007.12.007.
- Skasko M., Weiss K.K., Reynolds H.M., Jamburuthugoda V., Lee K., Kim B. Mechanistic differences in RNA-dependent DNA polymerization and fidelity between murine leukemia virus and HIV-1 reverse transcriptases. *J. Biol. Chem.* 2005;280(13):12190-12200. DOI 10.1074/jbc.M412859200.
- Wulf M.G., Maguire S., Humbert P., Dai N., Bei Y., Nichols N.M., Corrêa I.R., Jr, Guan S. Non-templated addition and template switching by Moloney murine leukemia virus (MMLV)-based reverse transcriptases co-occur and compete with each other. *J. Biol. Chem.* 2019;294:18220-18231. DOI 10.1074/jbc.RA119.010676.

---

**ORCID ID**

I.A. Pyshnaya orcid.org/0000-0002-7559-2376


**Acknowledgements.** This research was funded by Russian Foundation for Basic Research No. 19-34-90132.

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

Received September 1, 2021. Revised October 7, 2021. Accepted October 8, 2021.


Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

# Cancer-associated fibroblasts and their role in tumor progression

M.S. Ermakov<sup>1</sup>, A.A. Nushtaeva<sup>1</sup>, V.A. Richter<sup>1</sup>, O.A. Koval<sup>1, 2</sup> 

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

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


 o\_koval@ngs.ru

**Abstract.** The stromal elements of a malignant tumor can promote cancer progression and metastasis. The structure of the tumor stroma includes connective tissue elements, blood vessels, nerves, and extracellular matrix (ECM). Some of the cellular elements of the tumor stroma are cancer-associated fibroblasts (CAFs). The origin and function of CAFs have been actively studied over the past thirty years. CAFs produce collagen, the main scaffold protein of the extracellular matrix. Collagen in the tumor stroma stimulates fibrosis, enhances the rigidity of tumor tissue, and disrupts the transmission of proliferation and differentiation signaling pathways. CAFs control tumor angiogenesis, cell motility, tumor immunogenic properties, and the development of resistance to chemo- and immunotherapy. As a result of metabolic adaptation of rapidly growing tumor tissue to the nutrients and oxygen deprivation, the main type of energy production in cells changes from oxidative phosphorylation to anaerobic glycolysis. These changes lead to sequential molecular alterations, including the induction of specified transcriptional factors that result in the CAFs activation. The molecular phenotype of activated CAFs is similar to fibroblasts activated during inflammation. In activated CAFs, alpha-smooth muscle actin ( $\alpha$ -SMA) is synthesized *de novo* and various proteases and fibronectin are produced. Since CAFs are found in all types of carcinomas, these cells are potential targets for the development of new approaches for anticancer therapy. Some CAFs originate from resident fibroblasts of the organs invaded by the tumor, while others originate from epithelial tumor cells, which are undergoing an epithelial-mesenchymal transition (EMT). To date, many molecular and metabolic inducers of the EMT have been discovered including the transforming growth factor-beta (TGF- $\beta$ ), hypoxia, and inflammation. This review classifies modern concepts of molecular markers of CAFs, their functional features, and discusses the stages of epithelial-mesenchymal transition, and the potential of CAFs as a target for antitumor therapy.

Key words: cancer-associated fibroblasts; epithelial-to-mesenchymal transition; carcinoma; hypoxia.


**For citation:** Ermakov M.S., Nushtaeva A.A., Richter V.A., Koval O.A. Cancer-associated fibroblasts and their role in tumor progression. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):14-21. DOI 10.18699/VJGB-22-03

# Опухоль-ассоциированные фибробласты и их роль в опухолевой прогрессии

М.С. Ермаков<sup>1</sup>, А.А. Нуштаева<sup>1</sup>, В.А. Рихтер<sup>1</sup>, О.А. Коваль<sup>1, 2</sup> 

<sup>1</sup> Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия

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

 o\_koval@ngs.ru

**Аннотация.** Стромальные элементы опухоли могут стимулировать прогрессию опухолевого роста и метастазирование. В структуру опухолевой стромы входят соединительнотканые элементы, сосуды, нервы и внеклеточный матрикс. Одним из клеточных элементов стромы опухоли являются опухоль-ассоциированные фибробласты (ОАФ), происхождение и функции которых активно изучают на протяжении последних тридцати лет. ОАФ продуцируют основной каркасный белок внеклеточного матрикса – коллаген, избыток которого в строме опухоли стимулирует фиброз, повышает жесткость опухолевой ткани и нарушает передачу сигналов пролиферации и дифференцировки. ОАФ контролируют ангиогенез в опухоли, подвижность опухолевых клеток, иммуногенные свойства опухоли и развитие резистентности к химиопрепаратам и иммунотерапии. В результате метаболической адаптации быстрорастущей опухолевой ткани к нехватке питательных веществ и кислорода, при конверсии основного типа производства энергии в клетках с окислительного фосфорилирования на анаэробный гликолиз инициируются молекулярные изменения, в том числе индукция определенных транскрипционных факторов, которые обуславливают активацию ОАФ. Молекулярный фенотип активированных ОАФ имеет сходство с фибробластами, активированными в процессе воспаления. В активированных ОАФ происходит *de novo* синтез альфа-актина гладкой мускулатуры ( $\alpha$ -SMA), продукция

различных протеаз и фибронектина. Поскольку ОАФ обнаружены во всех типах карцином, эти клетки могут быть мишенями для разработки новых подходов противоопухолевой терапии. Часть ОАФ происходит из резидентных фибробластов пораженного опухолью органа, другие берут свое начало из опухолевых эпителиоцитов, претерпевая эпителиально-мезенхимальный переход. На сегодняшний день обнаружено большое количество молекулярных и метаболических индукторов эпителиально-мезенхимального перехода, к которым относят трансформирующий фактор роста бета (TGF- $\beta$ ), гипоксию и воспаление. В данном обзоре систематизированы современные представления о молекулярных маркерах ОАФ, их функциональных особенностях, этапах эпителиально-мезенхимального перехода и обсуждается возможность использования ОАФ в качестве мишеней для противоопухолевой терапии.

Ключевые слова: опухоль-ассоциированные фибробласты; эпителиально-мезенхимальный переход; карцинома; гипоксия.

## The biology of cancer-associated fibroblasts

Modern concept of tumor morphology postulates that solid tumors are formed by epithelial and stromal cells, such as fibroblasts, endothelial cells, and immune cells (Wang et al., 2017). Stromal cells with a fibroblast-like phenotype, the so-called cancer-associated fibroblasts (CAFs), in contrast to normal fibroblasts, contain various chromosomal abnormalities, such as duplications, multiple rearrangements, and even the loss of entire chromosomes (Hosein et al., 2010). CAFs control tumor angiogenesis, motility and metastasis of cancer cells, tumor immunogenic properties, and the development of resistance to chemotherapy and immunotherapy (Tripathi et al., 2012; Alkasalias et al., 2018; Nushtaeva et al., 2018).

A meta-analysis of the clinical relevance of the tumor stroma has demonstrated the association of high CAFs content with the advanced stages of tumor progression, as well as with the high risk of local recurrence after tumor resection (Knops et al., 2020).

## Heterogeneity of CAFs' precursors cells

In 1995, a heterogeneous origin of CAFs was hypothesized by Rønnov-Jessen and colleagues, who showed that breast cancer CAFs can originate from resident fibroblasts, vascular smooth muscle cells, and pericytes (Rønnov-Jessen et al., 1995). To date, it has been shown that precursors of mesenchymal cells from the red bone marrow, endothelial and epithelial cells, resident fibroblasts of the affected tissue, adipocytes and vascular adventitia cells can be sources of CAFs (Puré, Hingorani, 2018; Yin et al., 2019). For the initiation of the CAFs phenotype in some progenitor cells, additional stimulation with cytokines and growth factors, such as transforming growth factor beta (TGF- $\beta$ ), fibroblast growth factor (FGF), and other signaling molecules is required (Table 1) (Bordignon et al., 2019).

Tumor epithelial cells can undergo transformation into CAFs via the epithelial-mesenchymal transition (EMT) (Fig. 1). EMT is a dynamic process of transdifferentiation of epithelial cells into fibroblast-like cells. The EMT plays an important role not only in cancer, but also in embryogenesis and regeneration. In particular, EMT occurs in embryonic stem cells producing mesoderm and neural crest, and in skin cells during wound healing (Kim et al., 2017). Dynamic changes in cell morphology during EMT are caused by changes in the regulatory genes' expression with production of certain proteins. These proteins are considered as EMT markers (see Fig. 1). Among these markers, the most significant are N-cadherin and vimentin, which are responsible for the rear-

range of the cytoskeleton and the change in the shape of the cell, as well as the change in cell-to-cell and cell-to-extracellular matrix (ECM) interactions (Massagué, 2008; Ye, Weinberg, 2015).

In addition to molecular inducers of EMT, the important role of hypoxic conditions has been shown. Hypoxia activates EMT via the binding of the hypoxia-inducible factor (HIF-1) to the promoters of genes responsible for EMT activation. HIF-1 has been shown to increase the expression of the transcription factors genes of the zinc finger motif family such as ZEB1, Snail and SLUG. Overexpression of these factors is associated with the mesenchymal phenotype and a decrease in the abundance of epithelial cell markers – E-cadherin and type 1 tight junction protein (TJP1 or ZO-1) (Nushtaeva et al., 2019; Tam et al., 2020).

Endothelial cells of tumor vessels can undergo an endothelial-mesenchymal transition (EndMT) and acquire the phenotype and functional features of CAFs with the loss of endothelial cells molecular markers, such as the endothelial cell/platelet adhesion molecule (CD31), and the acquisition of markers specific for mesenchymal cells, such as  $\alpha$ -SMA and fibroblast specific protein 1 (FSP-1) (Zeisberg et al., 2007).

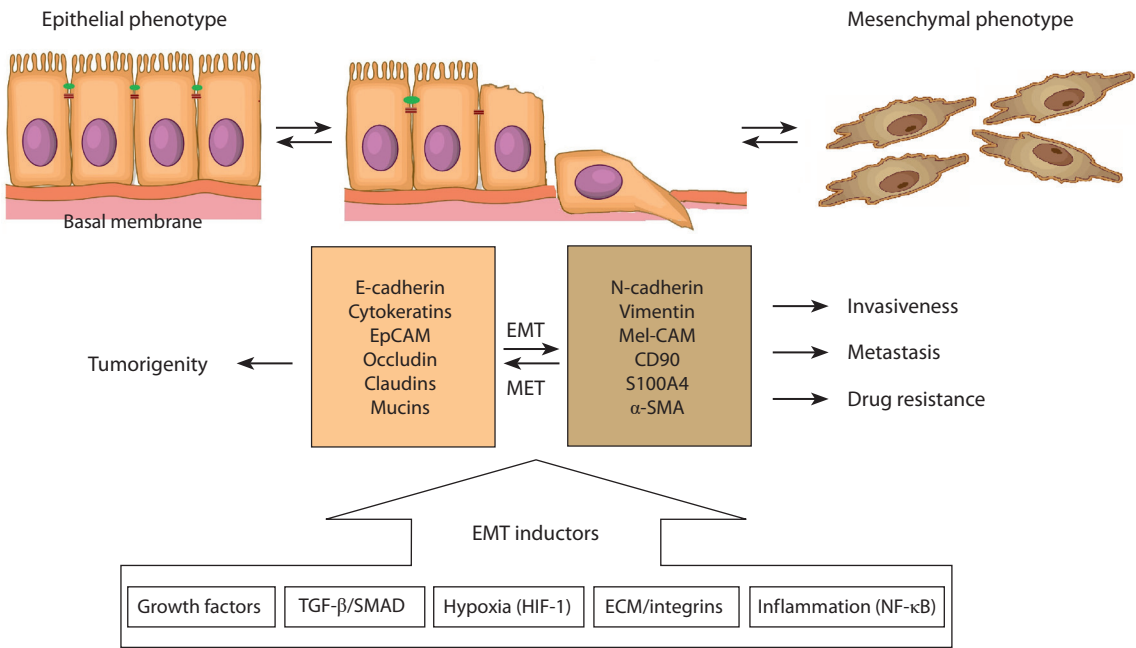
An important component of breast cancer stroma are adipose cells, which can transform into tumor-associated adipocytes, and then into CAFs. Such changes are accompanied by an increase in the expression of molecular markers of mesenchymal cells, including, PPARG (receptors induced by peroxisome activators gamma), RUNX-2 (transcription factor containing the Runt type 2 DNA-binding domain), and SOX9 (transcription factor of the HMG family DNA-binding proteins) (Bochet et al., 2013; Liu et al., 2021).

Using the model of prostate cancer, it was shown that mesenchymal stem cells (MSCs) can differentiate into CAFs after the activation of the chemokine receptor type 6 (CXCR6) by its ligand CXCL16. Moreover, the activation of CXCR6 results in the secretion of stromal factor-1 (CXCL12) involved in EMT (Jung et al., 2013). Weber and colleagues also showed that the extracellular structural protein osteopontin (OPN), which plays a key role in bone formation, activates TGF- $\beta$  gene expression in integrin-dependent MSCs to maintain the phenotype of CAFs in breast cancer. Interestingly, even specialized cells such as Ito cells in the liver, pancreatic stellate cells, and mammary myofibroblasts can acquire the phenotype of CAFs (Weber et al., 2015). These examples illustrate a wide range of cells that, responding to the molecular changes in a tumor, are able to acquire the CAFs' phenotype and, as a consequence, be involved in tumor homeostasis.

**Table 1.** The inducers of CAFs phenotype

Inductor	Activation pathway	References
TGF-β	CLIC4/Arf6	Calon et al., 2014
FGF	RAS/MAPK, PI3k/AKT	Bordignon et al., 2019; Mossahebi-Mohammadi et al., 2020
PDGF	PI3k/AKT	Yoshida, 2020
HSF-1	TRiC/CCT	Scherz-Shouval et al., 2014; Grantham, 2020

Note. PDGF – platelet derived growth factor; HSF-1 – heat shock factor 1; CLIC4/Arf6 – intracellular chloride channel 4, ADP-ribosylation factor 6; RAS/MAPK – mitogen activated protein kinase; PI3k/AKT – phosphoinositide 3-kinase, alpha serine/threonine protein kinase; TRiC/CCT is a chaperonin containing TCP-1.



**Fig. 1.** Cellular and molecular changes during EMF and MET.  
EpCAM – epithelial adhesion molecule; Mel-CAM – melanoma adhesion molecule; α-SMA – alpha smooth muscle actin; HIF-1 – hypoxia inducible factor 1.

**Markers of cancer-associated fibroblasts**

The involvement of CAFs in carcinogenesis and tumor progression makes them a potential target for the development of novel therapeutic approaches. A potentially clinically significant marker of CAF is the transmembrane mucin-like protein podoplanin (PDPN) (Table 2); to date, PDPN has been described as a marker of lymphoid capillary progenitor cells and CAFs in lung cancers. Expression of podoplanin was showed in 54 (30.5 %) out of 177 CAFs’ populations studied in the work of Yurugi et al. Interestingly, all podoplanin-positive CAFs correlated with invasiveness of adenocarcinomas, while a podoplanin-negative phenotype was shown only in non-invasive adenocarcinomas (Yurugi et al., 2017).

Platelet-derived growth factor receptors α/β are important markers of CAFs. PDGFRα/β belong to the 3rd class of tyrosine kinases and are activated by interaction with the PDGF ligand. PDGFR regulates the organogenesis of various systems during embryogenesis; however, the significance

of the PDGFRα and -β receptors activation in tumors is still poorly understood. It has been shown that the expression of the PDGFRβ receptor is increased in the tumor microenvironment cells, where platelet growth factor activates CAFs and, probably, stimulates cancer progression (Anderberg et al., 2009). PDGFRα-positive CAFs have been found in the stroma of melanoma, suggesting that these CAFs originate from resident fibroblasts as a result of their activation (Lynch, Watt, 2018). Serum amyloid A (SAA-1) protein is one of the potential targets of CAFs; its expression and involvement in tumor progression has been shown in CAFs from gastric tumors (Yasukawa et al., 2021).

In the search for specific markers of the tumor stroma cells, among the CAFs of prostate adenocarcinoma, an increased content of the surface protein with a single V-domain of immunoglobulin (CD90), initially found on T cells and neurons, was identified as a specific marker. The high level of CD90 on the cell surface differentiates the tumor-associated stroma and



**Table 2.** Potential molecular markers of CAFs

CAFs' markers	Presence in tumors	References
$\alpha$ -SMA	Breast cancer	Yang et al., 2021
SAA-1	Stomach cancer	Yasukawa et al., 2021
Osteopontin	Breast cancer	Weber et al., 2015
Caveolin-1	Breast cancer: luminal (overexpression), basal (poor expression)	Witkiewicz et al., 2009; Goetz et al., 2011
PDGFR $\alpha/\beta$	Breast cancer, melanoma	Jansson et al., 2018; Lynch, Watt, 2018
CD90	Prostate cancer, breast cancer, lung cancer	True et al., 2010; Lobba et al., 2018
Podoplanin	Lung cancer	Yurugi et al., 2017
Metalloprotease	Head-and-neck cancer	Glentis et al., 2017
S100A4	Breast cancer	Grum-Schwensen et al., 2005

Note.  $\alpha$ -SMA – alpha-smooth muscle actin; PDGFR $\alpha/\beta$  – platelet-derived growth factor receptors  $\alpha/\beta$ .

“benign” stroma. Since CD90 expression was shown only in tumor associated fibroblasts, this marker is a potential target for therapy (True et al., 2010).

Certain CAFs proteins can be prognostic markers of tumor invasiveness. One of these markers is a protein from the family of low molecular weight calcium-binding proteins of the S100 – S100A4 family (Fei et al., 2017). S100 family proteins have both intracellular and extracellular activity due to maintaining the calcium balance and  $\text{Ca}^{2+}$ -dependent processes. S100A4 activates a cascade of reactions associated mainly with the secretion of pro-inflammatory cytokines and the expression of growth factors, extracellular matrix proteins, metalloproteinases, and others. Intracellular activity of S100A4 is of particular interest, and is associated with the enhancement of the invasive capabilities of tumor cells, their escape from apoptosis, and the stem phenotype of the cells (Ambartsumian et al., 2019). During the study of the role of S100A4 in tumor progression, it was shown that suppression of S100A4 decreased tumor growth (Joyce, Pollard, 2009; Grum-Schwensen et al., 2015). The role of stromal cells that secrete S100A4 was shown in the MMTV-PyVmT mouse model with the S100A4 knocked-out gene during orthotopic co-transplantation of CSML100 mouse mammary adenocarcinoma cells and MEF mouse embryonic fibroblasts. MEF cell lines were obtained by spontaneous immortalization of primary embryonic fibroblasts from mouse embryos with the S100A4<sup>+</sup> and S100A4<sup>-</sup> phenotypes. Upon co-transplantation of tumor cells and fibroblasts with the S100A4<sup>-</sup> phenotype in syngeneic mice, no metastases were formed, however, upon transplantation of S100A4<sup>+</sup> fibroblasts, the metastatic potential of tumor cells returned. S100A4<sup>+</sup> fibroblasts were characterized by increased mobility and invasiveness compared to S100A4<sup>-</sup> fibroblasts, as well as the ability to secrete S100A4 into the tumor microenvironment (Grum-Schwensen et al., 2005).

To date, it is clear that the expression of certain CAFs markers does not unambiguously predict the aggressiveness of the tumor. For example, the loss of caveolin-1 in breast cancer CAFs has been shown to be associated with a poor prognosis because the population of these cells stimulates the

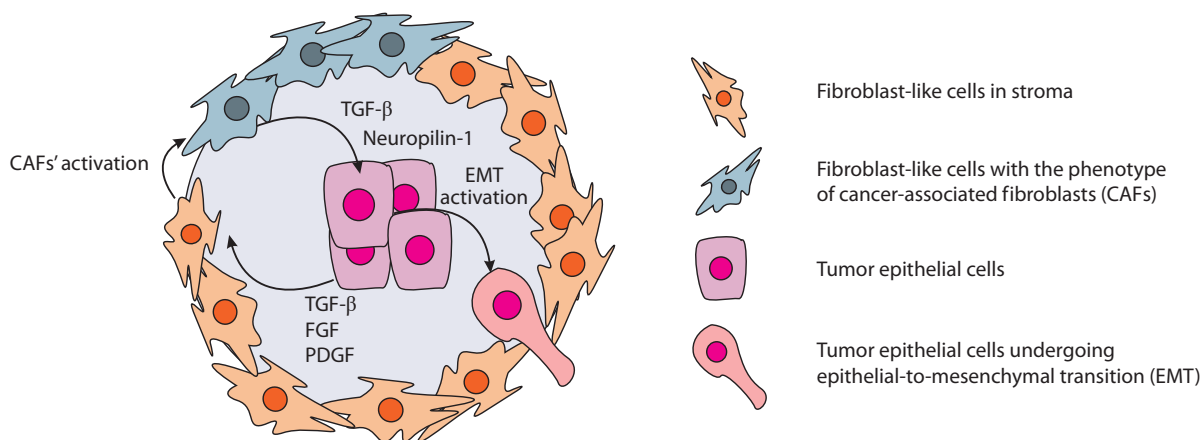
growth of triple negative (ER-/PR-/HER2-) breast cancer cells (Witkiewicz et al., 2009). In a parallel study, the expression of caveolin-1 in the breast cancer CAFs stimulated the remodeling of the tumor microenvironment, thereby facilitating the invasion of malignant cells and an increased invasiveness level correlated with the metastatic potential of the tumor (Witkiewicz et al., 2009; Goetz et al., 2011). These contradicting results indicate the diverse role of caveolin-1 in histologically different tumors. More studies should be made to determine caveolin-1 as a tumor prognostic marker.

### Role of cancer-associated fibroblasts in tumor progression

CAFs-dependent stimulation of the tumor cells proliferation and their invasion is of particular interest in the study of tumor. This interest is primarily due to the fact that even in the precancerous phenotype of epithelial cells, some resident fibroblasts are already transformed into CAFs (Liotta, Kohn, 2001). Fibroblasts from intestinal tumors and polyps were a good model to confirm the contribution of stromal cells to tumor growth and progression. These fibroblasts were shown to stimulate the proliferation of tumor and polyp cells (Mukaida, Sasaki, 2016).

The interaction of tumor epithelial cells with CAFs was analyzed by comparing the histological picture of various types of gastric cancer. In a study by Orimo and Weinberg, it was demonstrated that in the case of diffuse gastric cancer, CAFs and epithelial cells are more closely spaced, while in the intestinal type, CAFs form a stroma-like matrix, due to which tumor epithelial cells retain their glandular structure (Orimo, Weinberg, 2006).

Using the model of heterogeneous 3D spheroids, consisting of breast cancer epithelial cells and fibroblasts, Dang and colleagues showed that CAFs stimulated the migration of tumor cells of basal breast cancer (ER-/PR-/HER2-). Interestingly, this effect was not observed in the models of luminal breast cancer types (ER+/PR+/HER2+, ER+/PR+/HER2-) (Dang et al., 2011). These data are consistent with clinical observations indicating a higher percentage of metastasis in triple-negative breast cancer compared with other types of breast



**Fig. 2.** The interactions of cancer epithelial cells and stromal elements in the tumor.

TGF-β – transforming growth factor-β; FGF – fibroblast growth factor; PDGF – platelet growth factor.

tumors. However, which factors make cells of basal breast cancer sensitive to CAFs stimulation remain undiscovered (Al-Mahmood et al., 2018).

In order to reach the blood and lymphatic streams, tumor cells should pass through the basement membrane (BM), separating them from connective tissue and vessels. Thus, CAFs are able to synthesize metalloproteinases – endopeptidases capable of destroying proteins of all types of the BM extracellular matrix (Gonzalez-Avila et al., 2019). In 2017, Glentis and colleagues revealed a metalloproteinase-independent CAFs-supported overcome of BM by tumor cells. They demonstrated the ability of CAFs to stretch BM with the formation of pores, and through these pores, epithelial tumor cells and CAFs can migrate into the bloodstream and form metastases in distant organs. Interestingly, BM regions with low expression of laminin and type IV collagen exhibited the highest tendency for stretching (Glentis et al., 2017). This alternative CAFs-dependent migration pathway explains the ineffectiveness of the metalloproteinase inhibitors application in patients with head and neck tumors.

The paracrine secretion of IL-1α by CAFs in bladder cancer with further activation of the Wnt pathway in tumor cells is the perfect illustration of the pro-carcinogenic role of CAFs (Yang et al., 2021). Moreover, bladder cancer CAFs secreting IL-8 are able to stimulate the secretion of neuropilin-1, which enhances the proliferation of tumor cells and is one of the potential prognostic markers of malignancy (Chen C. et al., 2020). Interestingly, recent studies have shown that neuropilin-1 may be a co-factor in the induction of EMT (Chen Z. et al., 2020).

Paracrine stimulation of the epithelial-mesenchymal transition in tumor epithelial cells by CAFs is an important factor contributing to tumor progression. The central mechanism of EMT in the tumor is the TGF-β/Smad pathway activation, induced by TGF-β from stromal fibroblasts (Fig. 2) (Yu et al., 2014). The Smad factor is a transcription factor that controls the expression of EMT genes. Vered and colleagues showed that cells with EMT markers are found in primary foci of squamous cell carcinoma of the tongue as well as in regional lymph nodes metastases. This confirms the importance of

CAFs in the induction of metastasis and in the formation of a secondary tumor node (Vered et al., 2010).

The central mechanism of the EMT activation in ovarian cancer is the induction of the CXCR4/Wnt/β-catenin pathway in tumor epithelial cells. CAFs secreting stromal growth factor-1 (SDF-1 or CXCL12) have been shown to be the major players in this process. Moreover, SDF-1 is also interlinked with the resistance of tumor cells to chemotherapeutic agents such as cisplatin (Zhang et al., 2020).

In 2020, Franzè and colleagues demonstrated the activation of CAFs phenotype in normal fibroblasts from rectal polyps co-cultured with CAFs. They showed that CAFs derived from colorectal tumors can secrete IL-34, which in normal fibroblasts activates the expression of CAFs markers such as α-SMA, vimentin, and fibroblast activating protein (FAP) (Franzè et al., 2020).

Even though most of the described functions of CAFs are associated with the stimulation of tumor progression, some authors also describe CAFs as tumor-suppressing players. For example, the subpopulation of CAFs expressing the melanoma adhesion molecule (CD146) in breast tumors correlated with a retarded cell proliferation in estrogen-dependent types of breast cancer (Brechtbuhl et al., 2017). Since CAFs secrete cytokines involved in the recruitment and maturation of macrophages, T-lymphocytes and natural killer cells (IL-10, TGF-β, TNF, IFN-γ and IL-6), they increase the availability of the tumor to immune cells and promote antitumor immune response (Marlow et al., 2008). In the study of oral cancer, it was also shown that CAFs can suppress the proliferation of tumor cells. In particular, the population of CAFs secreting Bone Morphogenetic Protein 4 (BMP4) and expressing α-SMA inhibited the proliferation of cancer stem cells (CSC) (Patel et al., 2018). Using the mice model with a predisposition for the development of pancreatic cancer, Rhim and colleagues exhibited the role of CAFs in cancer progression. They excluded the α-SMA-positive population of CAFs or CAFs with inhibited Hedgehog signaling pathway. These modifications suppress the growth of pancreatic ductal adenocarcinoma. Histological analysis of tumors revealed abnormalities in the vessel's formation (Rhim et al., 2014).

These examples demonstrate the tumor-suppressing function of CAFs only in high differentiated cancers with no similarity in undifferentiated ones. It can be assumed that apart from origin and tissue of the affected organ, the function of CAFs is determined by the differentiation stage of cancer cells (Bu et al., 2019).

### Isolation of CAFs for research purposes

Cell cultures obtained from patients' tumors after surgery are most often used to study the properties of CAFs. It is necessary to establish new CAFs cell cultures, since *in vitro* CAFs tend to age rapidly, and the possibilities of their use are limited by early passages (Taddei et al., 2014). Moreover, in commercially available cell collections American Type Culture Collection (ATCC), European Collection of Authenticated Cell Cultures (ECACC), Russian collection of cell cultures, etc. cell lines with the CAFs phenotype are limited. For instance, in ATCC, only one CAFs cell line is available. This cell culture originates from the prostate adenocarcinoma and is modified by the introduction of the telomerase transgene under the control of the constitutive promoter of the polyoma virus SV40 hTERT PF179T CAF. This modification of fibroblasts is aimed at maintaining the proliferative properties of CAFs (Madar et al., 2009).

To obtain cell cultures from tumor tissue, mechanical disaggregation, enzymatic dissociation, chelation and their combination are used. Trypsin and type IV collagenase are most often used to destroy the stroma of tumor tissue. The choice of what technique to use should consider the histological origin of the tissue of interest. When tissue disaggregated, CAFs can represent a small population of cells and obtaining a monoculture requires an additional stage of their separation from the total cell mass. In order to isolate a particular population of CAFs, magnetic separation, or FACS of cells with immunostaining of specific CAFs markers such as FAP or  $\alpha$ -SMA are used (Sharon et al., 2013; Huang et al., 2017; Sha et al., 2018). The main difficulty in isolating CAFs lies in adapting protocols for vital staining of intracellular markers such as  $\alpha$ -SMA, FAP, and vimentin. Therefore, it is highly desirable to include surface markers such as CD90 in the analysis.

### Conclusion

Clear understanding the tumor microenvironment role is crucial for the development of new approaches in cancer diagnostics and treatment. The multifaceted influence of CAFs on tumor progression makes them an important object for the study of carcinogenesis and the development of new antitumor agents. The use of drugs targeted to the components of the tumor microenvironment has not demonstrated efficacy for anti-metalloproteinase compounds and angiogenesis inhibitors as well as T-cell immunity checkpoints inhibitors in some types of cancer (Wang-Gillam, 2019). The heterogeneity of the molecular phenotypes of CAFs can be an important factor in the failure of CAF-targeted cancer treatment. The scientific community should develop more detailed classifications of various subtypes of CAFs considering their involvement in tumor progression.

Thus, a detailed classification of CAFs and a study of the functions of each phenotypic subgroup may provide im-

portant knowledge for the development of new methods for the CAF-related treatment and diagnosis of oncological diseases.

### References

- Alkasalias T., Moyano-Galceran L., Arsenian-Henriksson M., Lehti K. Fibroblasts in the tumor microenvironment: shield or spear? *Int. J. Mol. Sci.* 2018;19(5):1532. DOI 10.3390/ijms19051532.
- Al-Mahmood S., Sapiezynski J., Garbuzenko O.B., Minko T. Metastatic and triple-negative breast cancer: challenges and treatment options. *Drug Deliv. Transl. Res.* 2018;8(5):1483-1507. DOI 10.1007/s13346-018-0551-3.
- Ambartsumian N., Klingelhöfer J., Grigorian M. The multifaceted S100A4 protein in cancer and inflammation. In: Heizmann C. (Ed.). Calcium-Binding Proteins of the EF-Hand Superfamily: From Basics to Medical Applications. (Ser. Methods in Molecular Biology). New York: Humana Press, 2019;1929:339-365. DOI 10.1007/978-1-4939-9030-6\_22.
- Anderberg C., Li H., Fredriksson L., Andrae J., Betsholtz C., Li X., Eriksson U., Pietras K. Paracrine signaling by platelet-derived growth factor-CC promotes tumor growth by recruitment of cancer-associated fibroblasts. *Cancer Res.* 2009;69(1):369-378. DOI 10.1158/0008-5472.CAN-08-2724.
- Bochet L., Lehuède C., Dauvillier S., Wang Y.Y., Dirat B., Laurent V., Dray C., Guet R., Maridonneau-Parini I., Le Gonidec S., Couderc B., Escourrou G., Valet P., Muller C. Adipocyte-derived fibroblasts promote tumor progression and contribute to the desmoplastic reaction in breast cancer. *Cancer Res.* 2013;73(18):5657-5668. DOI 10.1158/0008-5472.CAN-13-0530.
- Bordignon P., Bottoni G., Xu X., Popescu A.S., Truan Z., Guenova E., Kofler L., Jafari P., Ostano P., Röcken M., Neel V., Dotto G.P. Dualism of FGF and TGF- $\beta$  signaling in heterogeneous cancer-associated fibroblast activation with ETV1 as a critical determinant. *Cell Rep.* 2019;28(9):2358-2372.e6. DOI 10.1016/j.celrep.2019.07.092.
- Brechbuhl H.M., Finlay-Schultz J., Yamamoto T.M., Gillen A.E., Citelley D.M., Tan A.-C., Sams S.B., Pillai M.M., Elias A.D., Robinson W.A., Sartorius C.A., Kabos P. Fibroblast subtypes regulate responsiveness of luminal breast cancer to estrogen. *Clin. Cancer Res.* 2017;23(7):1710-1721. DOI 10.1158/1078-0432.CCR-15-2851.
- Bu L., Baba H., Yoshida N., Miyake K., Yasuda T., Uchihara T., Tan P., Ishimoto T. Biological heterogeneity and versatility of cancer-associated fibroblasts in the tumor microenvironment. *Oncogene.* 2019;38(25):4887-4901. DOI 10.1038/s41388-019-0765-y.
- Calon A., Tauriello D.V.F., Batlle E. TGF-beta in CAF-mediated tumor growth and metastasis. *Semin. Cancer Biol.* 2014;25:15-22. DOI 10.1016/j.semcancer.2013.12.008.
- Chen C., Zhang R., Ma L., Li Q., Zhao Y., Zhang G., Zhang D., Li W., Cao S., Wang L., Geng Z. Neuropilin-1 is up-regulated by cancer-associated fibroblast-secreted IL-8 and associated with cell proliferation of gallbladder cancer. *J. Cell. Mol. Med.* 2020;24(21):12608-12618. DOI 10.1111/jcmm.15825.
- Chen Z., Gao H., Dong Z., Shen Y., Wang Z., Wei W., Yi J., Wang R., Wu N., Jin S. NRP1 regulates radiation-induced EMT via TGF- $\beta$ /Smad signaling in lung adenocarcinoma cells. *Int. J. Radiat. Biol.* 2020;96(10):1281-1295. DOI 10.1080/09553002.2020.1793015.
- Dang T.T., Prechtel A.M., Pearson G.W. Breast cancer subtype-specific interactions with the microenvironment dictate mechanisms of invasion. *Cancer Res.* 2011;71(21):6857-6866. DOI 10.1158/0008-5472.CAN-11-1818.
- Fei F., Qu J., Zhang M., Li Y., Zhang S. S100A4 in cancer progression and metastasis: a systematic review. *Oncotarget.* 2017;8(42):73219-73239. DOI 10.18632/oncotarget.18016.
- Franzè E., Di Grazia A., Sica G.S., Biancone L., Laudisi F., Monteleone G. Interleukin-34 enhances the tumor promoting function of colorectal cancer-associated fibroblasts. *Cancers.* 2020;12(12):3537. DOI 10.3390/cancers12123537.



- Glentis A., Oertle P., Mariani P., Chikina A., El Marjou F., Attieh Y., Zaccarini F., Lae M., Loew D., Dingli F., Sirven P., Schoumacher M., Gurchenkov B.G., Plodinec M., Vignjevic D.M. Cancer-associated fibroblasts induce metalloprotease-independent cancer cell invasion of the basement membrane. *Nat. Commun.* 2017;8(1):924. DOI 10.1038/s41467-017-00985-8.
- Goetz J.G., Minguet S., Navarro-Lérida I., Lazcano J.J., Samaniego R., Calvo E., Tello M., Osteso-Ibáñez T., Pellinen T., Echarri A., Cerezo A., Klein-Szanto A.J.P., Garcia R., Keely P.J., Sánchez-Mateos P., Cukierman E., Del Pozo M.A. Biomechanical remodeling of the microenvironment by stromal caveolin-1 favors tumor invasion and metastasis. *Cell.* 2011;146(1):148-163. DOI 10.1016/j.cell.2011.05.040.
- Gonzalez-Avila G., Sommer B., Mendoza-Posada D.A., Ramos C., Garcia-Hernandez A.A., Falfán-Valencia R. Matrix metalloproteinases participation in the metastatic process and their diagnostic and therapeutic applications in cancer. *Crit. Rev. Oncol. Hematol.* 2019; 137:57-83. DOI 10.1016/j.critrevonc.2019.02.010.
- Grantham J. The molecular chaperone CCT/TRiC: an essential component of proteostasis and a potential modulator of protein aggregation. *Front. Genet.* 2020;11:172. DOI 10.3389/fgene.2020.00172.
- Grum-Schwensen B., Klingelhöfer J., Beck M., Bonfeld C.M., Hammerlik P., Guldberg P., Grigorian M., Lukanidin E., Ambartsumian N. S100A4-neutralizing antibody suppresses spontaneous tumor progression, pre-metastatic niche formation and alters T-cell polarization balance. *BMC Cancer.* 2015;15(1):44. DOI 10.1186/s12885-015-1034-2.
- Grum-Schwensen B., Klingelhofer J., Berg C.H., El-Naaman C., Grigorian M., Lukanidin E., Ambartsumian N. Suppression of tumor development and metastasis formation in mice lacking the S100A4(mts1) gene. *Cancer Res.* 2005;65(9):3772-3780. DOI 10.1158/0008-5472.CAN-04-4510.
- Hosein A.N., Wu M., Arcand S.L., Lavallée S., Hébert J., Tonin P.N., Basik M. Breast carcinoma-associated fibroblasts rarely contain p53 mutations or chromosomal aberrations. *Cancer Res.* 2010; 70(14):5770-5777. DOI 10.1158/0008-5472.CAN-10-0673.
- Huang Yingying, Zhou S., Huang Yong, Zheng D., Mao Q., He J., Wang Y., Xue D., Lu X., Yang N., Zhao Y. Isolation of fibroblast-activation protein-specific cancer-associated fibroblasts. *BioMed Res. Int.* 2017;2017:4825108. DOI 10.1155/2017/4825108.
- Jansson S., Aaltonen K., Bendahl P.-O., Falck A.-K., Karlsson M., Pietras K., Rydén L. The PDGF pathway in breast cancer is linked to tumour aggressiveness, triple-negative subtype and early recurrence. *Breast Cancer Res. Treat.* 2018;169(2):231-241. DOI 10.1007/s10549-018-4664-7.
- Joyce J.A., Pollard J.W. Microenvironmental regulation of metastasis. *Nat. Rev. Cancer.* 2009;9(4):239-252. DOI 10.1038/nrc2618.
- Jung Y., Kim J.K., Shiozawa Y., Wang Jingcheng, Mishra A., Joseph J., Berry J.E., McGee S., Lee E., Sun H., Wang Jianhua, Jin T., Zhang H., Dai J., Krebsbach P.H., Keller E.T., Pienta K.J., Taichman R.S. Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. *Nat. Commun.* 2013;4(1):1795. DOI 10.1038/ncomms2766.
- Kim D., Xing T., Yang Z., Dudek R., Lu Q., Chen Y.-H. Epithelial mesenchymal transition in embryonic development, tissue repair and cancer: a comprehensive overview. *J. Clin. Med.* 2017;7(1):1. DOI 10.3390/jcm7010001.
- Knops A.M., South A., Rodeck U., Martinez-Outschoorn U., Harshyne L.A., Johnson J., Luginbuhl A.J., Curry J.M. Cancer-associated fibroblast density, prognostic characteristics, and recurrence in head and neck squamous cell carcinoma: a meta-analysis. *Front. Oncol.* 2020;10:565306. DOI 10.3389/fonc.2020.565306.
- Liotta L.A., Kohn E.C. The microenvironment of the tumour-host interface. *Nature.* 2001;411(6835):375-379. DOI 10.1038/35077241.
- Liu B., Pan S., Liu J., Kong C. Cancer-associated fibroblasts and the related Runt-related transcription factor 2 (RUNX2) promote bladder cancer progression. *Gene.* 2021;775:145451. DOI 10.1016/j.gene.2021.145451.
- Lobba A.R.M., Carreira A.C.O., Cerqueira O.L.D., Fujita A., DeOce-sano-Pereira C., Osorio C.A.B., Soares F.A., Rameshwar P., So-gayar M.C. High CD90 (THY-1) expression positively correlates with cell transformation and worse prognosis in basal-like breast cancer tumors. *PLoS One.* 2018;13(6):e0199254. DOI 10.1371/journal.pone.0199254.
- Lynch M.D., Watt F.M. Fibroblast heterogeneity: implications for human disease. *J. Clin. Invest.* 2018;128(1):26-35. DOI 10.1172/JCI93555.
- Madar S., Brosh R., Buganim Y., Ezra O., Goldstein I., Solomon H., Kogan I., Goldfinger N., Klocker H., Rotter V. Modulated expression of WFDC1 during carcinogenesis and cellular senescence. *Carcinogenesis.* 2009;30(1):20-27. DOI 10.1093/carcin/bgn232.
- Marlow R., Strickland P., Lee J.S., Wu X., Pebenito M., Binnewies M., Le E.K., Moran A., Macias H., Cardiff R.D., Sukumar S., Hinck L. SLITs suppress tumor growth *in vivo* by silencing Sdf1/Cxcr4 within breast epithelium. *Cancer Res.* 2008;68(19):7819-7827. DOI 10.1158/0008-5472.CAN-08-1357.
- Massagué J. TGFβ in cancer. *Cell.* 2008;134(2):215-230. DOI 10.1016/j.cell.2008.07.001.
- Mossahebi-Mohammadi M., Quan M., Zhang J.-S., Li X. FGF signaling pathway: a key regulator of stem cell pluripotency. *Front. Cell Dev. Biol.* 2020;8:79. DOI 10.3389/fcell.2020.00079.
- Mukaida N., Sasaki S. Fibroblasts, an inconspicuous but essential player in colon cancer development and progression. *World J. Gastroenterol.* 2016;22(23):5301. DOI 10.3748/wjg.v22.i23.5301.
- Nushtaeva A.A., Karpushina A.A., Ermakov M.S., Gulyaeva L.F., Gerasimov A.V., Sidorov S.V., Gayner T.A., Yunusova A.Y., Tkachenko A.V., Richter V.A., Koval O.A. Establishment of primary human breast cancer cell lines using “pulsed hypoxia” method and development of metastatic tumor model in immunodeficient mice. *Cancer Cell Int.* 2019;19(1):46. DOI 10.1186/s12935-019-0766-5.
- Nushtaeva A.A., Stepanov G.A., Semenov D.V., Juravlev E.S., Bala-honova E.A., Gerasimov A.V., Sidorov S.V., Savelyev E.I., Kuligina E.V., Richter V.A., Koval O.A. Characterization of primary normal and malignant breast cancer cell and their response to chemotherapy and immunostimulatory agents. *BMC Cancer.* 2018;18(1):728. DOI 10.1186/s12885-018-4635-8.
- Orimo A., Weinberg R.A. Stromal fibroblasts in cancer: a novel tumor-promoting cell type. *Cell Cycle.* 2006;5(15):1597-1601. DOI 10.4161/cc.5.15.3112.
- Patel A.K., Vipparthi K., Thatikonda V., Arun I., Bhattacharjee S., Sharan R., Arun P., Singh S. A subtype of cancer-associated fibroblasts with lower expression of alpha-smooth muscle actin suppresses stemness through BMP4 in oral carcinoma. *Oncogenesis.* 2018; 7(10):78. DOI 10.1038/s41389-018-0087-x.
- Puré E., Hingorani S.R. Mesenchymal cell plasticity and perfidy in epithelial malignancy. *Trends Cancer.* 2018;4(4):273-277. DOI 10.1016/j.trecan.2018.02.007.
- Rhim A.D., Oberstein P.E., Thomas D.H., Mirek E.T., Palermo C.F., Sastra S.A., Dekleva E.N., Saunders T., Becerra C.P., Tattersall I.W., Westphalen C.B., Kitajewski J., Fernandez-Barrena M.G., Fernandez-Zapico M.E., Iacobuzio-Donahue C., Olive K.P., Stanger B.Z. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell.* 2014;25(6):735-747. DOI 10.1016/j.ccr.2014.04.021.
- Rønnov-Jessen L., Petersen O.W., Koteliansky V.E., Bissell M.J. The origin of the myofibroblasts in breast cancer. Recapitulation of tumor environment in culture unravels diversity and implicates converted fibroblasts and recruited smooth muscle cells. *J. Clin. Invest.* 1995;95(2):859-873. DOI 10.1172/JCI117736.
- Scherz-Shouval R., Santagata S., Mendillo M.L., Sholl L.M., Ben-Aharon I., Beck A.H., Dias-Santagata D., Koeva M., Stemmer S.M., Whitesell L., Lindquist S. The reprogramming of tumor stroma by HSF1 is a potent enabler of malignancy. *Cell.* 2014;158(3):564-578. DOI 10.1016/j.cell.2014.05.045.
- Sha M., Jeong S., Qiu B., Tong Y., Xia L., Xu N., Zhang J., Xia Q. Isolation of cancer-associated fibroblasts and its promotion to the



- progression of intrahepatic cholangiocarcinoma. *Cancer Med.* 2018; 7(9):4665-4677. DOI 10.1002/cam4.1704.
- Sharon Y., Alon L., Glanz S., Servais C., Erez N. Isolation of normal and cancer-associated fibroblasts from fresh tissues by Fluorescence Activated Cell Sorting (FACS). *J. Vis. Exp.* 2013;71:4425. DOI 10.3791/4425.
- Taddei M.L., Cavallini L., Comito G., Giannoni E., Folini M., Marini A., Gandellini P., Morandi A., Pintus G., Raspollini M.R., Zaffaroni N., Chiarugi P. Senescent stroma promotes prostate cancer progression: the role of miR-210. *Mol. Oncol.* 2014;8(8):1729-1746. DOI 10.1016/j.molonc.2014.07.009.
- Tam S.Y., Wu V.W.C., Law H.K.W. Hypoxia-induced epithelial-mesenchymal transition in cancers: HIF-1 $\alpha$  and beyond. *Front. Oncol.* 2020;10:468. DOI 10.3389/fonc.2020.00486.
- Tripathi M., Billet S., Bhowmick N.A. Understanding the role of stromal fibroblasts in cancer progression. *Cell Adh. Migr.* 2012;6(3): 231-235. DOI 10.4161/cam.20419.
- True L.D., Zhang H., Ye M., Huang C.-Y., Nelson P.S., von Haller P.D., Tjoelker L.W., Kim J.-S., Qian W.-J., Smith R.D., Ellis W.J., Liebeskind E.S., Liu A.Y. CD90/THY1 is overexpressed in prostate cancer-associated fibroblasts and could serve as a cancer biomarker. *Mod. Pathol.* 2010;23(10):1346-1356. DOI 10.1038/modpathol.2010.122.
- Vered M., Dayan D., Yahalom R., Dobriyan A., Barshack I., Bello I.O., Kantola S., Salo T. Cancer-associated fibroblasts and epithelial-mesenchymal transition in metastatic oral tongue squamous cell carcinoma. *Int. J. Cancer.* 2010;127(6):1356-1362. DOI 10.1002/ijc.25358.
- Wang M., Zhao J., Zhang L., Wei F., Lian Y., Wu Y., Gong Z., Zhang S., Zhou J., Cao K., Li X., Xiong W., Li G., Zeng Z., Guo C. Role of tumor microenvironment in tumorigenesis. *J. Cancer.* 2017;8(5): 761-773. DOI 10.7150/jca.17648.
- Wang-Gillam A. Targeting stroma: a tale of caution. *J. Clin. Oncol.* 2019;37(13):1041-1043. DOI 10.1200/JCO.19.00056.
- Weber C.E., Kothari A.N., Wai P.Y., Li N.Y., Driver J., Zapf M.A.C., Franzen C.A., Gupta G.N., Osipo C., Zlobin A., Syn W.K., Zhang J., Kuo P.C., Mi Z. Osteopontin mediates an MZF1-TGF- $\beta$ 1-dependent transformation of mesenchymal stem cells into cancer-associated fibroblasts in breast cancer. *Oncogene.* 2015;34(37):4821-4833. DOI 10.1038/onc.2014.410.
- Witkiewicz A.K., Dasgupta A., Sotgia F., Mercier I., Pestell R.G., Sabel M., Kleer C.G., Brody J.R., Lisanti M.P. An absence of stromal caveolin-1 expression predicts early tumor recurrence and poor clinical outcome in human breast cancers. *Am. J. Pathol.* 2009;174(6): 2023-2034. DOI 10.2353/ajpath.2009.080873.
- Yang F., Guo Z., He C., Qing L., Wang H., Wu J., Lu X. Cancer-associated fibroblasts promote cell proliferation and invasion via paracrine Wnt/IL1 $\beta$  signaling pathway in human bladder cancer. *Neoplasma.* 2021;68(1):79-86. DOI 10.4149/neo\_2020\_200202N101.
- Yasukawa Y., Hattori N., Iida N., Takeshima H., Maeda M., Kiyono T., Sekine S., Seto Y., Ushijima T. SAA1 is upregulated in gastric cancer-associated fibroblasts possibly by its enhancer activation. *Carcinogenesis.* 2021;42(2):180-189. DOI 10.1093/carcin/bgaa131.
- Ye X., Weinberg R.A. Epithelial-mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol.* 2015;25(11):675-686. DOI 10.1016/j.tcb.2015.07.012.
- Yin Z., Dong C., Jiang K., Xu Z., Li R., Guo K., Shao S., Wang L. Heterogeneity of cancer-associated fibroblasts and roles in the progression, prognosis, and therapy of hepatocellular carcinoma. *J. Hematol. Oncol.* 2019;12(1):101. DOI 10.1186/s13045-019-0782-x.
- Yoshida G.J. Regulation of heterogeneous cancer-associated fibroblasts: the molecular pathology of activated signaling pathways. *J. Exp. Clin. Cancer Res.* 2020;39(1):112. DOI 10.1186/s13046-020-01611-0.
- Yu Y., Xiao C.-H., Tan L.-D., Wang Q.-S., Li X.-Q., Feng Y.-M. Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through paracrine TGF- $\beta$  signalling. *Br. J. Cancer.* 2014;110(3):724-732. DOI 10.1038/bjc.2013.768.
- Yurugi Y., Wakahara M., Matsuoka Y., Sakabe T., Kubouchi Y., Haruki T., Nosaka K., Miwa K., Araki K., Taniguchi Y., Shiomi T., Nakamura H., Umekita Y. Podoplanin expression in cancer-associated fibroblasts predicts poor prognosis in patients with squamous cell carcinoma of the lung. *Anticancer. Res.* 2017;37(1):207-214. DOI 10.21873/anticancer.11308.
- Zeisberg E.M., Potenta S., Xie L., Zeisberg M., Kalluri R. Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. *Cancer Res.* 2007;67(21):10123-10128. DOI 10.1158/0008-5472.CAN-07-3127.
- Zhang F., Cui J., Gao H., Yu H., Gao F., Chen J., Chen L. Cancer-associated fibroblasts induce epithelial-mesenchymal transition and cisplatin resistance in ovarian cancer via CXCL12/CXCR4 axis. *Future Oncol.* 2020;16(32):2619-2633. DOI 10.2217/fon-2020-0095.

#### ORCID ID

M.S. Ermakov orcid.org/0000-0001-7107-4187  
A.A. Nushaeva orcid.org/0000-0001-9367-807X  
O.A. Koval orcid.org/0000-0001-7788-2249

**Acknowledgements.** This research was funded by the Russian Science Foundation, grant number 20-74-10039 and Russian State funded budget project of ICBFM SB RAS No. AAAA-A17-117020210023-1.

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

Received July 19, 2021. Revised November 10, 2021. Accepted November 12, 2021.

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

## Specific shoot formation in *Miscanthus sacchariflorus* (Poaceae) under different environmental factors and DNA passportization using ISSR markers

O.V. Dorogina<sup>1,3</sup>✉, N.S. Nuzhdina<sup>1</sup>, G.A. Zueva<sup>1</sup>, Yu.A. Gismatulina<sup>2</sup>, O.Yu. Vasilyeva<sup>1</sup>

<sup>1</sup> Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Institute for Problems of Chemical and Energetic Technologies of the Siberian Branch of the Russian Academy of Sciences, Biysk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

✉ [olga-dorogina@yandex.ru](mailto:olga-dorogina@yandex.ru)

**Abstract.** The generic complex *Miscanthus* Anderss. (Poaceae) is a unique example among herbaceous plants characterized by high values of growth of aboveground vegetative mass and practical use as a valuable source of alternative energy. *Miscanthus* is one of the most efficient solar energy accumulators, and since phytomeliorative use implies the cultivation of these resource plants in inconvenient and semi-shady areas, the question about the effect of insufficient lighting on the productivity of *Miscanthus* arises. As a result of a long-lasting introduction effort, the Central Siberian Botanical Garden SB RAS created a population of *Miscanthus sacchariflorus* (Maxim.) Benth., which has good prospects for growing under the conditions of the forest-steppe area in Western Siberia. The goals of our study were: (1) to determine the peculiarities of shoot formation, (2) to assess the cellulose and lignin accumulation in *M. sacchariflorus* populations under different lighting conditions and (3) to perform a DNA passportization of the *Miscanthus* population by ISSR marking. Evaluation of shoot formation and the amount of accumulated cellulose and lignin in plants was carried out under different degrees of illumination: one variant was grown in a sunny area, and the other, in partial shade. As a result of analysis of variance, it was found that the number of shoots does not depend on environmental conditions, but on the age of the plant, while environmental conditions have a significant effect on plant height. Although the samples of both *M. sacchariflorus* variants were characterized by different rates of creation of a continuous projective cover, plants in semi-shaded areas formed up to 89.34 % of shoots compared to their peers in illuminated areas, which did not affect significantly the size of the aboveground mass and the cellulose content in it. As a result of ISSR-analysis of genomic DNA in the *M. sacchariflorus* population, unique molecular polymorphic fragments were identified, which can be used for identification and DNA passportization at the inter-population level. Thus, the complex use of *M. sacchariflorus* as a valuable meliorative and bioenergetic culture is due to the high adaptive potential of this species. It was found that the illumination factor has virtually no effect on the amount of the cellulose content in the shoot, and a reduced content of the technologically undesirable lignin was observed in plants growing in the partial shade conditions. Key words: *Miscanthus*; bioenergy; cellulose; lignin; shoot formation; DNA passportization; ISSR markers.

**For citation:** Dorogina O.V., Nuzhdina N.S., Zueva G.A., Gismatulina Yu.A., Vasilyeva O.Yu. Specific shoot formation in *Miscanthus sacchariflorus* (Poaceae) under different environmental factors and DNA passportization using ISSR markers. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):22-29. DOI 10.18699/VJGB-22-04

## Особенности побегообразования в популяциях *Miscanthus sacchariflorus* (Poaceae) под влиянием экологических факторов и паспортизация с помощью ISSR-маркеров

О.В. Дорогина<sup>1,3</sup>✉, Н.С. Нуждина<sup>1</sup>, Г.А. Зуева<sup>1</sup>, Ю.А. Гисматулина<sup>2</sup>, О.Ю. Васильева<sup>1</sup>

<sup>1</sup> Центральный сибирский ботанический сад Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Институт проблем химико-энергетических технологий Сибирского отделения Российской академии наук, Бийск, Россия

<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ [olga-dorogina@yandex.ru](mailto:olga-dorogina@yandex.ru)

**Аннотация.** Уникальным примером травянистых растений, характеризующихся высокими значениями нарастания надземной вегетативной массы и практическим применением в качестве источника альтернативной энергии, является родовой комплекс *Miscanthus* Anderss. (Poaceae). Мискантус относится к числу наиболее эффективных аккумуляторов солнечной энергии, и поскольку фитомелиоративное использование подразумевает выращивание этих ресурсных видов на неудобьях и полутенистых участках, то встает вопрос о влиянии недо-

статочного освещения на показатели продуктивности мискантуса. В результате длительной интродукции в Центральном сибирском ботаническом саду СО РАН создана перспективная для условий лесостепи Западной Сибири популяция *Miscanthus sacchariflorus* (Maxim.) Benth. Целью данного исследования было изучение особенностей побегообразования, оценка целлюлозы и лигнина в популяциях *M. sacchariflorus* при различных условиях освещенности и паспортизация перспективной популяции с помощью ISSR-маркеров. Оценка побегообразования и количества накапливаемой целлюлозы и лигнина у растений проводилась в зависимости от освещенности (один вариант выращивался на солнечном участке, а другой – в полутени). В результате дисперсионного анализа установлено, что число побегов зависит не от экологических условий, а от возраста растения, в то время как на высоту растений существенно воздействуют экологические условия. Несмотря на то что для образцов обоих вариантов *M. sacchariflorus* была характерна различная скорость создания сплошного проективного покрытия, растения на полутенистых участках образовывали до 89.34 % побегов в сравнении с растениями на освещенных участках, что не оказывало существенного влияния на величину надземной массы и содержание в ней целлюлозы. В результате электрофореза геномной ДНК в популяции *M. sacchariflorus* при амплификации с пятью ISSR-праймерами выявлены уникальные молекулярные полиморфные фрагменты, которые были применены для идентификации и паспортизации данной популяции. Таким образом, комплексное использование *M. sacchariflorus* в качестве средоулучшающей и биоэнергетической культуры обусловлено высоким адаптивным потенциалом вида. Обнаружено, что фактор освещенности практически не влияет на количество целлюлозы в стебле, а пониженное содержание технологически нежелательного компонента, лигнина, отмечено при выращивании в условиях полутени.

Ключевые слова: род *Miscanthus*; биоэнергетика; целлюлоза; лигнин; побегообразование; паспортизация; ISSR-маркеры.

## Introduction

Over the past two decades, species of the genus *Miscanthus*, also known as elephant grass, have become one of the plant objects that are practically inexhaustible sources of renewable raw materials for the production of glucose, which is a basic product for many developments in the field of alternative energy (Slynko et al., 2013). *Miscanthus* is one of the most efficient solar energy accumulators on the planet (Dohleman, Long, 2009). According to physiological researches, *Miscanthus* species have high productivity potential. The production of up to 40 tons of dry biomass per hectare is associated with the C<sub>4</sub> type of photosynthesis, which is characteristic for these species. Unlike most traditionally cultivated C<sub>4</sub> plants, such as sugarcane and corn, *Miscanthus* is able to maintain a high rate of photosynthesis even under relatively low temperatures (Naidu et al., 2003; Anisimov et al., 2016). This explains the high productivity of this grass grown in more severe than natural climatic conditions for the purpose of economic use as a technical (bioenergetic) crop in the forest-steppes of Western Siberia. Phytomeliorative use implies the cultivation of a resource species in inconvenient and semi-shady areas. In particular, this also applies to *M. sacchariflorus* plants for photosynthesis, which requires a significant influx of photosynthetically active radiation. There is almost no data concerned with *M. sacchariflorus* usage as resource plants in semi-shady areas.

Currently, the problem of genetic identification of wild plant species and their populations is extremely urgent and is at the initial stage of development, although genetic passportization is believed to be an important stage required for registration and certification of new varieties (Kalayev et al., 2012).

The basis of selecting forms or varieties for the purpose of genetic passportization is to mark genetically determined characteristics using molecular methods. Some protein molecules like storage proteins or isozymes, or specific DNA loci can be used as molecular markers (Naeem et al., 2014; Chelyustnikova et al., 2019). Passportization of varieties and hybrids of many agricultural plants and crops was carried out using

ISSR and IRAP markers (Sukhareva, Kuluev, 2018), as well as molecular certification of rare and endemic plant species and their natural populations like two species of *Adonis*, *A. vernalis* and *A. sibirica* (Boronnikova, 2009). On the basis of this technique, I. V. Boboshina, S. V. Boronnikova received a patent for the invention “Method of molecular-genetic identification of woody plant species populations” (Boboshina, Boronnikova, 2014). To identify raw materials of medicinal plants by roots or other plant tissues, DNA and chemical markers are used, since they are not tissue-specific and have a high resolution power and accuracy (Wallerger et al., 2012; Ganiea et al., 2015). However, we found no literature data on the molecular certification of promising populations and varieties of *Miscanthus*.

Consequently, the goals of the present study were: (1) to determine the specific shoot formation in *M. sacchariflorus* populations under different environment conditions, (2) to assess the cellulose and lignin accumulation in various lighting conditions and (3) DNA passportization of *M. sacchariflorus* population by ISSR marking.

## Materials and methods

The experimental plots of the Central Siberian Botanical Garden (CSBG SB RAS, Novosibirsk, Russia) are located in the forest-steppe part of Western Siberia, which belongs to the IV lighting zone and, in terms of the total number of hours of sunshine, it is close to Krasnodar and Yalta. The object of the study was a selective population of *M. sacchariflorus*, identified as a result of many years of introduction experiments, which was formed from material collected in the Khasansky district of Primorsky Krai. One sample from this population was grown in partial shade (sample 1), and the other sample (sample 2) was grown in an open, well-lit area. The control (sample 3) was the introduction population of *M. sacchariflorus*, from which selections were made to study the features of shoot formation.

Experimental individuals were planted in 2017 on plots 2×2 m in size in four replicates. Sample 1 was placed in the

penumbra (factor A<sub>0</sub>), and sample 2 – in an open, well-lit place (factor A<sub>1</sub>). On the plots, 1 rhizome (rhizome piece) with 5 shoots was planted in each of the staggered 16 holes (4 luns/m<sup>2</sup>). Thus, the total number of shoots during planting was 20 shoots/m<sup>2</sup>. Sample 3 (control) represented a continuous perennial clump located in an illuminated place. In autumn, at the end of the vegetation season, we carried out continuous pruning of the plants, leaving the height of the shoots 15 cm from the soil level.

Subsequently, at the end of the growing seasons, the number of shoots was counted for the plants generated from rhizomes: for 2-year-old plants in 2018 (factor B<sub>0</sub>), and for 3-year-old plants in 2019 (factor B<sub>1</sub>). The results of the two-factor experiment were processed by the method of variance described by Dospekhov (1985).

The dynamics of growth and shooting of *M. sacchariflorus* was studied by the method of phenological observations, carrying out biometric measurements and counting the number of shoots formed during three growing seasons.

The chemical composition was studied in 2019 in the aerial part of the *Miscanthus*, cut off at a distance of 10–15 cm from the ground. Before performing the chemical analysis, the raw material was dried in air to minimize the moisture content (less or equal 8 %) and grinded up to a size of 5–10 mm. The chemical composition of plant materials was determined by standard analytical methods using the equipment of the Biysk Regional Center for Collective Use (Institute for Problems of Chemical and Energetic Technologies SB RAS, Biysk, Russia). The determination of the mass fraction of cellulose was carried out using the Kurschner method (in terms of absolutely dry raw material – a.d.m.), with the determination of the mass fraction of acid-insoluble lignin (a.d.m.), the mass fraction of pentosans (a.d.m.), ash content (a.d.m.), the mass fraction of extractives – fatty wax fraction (FWF) (extractant – dichloromethane, a.d.m.), according to the standard analysis of plant raw materials (Obolenskaya et al., 1991).

Extraction of genomic DNA from dried leaves was performed by the CTAB method (Doyle J.J., Doyle J.L., 1990). DNA concentration was determined spectrophotometrically using a BioSpectrometer kinetic and a µCuvette G1.0 microcuvette (Eppendorf, Germany).

For molecular analysis of populations, 16 ISSR (inter simple sequence repeats) oligonucleotides (primers) were tested. The most polymorphic five oligonucleotides were selected to obtain molecular genetic formulas (Table 1).

PCR was carried out under the following conditions: (1) DNA denaturation: 90 s at 94 °C; (2) 35 amplification cycles: 40 s at 94 °C, 45 s at 41–56 °C (primer annealing) and 90 s at 72 °C; (3) elongation: 5 min at 72 °C. The PCR mixture with a volume of 25 µL consisted of 2.7 mM MgCl<sub>2</sub>, 1.25 mM primer, 0.4 mM dNTPs, 2.5 µL 10× PCR buffer, 1 unit of Taq DNA polymerase and 20 ng genomic DNA. PCR was performed on a Thermal Cycler C1000 amplifier (Bio-Rad, USA). Electrophoretic analysis of ISSR-PCR products was carried out in 1 % agarose gel. The amplified fragments were stained with SYBR-Green (ThermoFischer Scientific). Visualization and recording of the separated PCR fragments was carried out using the Gel-Doc XR+ gel documentation system and the ImageLab Software Imaging System (Bio-Rad).

**Table 1.** Characteristics of ISSR primers tested and selected (in bold) for the study of genetic polymorphism of *M. sacchariflorus* population

No.	Nucleotide sequences, 5'–3'	Temperature of annealing, °C
1	<b>(CA)<sub>6</sub>GT</b>	42
2	<b>(CA)<sub>6</sub>GG</b>	42
3	<b>(CA)<sub>6</sub>AG</b>	47
4	<b>(CT)<sub>8</sub>GC</b>	48
5	<b>(CT)<sub>8</sub>TG</b>	51
6	(AC) <sub>8</sub> YG	55
7	(CT) <sub>8</sub> AC	48
8	(AC) <sub>8</sub> CG	47
9	(AG) <sub>10</sub> G	64
10	(CA) <sub>6</sub> RG	49
11	(CTC) <sub>3</sub> GC	42
12	(CA) <sub>6</sub> AC	42
13	(CAC) <sub>3</sub> GC	41
14	(GACA) <sub>4</sub>	45
15	(GT) <sub>6</sub> GG	48
16	(GAA) <sub>6</sub>	48

Molecular genetic formulas for the passportization of the *M. sacchariflorus* population were drawn up according to the principle proposed by A.A. Novikova and co-authors (Novikova et al., 2012). Statistical analysis was carried out using the MS Excel program.

Results

Under experimental conditions, *M. sacchariflorus* plants regrowth and further development was observed in the third decade of May – first decade of June, 2018. No active growth of the vegetative mass was noted in the third decade of May since the air temperature did not exceed 9.6 °C (Fig. 1). Starting from the second decade of June, with an increase in temperature, the number of shoots rose due to active tillering and intensive growth rates.

From the meteorological point of view, 2019 was favorable for the elephant grass. The average temperature in May was 10.8 °C, which contributed to active vegetation. Further, in plants with a well-developed and successfully overwintered underground shoot system, an aboveground part was formed, resembling a clearly polycentric biomorph: diaspores are formed on the plagiotropic shoots of this species, at the moment of the appearance of their own root system they are fixed, maintaining a connection with the mother plant.

In June, the temperature slowly increased without drops (see Fig. 1), the tillering process took place from July (especially during the period of maximum precipitation) to the beginning



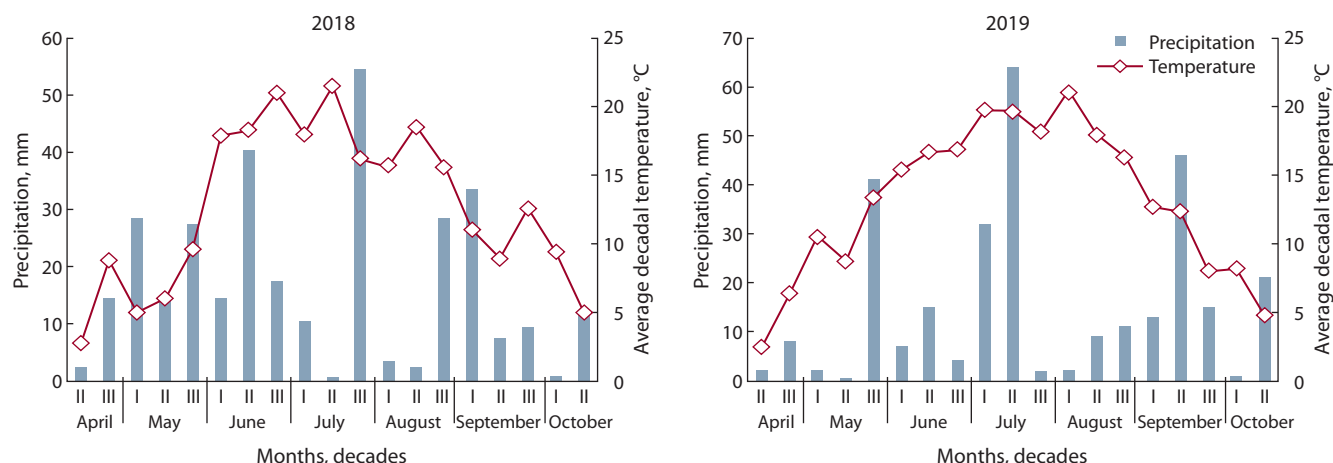


Fig. 1. Hydrothermal conditions of the growing seasons 2018–2019.

**Table 2.** Results of the variance analysis of a two-factor experiment to study the influence of environmental conditions and the age of rhizomes on the number of shoots of *M. sacchariflorus*

Lighting conditions (A)	Age of rhizomes (B)	Number of shoots, X			
		I	II	III	IV
A <sub>0</sub>	B <sub>0</sub>	8.3	12.0	8.0	8.8
	B <sub>1</sub>	22.8	26.0	27.5	32.0
A <sub>1</sub>	B <sub>0</sub>	6.8	8.3	15.0	14.3
	B <sub>1</sub>	20.5	29.3	28.3	56.3

Dispersion	Sum of squares	Freedom level	Medium square	F <sub>fact</sub>	F <sub>0.5</sub>
Common dispersion	2555.3	15	—	—	—
Lighting conditions (A)	69.7	1	69.7	1.0	4.75
Age of rhizomes (B)	1624.1	1	1624.1	22.7	4.75
AB interaction	4.1	1	4.1	0.1	4.75
Remain (mistakes)	857.4	12	71.5	—	—

of August. Plants in all variants formed a greater number of shoots during the 2019 growing season than in 2018. In the second half of August, the activity of the tillering process correlated with the temperature and humidity level (22 mm of precipitation – 33 % of the norm). Plants reduced the productivity of the vegetative mass and started generative processes. At this time, active elongation of shoots was observed due to the increase in a hollow peduncle, the formation of an upper “flag” leaf and the appearance of panicle inflorescences, which lead to growth of the vegetative mass of shoots.

To identify the effect of ecological conditions and plant age on the shoot formation of *M. sacchariflorus* (in 2018 and 2019), a two-factor analysis of variance was carried out (Tables 2 and 3). As can be seen from Table 2, the number of formed shoots was significantly influenced by the age of

**Table 3.** Results of the variance analysis of a two-factor experiment studying the influence of environmental conditions and the age of rhizomes on the height of *M. sacchariflorus* plants

Lighting conditions (A)	Age of rhizomes (B)	Height of plants, X			
		I	II	III	IV
A <sub>0</sub>	B <sub>0</sub>	96.5	116.0	107.8	136.3
	B <sub>1</sub>	115.5	140.3	172.3	175.0
A <sub>1</sub>	B <sub>0</sub>	112.3	138.5	129.5	130.3
	B <sub>1</sub>	139.0	166.3	182.0	190.8

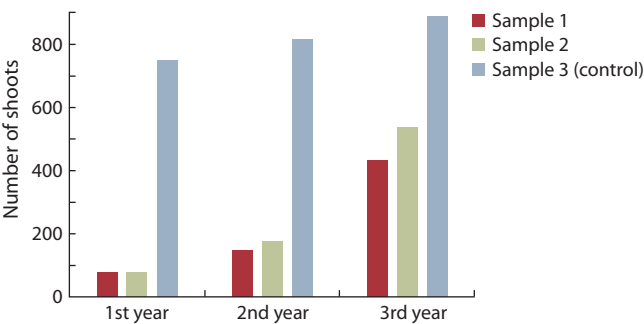
  

Dispersion	Sum of squares	Freedom level	Medium square	F <sub>fact</sub>	F <sub>0.5</sub>
Common dispersion	12 394.4	15	—	—	—
Lighting conditions (A)	6162.2	1	6162.2	14.3	4.75
Age of rhizomes (B)	1040.1	1	1040.1	2.4	4.75
AB interaction	27.6	1	27.6	0.1	4.75
Remain (mistakes)	5164.5	12	430.4	—	—

rhizomes (B); at the same time, the influences of environmental conditions (A) and the interaction of factors (AB) were insignificant. The factor of ecological conditions (A) has an influence on plant height (see Table 3), but the effects of the age of rhizomes (B) and the interaction of factors (AB) were insignificant.

Thus, as a result of a variance analysis (see Table 2, at  $F = 0.5$ ), it was revealed that the factor of the age of rhizomes ( $B 22.7 > 4.75$ ) has a significant effect on the number of shoots, but not the ecological conditions. Meanwhile, the height of the plants largely depends on environmental conditions ( $A 14.3 > 4.75$ ), but not on the age (see Table 3).

The control population consisted of perennial plants located in a well-lighted place. As is shown on Figure 2, there is a little increase in the number of shoots in control plants that



**Fig. 2.** Features of shoot formation of *M. sacchariflorus* specimens under various ecological conditions.  
Sample 1 – semi-shade area, sample 2 – well-lighted area, sample 3 (control) – perennial plants.

have previously formed a dense curtain, and, on the contrary, intensive shoot formation in young plants (initially represented by rhizomes) at the 2nd and 3rd years of life. It was revealed that *M. sacchariflorus* plants in semi-shady areas at 2 years of age formed up to 89.34 % of shoots in comparison with illuminated areas.

However, sample 1, which has grown in the shade, in terms of the number of shoots developed during the second (148 shoots) and third (433) years of life, lags behind sample 2 with 177 and 537 shoots, respectively. The effect of higher illumination stimulates tillering in sample 2, which leads to quick closing of separately located plants.

Less intense tillering was observed in sample 1 in the penumbra; however, in this case, too, the projective vegetation cover in the third year of experiment was quite high – from 65 to 75 %. Sample 3 has been growing in one place for more

than 15 years. It is noted that plants actively grew and developed annually, no degradation phenomena were observed. The projective cover was over 70 %. However, the possibilities of intensive shoot formation were practically exhausted, therefore, in 2019, the increase in the number of shoots compared to 2018 was only 9.06 % (816 and 890 shoots). For sample 1 located in the semi-shade area, this increase was 192.57 % (148 and 433), and for sample 2 located in the well-lighted area – 203.40 % (177 and 537).

Chemical analysis of these three samples, carried out on the material of *M. sacchariflorus* in 2019, separately on the stems (since the stem cellulose is valued higher) and leaves, showed that the cellulose content in the penumbra (50.9 %) was higher than in the illuminated area (50.1 %). Reduced (by 12 %) content of a technologically undesirable component lignin was noted in the least economically valuable semi-shady area (Table 4). This could be caused by the fact that tissue differentiation of shoots, including lignification, occurs more intensively in sufficient illumination.

The mass fraction of cellulose in the leaf regardless of the light conditions (40.2 % in the partial shade area and 42.2 % in the sunny area) is significantly lower than in the stem, which is in good agreement with the previously obtained results (Gismatulina et al., 2019). Similarly to the stem, the mass fraction of lignin is 7.6 % lower in the partial shade area than in the sunny area. The mass fractions of pentosans, FWF, and ash are approximately at the same level both in the stem and in the leaf, regardless of the lighting conditions of the plantation.

As a result of electrophoresis of PCR products of genomic DNA in the *M. sacchariflorus* population obtained by amplification with five selected ISSR primers, a high genetic polymorphism of the studied objects was found (Fig. 3).

**Table 4.** Chemical composition of three samples of *M. sacchariflorus* in 2019

Chemical and technological indicators	Sample 1, Semi-shade area	Sample 2, Well-lighted area	Sample 3 (control), perennial plants
Stem			
Weight, g	339	295	294
Humidity, %	5.7 ± 0.1	5.8 ± 0.1	5.7 ± 0.1
Ash, %	1.91 ± 0.05	1.60 ± 0.05	1.48 ± 0.05
Lignin, %	18.7 ± 0.1	21.3 ± 0.1	25.5 ± 0.1
Cellulose according to Kurschner, %	50.9 ± 0.1	50.1 ± 0.1	52.0 ± 0.1
Pentosans, %	23.2 ± 0.1	23.9 ± 0.1	22.1 ± 0.1
FWF, %	0.8 ± 0.1	1.9 ± 0.1	0.9 ± 0.1
Leaf			
Weight, g	180	134	134
Humidity, %	9.3 ± 0.1	7.5 ± 0.1	7.5 ± 0.1
Ash, %	4.74 ± 0.05	4.83 ± 0.05	4.83 ± 0.05
Lignin, %	20.1 ± 0.1	21.7 ± 0.1	21.7 ± 0.1
Cellulose according to Kurschner, %	40.2 ± 0.1	42.2 ± 0.1	42.2 ± 0.1
Pentosans, %	23.8 ± 0.1	23.3 ± 0.1	23.3 ± 0.1
FWF, %	1.9 ± 0.1	1.8 ± 0.1	1.8 ± 0.1

Note. FWF – fatty wax fraction; the half-width of the confidence interval was determined at the significance level of 0.05.

From one to four specific molecular markers (unique PCR fragments) were identified (Table 5). As follows from Table 5, the length of polymorphic fragments in ISSR analysis ranged from 660 to 2000 bp. The identified unique molecular polymorphic fragments representing sequences of a certain length were selected for passportization of *M. sacchariflorus* population.

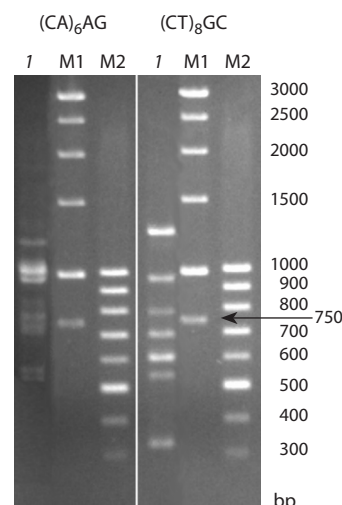
Thus, taking into account the genetic formula proposed by A.A. Novikova with co-authors for *Rhododendron canadense* (Novikova et al., 2012), the genetic formula for the *M. sacchariflorus* population will be the following: ISSR/(CA)<sub>6</sub>AG-925,980/(CT)<sub>8</sub>GC-600,690,780,940/(CT)<sub>8</sub>TG-1060/(CT)<sub>8</sub>AC-690,800,1030,1390/(AC)<sub>8</sub>YG-650,975,1470,2000.

## Discussion

The study of the specificity of shoot formation in *M. sacchariflorus* introduced into CSBG under the conditions of the continental climate of Western Siberia showed that early generative development of plants is undesirable for growing this species as a bioenergetic culture, since the accumulation of biomass stops. It was found that this species has a rather long period of active growth. It should be taken into account that plants of *M. sacchariflorus* start growing only after the air warms up to 25 °C. In experimental 3-year-old plants the projective vegetation cover in triplicate varied from 70 to 80 %.

Based on the results of variance analysis we can conclude that the number of shoots depends on the age of the plants and the influence of environmental conditions, and the interaction of these factors on the number of shoots is insignificant. At the same time, ecological conditions have a significant effect on plant height and age, and the interaction of these factors practically do not affect plant height. In this regard, an important issue in the study of the shoot formation of *M. sacchariflorus* is the initiation of tillering period, which is associated either with the beginning of the growth of lateral shoots in the zone of shortened internodes (Langer, 1963; Smelov, 1966), or with intensive growth of this zone (Dobrynin, 1969; Gorchakova, 2003).

It should be noted that the mass fraction in the stem of the technologically significant component cellulose does not change depending on the lightning conditions (50.9 % in partial shade area and 50.1 % in the open area). At the same



**Fig. 3.** PCR-ISSR profile of *M. sacchariflorus* generated from amplification with the primers (CA)<sub>6</sub>AG and (CT)<sub>8</sub>GC.

The track numbers correspond to the samples: 1 – *M. sacchariflorus* (No. 1), M1 and M2 – molecular weight markers.

time, the mass fraction of lignin, which adversely affects technological processes, was 12 % lower in the penumbra.

Thus, it was found that the adaptive potential of *M. sacchariflorus*, the high content of cellulose (52.04 %) with a relatively low content of lignin (21.3 %) allows to suggest the population as an environment-improving and bioenergetic culture.

For genetic passportization of the population, five ISSR markers were selected. Based on our studies and the results obtained by other authors, we can conclude that ISSR primers which have been used are polymorphic and can be recommended for identifying other samples, populations and species, as well as for composing genetic formulas and passports for the genus *Miscanthus* (Boronnikova, 2009; Artyukhova et al., 2011; Novikova et al., 2012; Lebedev et al., 2014). I.A. Klimenko with co-authors carried out identification and certification clover varieties using SSR and SRAP markers and proposed a set of DNA markers (Klimenko et al., 2020). The data obtained using DNA analysis are the most objective for describing plant varieties and species, since they are not susceptible to genotypic variability and mostly have a co-

**Table 5.** ISSR markers used in the study to identify the molecular formula of *M. sacchariflorus* population

Primer, 5'–3'	Number of specific markers*	Length of polymorphic phragments, bp
(CA) <sub>6</sub> AG	2/8 (25 %)	(CA) <sub>6</sub> AG <sub>925</sub> , (CA) <sub>6</sub> AG <sub>980</sub>
(CT) <sub>8</sub> GC	4/7 (57 %)	(CT) <sub>8</sub> GC <sub>600</sub> , (CT) <sub>8</sub> GC <sub>690</sub> , (CT) <sub>8</sub> GC <sub>780</sub> , (CT) <sub>8</sub> GC <sub>940</sub>
(CT) <sub>8</sub> TG	1/3 (33 %)	(CT) <sub>8</sub> TG <sub>1060</sub>
(CT) <sub>8</sub> AC	4/7 (57 %)	(CT) <sub>8</sub> AC <sub>690</sub> , (CT) <sub>8</sub> AC <sub>800</sub> **, (CT) <sub>8</sub> AC <sub>1030</sub> , (CT) <sub>8</sub> AC <sub>1390</sub>
(AC) <sub>8</sub> YG	4/13 (31 %)	(AC) <sub>8</sub> YG <sub>650</sub> , (AC) <sub>8</sub> YG <sub>975</sub> , (AC) <sub>8</sub> YG <sub>1470</sub> **, (AC) <sub>8</sub> YG <sub>2000</sub>

\* Total number of markers (denominator), number of unique markers (numerator) and their percentage (in brackets).

\*\* ISSR markers with a weak fluorescent signal.

dominant type of inheritance (Ramazanova, Kolomytseva, 2020).

The genetic passport of *M. sacchariflorus*, presented as a genetic formula generated by amplified DNA, contains information about the method used, oligonucleotide sequences, and specific amplified DNA fragments lengths. If necessary, it is possible to improve the form of recording the molecular genetic formula indicating the specificity level of a fragment (generic, species, polymorphic), as it was suggested by S.V. Boronnikova (2009). In general, the molecular genetic formula population makes it possible to determine the belonging of the *Miscanthus* individuals not only to the species and variety, but also to a specific population.

## Conclusion

The results obtained during the study allow concluding that *M. sacchariflorus* can be successfully grown in semi-shady forest steppes of Western Siberia, and the lignin content in raw plant material will be reduced by the time of harvesting in case of growing at the local microecological conditions.

The high projective vegetation cover under various environmental conditions, as well as the longevity of the clones, indicate the prospects for the phytomeliorative use of selected forms of *M. sacchariflorus* in the continental climate of the forest-steppe of Western Siberia. The content of cellulose in the stem, the most important component in technical plant material, varies slightly depending on the lighting conditions. At the same time, the content of lignin, which negatively affects technological processes, turned out to be lower in plants grown in partial shade.

The obtained molecular genetic formulas for the *M. sacchariflorus* population make it possible to determine the belonging of individuals of *Miscanthus* not only to the species and variety, but also to a specific population.

Genetic passportization based on molecular data of promising forms of *Miscanthus*, the development of scientific and practical recommendations and a set of cultivation techniques will make it possible to use the representatives of this genus in breeding under the conditions of the continental climate of Western Siberia.

The selective work with the varieties obtained by vegetative reproduction of the perspective individuals and their molecular DNA identification make it possible to recommend the genus *Miscanthus* as an environmentally friendly technical crop and as a renewable source of plant material promising for the implementation of an alternative bioenergy program in Western Siberia.

## References

- Anisimov A.A., Khokhlov N.F., Tarakanov I.G. Photoperiodic regulation of ontogenesis in different *Miscanthus* species (*Miscanthus* spp.). *Izvestiya Timiryazevskoy Selskokhozyaystvennoy Akademii* = *Izvestiya of Timiryazev Agricultural Academy*. 2016;6:56-72. (in Russian)
- Artyukhova A.V., Grishin S.Yu., Lukashevich M.I., Zayakin V.V., Nam I.Ya. Development of a method for *Lupinus* variety certification. *Vestnik Bryanskogo Gosudarstvennogo Universiteta* = *Herald of the Bryansk State University*. 2010;4:81-84. (in Russian)
- Boboshina I.V., Boronnikova S.V. Method of molecular-genetic identification of woody plant species populations. Patent for invention RU 2505956 C2. Publ. 10.02.2014. Bull 4. (in Russian)
- Boronnikova S.V. Genetic certification of populations of rare *Adonis* species using ISSR and IRAP markers. *Izvestiya Timiryazevskoy Selskokhozyaystvennoy Akademii* = *Proceedings of the Timiryazev Agricultural Academy*. 2009;1:82-88. (in Russian)
- Chelyustnikova T.A., Guchetl S.Z., Antonova T.S. Microsatellite loci for identification of oil flax varieties of the breeding of V.S. Pustovoi All-Russian Research Institute of Oil Crops: selection of informative primers and optimal conditions for DNA PCR. *Maslichnye Kultury* = *Oil Crops*. 2019;2:41-46. (in Russian)
- Dobrynin G.M. Growth and Formation of Bread and Fodder Cereals. Leningrad: Kolos Publ., 1969. (in Russian)
- Dohleman F.G., Long S.P. More productive than maize in the mid-west: how does *Miscanthus* do it? *Plant. Physiol.* 2009;50(4):2104-2115.
- Dospekhov B.A. Methodology of Field Experiments with the Fundamentals of Statistical Processing of Results, 5th revised edition. Moscow: Agropromizdat Publ., 1985. (in Russian)
- Doyle J.J., Doyle J.L. Isolation of plant DNA from fresh tissue. *Focus*. 1990;12:13-15.
- Ganiea S.H., Upadhyaya P., Maheshwer S.D., Sharmab P. Authentication of medicinal plants by DNA markers. *Plant Gene*. 2015;4: 83-99. DOI 10.1016/j.plgene.2015.10.002.
- Gismatulina Y.A., Budaeva V.V. Chemical composition of five *Miscanthus sinensis* harvests and nitric-acid cellulose therefrom. *Ind. Crops Prod.* 2017;109:227-232. DOI 10.1016/j.indcrop.2017.08.026.
- Gismatulina Yu.A., Budaeva V.V., Sakovich G.V., Vasilyeva O.Yu., Zueva G.A., Gusar A.S., Dorogina O.V. Features of the resource species *Miscanthus sacchariflorus* (Maxim.) Hack. when introduced in West Siberia. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2019;23(7):933-940. DOI 10.18699/VJ19.569.
- Gorchakova A.Yu. New about the branching of cereals. *Byulleten Botanicheskogo Sada Kubanskogo Gosudarstvennogo Agronomicheskogo Universiteta* = *Bulletin of the Botanical Garden of the Kuban State Agrarian University*. 2003;21:39-40 (in Russian)
- Kalayev V.N., Zemlyanukhina O.A., Karpechenko I.Yu., Karpechenko K.A., Kondratyeva A.M., Veprintsev V.N., Karpechenko N.A., Karpova S.S., Moiseeva E.V., Baranova T.V. The use of molecular genetic analysis methods to study DNA polymorphism in plants of the genus *Rhododendron* for the purpose of their certification. *Fundamentalnye Issledovaniya* = *Fundamental Research*. 2012;6(2): 323-328. (in Russian)
- Klimenko I.A., Kozlov N.N., Kostenko S.I., Shamustakimova A.O., Mavlyutov Yu.M. Identification and Certification of Forage Grass Varieties Based on DNA Markers: Recommended Practice. Moscow: Ugreshat Publ., 2020. (in Russian)
- Langer R.H.M. Tillering in herbage grasses. *Herb. Abstr.* 1963;33(3): 141-148.
- Lebedev V.G., Subbotina N.M., Kirkach V.V., Vidyagina E.O., Pozdnyakova I.A., Shestibratov K.A. Analysis of microsatellite loci as the first step in the marker-assisted selection of raspberry and strawberry. *Seleksiya i Sortorazvedenie Sadovykh Kultur* = *Breeding and Variety Cultivation of Horticultural Crops*. 2018;5(1):65-68. (in Russian)
- Naeem R. Molecular markers in plant genotyping. *J. Bio-Mol. Sci.* 2014;2(3):78-85.
- Naidu S.L., Moose S.P., Al-Shoaibi A.K., Raines C.A., Long S.P. Cold tolerance of *C<sub>4</sub>* photosynthesis in *Miscanthus × giganteus*: adaptation in amounts and sequence of *C<sub>4</sub>* photosynthetic enzymes. *Plant Physiol.* 2003;132:1688-1697. DOI 10.1104/pp.103.021790.
- Novikova A.A., Sheikina O.V., Novikov P.S., Doronina G.U. Assessment of the possibility of using ISSR markers for the systematization and genetic certification of plants of the genus *Rhododendron*. *Politematicheskij Setevoy Elektronnyy Nauchnyy Zhurnal Kubanskogo Gosudarstvennogo Agrarnogo Universiteta* = *Polythematic Scientific Online Journal of the Kuban State Agrarian University*. 2012;82:79-89. (in Russian)



- Obolenskaya A.V., Elnitskaya Z.P., Leonovich A.A. Laboratory Work on the Chemistry of Wood and Cellulose: Textbook for universities. Moscow: Ecology Publ., 1991. (in Russian)
- Ramazanov S.A., Kolomytseva A.S. Optimization of soybean genotyping process using analysis of a polymorphism of SSR-loci in DNA. *Maslichnye Kultury = Oil Crops*. 2020;1:42-48. (in Russian)
- Slyn'ko N.M., Goryachkovskaya T.N., Shekhovtsov S.V., Bannikova S.V., Burmakina N.V., Starostin K.V., Rozanov A.S., Nechiporenko N.N., Veprev S.G., Shumny V.K., Kolchanov N.A., Pel'tek S.E. The biotechnological potential of the new crops, *Miscanthus* cv. Soranovskii. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/1):765-771. (in Russian)
- Smelov S.P. Theoretical Foundations of Meadow Farming. Moscow: Kolos Publ., 1966. (in Russian)
- Sukhareva A.S., Kuluev B.R. DNA markers for genetic analysis of crops. *Biomika = Biomics*. 2018;10(1):69-84. DOI 10.31301/2221-6197.bmcs.2018-15. (in Russian)
- Wallinger C., Juen A., Staudacher K., Schallhart N., Mitterrutzner E., Steiner E.-M., Thalinger B., Traugott M. Rapid plant identification using species- and group-specific primers targeting chloroplast DNA. *PLoS One*. 2012;7(1):e29473. DOI 10.1371/journal.pone.0029473.

---

#### ORCID ID

O.V. Dorogina orcid.org/0000-0001-5729-3594  
N.S. Nuzhdina orcid.org/0000-0002-0634-8820

G.A. Zueva orcid.org/0000-0001-7950-054X  
Yu.A. Gismatulina orcid.org/0000-0001-5480-7449  
O.Yu. Vasilyeva orcid.org/0000-0003-0730-3365

**Acknowledgements.** The study was carried out according to the complex project of fundamental research program of CSBG SB RAS "Analysis of biodiversity, conservation and restoration of rare and resource plant species using experimental methods" (No. AAAA-A21-121011290025-2). Materials from the bioresource repository "Collections of living plants in open and protected ground" (USU 440534) were used. The study was supported by project FSUS-2021-0012 "Ecosystems of grass pine and small-leaved forests as regulators of the nitrogen and carbon turnovers in the forest-steppe landscape of West Siberia".

Research on the chemical composition was carried out within the framework of the state assignment of IPCET SB RAS "Fundamentals of the creation of an integrated technology for processing easily renewable non-food plant raw materials into products demanded by the economy of the Russian Federation".

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

Received April 15, 2021. Revised July 26, 2021. Accepted July 26, 2021.

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

## Dynamic changes in betanin content during the growing season of table beet: their interplay with abiotic factors

D.V. Sokolova

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

✉ [dianasokol@bk.ru](mailto:dianasokol@bk.ru)

**Abstract.** The table beet, a widespread edible root crop known for its medicinal and antioxidant properties, early maturation, good shelf life, and high contents of bioactive compounds, vitamins and minerals, is used for the production of a natural red food dye. The relevance of this study is dictated by the lack of knowledge about the dynamic changes in the content of betanin during the growing season when developing table beet cultivars with a focus on pigment extraction. The article presents the results of a study of 29 red-colored table beet accessions from the collection of the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR). Dynamic changes in the content of the pigment during the growing season were observed on two beet accessions, cvs. 'Russkaya odnosemyannaya' and 'Bordo odnosemyannaya'. Four pH versions of the buffer solution were tested, and the test results are presented. A buffer solution with pH 6.5 is recommended for research purposes. The amplitude of variability in the content of betanin in the peel (39.9–239.2 mg/100 g) and flesh (14.4–127.5 mg/100 g) of beets was determined. It was confirmed that the content of betanin in the peel exceeded that in the flesh in all samples. A positive relationship between these indicators was revealed ( $r = 0.74$ ,  $p \leq 0.05$ ). It was found that betanin accumulation did not occur in beet roots during the growing season. The pigment showed considerable fluctuations associated with abiotic environmental factors. Correlation analysis showed a significant positive relationship between air temperature and betanin content in the root flesh ( $r = 0.32$ – $0.31$ ,  $p \leq 0.05$ ). A negative impact of environmental temperature on betanin content in the peel manifested itself on the third day ( $r = -0.34$ ... $-0.35$ ,  $p \leq 0.05$ ). The negative response to precipitation was less expressed in cv. 'Bordo odnosemyannaya' due to the genotype's more active metabolism and plasticity. Structural morphological features of the photosynthetic apparatus were described for the tested accessions, and their interrelations with the studied character were specified. Recommendations are given concerning the choice of a planting pattern and the timing of table beet harvesting for pigment extraction.

Key words: betanin; natural food coloring; dynamics; peel; flesh; *Beta vulgaris* L.; environmental factors.

**For citation:** Sokolova D.V. Dynamic changes in betanin content during the growing season of table beet: their interplay with abiotic factors. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):30-39. DOI 10.18699/VJGB-22-05

## Динамические изменения содержания бетанина в столовой свекле в течение вегетационного периода: их взаимодействие с абиотическими факторами

Д.В. Соколова

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

✉ [dianasokol@bk.ru](mailto:dianasokol@bk.ru)

**Аннотация.** Для производства натурального пищевого красителя красного цвета используется широко распространенная корнеплодная культура – столовая свекла, характеризующаяся лечебными, антиоксидантными свойствами, скороспелостью, длительной сохранностью корнеплодов, высоким содержанием биологически активных веществ, витаминов и минеральных элементов. Актуальность исследования продиктована недостатком знаний о динамических изменениях в содержании бетанина в течение вегетации при создании сортов свеклы, ориентированных на получение красителя. В работе приведены результаты изучения 29 краснокрашенных образцов коллекции столовой свеклы Всероссийского института генетических ресурсов растений им. Н.И. Вавилова. Наблюдение за динамическими изменениями содержания пигмента в процессе вегетации выполнено на двух образцах свеклы: 'Русская односемянная' и 'Бордо односемянная'. Приведены результаты тестирования четырех вариантов pH буферного раствора. Рекомендован буферный раствор с pH 6.5. В результате исследования определена амплитуда изменчивости содержания бетанина в кожице (39.9–239.2 мг/100 г) и мякоти (14.4–127.5 мг/100 г) корнеплодов столовой свеклы. Подтверждено, что содержание бетанина в кожице у всех образцов превышало его содержание в мякоти. Выявлена положительная взаимосвязь этих показате-

лей ( $r = 0.74$ ,  $p \leq 0.05$ ). Установлено, что в процессе вегетации значимой аккумуляции бетанина в корнеплодах не происходит. Показаны значительные колебания пигмента, сопряженные с абиотическими факторами среды. Определены корреляционные связи между температурой воздуха и бетанином в мякоти корнеплода ( $r = 0.32-0.31$ ,  $p \leq 0.05$ ). Отрицательный эффект температуры среды на бетанин в кожце проявлялся на третьи сутки ( $r = -0.34...-0.35$ ,  $p \leq 0.05$ ). Негативная реакция на осадки была менее выражена у сорта 'Бордо односемянная', благодаря более активному метаболизму и пластичности генотипа. Описаны морфологические особенности строения фотосинтетического аппарата опытных образцов, отмечены взаимосвязи с изучаемым признаком. Даны рекомендации по выбору схемы посадки и сроков уборки урожая столовой свеклы для выделения красителя.

Ключевые слова: бетанин; натуральный пищевой краситель; динамика; кожа; мякоть; *Beta vulgaris* L.; факторы среды.

## Introduction

Red color shades for food production are mainly supplied by two groups of plant pigments: anthocyanins and betalains. Most flowering plants generate purple pigments, called anthocyanins (Yudina et al., 2021). The exceptions are representatives of several families in the order *Caryophyllales*: they synthesize betalains.

Betanin (betanidine-5-O- $\beta$ -glucoside) is the main pigment (70–95 %) in the betalain group (von Elbe, 2001; Sawicki et al., 2016). It is a glycoside: its saccharide part is glucose, and its aglycone is betanidine. Betanin is a nontoxic compound that exhibits pronounced anti-inflammatory, anticarcinogenic and antioxidant properties, which is the reason why the interest in it is growing not only among food producers but also in the pharmaceutical and cosmetic sectors (Jiratanan, Liu, 2004; Tesoriere et al., 2004; Stintzing, Carle, 2007). An important advantage of betanin over anthocyanins is its stability in the pH range from 3 to 7, which makes it possible to use it as a color additive with both acidic and neutral media (Herbach et al., 2006). At the same time, a significant drawback of betanin is the degradation of the pigment upon heating – as a result of decarboxylation, the molecule of betanin loses its properties and is converted into neobetainin (Aztatzi-Rugiero et al., 2019).

The main sources of betanin are roots of table beet (*Beta vulgaris* L. ssp. *vulgaris* var. *conditiva* Alef.), fruits of prickly pear (*Opuntia vulgaris* Mill.), and red-colored forms of amaranth (*Amaranthus* L.) (Cai et al., 1998; Castellanos-Santiago, Yahia, 2008). The dominant place among them is occupied by the cultivated beet; the other sources of this pigment cannot compete with beet due to its high yield (50–60 t/ha), environmental plasticity, and high produce of betanin (Stintzing et al., 2000; Sokolova, 2018).

The history of the use of betanin for coloring food products began in the early 20th century: the pigment was added to pastries, dry mixes, and dairy and meat products. The dye is known to be used in the form of a juice concentrate and a dry powder obtained through freeze- or spray-drying (Nemzer et al., 2011). Publications analyzing the effect of the processing of beet raw materials on the pigment content include a mandatory peeling of the beetroot in their guidelines (Azeredo et al., 2009; Burak, Zavaley, 2020), which is certainly necessary for the preparation of juices, for baby food and healthy diets. For dye production, however, the unpeeled roots undergo blanching (Frolov, Chizhik, 1997).

Generation and accumulation of betalain pigments, including betanin, in table beet plants are considered to be a dynamic process that depends not only on the specific genotype and

the phase of ontogenesis but also on various environmental factors, the maturity of roots, their size, agricultural practice, and soil fertility (Mglinets, Osipova, 2010; Vulić et al., 2013). Betalain stability is affected by numerous external and internal factors: temperature, acidity, presence/absence of light, oxygen, enzymes, nitrogen, metal cations, and the degree of glycosylation and acylation. Such data, in most cases, were obtained when studying the pigment extracted from mature roots (Saguy et al., 1978; Saguy, 1979; Schliemann, Strack, 1998; Herbach et al., 2006).

The objective of this study was to trace the dynamic changes in the pigment content during the growing season separately for the peel and flesh of table beet roots and measure the effect size of environmental factors.

## Materials and methods

A set of 29 accessions of red-colored table beet (*Beta vulgaris* L. var. *conditiva* Alef.) from the collection of the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR) served as the material for the research. Dynamic changes in betanin content during the growing season were observed on two beet accessions: cvs. 'Russkaya odnosemyannaya' (k-3698) and 'Bordo odnosemyannaya' (k-3151), identified as the most promising during the screening and ecogeographic study of the VIR collection in 2008–2018 (Sokolova, 2019; Sokolova, Solovieva, 2019).

Field experiments were conducted according to a unified methodology (Burenin, 1989) in 2020 at Pushkin and Pavlovsk Laboratories of VIR (Pushkin, St. Petersburg, Russia). Soils in Pushkin are mainly sod podzols and sandy loams. The accessions were cultivated under natural conditions, without fertilizers or pesticides. The area of record plots was 24 m<sup>2</sup>. Sowing was done manually on May 28, and harvesting on September 24. The distance between rows was 70 cm, with 6–8 cm between plants in a row.

Observations of the weather conditions during the growing season were conducted at the VIR Hydrometeorological Station. Weather conditions in 2020 were favorable for table beet cultivation, with moderate air temperatures throughout the entire growing period. The sum of active temperatures ( $>+10$  °C) from May 20 to September 24 was 2009 °C; the precipitation amount for the same period was 290 mm, i. e., 36 mm lower than the mean values for many years.

For the analysis of dynamic changes in betanin content, roots of cvs. 'Russkaya odnosemyannaya' (k-3698) and 'Bordo odnosemyannaya' (k-3151) were collected twice a week from July 13 to September 24, 2020. The pigment content was

analyzed using the peel (cut with a knife at a depth of 1–2 mm) and the flesh of ten roots separately. All measurements were taken within 3 hours after removing the roots from the soil. The juice was squeezed from the test samples using a Bork JU CUP 21085 WT juicer (Germany) and filtered through a membrane filter (0.45 microns). The filtered juice (1–1.5 g per sample) was weighed and diluted with a phosphate buffer (pH 6.5) up to the mark of 100 ml. Betanin content was measured spectrophotometrically at a wavelength of 538 nm. Measurements at 600 nm were used to correct for impurities. The absorption peak at 538 nm reflects the structure and is used to analyze betanin without isolating specific pigments. The filtrate from the roots was studied on a Shimadzu UV-1800 double-beam spectrophotometer (Japan). Betanin concentrations ( $C_B$ ) were measured according to Nilsson (1970) as follows:

$$C_B = \frac{(A_{538} - A_{600}) \times V}{1120 \times m},$$

where:  $A_{538}$  is the optical density at 538 nm;  $A_{600}$  is the optical density at 600 nm;  $V$  is the dilution ratio;  $m$  is the mass of the sample, g; 1120 is the specific absorption of 1 betanin solution in a 1 cm cuvette.

The data were statistically processed using the Excel and Statistica 8.0 software. The variability in the structure of relationships among characters was assessed using factor analysis. Factor loadings were calculated using the principal component method. The values of the Pearson correlation coefficient at  $r < 0.3$  were considered as weak,  $0.3 < r < 0.5$  as moderate,  $0.5 < r < 0.7$  as conspicuous,  $0.7 < r < 0.9$  as strong, and  $r > 0.9$  as very strong.

## Results and discussion

Betanin has the maximum light absorption in the visible spectrum range at wavelengths from 535 to 540 nm. Four pH versions were tested to select the optimal pH of the buffer solution using the example of cv. 'Russkaya odnosemyannaya' (k-3698, Russia) (Fig. 1). The spectra of betanin solutions with the same initial concentration manifested changes in the pH range from 3 to 7.5: there was a slight hypochromic shift at

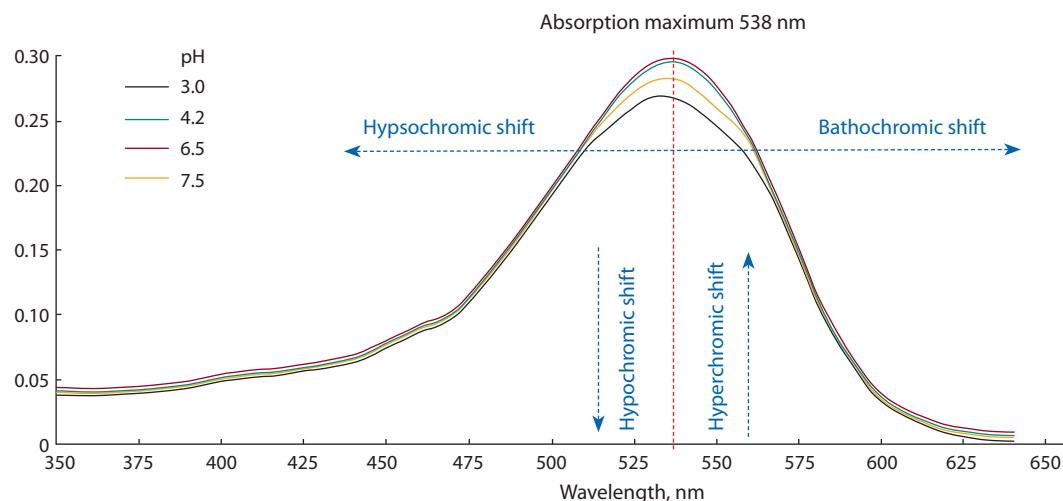
pH 3 and 7.5. No displacement of the absorption maximum was observed. A phosphate buffer solution with a pH of 6.5 was used in this study.

The phenotypic diversity of table beet is usually grouped into cultivar types. Such grouping is based on the similarity of morphological parameters (Fig. 2). This study included table beet accessions of six cultivar types.

The group of accessions representing the Crosby cultivar type, characterized by the rounded shape of the taproot, demonstrated the highest yield. The testing of 29 red-colored table beet accessions showed that on September 24 the average yield was 16.5 kg/10 m<sup>2</sup> (Table 1). The yield indicator varied significantly depending on the genotype of an accession ( $p < 0.05$ ). The average weight of one taproot was 127.9 g. Variation of this indicator within each cultivar was insignificant (coefficient of variation:  $CV < 33.3\%$ ), attesting to the alignment of the populations. Cv. 'Russkaya odnosemyannaya' (k-3698) and the local cultivar population from Kazakhstan (k-3885) were the exceptions: their coefficients of variation were 35.1 and 40.1 %, respectively.

The betanin content at the time of harvesting varied significantly, and the range of variations was wide. The average value was 116.9 mg/100 g in the peel, and 58.9 mg/100 g in the root flesh. The pigment content was observed to depend on the morphological type of the root and the intensity of its color. Thus, our earlier conclusions about the preference of the rounded or roundish oval root shape for breeding for high betanin content (Sokolova, Solovieva, 2019) were confirmed. It is worth mentioning that there were exceptions among the flattish round accessions, which, as a rule, represented early maturing forms with relatively low betanin content (Sokolova, 2019). For example, the cultivar of American origin 'Crosby Egyptian' (k-1587) and the local cultivar from Kazakhstan 'Svekla mestnaya' (k-3885) also demonstrated high levels of betanin content in their peel: 191.93 and 174.62 mg/100 g, respectively.

The highest betanin content in the peel of the root was recorded in the accessions of domestic origin: 'Bordo odnosemyannaya' (k-3151) and 'Russkaya odnosemyannaya'



**Fig. 1.** Betacyanin solution spectra of *Beta vulgaris* L. at different pH of the buffer solution (cv. 'Russkaya odnosemyannaya', k-3698).



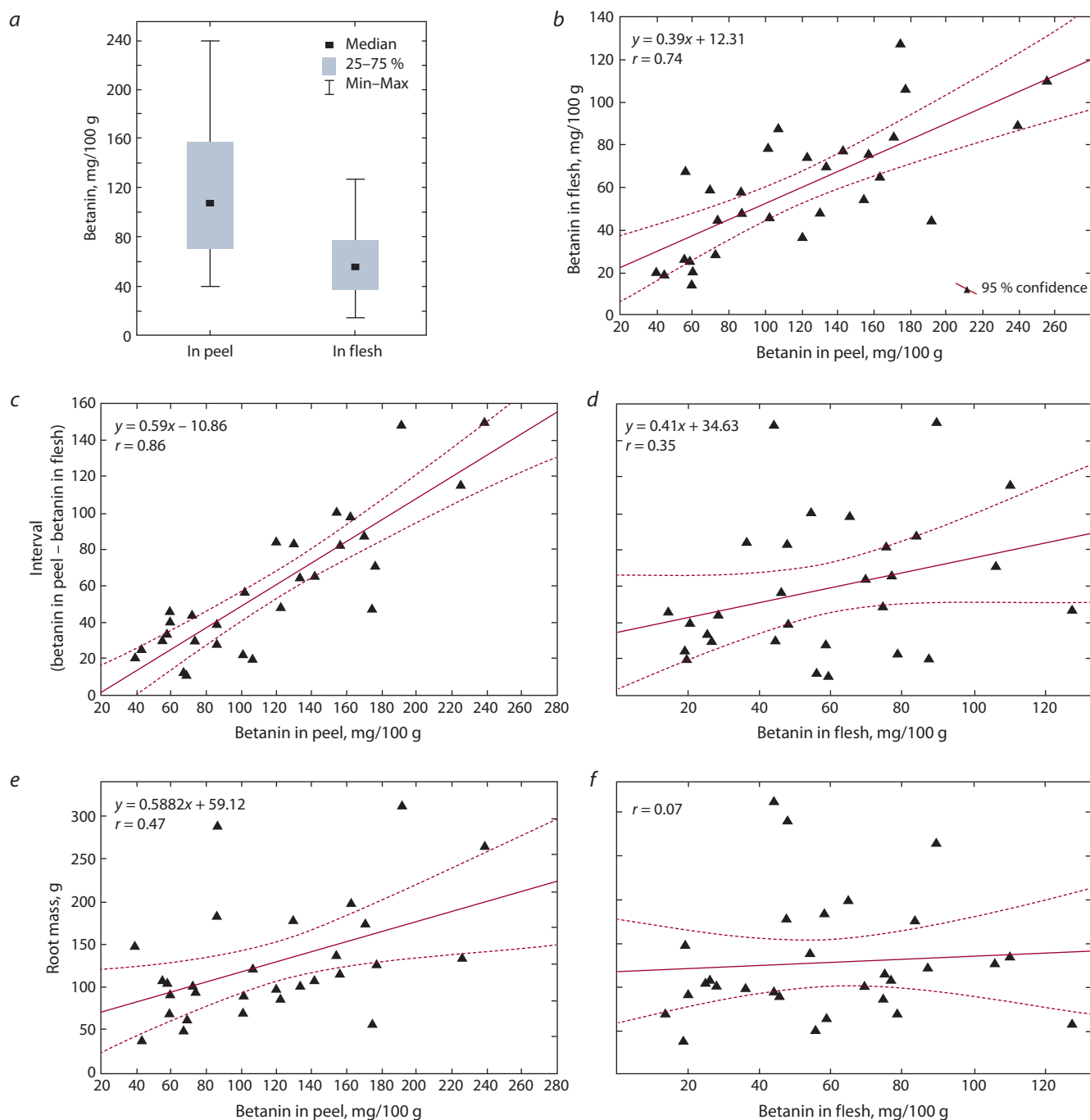


**Fig. 2.** Cultivar types of table beet accessions in the experiment.

**Table 1.** Yield and betanin content of the 29 tested table beet accessions

VIR catalogue No.	Accession name	Origin	Cultivar type	Yield, kg/10 m <sup>2</sup>	Root weight, g		Betanin*, mg/100 g	
					Mean ± SE	CV, %	Peel	Flesh
7	Rouge naine de Dell	France	Bordo	11.92	101.0 ± 6.5	17.1	72.20	28.33
10	Rouge ronde precoce	France	Crosby	8.14	68.6 ± 4.8	18.6	59.77	14.38
478	Breck's beat's all	USA	Crosby	17.46	148.3 ± 7.1	12.6	39.89	19.66
1297	Rannyaya kruglaya ploskaya	Russia	Bordo	12.27	103.9 ± 5.2	13.1	58.52	25.54
1561	Ranneye chudo	Russia	Crosby	12.63	106.9 ± 5.8	14.5	55.74	26.57
1587	Crosby Egyptian	USA	Egyptian flat	36.82	312.4 ± 17.2	14.6	191.93	44.23
1671	Donskaya ploskaya 367	Russia	Egyptian flat	7.43	62.7 ± 6.8	28.7	69.44	59.12
1757	Detroit dark red improved	France	Crosby	14.28	120.7 ± 7.1	15.5	107.22	87.42
1937	Burpee's Red Ball	USA	Crosby	16.17	137.4 ± 3.4	6.5	154.66	54.41
1938	Forma nova	USA	Cylindrical	10.62	90.3 ± 8.7	25.4	102.10	46.01
2040	Avonearly	Italy	Crosby	12.74	108.1 ± 5.4	13.3	142.45	77.17
2227	Vitenu Bordo	Lithuania	Crosby	10.03	84.6 ± 5.3	16.5	122.77	74.63
2237	Slavyanka	Russia	Crosby	10.74	90.9 ± 3.8	11.1	60.24	20.47
3012	Uniball	Netherlands	Egyptian flat	4.37	37.4 ± 4.0	28.3	43.72	19.15
3110	Regala	Netherlands	Crosby	11.92	100.7 ± 5.2	13.7	133.63	69.86
3119	84/49	Russia	Bordo	8.40	69.9 ± 4.0	15.0	101.39	78.66
3151	Bordo odnosemyannaya	Russia	Crosby	31.20	264.6 ± 14.1	14.1	<b>239.20</b>	89.69
3595	Green top	Canada	Green-leaved	13.72	98.3 ± 9.5	25.6	120.30	36.45
3600	Dwergina	Netherlands	Crosby	27.86	198.7 ± 8.5	11.3	163.11	65.30
3627	Detroit Dark Red Morse	Mexico	Bordo	7.00	50.1 ± 3.2	17.1	67.92	55.93
3698	Russkaya odnosemyannaya	Russia	Crosby	28.90	133.4 ± 17.7	35.1	<b>225.57</b>	110.35
3718	Baby Beat	Netherlands	Crosby	15.41	115.0 ± 4.6	10.7	157.02	75.39
3720	Zeppo	Netherlands	Crosby	23.45	174.9 ± 9.8	14.8	171.25	84.05
3721	Bulls Blood	Netherlands	Red-leaved	12.60	94.3 ± 7.7	21.5	73.96	44.56
3762	Bridzhit F1	Netherlands	Crosby	24.66	184.1 ± 13.4	19.3	86.53	58.57
3806	Kreolka	Russia	Crosby	23.85	178.4 ± 6.9	10.2	130.29	47.74
3879	Manolo F1	Netherlands	Crosby	38.86	289.7 ± 13.5	12.3	86.62	48.06
3885	Svekla mestnaya	Kazakhstan	Egyptian flat	7.64	57.0 ± 8.6	40.1	174.62	127.54
3887	Svekla bordovaya	Kazakhstan	Crosby	16.88	126.0 ± 7.7	16.1	177.09	106.46
Mean ± SE				16.5 ± 1.7	127.9 ± 12.8		116.9 ± 10.7	58.9 ± 5.5
LSD <sub>05</sub>				6.65				25.0

\* Values are obtained from 10 roots per accession.



**Fig. 3.** Mean betanin content in the peel and flesh of table beet roots (a). Linear correlation analysis: correlation between betanin content in the peel and that in the flesh of table beet (b); betanin in the peel and the ratio of the two parameters (betanin in the peel to betanin in the flesh) (c); betanin in the flesh and the ratio of the two parameters (d); betanin in the peel and the root weight (e); betanin in the flesh and the root weight (f).

(k-3698) – 239.20 mg/100 g and 225.57 mg/100 g, respectively. These genotypes are of interest for further breeding practice aimed at increasing the pigment content.

Separate analyses of betanin content in the peel and flesh of beet roots revealed statistically significant differences. The pigment content in the peel of all the studied accessions exceeded this parameter in the flesh (Fig. 3, a). The smallest difference was observed in ‘Donskaya ploskaya 367’ (k-1671): 10.3 mg/100 g. The greatest difference was demonstrated by cv. ‘Bordo odnosemyannaya’: 149.5 mg/100 g, with the ratio

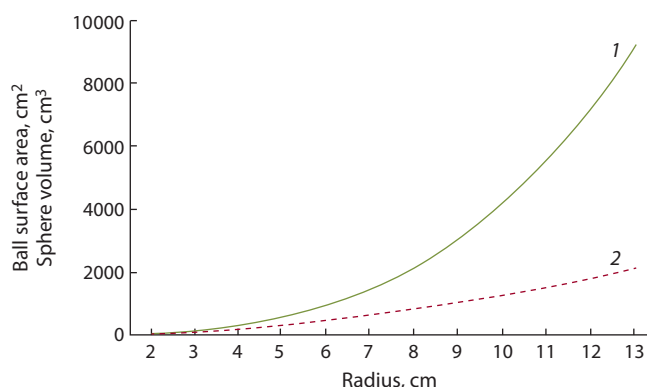
73 % of betanin in the peel to 27 % in the root flesh. Correlation analysis showed the presence of a significant positive relationship between the pigment contents in the peel and flesh ( $r = 0.74$ ) (see Fig. 3, b). At the same time, when betanin content in the peel was higher, the span between the parameters also increased ( $r = 0.86$ ) (see Fig. 3, c). Unlike betanin in the peel, no significant correlation was found between the pigment content in the flesh and the variations between the parameters, which attested to its higher stability in the flesh during the growing season (see Fig. 3, d). Thus, an increase

in the total betanin content depends, first of all, on the content of the pigment in the peel.

No significant correlation ( $r > 0.5$ ) was found between the root weight and the betanin content in the peel and flesh (see Fig. 3, e, f).

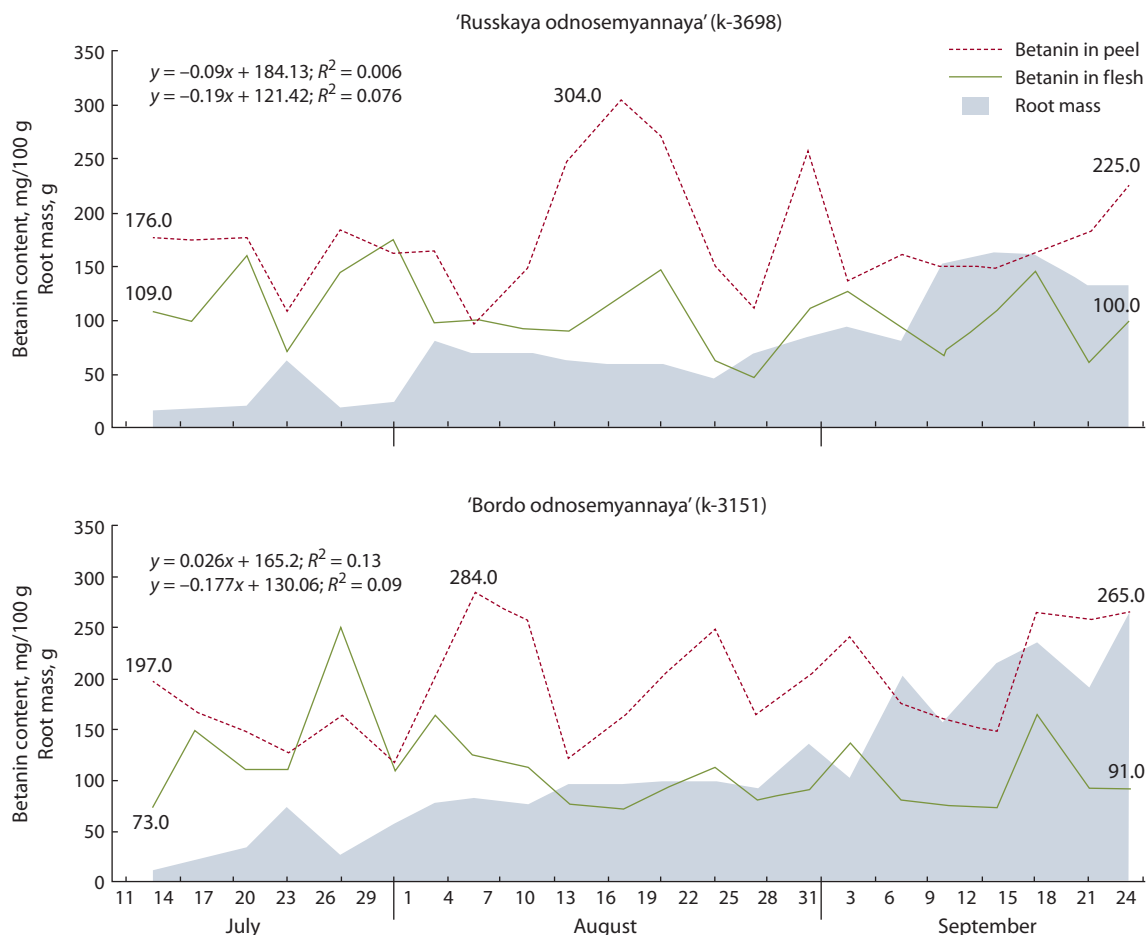
It is important for the production of dye from beet roots that the pigment content in the total biomass is high. If the shape of the table beet root is schematically regarded as a sphere, it is possible to calculate how the ratio between its volume and surface area will change depending on the radius. The diagram (Fig. 4) demonstrates that the increase in the volume of the sphere has a cubic dependence, while the expansion of the surface area has a quadratic one. Extrapolating the theoretical layout to the object of this study, it can be argued that, assuming that the thickness of the peel is constant in mature, ready-to-harvest beet roots, the portion of the flesh significantly exceeds that of the peel, which will ultimately have a negative effect on the output of the pigment. Meanwhile, the share of the peel in smaller table beet roots tends to be higher than that in large ones. It can be therefore assumed that the total yield of betanin extracted from smaller roots will be higher, provided that the pigment content is high at the time of harvesting.

To understand the dynamics of root growth and pigment accumulation separately in the peel and flesh, we conducted



**Fig. 4.** Dependence of the volume (1) and surface area (2) of a sphere on its radius.

a comparative analysis of two promising table beet accessions: 'Russkaya odnosemyannaya' (k-3698) and 'Bordo odnosemyannaya' (k-3151). The material for the analysis was collected 22 times at regular intervals from the moment when the roots exceeded the weight of 10 g. No significant accumulation of betanin in the peel or flesh of both cultivars was recorded during the growing season (Fig. 5), which confirmed the results of the earlier ecogeographic study (Sokolova, 2019). At



**Fig. 5.** Dynamic changes in betanin content and root weight in the table beet cultivars 'Russkaya odnosemyannaya' and 'Bordo odnosemyannaya'.



**Table 2.** Metric parameters of the leaf surface area of the tested table beet accessions at the time of harvesting

Metric parameters	'Bordo odnosemyannaya', k-3151	'Russkaya odnosemyannaya', k-3698
Leaf		
Number, pcs.	14	12
Length, cm	15	11
Width, cm	6	5
Area of one leaf, cm <sup>2</sup>	67.5	41.2
Total leaf surface area, cm <sup>2</sup>	945.0	494.2
Petiole length, cm	14	12

the same time, the range of variations in individual periods of development was statistically significant ( $p < 0.05$ ). The dynamics had a wavelike pattern and was fairly synchronous for both cultivars, both in the peel and in the root flesh. It can be noted that abrupt upward movements of betanin content and root weight increases were often directed differently. Besides, the pigment content in the flesh changed more conservatively than in the peel.

The pathway of betanin biosynthesis is known to be quite flexible and strongly influenced by exogenous factors (Tzin, Galili, 2010; Cabanes et al., 2014; Sakuta, 2014; Esatbeyoglu et al., 2015). An excess in the content of betanin in the flesh over its content in the peel for both cultivars was observed only once during the entire growing season (the period from July 27 to July 31). Concurrently, its maximum in the flesh

for the entire growing season was recorded. Abundant continuous rainfall that settled from July 21, together with mean daily air temperatures above +15 °C, caused inhibition of root development, which most likely led to an increase in betanin content in the root flesh.

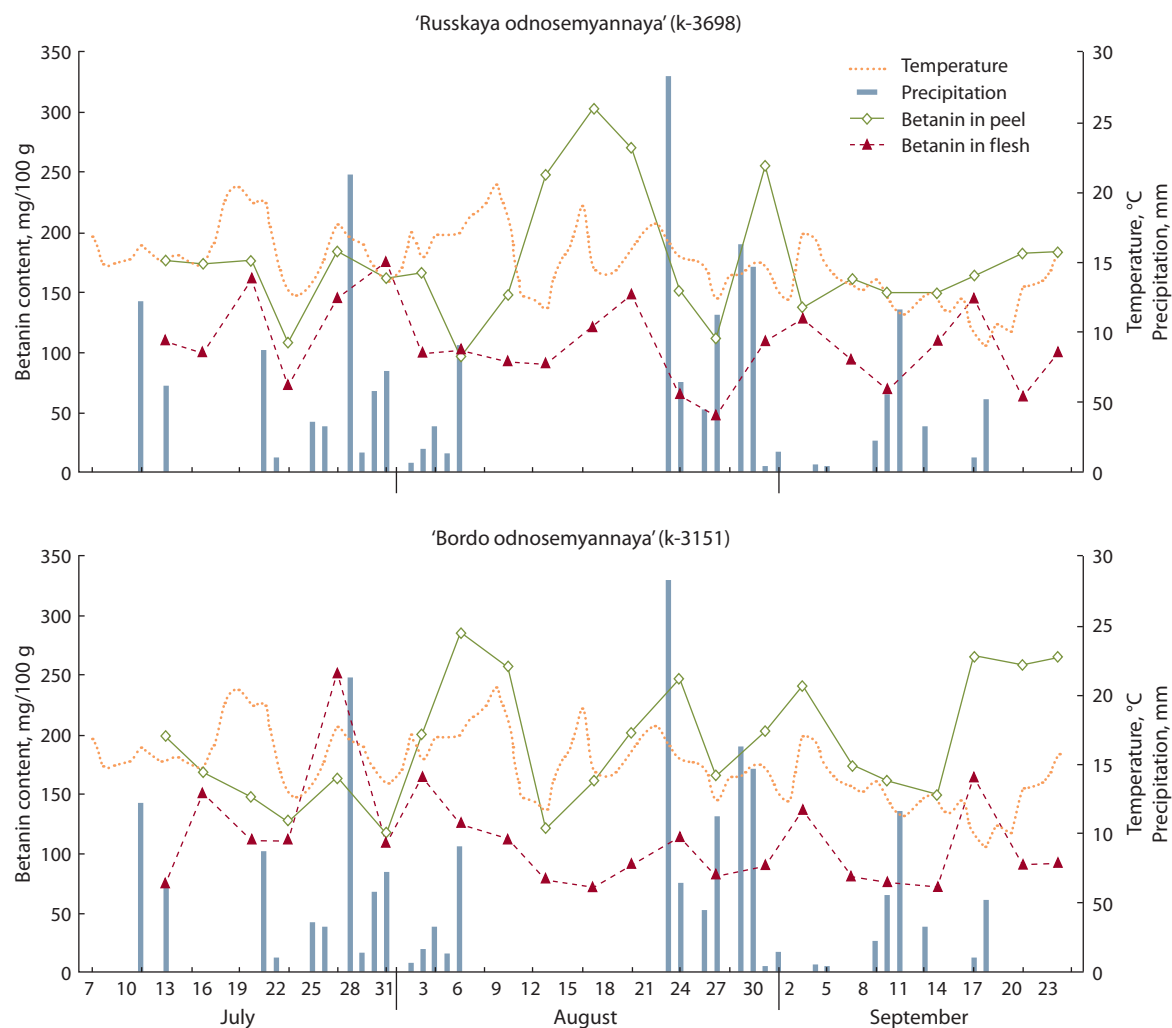
Unlike the pigment content, an increase in the root biomass positively correlated with the duration of the growing season ( $R^2 = 0.76–0.86$ ).

The activity of photosynthesis and, as a consequence, plant metabolism is associated with the surface area of the plant's leaf biomass. The plants of cv. 'Russkaya odnosemyannaya' had small, pigmented and slightly wavy leaf blades with shorter petioles (Table 2, Fig. 6). The number of leaves per plant was less than that of cv. 'Bordo odnosemyannaya'. At the time of harvesting, the leaf surface area of 'Russkaya



**Fig. 6.** Anatomical and morphological structure and appearance of field crops of experimental table beet accessions 'Russkaya odnosemyannaya' (k-3698) and 'Bordo odnosemyannaya' (k-3151).





**Fig. 7.** Dynamic changes in the characteristics of weather conditions and the content of betanin in the table beet cultivars 'Russkaya odnosemyannaya' and 'Bordo odnosemyannaya'.

odnosemyannaya' was 494.2 cm<sup>2</sup>, which was almost twice less than that of 'Bordo odnosemyannaya'. These characteristics are important for understanding the differences in the cultivars' response to abiotic environmental factors.

Mean daily air temperature and rainfall are two of the most influential and constantly fluctuating environmental factors. Table beet is a fairly adaptive crop: it is capable of producing harvests both in the southern regions of Russia and under the conditions of the north (Burenin et al., 2013). It was this quality that made the beet crop widespread. The study has shown that the high plasticity of the genotype is a valuable property that allows the crop to safely endure unfavorable periods. Figure 7 presents the dynamics of the pigment content with varying rainfall and air temperature values. The growing season in 2020 was characterized by two stressful periods with complete absence of precipitation: from July 14 to 20, and from August 7 to 22. There were also two periods with prolonged rainfall: from July 21 to August 6, and from August 23 to September 1. Significant precipitation amounts were registered within one day on July 28 and August 23 (21.2 and 28.2 mm, respectively). The same weather held on for 9 days, although with less intensive rains. It negatively

affected betanin content in the table beet peel and flesh. The response of both cultivars in terms of their pigment content was practically similar, but there were some differences. The graph also shows a lag of 7 to 10 days in the response of 'Russkaya odnosemyannaya' (see Fig. 7).

To clarify the relationships between the studied parameters, a correlation analysis of the cultivars was conducted in dynamics (3-day and 6-day shifts of environmental parameters) (Table 3). Both cultivars demonstrated a moderate negative correlation between root weight and air temperature, which persisted for 6 days. Higher temperatures increased the transpiration of plants, arresting the growth of roots and raising the pigment content in the root flesh ( $r = 0.32-0.31$ ). Three days later, the effect of an increase in the pigment concentration in response to temperature was no longer manifested.

Betanin content in the peel also negatively correlated with air temperature increases: this effect did not appear immediately, but on the third day. This correlation was retained by cv. 'Bordo odnosemyannaya' until the 6th day ( $r = -0.48$ ).

Precipitation in general was characterized by weak negative correlations with all the studied characters. Rainfall had a weak correlation with the pigment content ( $r < 0.3$ ). Betanin

**Table 3.** The correlation matrix for the indicators of the table beet cultivars

Parameter	Root weight	Betanin in peel	Betanin in flesh
Same day			
Temperature	<b>-0.50/-0.47*</b>	0.02/0.27	<b>0.32/0.31</b>
Precipitation	-0.16/-0.23	<b>-0.48/0.05</b>	-0.26/-0.16
Root weight		-0.08/ <b>0.36</b>	-0.28/-0.30
Betanin in peel			<b>0.31/0.15</b>
After 3 days			
Temperature	<b>-0.47/-0.39</b>	<b>-0.34/-0.35</b>	-0.07/-0.10
Precipitation	-0.14/-0.26	-0.05/0.08	-0.19/-0.12
Root weight		-0.08/ <b>0.36</b>	-0.28/-0.30
Betanin in peel			<b>0.31/0.15</b>
After 6 days			
Temperature	<b>-0.40/-0.54</b>	0.10/ <b>-0.48</b>	-0.13/0.21
Precipitation	-0.12/-0.17	-0.01/0.12	0.17/0.19
Root weight		-0.08/ <b>0.36</b>	-0.28/-0.30
Betanin in peel			<b>0.31/0.15</b>

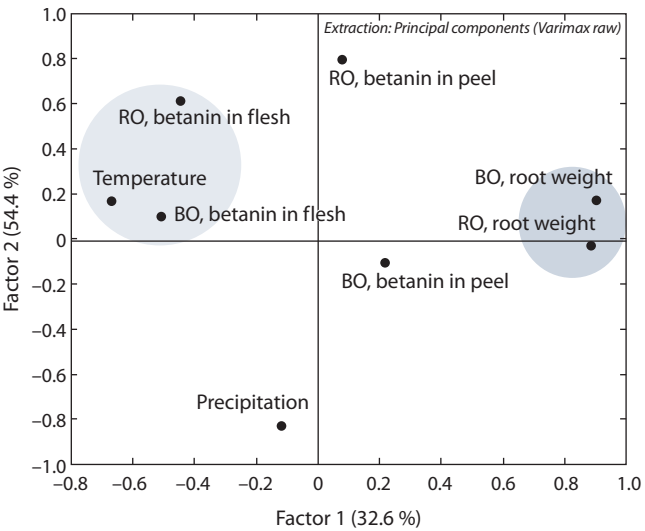
\*‘Russkaya odnosemyannaya’/‘Bordo odnosemyannaya’.

biosynthesis in the peel of ‘Russkaya odnosemyannaya’ reacted negatively to precipitation ( $r = -0.48$ ). Three days later, however, the plants adapted and this correlation was not observed. The differences in the response of the cultivars may be explained by the difference in the photosynthetic surface area of the genotypes.

Research into the structure of relationships among the studied parameters (PCA) disclosed two unbalanced factors (Fig. 8). The components of Factors 1 and 2 together explain 87 % of the total variance: 32.6 and 54.4 %, respectively. Factor 1 included the main parameter of productivity – the weight of one root. Factor 2 combined the indicators “temperature” and “betanin in flesh”, confirming the results of the correlation analysis. The indicators within Factor 2 had the closest relationship, i.e., they accounted for most of the variance. At the same time, other variables also had a fairly high variance, which confirms the significant contribution of each of them. Thus, we can conclude that the process of betanin biosynthesis in table beet is extremely flexible and strongly affected by environmental factors.

Conclusion

The results of the study make it possible to conclude that during the growing season there was no accumulation of betanin in table beet roots. The pigment content was associated primarily with the original genotype of an accession. Mean values of betanin content in the studied accessions were: 116.9 mg/100 g in the peel, and 58.9 mg/100 g in the flesh. The pigment content in the peel exceeded its content in the flesh in all accessions. With a positive correlation be-



**Fig. 8.** The scatter chart showing the contribution of principal components in the factor analysis to generalized variance. RO – ‘Russkaya odnosemyannaya’, BO – ‘Bordo odnosemyannaya’.

tween these indicators ( $r = 0.74$ ), an increase in the span of the ratio between them depended precisely on betanin in the peel ( $r = 0.86$ ).

The pigment content was subject to significant fluctuations during the growing season. The activity of metabolic processes that depended on the photosynthetic surface of plants was of great importance for the response of table beet genotypes to various environmental factors.

The pigment biosynthesis process was extremely sensitive to weather conditions, especially air temperature. An increase in betanin content in the peel in response to an increase in air temperature manifested itself on the third day. An increase in the mass of one root negatively correlated with the content of betanin in the root flesh, which, in its turn, was closely associated with air temperature.

Table beet cultivars with a round or flattish round shape of their root are the most suitable for breeding targeted at higher betanin content. Smaller beet roots are better suited for obtaining the maximum betanin yield. Denser planting patterns or earlier harvesting are recommended for their cultivation, as well as optimal timing. When choosing a specific date for harvesting, it is necessary to focus on the characteristics of weather conditions. The domestic table beet cultivars ‘Bordo odnosemyannaya’ (k-3151) and ‘Russkaya odnosemyannaya’ (k-3698) are recommended for extracting the betanin dye, because under favorable conditions they are capable of yielding high amounts of the pigment.

References

Azeredo H.M.C., Pereira A.C., de Souza A.C.R., Gouveia S.T., Mendes K.C.B. Study on efficiency of betacyanin extraction from red beetroots. *Int. J. Food Sci. Technol.* 2009;44(12):2464-2469. DOI 10.1111/j.1365-2621.2009.02037.x.

Aztatzi-Rugiero L., Granados-Balbuena S.Y., Zainos-Cuapio Y., Ocaranza-Sánchez E., Rojas-López M. Analysis of the degradation of betanin obtained from beetroot using Fourier transform infrared spectroscopy. *J. Food Sci. Technol.* 2019;56(8):3677-3686. DOI 10.1007/s13197-019-03826-2.

- Burak L.Ch., Zavaley A.P. Influence of the method of processing table beets on the antioxidant activity of beetroot and mixed juices. *Tekhnologii Pishchevoy i Pererabatyvayushchey Promyshlennosti APK – Produkty Zdorovogo Pitaniya = Technologies of Food and Processing Industry of AIC: Healthy Food*. 2020;4:51-61. (in Russian)
- Burenin V.I. (Ed.). Methodological Guidelines for the Study and Maintenance of the World Collection of Root Crops. Leningrad: VIR Publ., 1989. (in Russian)
- Burenin V.I., Sokolova D.V., Emelyanov A.V. Variability of quantitative traits of red beet varieties in an ecogeographical study. *Sakhar-naya Svekla = Sugar Beet*. 2013;8:16-19. (in Russian)
- Cabanes J., Gandia-Herrero F., Escribano J., Garcia-Carmona F., Jimenez-Atienzar M. One-step synthesis of betalains using a novel betalamic acid derivatized support. *J. Agric. Food Chem.* 2014; 62(17):3776-3782. DOI 10.1021/jf500506y.
- Cai Y., Sun M., Wu H., Huang R., Corke H. Characterization and quantification of betacyanin pigments from diverse *Amaranthus* species. *J. Agric. Food Chem.* 1998;46(6):2063-2070. DOI 10.1021/jf9709966.
- Castellanos-Santiago E., Yahia E.M. Identification and quantification of betalains from the fruits of 10 Mexican prickly pear cultivars by high-performance liquid chromatography and electrospray ionization mass spectrometry. *J. Agric. Food Chem.* 2008;56(14):5758-5764. DOI 10.1021/jf800362t.
- Esatbeyoglu T., Wagner A.E., Schini-Kerth V.B., Rimbach G. Betanin – a food colorant with biological activity. *Mol. Nutr. Food Res.* 2015; 59(1):36-47. DOI 10.1002/mnfr.201400484.
- Frolov V.L., Chizhik Ju.L. Method of Preparing Food Dye from Beet. Invention patent. RU 2081136 C1. Publ. 10.06.1997. (in Russian)
- Herbach K.M., Stintzing F.C., Carle R. Stability and color changes of thermally treated betanin, phyllocactin, and hylocerenin solutions. *J. Agric. Food Chem.* 2006;54(2):390-398. DOI 10.1021/jf051854b.
- Jiratanan T., Liu R.H. Antioxidant activity of processed table beets (*Beta vulgaris* var. *conditiva*) and green beans (*Phaseolus vulgaris* L.). *J. Agric. Food Chem.* 2004;52(9):2659-2670. DOI 10.1021/jf034861d.
- Mglinets A.V., Osipova Z.A. Formation of root color and its genetic control in fodder beet. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2010;14(4):720-728. (in Russian)
- Nemzer B., Pietrkowski Z., Spórna A., Stalica P., Thresher W., Michałowski T., Wybranie S. Betalainic and nutritional profiles of pigment-enriched red beet root (*Beta vulgaris* L.) dried extracts. *Food Chem.* 2011;127(1):42-53. DOI 10.1016/j.foodchem.2010.12.081.
- Nilsson T. Studies into the pigments in beetroot (*Beta vulgaris* L. ssp. *vulgaris* var. *rubra* L.). *Lantbrukshogskolans Annaler*. 1970;36: 179-219.
- Saguy I. Thermostability of red beet pigments (betanine and vulgaxanthin-I): influence of pH and temperature. *J. Food Sci.* 1979;44(5): 1554-1555. DOI 10.1111/j.1365-2621.1979.tb06488.x.
- Saguy I., Kopelman I.J., Mizrahi S. Thermal kinetic degradation of betanin and betalamic acid. *J. Agric. Food Chem.* 1978;26(2):360-362. DOI 10.1021/jf60216a052.
- Sakuta M. Diversity in plant red pigments: anthocyanins and betacyanins. *Plant Biotechnol Rep.* 2014;8:37-48. DOI 10.1007/s11816-013-0294-z.
- Sawicki T., Bączek N., Wiczowski W. Betalain profile, content and antioxidant capacity of red beetroot dependent on the genotype and root part. *J. Funct. Foods*. 2016;27:249-261. DOI 10.1016/j.jff.2016.09.004.
- Schliemann W., Strack D. Intramolecular stabilization of acylated betacyanins. *Phytochemistry*. 1998;49(2):585-588. DOI 10.1016/S0031-9422(98)00047-8.
- Sokolova D.V. The evaluation of genotype-environment interaction in red beet varieties of VIR collection. *Ovoshchi Rossii = Vegetable Crops of Russia*. 2018;6:26-30. DOI 10.18619/2072-9146-2018-6-26-30. (in Russian)
- Sokolova D.V. Environmental and geographic study of betanin accumulation in promising red beet accessions from the VIR collection. *Trudy po Prikladnoy Botanike, Genetike i Seleksii = Proceedings on Applied Botany, Genetics, and Breeding*. 2019;180(4):66-74. DOI 10.30901/2227-8834-2019-4-66-74. (in Russian)
- Sokolova D.V., Solovieva A.E. Promising starting material for selection of beet varieties with a high content of betanin. *Agrarnaya Rossiya = Agricultural Russia*. 2019;8:26-32. DOI 10.30906/1999-5636-2019-8-26-32. (in Russian)
- Stintzing F.C., Carle R. Betalains – emerging prospects for food scientists. *Trends Food Sci. Technol.* 2007;18(10):514-525. DOI 10.1016/j.tifs.2007.04.012.
- Stintzing F.C., Schieber A., Carle R. Rote Bete als färbendes Lebensmittel – eine Bestandsaufnahme. *Obst- Gemüse- Kartoffelverarb.* 2000;85:196-204.
- Tesorieri L., Allegra M., Butera D., Livrea M.A. Absorption, excretion, and distribution of dietary antioxidant betalains in LDLs: potential health effects of betalains in humans. *Am. J. Clin. Nutr.* 2004;80(4): 941-945. DOI 10.1093/ajcn/80.4.941.
- Tzin V., Galili G. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Mol. Plant*. 2010;3(6):956-972. DOI 10.1093/mp/ssq048.
- von Elbe J.H. Betalains. *Curr. Protoc. Food Anal. Chem.* 2001;F3.1.1–F3.1.7. DOI 10.1002/0471142913.faf0301s00.
- Vulić J.J., Čebović T.N., Čanadanović V.M., Četković G.S., Djilas S.M., Čanadanović-Brunet J.M., Velićanski A.S., Cvetković D.D., Tumbas V.T. Antiradical, antimicrobial and cytotoxic activities of commercial beetroot pomace. *Food Funct.* 2013;4(5):713-721. DOI 10.1039/C3FO30315B.
- Yudina R.S., Gordeeva E.I., Shoeva O.Y., Tikhonova M.A., Khlestkina E.K. Anthocyanins as functional food components. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2021;25(2):178-189. DOI 10.18699/VJ21.022. (in Russian)

## ORCID ID

D.V. Sokolova orcid.org/0000-0002-9967-7454

**Conflict of interest.** The author declares no conflict of interest.

Received July 19, 2021. Revised October 10, 2021. Accepted November 18, 2021.

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

# The relationship of lamins with epigenetic factors during aging

R.N. Mustafin<sup>1</sup>✉, E.K. Khusnutdinova<sup>2</sup>

<sup>1</sup> Bashkir State Medical University, Ufa, Russia

<sup>2</sup> Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia

✉ ruji79@mail.ru

**Abstract.** The key factor of genome instability during aging is transposon dysregulation. This may be due to senile changes in the expression of lamins, which epigenetically modulate transposons. Lamins directly physically interact with transposons. Epigenetic regulators such as SIRT7, BAF, and microRNA can also serve as intermediaries for their interactions. There is also an inverse regulation, since transposons are sources of miRNAs that affect lamins. We suggest that lamins can be attributed to epigenetic factors, since they are part of the NURD, interact with histone deacetylases and regulate gene expression without changing the nucleotide sequences. The role of lamins in the etiopathogenesis of premature aging syndromes may be associated with interactions with transposons. In various human cells, LINE1 is present in the heterochromatin domains of the genome associated with lamins, while SIRT7 facilitates the interaction of this retroelement with lamins. Both retroelements and the nuclear lamina play an important role in the antiviral response of organisms. This may be due to the role of lamins in protection from both viruses and transposons, since viruses and transposons are evolutionarily related. Transposable elements and lamins are secondary messengers of environmental stressors that can serve as triggers for aging and carcinogenesis. Transposons play a role in the development of cancer, while the microRNAs derived from them, participating in the etiopathogenesis of tumors, are important in human aging. Lamins have similar properties, since lamins are dysregulated in cancer, and microRNAs affecting them are involved in carcinogenesis. Changes in the expression of specific microRNAs were also revealed in laminopathies. Identification of the epigenetic mechanisms of interaction of lamins with transposons during aging can become the basis for the development of methods of life extension and targeted therapy of age-associated cancer.

Key words: lamins; microRNAs; transposons; epigenetic factors.

**For citation:** Mustafin R.N., Khusnutdinova E.K. The relationship of lamins with epigenetic factors during aging. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):40-49. DOI 10.18699/VJGB-22-06

## Взаимосвязь ламинов с эпигенетическими факторами при старении

Р.Н. Мустафин<sup>1</sup>✉, Э.К. Хуснутдинова<sup>2</sup>

<sup>1</sup> Башкирский государственный медицинский университет, Уфа, Россия

<sup>2</sup> Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия

✉ ruji79@mail.ru

**Аннотация.** Дисбаланс активации транспозонов – один из важнейших факторов нестабильности генома при старении. Причинами этого явления могут быть ассоциированные с возрастом изменения экспрессии ламинов, которые влияют на эпигенетическую регуляцию мобильных генетических элементов. Взаимосвязь ламинов и транспозонов может быть обусловлена специфическим физическим контактом между молекулами, а также опосредована эпигенетическими регуляторами, такими как SIRT7, BAF и микроРНК. Характерна взаиморегуляция ламинов с мобильными элементами, которые являются источниками микроРНК, влияющими на ламины. Ламины входят в состав NURD (nucleosome remodeling deacetylase complex), взаимодействуют с гистоновыми деацетилазами и регулируют экспрессию генов без изменения структуры нуклеотидных последовательностей. Роль ядерной ламины в этиопатогенезе синдромов преждевременного старения может быть обусловлена взаимодействием с транспозонами, так как истощение ламинов приводит к активации мобильных генетических элементов. В различных клетках человека LINE1 представлены в связанных с ламинами гетерохроматиновых доменах генома, при этом SIRT7 способствует взаимодействию данного ретроэлемента с ядерной ламиной. В противовирусном ответе организмов важную роль играют как ретроэлементы, так и ядерная ламина. Это согласуется с ролью ламинов в защите как от вирусов, так и от транспозонов, которые характеризуются филогенетическим родством. Мобильные генетические элементы и ламины – вторичные мессенджеры средовых стрессорных воздействий, которые могут служить пусковыми факторами для старения и канцерогенеза. Транспозоны играют роль в развитии злокачественных новообразований, при этом происходящие от них микроРНК,



участвующие в этиопатогенезе опухолей, имеют значение в старении человека. Сходные свойства типичны для ламинов, поскольку при злокачественных новообразованиях выявлена дисрегуляция ламинов, а влияющие на них микроРНК участвуют в канцерогенезе. Изменение экспрессии специфических микроРНК отмечено также при ламинопатиях. Определение точных эпигенетических механизмов взаимодействия ламинов с мобильными генетическими элементами при старении может стать основой для разработки методов продления жизни и таргетной терапии ассоциированных с возрастом злокачественных новообразований.

Ключевые слова: ламины; микроРНК; транспозоны; эпигенетические факторы.

## Introduction

Nuclear lamina (NL) is a protein network connected to the inner side of the cell's nuclear envelope. Nuclear lamina performs structural, signaling and regulatory functions. The study of the evolution of NL components made it possible to discover a wide variety of domains and sequence architectures that go beyond the classical alpha-helices (Kollmar, 2015). Nuclear lamina is composed of lamins and associated proteins (Lemaitre, Bickmore, 2015). The main components of NL are lamins (Cibulka et al., 2012). Their genes were present at the earliest stages of eukaryotic evolution. Lamins have been identified in multicellular organisms, amoebas, and primitive opisthokonts such as *Ichthyosporea* and *Choanoflagellates* (Kollmar, 2015). Lamins are required not only to maintain the shape of the nucleus, but also to control replication and transcription. Lamins are proteins of the family of intermediate filaments (class V) with a specific structure. In mammalian cells, 4 types of lamins are expressed: A and C (splicing isoforms of one gene), B1 and B2 (Cibulka et al., 2012). Minor isoforms A $\delta$ 10 and C2 are also known. The protein precursor of lamin A is prelamin A. Its C-terminal region is post-translationally modified: farnesylated, carboxymethylated. Proteolysis of the CaaX motif of prelamin A also occurs using the metalloprotease ZMPSTE24 (encodes a metalloproteinase that metabolizes lamin A) (Wang et al., 2016). Lamins A, C, A $\delta$ 10 and C2 are encoded by *LMNA* gene (Turgay et al., 2017).

Proteins that interact with lamins and are closely associated with them functionally include LBR (lamin B receptor), BAF (barrier to autointegration factor), SUN1, SUN2, nesprin. They are involved in the structural organization of the nucleus and in the regulation of nuclear processes. LBR and endoprotease prelamin A have enzymatic properties (Meinke et al., 2014).

At the periphery of the nucleus, genomic DNA binds to lamins A and B, forming heterochromatic domains. Lamin A also binds to chromatin inside the nucleus – the nucleoplasmic environment is represented mainly by euchromatin, which indicates the role of lamin A in the regulation of gene expression of the entire genome (Briand et al., 2018). Lamin B interacts with genomic domains that are relatively gene-poor and transcriptionally inactive (Guelen et al., 2008). It is assumed that in the last common predecessor of all eukaryotes (LECA – last eukaryotic common ancestor), the configurations of the nuclear envelope and the associated NL played an important role in determining the activity of the nucleus.

In the subsequent evolution, there were changes in the mechanisms of regulation of gene expression of lamins in different taxa. The protein diversity of lamins in plants and trypanosomes is taxonomically limited, while in multicellular animals it is manifested in a wider range. The phylogenetic tree of lamina genes is characterized by vertical evolution. For example, two lamins in protists from strongly diverging taxa have targets in mammalian cell nuclei and polymerize into filamentous structures, which indicates the functional preservation of distant homologues of lamins. In certain groups of eukaryotes, a pronounced evolutionary plasticity of the structures of the NL was determined by the mechanisms of chromatin binding and epigenetic control due to a high level of divergence of lamins' homologues (Korney, Field, 2016).

Lamin changes occur with aging and in age-associated diseases such as malignant neoplasms. For example, in lung cancer, significant reductions in lamin B1 levels are found (Garvalov et al., 2019). Lamin A regulates the activity of mTOR (mechanistic target of rapamycin), the low activity of which contributes to an increase in life expectancy. Inhibition of mTOR (for example, via rapamycin) leads to the degradation of defective molecules and organelles that accumulate in cells during aging, since low autophagy activity is a characteristic feature of aging-related diseases (Cenni et al., 2020).

Lamins can be attributed to epigenetic factors, since they ensure the inheritance of the functional status of the gene. For example, the attachment of lamin B1 to specific genome loci leads to transcription suppression of genes located in them. Most differentially expressed genes are activated due to the loss of lamin B1. These include the genes of the *RET* oncogene and its corepressor GFR $\alpha$ 1. Therefore, when lamin B1 is depleted with aging, the risk of tumors increases. Another mechanism of *RET* activation is H3K27me3 histone methylation (Garvalov et al., 2019).

Epigenetic factors include chromatin remodeling, modifications of histones and DNA nucleotides, and the influence of non-coding RNAs (ncRNAs). During cellular senescence in mammals, H3K56ac and H4K16ac levels decrease, which contributes to altered gene expression, genomic instability, and telomere damage. Aging is accompanied by a global decrease in DNA methylation levels, coinciding with age-related loss of heterochromatin. As a result, the regulation of gene expression and activation of transposable elements (TEs) is disrupted. At the same time, the participation of lamins in epigenetic regulation associated with aging oc-

curs through direct interaction with chromatin in regions of specific DNA sequences called lamin-associated domains (LADs).

Lamins also promote epigenetic changes during aging through functional interactions with sirtuins. For example, lamin A enhances the deacetylase activity of SIRT1, stimulates the functioning of SIRT6 in DNA repair, and recruits histone deacetylase 2 (HDAC2). Lamins A/C are also direct participants in epigenetic regulation, since they serve as components of the NURD (nucleosome remodeling deacetylase complex). This complex also includes HDAC1, RBBP4, RBBP7 (Cenni et al., 2020). Lamins A/C interact with histone deacetylase 2 (HDAC2) and with PCAF acetyltransferase (p300-CBP associated factor) (Santi et al., 2020). Lamin B1 recruits the PRC2 (Polycomb repressive complex 2) complex, through which the H3K27me3 landscape changes with repression of specific genes involved in signaling and cell migration (Jia et al., 2019).

### The role of lamins in premature aging syndromes

The development of progeria in patients with germinal mutations in the genes of lamins proves the role of lamins in regulation of life expectancy. Progeroid laminopathies (PLs) are characterized by premature aging and death from complications of atherosclerosis such as myocardial infarction, stroke, or heart failure. PLs are usually not inherited because patients do not survive to puberty.

An important genetic feature of progeria is telomere shortening with each replication cycle (Ahmed et al., 2018). The most famous of these diseases is Hutchinson–Guilford progeria (HGP), which is characterized by significant telomere shortening. The disease is diagnosed from two years of age, when noticeable symptoms of premature aging are determined. Life expectancy in HGP patients is 10–20 years. In 90 % of HGP patients, mutations in the *LMNA* gene are determined (Ahmed et al., 2018). The frequency of occurrence of HGP is 1 in 8 million newborns (Burla et al., 2018). The most common mutation in *LMNA* gene is C1824T, which leads to accumulation of progerin (the dominant negative form of lamin A).

Interestingly, progerin also accumulates in cells during physiological aging as one of the rare splicing forms of lamin A transcripts. However, in comparison with physiological aging, more severe epigenetic changes occur in HGP, such as covalent modifications of histones with a tendency to loss of separation into hetero- and euchromatin. These epigenetic changes lead to changes in the spatial compartmentalization and conformation of chromatin in the nucleus. These processes involve microRNAs, such as miR-9, which can be used as a target for brain protection in HGP patients (Arancio et al., 2014). MiR-9 interacts with the 3'-untranslated region (UTR) of lamin A mRNA without affecting lamin C. 3'-UTR region of prelamin A contains an additional binding site for miR-9. In experiments on mice and human HeLa cell cultures, it was proved that miR-9 expression significantly reduces the levels of lamin A (Jung et al., 2012).

As with normal aging, abnormal accumulation of progerin suppresses the interactions of lamin A with SIRT1, HDAC2, and SIRT6. In addition, in patients with HGP, the regulation of the heterochromatin protein HP1 is impaired, the levels of H3K9me3 decrease, and the NURD function is suppressed (of which lamin A/C is a component). At the same time, the relationship of lamin A/C with HDAC2 causes the activation of the *CDKN1A* gene, which is the most important determinant of cellular senescence (Cenni et al., 2020).

Mutations in *LMNA* gene also cause atypical Werner's syndrome (AWS), mandibuloacral dysplasia type A (MADA), atypical progeria syndrome (APS) (Burla et al., 2018). They differ in disease-specific mutations in *LMNA* gene and in clinical manifestations. Dunnigan-type familial partial lipodystrophy is caused by an R482W missense mutation in lamin A, leading to a pathological change in its effect on the periphery and inside the nucleus on a three-dimensional rearrangement of chromatin (Briand et al., 2018). MADA can be caused by mutations in the *LMNA* or *ZMPSTE24* genes. It is a rare autosomal recessive disorder characterized by bone abnormalities with localized osteolysis and generalized osteoporosis, skin pigmentation, lipodystrophy, and accelerated aging. These mutations cause accumulation of prelamin A, which leads to disruption of chromatin dynamics. Similar changes are found during physiological aging (Cenni et al., 2018).

Progeroid laminopathies are also caused by mutations in genes that interact with lamins. For example, Nestor–Guillermo progeria (NGPS) is caused by mutations in the *BAF* gene (Loi et al., 2016). Premature aging syndromes can be caused not only by mutations in the genes of lamins and proteins interacting with them, but also by mutations in the genes of DNA repair and maintenance enzymes. For example, mutations in helicase genes cause Rothmund–Thomson syndrome (*RECQL4* gene), Bloom syndrome (*BLM* gene), Werner syndrome (*WRN* gene). Mutations in DNA repair genes cause Dyskeratosis congenita (*DKC1*, *TERC*, *TERT* genes) and Cockayne syndrome (*ERCC8* or *ERCC6* genes). In contrast to these premature aging syndromes, PLs are characterized by an early onset, more severe manifestations of aging and a lack of predisposition to cancer (Burla et al., 2018). All PLs are characterized by generalized osteoporosis and osteolysis, crowding of teeth with malocclusion. Progeroid laminopathies with a specific lesion of the musculoskeletal system include heart-hand syndrome of Slovenian (HHS-S) (Gargiuli et al., 2018). Mutations in the *ZMPSTE24* gene are causes of the development of restrictive dermopathy (RD) and mandibuloacral dysplasia type B (MADB) (Burla et al., 2018).

In ataxia-telangiectasia patients, accumulation of lamin B1 causes changes in the shape of the nucleus and cell aging. Oxidative stress in this syndrome increases lamin B1 levels by stimulating mitogen-activated protein kinase p38 (p38 MAPK). In cell experiments using the MAPK activator anisomycin and the MAPK inhibitor SB203580 to determine the effect of p38 MAPK on lamin B1 levels, p38 MAPK activation was shown to significantly increase

lamin B1 levels. At the same time, inhibition of p38 MAPK decreases lamin B1 levels. Using the PLA method (proximity ligation assay), it was proved that p38 MAPK interacts with lamin B1, causing its phosphorylation (Barascu et al., 2012). These changes also occur during natural aging. Also, PLs are characterized by decrease in perinuclear heterochromatin and increase in cellular senescence – two conditions that correlate with dysregulation of TEs (Andrenacci et al., 2020). It can be assumed that in evolution, lamins arose as one of the protective mechanisms aimed at silencing TEs to protect hosts from genomic instability (Cavaliere et al., 2020). Therefore, laminopathies that cause premature aging and aging-associated pathology are most likely to cause activation of TEs as a key mechanism for the development of diseases (Andrenacci et al., 2020). Indeed, in an experiment on cell models expressing progerin characteristic of HGP, a pronounced increase in the expression of LINE, SINE, HERV and DNA transposons was revealed (Arancio, 2019).

### The relationship between lamins and transposons

Transposable elements are elements capable of transposition within the genome. There are two classes of TEs: retroelements (REs), transposed by reverse transcription and intermediate RNA; DNA-TEs, transposed using transposase by the mechanism of excision and insertion (Andrenacci et al., 2020). All TEs are subdivided into autonomous (contain genes of transposition enzymes) and non-autonomous (use protein products of other TEs for their transpositions). Retroelements are classified into those containing LTR (long terminal repeats) – LTR-REs and non-LTR-REs. The most common autonomous non-LTR-REs in the human genome are LINE1 (long interspersed nuclear elements), and non-autonomous ones are SINE (small interspersed nuclear elements). LTR-REs includes human endogenous retroviruses (HERV) (Arancio, 2019). Silencing of TEs is normally ensured by the degradation of their RNA and the formation of heterochromatin. Therefore, TEs are characterized by activation during aging of organisms as a result of age-associated deregulation of heterochromatin and microRNA (Cavaliere et al., 2020).

The relationship between TEs activation and aging was found in termites. Reproductive queens live for tens of years without a significant increase in TEs expression, while working termites live for only a few weeks, which is due to TEs dysregulation in their genomes due to changes in piRNA expression (Elsner et al., 2018). Interestingly, in the somatic cells of *Cnidaria*, which are characterized by almost unlimited regeneration and immortality, piRNA and Piwi proteins are also expressed. It is accompanied by low levels of TEs activity in their genomes. These properties were revealed in *Hydra vulgaris* (Juliano et al., 2014). Piwi homologues named Cniwi are found at all stages of development of *Podocoryne carnea* not only in reproductive cells, but also in differentiated somatic cells (Seipel et al., 2004). Since the DNA repair systems in *Cnidaria* do not have a specific efficiency that distinguishes them from other animals, it

can be assumed that the expression of piRNA in their organisms is the cause of delayed aging processes (Juliano et al., 2014).

During human aging, there is a progressive activation of LINE1, which leads to the activation of the interferon response due to the accumulation of LINE1 cDNA. These changes are the cause of aseptic inflammation and the activation of interferon observed during aging (De Cecco et al., 2019). At the same time, an increase in the level of proinflammatory cytokines in myeloid immune cells in the thymus during aging is accompanied by a gradual decrease in the level of lamin B1, which specifically functions for the correct organogenesis of the thymus by maintaining the expression of epithelial cell genes. It was found that a decrease in the level of lamin B1 leads to an increase in transcription of 533 genes and suppression of expression of 778 genes. Analysis of these genes showed their participation in cell adhesion, development of the immune system, differentiation of T-lymphocytes and cytokine production (Yue et al., 2019). Interestingly, chronic inflammation is also characteristic of PLs. In this case, the STAT1-regulated interferon-like response is induced by DNA hybrids: RNA through the signaling pathways cGAS (cyclic GMP-AMP synthase) and STING (stimulator of interferon genes) (Kreienkamp et al., 2018).

Transposable elements are the most important sensors of environmental stressors, exerting an adaptive regulatory effect on protein-coding genes (Mustafin, Khusnutdinova, 2019). According to the theory of “buffering function of the tail of lamins”, NL acts as an intracellular sensor of the reactive oxygen species (through conservative changes in the cysteine residue in the tail domain of the lamin). Throughout mammalian phylogeny, three cysteine residues (C522, C588, C591) at the C-terminus of lamin A are conserved. Of these, C588 and C591 are also characteristic of other vertebrates, and C522 is absent in non-mammalian animals.

Experiments on human skin fibroblasts showed that these amino acids in the functional tail domain of lamin A increase the sensitivity to reactive oxygen species, and their replacement with alanine promotes cellular aging (Pekovic et al., 2011). Thus, lamins and TEs are sensors for stressors that can cause their dysregulation. Moreover, abnormalities in the expression of TEs and lamins can potentiate each other, contributing to aging. Indeed, in experiments on *Drosophila*, the relationship between the age-related decrease in lamin B and the activation of various REs was proved (Chen H. et al., 2016).

Aging can be induced not only by decreased levels of lamins, but also by their increased production (Dreesen et al., 2013). REs are also characterized by similar properties. For example, in the study of 111 known REs in young and old *Drosophila*, not only an increase in the expression of 18 specific REs was determined, but also a decrease in the levels of other 18 REs (Chen H., et al., 2016). That is, with aging, there is not just an activation of TEs or a decrease in the levels of lamins, but a variety of changes in their expression, which potentiate each other.



The question arises as to what is primary – dysregulation of lamins or TEs. Investigation of the relationship between lamins and TEs could provide a basis for identifying the key mechanisms of physiological aging and for the development of new ways to prolong life. It is possible that changes in the levels of lamins and TEs expression during aging is a correlation rather than a causal relationship; however, the available data on their interconnections suggest a mutual potentiation of these structures during aging.

In different cells, LINE1s in the human genome are abundant in the lamin-associated domains of heterochromatin regions at the nuclear periphery. Sirtuin SIRT7 is involved in epigenetic transcriptional repression of LINE1 throughout the genome. The interaction of SIRT7 with A/C lamins plays an important role in this process. At the same time, SIRT7 provides deacetylation of H3K18, facilitating the interaction of LINE1s with NL. Deacetylation of H3K18 is insufficient to suppress LINE1s expression in the absence of lamins A/C, and depletion of lamins leads to transcriptional activation of LINE1s. That is, for the repression of LINE1s, their relationship with lamins A/C is necessary (Vazquez et al., 2019).

During the supercompactisation of differentiating human neutrophils genomes, centromeres, pericentromeres, and LINEs move to the NL region. This property is retained for LINE in LBR-deficient cells, but is lost for other heterochromatin regions (Zhu et al., 2017). Human retroelements LTR7/HERV-H, LTR5\_Hs, and L1HS form specific regulatory regions in the genome that selectively and site-specifically bind to lamin B1, as well as to NANOG homeobox, POU5F1, and CCCEC-binding factor (Glinsky, 2015).

There is also an inverse relationship where lamin B maintains the integrity of the nuclear genome by repressing TEs, which was found in experiments on *Drosophila* (Chen C.K. et al., 2016). Prelamin A is interconnected with TEs through BAF, which inhibits the integration of retroviruses. BAF is also found in protein complexes containing Sleeping Beauty transposase (Wang et al., 2014). BAF is required to modulate the effect of prelamin A on chromatin structure because it induces histone H3K9 trimethylation, as well as nuclear relocalization of the lamin-associated proteins LAP2- $\alpha$  and HP1- $\alpha$  (Loi et al., 2016).

### The relationship of lamins and transposable elements with viruses

Lamins (Pekovic et al., 2011) and transposons (Mustafin, Khusnutdinova, 2019) are sensors of environmental stress effects, characterized by their interconnection and possible potentiation of dysregulation during aging. Therefore, it is necessary to consider other possible mechanisms affecting their activity during aging. Environmental factors causing dysregulation of lamins and TEs include exogenous viruses. Determination of the role of lamins in viral infection can become the basis for identifying possible new ways of antiviral therapy, as well as the use of viral vectors for correcting aging processes by acting on specific lamins.

Induction of lamins A/C occurs in naive CD4<sup>+</sup> T cells upon antigen recognition. In this case, lamins A/C act as a link between the nucleus and the plasma membrane during the activation of T cells. In experiments on mice, the role of lamins A/C in the response to vaccinia virus was revealed (Toribio-Fernandez et al., 2018). Proteins hnRNP and lamins A/C serve as carriers and mediators for the movement of p17 protein of avian reovirus between the nucleus and the cytoplasm (Chiu et al., 2019). The latency of HIV-1 after integration is characterized by reversible silencing of transcription driven by the LTR of the HIV genome. The lamin-interacting protein SUN2 maintains repressive chromatin and inhibits the transcription of proviral DNA, which is regulated by HIV LTR through its association with lamins A/C. Lamins A/C bind SUN2 to nucleosomes and to HIV-1 5'-LTR, which causes blockages of viral initiation and elongation of transcription (Sun et al., 2018).

The release of herpesviruses from the nucleus is accompanied by a change in the architecture of the NL (Vu et al., 2016). Herpes simplex virus-1 induces phosphorylation and reorganization of lamins A/C through the virulence factor – the product of the *γ134.5* gene. It allows the virus to exit through the nucleus (Wu et al., 2016). In turn, lamins facilitate the access of capsid to the inner nuclear membrane and the curvature of its sections around the capsid of the herpes virus during budding (Vu et al., 2016).

TRIM E3 ligase controls the replication of HSV-1 herpesviruses by affecting the integrity of lamins through changes in the structure of the host cell centrosomes. TRIM43 ubiquitinates the centrosome protein pericentrin, causing its proteasome degradation. It leads to lamins changes that suppress the active state of viral chromatin (Full et al., 2019). The mechanism of this phenomenon is due to the specific interaction of lamin A/C with the genomic DNA of the virus at the nuclear periphery. In this case, lamin inhibits the formation of heterochromatin in the region of HSV promoters (Silva et al., 2008).

Caspase-6-dependent dephosphorylation of lamins A/C is essential for SV40 virus penetration (Butin-Israeli et al., 2011). *Baculovirus* promotes phosphorylation of lamin B and destruction of NL during infection (Zhang et al., 2017). *Canine parvovirus* in the late stage of infection reorganizes NL with decreasing the levels of lamins A/C in the apical part of the nucleus (Mantyla et al., 2015).

The interrelation of lamins with viruses may indicate the possible participation of NL in the interconversion of viruses with TEs, as well as in transposon-controlled regulatory networks of genomes. Like lamins, TEs are involved in antiviral response, which may be related to their phylogenetic relationship. In the course of evolution, LTR-REs became sources of exogenous retroviruses (Xiong, Eickbush, 1990). ERV of various mammalian species are capable of converting into infectious form of viruses and converting back to REs (Zhuo, Feschotte, 2015). Expression products of LTR-REs *env* genes cause restriction of exogenous retroviruses in animals (Malfavon-Borja, 2015). Retroelements enzymes



can also be used to integrate exogenous viruses into host genomes (Speiseder et al., 2014). Exogenous viruses can regulate TEs activity. For example, human cytomegalovirus causes tissue-specific activation of LTR-RE (Assinger et al., 2013). MicroRNAs can serve as possible mediators in the interactions of viruses with TE and lamins. Investigation of microRNAs is promising both for the development of methods for slowing aging processes and for antiviral therapy.

### Interrelation of lamins with transposons and miRNAs during aging

MicroRNA are involved in regulation of lamins (Sylvius et al., 2011; Toro et al., 2018) and TEs. It may be due to the origin of miRNAs in evolution from transposons. For the first time, evidence of the emergence of 55 different microRNAs from human TEs was obtained in 2007 (Piriyapongsa et al., 2007). In 2009, 73 human microRNAs derived from TEs were identified (Gu et al., 2009). In 2011, 191 microRNAs that arose from TE were identified (Filshtein et al., 2011). In the same year, another research group identified 226 human microRNAs derived from TE (MDTEs) (Yuan et al., 2011). In 2012, 235 human MDTEs were identified (Tempel et al., 2012), 409 MDTEs were identified in 2015 (Qin et al., 2015), 34 MDTEs were identified in 2020. In 2016, the MDTE database was created. This database contains information about microRNAs originated directly from TEs (Wei et al., 2016).

Aging is an important risk factor for the development of oncopathology. It may be due to common epigenetic mechanisms, since during aging, global DNA hypomethylation, TEs activation, and the development of genomic instability occur (Anwar et al., 2017). This assumption can be confirmed by data on the role of the same microRNAs in both carcinogenesis and aging. A lot of information has been accumulated on the association of various microRNAs with the development of tumors. Bioinformation database OncomiR was created ([www.oncomir.org](http://www.oncomir.org)) (Wong et al., 2018). We analyzed 410 different microRNAs featured in OncomiR using the MDTE database (Wei et al., 2016). As a result, we found that 94 of these microRNAs were derived from TEs. It indicates the role of these microRNAs in the development of malignant tumors and the possibility of their use as a target for antitumor therapy.

Like TEs, lamins also play a role in carcinogenesis. It was shown that a deficiency of lamins A/C can be used as an independent risk factor for the development of cervical cancer (Capo-chichi et al., 2016). Lamin B1 levels are significantly increased in patients with hepatocellular carcinoma (Abdelghany et al., 2018). Lamin B2 is highly expressed in non-small cell lung cancer and positively correlates with lymph node metastases. It is due to the interaction of lamin B2 with cyclin D1, which activates G9 $\alpha$  expression and increases H3K9me2 levels. As a result, H3K9me2 binds to the promoter region of the E-cadherin *CDH1* gene and stimulates cell migration (Zhang et al., 2020).

In gastric cancer, lamin B1 expression is reduced in tumor tissue, and low levels of lamin B1 are significantly

correlated with clinical stage severity, depth of invasion, and poor prognosis. In the experiment, inhibition of lamin B1 causes the proliferation and migration of gastric cancer cells, which is due to the activation of the PI3K/PTEN/Akt and MAPK/ERK pathways, as well as the suppression of p53/h21WAF1/CIP1 (Yu et al., 2020).

Low levels of lamins A/C expression in cancer are specific for poor prognosis. In breast cancer (BC), the loss or decrease in the expression of lamins A/C was significantly associated with large tumor sizes, poor prognosis and the development of long-term results (reduced survival time) (Alhudiri et al., 2019). Comparison of osteosarcoma and osteoblast cell lines showed a lower level of lamin A expression in osteosarcoma cells. Elevated lamin A levels reduce cancer cells' ability to migrate (Evagelisti et al., 2020). Thus, carcinogenesis is associated both with an increase and (most often) with a decrease in lamin expression, that is, it is explained by their dysregulation.

This is due to the global regulatory role of lamins in the expression of various genes. Therefore, it is important to determine the specific features of changes in the expression of lamins in each type of tumor in order to target them during antitumor therapy. Most tumors are characterized by decreased lamins expression, which may be one of the key mechanisms of initiation and maintenance of carcinogenesis, because lamins serve as host defense systems for TEs silencing. Therefore, the loss of control by the lamins leads to pathological activation of TEs with subsequent genomic instability and tumor development. The effect of increased levels of lamins on tumor development is probably due to the possibility of opposite regulatory effects of lamins on TEs, depending on localization of TEs in active or repressed genome regions (Cavaliere et al., 2020).

Like transposons, during aging and carcinogenesis, lamins are also associated with changes in microRNA levels. For example, the target of the oncogenic miR-129, the expression of which is increased in BC, is lamin A. The target of the oncogenic miR-218, the level of which decreases in BC, is lamin B1 (Setijono et al., 2018). *LMNA* expression in breast cancer cells is suppressed by miR-9, which leads to the progression of the disease due to changes in the deformability of the nucleus and an increase in the invasive ability of cells (Guinde et al., 2018). MiR-122 inhibits hepatocellular cancer cell proliferation by suppressing the expression of lamin B2 (Li et al., 2019). MiR-351-5p regulates lamin B1 expression and promotes phloxuridine-induced tumor cell apoptosis (Sato et al., 2020). In fibroblasts associated with BC, the expression of miR-222 is increased (compared to normal cells). The direct target of this microRNA is LBR. Therefore, miR-222 is considered for BC targeted therapy (Chatterje et al., 2019).

During human aging, the levels of *LMNB1* protein products in fibroblasts and keratinocytes of the skin decrease, which is mediated by the influence of miR-23a (Dreesen et al., 2013). An increase in *LMNB1* expression due to gene duplication in autosomal dominant leukodystrophy in adults leads to progressive brain demyelination. This is due to the

role of lamin B1 in regulation of myelin formation and maintenance during aging. The targets of miR-23a are also the *PTEN* gene transcript (phosphatase and tensin homolog on chromosome 10) and long noncoding RNA 2700046G09Rik (Lin et al., 2014). MiR-124-3p complementarily binds to 3'-UTR of *LMNA* gene, suppressing its expression (Bao et al., 2019). Activation of miR-141-3p during replicative senescence decreases the activity of HDAC1 and HDAC2. MiR-141-3p has a targeted regulatory effect on ZMPSTE24 endoprotease prelamin A (Yu et al., 2013).

Like TEs, miRNAs affect lamins according to the principle of mutual regulation, since miRNAs genes are located in genome regions associated with lamins. For example, in mice, aging-associated miRNA genes are located on the X chromosome in a cluster located in the lamin-associated domain (Elias et al., 2019). Long noncoding RNA Xist is involved in X chromosome inactivation by directly interacting with the LBR and recruiting it to the NL. It leads to remodeling of the three-dimensional structure of DNA, which allows Xist to distribute along the X chromosome and cause silencing of transcriptionally active genes (Chen C.K. et al., 2016).

In skeletal muscles of patients with muscular dystrophies caused by mutations in *LMNA* gene, pronounced dysregulation of sixteen different microRNAs was revealed, which are involved in Wnt-signaling pathways, MAPK, and the regulation of transforming growth factor  $\beta$ . Nine of these microRNAs are involved in regenerative processes (miR-100, -127-3p, -148a, -136 \*, -192, -335, -376c, -489, -502-3p) and are detected at high levels in fetal muscles (Sylvius et al., 2011). Expression of the mutant allele of *LMNA* gene (mutation R482W) is accompanied by an increase in levels of miR-335, which has antilipogenic properties (Briand et al., 2018). Significantly increased expression of let-7a-5p, miR-142-3p, miR-145-5p, and miR-454-3p is observed in patients with familial dilated cardiomyopathy associated with mutations in *LMNA* (Toro et al., 2018).

According to the MDTE database, among the miRNAs associated with lamina pathology, the origin from LINE2 was proven for miR-192 and miR-502, the origin from MIR (SINE-RE) was proven for miR-335 (Wei et al., 2016). It should be noted that miR-335 also plays a role in human physiological aging and age-associated neurological pathology (Raihan et al., 2018). It may be due to the role of lamins in protecting genomes from transpositions (Cavaliere et al., 2020), since TEs are involved in the epigenetic regulation of lamins. That is, the structural and functional relationship between lamins and TEs is mutually regulatory. On the one hand, lamins are involved in the control of TEs activity. On the other hand, TEs can exert an epigenetic regulatory effect on NL, which may be due to the influence of microRNAs derived from TEs.

## Conclusion

Lamins play an important role in driving gene expression and in TEs silencing, thereby preventing genome instability. Therefore, dysregulation of lamins, mainly associated with inactivating mutations in their genes or a decrease in their

expression, leads to TEs activation. Therefore, pathological activation of TEs and the resulting genomic instability play an important role in changing the levels of lamins in malignant neoplasms. The role of lamins in viral infection supports this assumption, since TEs are phylogenetically associated with viruses.

Transposable elements and lamins share common properties, such as dysregulation during aging and carcinogenesis, as well as interactions with microRNAs. Mutual regulation of lamins and TEs testifies to the mutual potentiation of their dysregulation in laminopathies, carcinogenesis and physiological aging. Identification of key changes in TEs and miRNAs derived from them can become the basis for targeted therapy of laminopathies, malignant neoplasms and for prolonging life. Investigation of the role of lamins in the interaction with TEs and viruses is also a promising direction in the development of antiviral therapy and vaccination.

## References

- Abdelghany A.M., Rezk N.S., Osman M.M., Hamid A.I., Al-Breedy A.M., Abdelsattar H.A. Using Lamin B1 mRNA for the early diagnosis of hepatocellular carcinoma: a cross-selectional diagnostic accuracy study. *F1000Res.* 2018;7:1339. DOI 10.12688/f1000research.14795.1.
- Ahmed M.S., Ikram S., Bibi N., Mir A. Hutchinson–Gilford progeria syndrome: a premature aging disease. *Mol. Neurobiol.* 2018;55: 4417–4427. DOI 10.1007/s12035-017-0610-7.
- Alhudiri I.M., Nolan C.C., Ellis I.O., Elzagheid A., Rakha E.A., Green A.R., Chapman C.J. Expression of Lamin A/C in early-stage breast cancer and its prognostic value. *Breast Cancer Res. Treat.* 2019;174(3):661–668. DOI 10.1007/s10549-018-05092-w.
- Andrenacci D., Cavaliere V., Lattanzi G. The role of transposable elements activity in aging and their possible involvement in laminopathies. *Ageing Res. Rev.* 2020;57:1000995. DOI 10.1016/j.arr.2019.100995.
- Anwar S.L., Wulaningsih W., Lehmann U. Transposable elements in human cancer: causes and consequences of deregulation. *Int. J. Mol. Sci.* 2017;18:974. DOI 10.3390/ijms1850974.
- Arancio W. Progerin expression induces a significant downregulation of transcription from human repetitive sequences in iPSC-derived dopaminergic neurons. *Geroscience.* 2019;41:39–49. DOI 10.1007/s11357-018-00050-2.
- Arancio W., Pizzolanti G., Genovese S.I., Pitrone M., Giordano C. Epigenetic involvement in Hutchinson–Gilford progeria syndrome: a mini-review. *Gerontology.* 2014;60:197–203. DOI 10.1159/000357206.
- Assinger A., Yaiw K.C., Gottesdorfer I., Leib-Mosch C., Soderberg-Naucler C. Human cytomegalovirus (HCMV) induces human endogenous retrovirus (HERV) transcription. *Retrovirology.* 2013; 10:132. DOI 10.1186/1742-4690-10-132.
- Bao H., Li H.P., Shi Q., Huang K., Chen X., Chen Y., Han Y., Xiao Q., Yao Q., Qi Y. Lamin A/C negatively regulated by miR-124-3p modulates apoptosis of vascular smooth muscle cells during cyclic stretch application in rats. *Acta Physiol. (Oxf.).* 2019;8:e13374. DOI 10.1111/apha.13374.
- Barascu A., Le Chalony C., Pennarun G., Genet D., Imam N., Lopez B., Bertrand P. Oxidative stress induces an ATM-independent senescence pathway through p38 MAPK-mediated lamin B1 accumulation. *EMBO J.* 2012;31:1080–1094. DOI 10.1038/emboj.2011.492.
- Briand N., Cahyani I., Madsen-Osterbye J., Paulen J., Ronningen T., Sorensen A.L., Collas P. Lamin A., chromatin and FPLD2: not just

- a peripheral menage-a-trois. *Front. Cell Dev. Biol.* 2018;6:73. DOI 10.3389/fcell.2018.00073.
- Burla R., La Torre M., Merigliano C., Verni F., Saggio I. Genomic instability and DNA replication defects in progeroid syndromes. *Nucleus*. 2018;9(1):368-379. DOI 10.1080/19491034.2018.1476793.
- Butin-Israeli V., Ben-nun-Shaul O., Kopatz I., Adam S.A., Shimi T., Goldman R.D., Oppenheim A. Simian virus 40 induces lamin A/C fluctuations and nuclear envelope deformation during cell entry. *Nucleus*. 2011;2(4):320-330. DOI 10.4161/nucl.2.4.16371.
- Capo-chichi C.D., Aguida B., Chabi N.W., Cai Q.K., Offrin G., Agos-sou V.K., Sanni A., Xu X. Lamin A/C deficiency is an independent risk factor for cervical cancer. *Cell. Oncol. (Dordr.)*. 2016;39:59-68. DOI 10.1007/s13402-015-0252-6.
- Cavaliere V., Lattanzi G., Andrenacci D. Silencing of euchromatic transposable elements as a consequence of nuclear lamina dysfunction. *Cells*. 2020;9(3):625. DOI 10.3390/cells9030625.
- Cenni V., Capanni C., Mattioli E., Schena E., Squarzone S., Bacalini M.G., Garagnani P., Salvioli S., Franceschi C., Lattanzi G. Lamin A involvement in aging processes. *Ageing Res. Rev.* 2020;62:101073. DOI 10.1016/j.arr.2020.101073.
- Cenni V., D'Apice M.R., Garagnani P., Colombaro M., Novelli G., Franceschi C., Lattanzi G. Mandibuloacral dysplasia: A premature ageing disease with aspects of physiological ageing. *Ageing Res. Rev.* 2018;42:1-13. DOI 10.1016/j.arr.2017.12.001.
- Chatterjee A., Jana S., Chatterjee S., Wastall L.M., Nargis N., Roy H., Hughes T., Bhattacharyya A. MicroRNA-222 reprogrammed cancer-associated fibroblasts enhance growth and metastasis of breast cancer. *Br. J. Cancer*. 2019;121(8):679-689. DOI 10.1038/s41416-019-0566-7.
- Chen C.K., Blanco M., Jackson C., Aznauryan E., Ollikainen N., Surka C., Chow A., Ceras A., McDonel P., Guttman M. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science*. 2016;354(6311):468-472. DOI 10.1126/science.aae0047.
- Chen H., Zheng X., Xiao D., Zheng Y. Age-associated de-repression of retrotransposons in the *Drosophila* fat body, its potential cause and consequence. *Ageing Cell*. 2016;15(3):542-552. DOI 10.1111/accel.12465.
- Chiu H.C., Huang W.R., Wang Y.Y., Li J., Liao T., Nielsen B.L., Liu H. Heterogeneous nuclear ribonucleoprotein A1 and lamin A/C modulate nucleocytoplasmic shuttling of avian reovirus p17. *J. Virol.* 2019;93(20):e00851-19. DOI 10.1128/JVI.00851-19.
- Cibulka J., Fraiberk M., Forstova J. Nuclear actin and lamins in viral infections. *Viruses*. 2012;4(3):325-347. DOI 10.3390/v4030325.
- De Cecco M., Ito T., Petrashe A.P., Elias A., Skvir N.J., Criscione S.W., Caligiana A., Broccoli G., Neretti N., Sedivy J.M. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature*. 2019;572(7767):E5. DOI 10.1038/s41586-019-1350-9.
- Dreesen O., Cojnowski A., Ong P.F., Zhao T.Y., Common J.E., Lunny D., Lane E.B., Vardy L.A., Stewart C.L., Colman A. Lamin B1 fluctuations have differential effects on cellular proliferation and senescence. *J. Cell. Biol.* 2013;200(5):605-617. DOI 10.1083/jcb.201206121.
- Elias A.E., Kun B., Sabula I.M.C. The mir-465 family is upregulated with age and attenuates growth hormone signaling in mouse liver. *Ageing Cell*. 2019;18(2):e12892. DOI 10.1111/accel.12892.
- Elsner D., Meusemann K., Korb J. Longevity and transposon defense, the case of termite reproductives. *Proc. Natl. Acad. Sci. USA*. 2018;115(21):5504-5509. DOI 10.1073/pnas.1804046115.
- Evangelisti C., Paganelli F., Giuntini G., Mattioli E., Cappellini A., Ramazzotti G., Faenza I., Maltarello M.C., Martelli A.M., Scotlandi K., Chiarini F., Lattanzi G. Lamin A and prelamin A counteract migration of osteosarcoma cells. *Cells*. 2020;9(3):774. DOI 10.3390/cells9030774.
- Filstein T.J., Mackenzie C.O., Dale M.D., Dela-Cruz P.S., Ernst D.M., Frankenberger E.A., He C., Heath K.L., Jones A.S., Jones D.K., King E.R., Maher M.B., Mitchell T.J., Morgan R.R., Sirobushanam S., Halkyard S.D., Tiwari K.B., Rubin D.A., Borchert G.M., Larson E.D. Origin-based identification of microRNA targets. *Mob. Genet. Elements*. 2012;2(4):184-192. DOI 10.4161/mge.21617.
- Full F., van Gent M., Sparrer K.M.J., Chiang C., Zurenski M.A., Scherer M., Brockmeyer N.H., Heinzerling L., Sturzl M., Korn K., Stamminger T., Ensser A., Gack M.U. Centrosomal protein TRIM43 restricts herpesvirus infection by regulating nuclear lamina integrity. *Nat. Microbiol.* 2019;4:164-176. DOI 10.1038/s41564-018-0285-5.
- Gargioli C., Schena E., Mattioli E., Greggi T., Lattanzi G. Lamins and bone disorders: current understanding and perspectives. *Oncotarget*. 2018;9(32):22817-22831. DOI 10.18632/oncotarget.25071.
- Garvalov B.K., Muhammad S., Dobrev G. Lamin B1 in cancer and aging. *Ageing (Albany NY)*. 2019;11:7336-7338. DOI 10.18632/ageing.102306.
- Glinsky G.V. Transposable elements and DNA methylation create in embryonic stem cells human-specific regulatory sequences associated with distal enhancers and noncoding RNAs. *Genome Biol. Evol.* 2015;7(6):1432-1454. DOI 10.1093/gbe/evv081.
- Gu T.J., Yi X., Zhao X.W., Zhao Y., Yin J.Q. Alu-directed transcriptional regulation of some novel miRNAs. *BMC Genomics*. 2009;10:563. DOI 10.1186/1471-2164-10-563.
- Guelen L., Pagie L., Brasset E., Meuleman W., Faza M.B., Talhout W., Eussen B.H., de Klein A., Wessels L., de Laat W., van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008;453(7197):948-951. DOI 10.1038/nature06947.
- Guinde J., Frankel D., Perrin S., Kaspi E. Lamins in lung cancer: Biomarkers and key factors for disease progression through miR-9 regulation? *Cells*. 2018;7(7):78. DOI 10.3390/cells7070078.
- Jia Y., Vong J.S., Asafova A., Garvalov B.K., Caputo L., Cordero J., Singh A., Boettger T., Gunther S., Fink L., Acker T., Barreto G., Seeger W., Braun T., Savai R., Dobrev G. Lamin B1 loss promotes lung cancer development and metastasis by epigenetic de-repression of RET. *J. Exp. Med.* 2019;216(6):1377-1395. DOI 10.1084/jem.20181394.
- Juliano C.E., Reich A., Liu N., Gotzfriend J., Zhong M., Uman S., Reenan R.A., Wessel G.M., Steele R.E., Lin H. PIWI proteins and PIWI-interacting RNAs function in *Hydra* somatic stem cells. *Proc. Natl. Acad. Sci. USA*. 2014;111(1):337-342. DOI 10.1073/pnas.1320965111.
- Jung H.J., Coffinier C., Choe Y., Beigneux A.P., Davies B.S.J., Yang S.H., Barnes R.H. 2nd, Hong J., Sun T., Pleasure S.J., Young S.G., Fong L.G. Regulation of prelamin A but not lamin C by miR-9, a brain-specific microRNA. *Proc. Natl. Acad. Sci. USA*. 2012;109(7):E423-E431. DOI 10.1073/pnas.1111780109.
- Kollmar M. Polyphyly of nuclear lamin genes indicates an early eukaryotic origin of the metazoan-type intermediate filament proteins. *Sci. Rep.* 2015;5:10652. DOI 10.1038/srep10652.
- Koreny L., Field M.C. Ancient eukaryotic origin and evolutionary plasticity of nuclear lamina. *Genome Biol. Evol.* 2016;8(9):2663-2671. DOI 10.1093/gbe/evw087.
- Kreienkamp R., Graziano S., Coll-Bonfill N., Bedia-Diaz G., Cybul-la E., Vindigni A., Dorsett D., Kubben N., Batista L.F.Z., Gonzalo S. A cell-intrinsic interferon-like response links replication stress to cellular ageing caused by progerin. *Cell Rep.* 2018;22(8):2006-2015. DOI 10.1016/j.celrep.2018.02.090.
- Lemaitre C., Bickmore W.A. Chromatin at the nuclear periphery and the regulation of genome functions. *Histochem. Cell Biol.* 2015;144(2):111-122. DOI 10.1007/s00418-015-1346-y.



- Li X.N., Yang H., Yang T. MiR-122 inhibits hepatocarcinoma cell progression by targeting LMNB2. *Oncol. Res.* 2019;28:41-49. DOI 10.3727/096504019X1561579.
- Lin S., Heng M.Y., Ptacek L.J., Fu Y. Regulation of myelination in the central nervous system by nuclear lamin B1 and non-coding RNAs. *Transl. Neurodegener.* 2014;3(1):4. DOI 10.1186/2047-9158-3-4.
- Loi M., Cenni V., Duchi S., Squarzone S., Lopez-Otin C., Foisner R., Lattanzi G., Capanni C. Barrier-to-autointegration factor (BAF) involvement in preamin A-related chromatin organization changes. *Oncotarget.* 2016;7(13):15662-15677. DOI 10.18632/oncotarget.6697.
- Malfavon-Borja R., Feschotte C. Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. *J. Virol.* 2015;89(8): 4047-4050. DOI 10.1128/JVI.03653-14.
- Mantyla E., Niskanen E.A., Ihalainen T.O., Vihinen-Ranta M. Reorganization of nuclear pore complexes and the lamina in late-stage parvovirus infection. *J. Virol.* 2015;89(22):11706-11710. DOI 10.1128/JVI.01608-15.
- Meinke P., Mattioli E., Haque F., Antoku S., Columbaro M., Straatman K.R., Worman H.J., Gundersen G.G., Lattanzi G., Wehnert M., Shackleton S. Muscular dystrophy-associated SUN1 and SUN2 variants disrupt nuclear-cytoskeletal connections and myonuclear organization. *PLoS Genet.* 2014;10(9):e1004605. DOI 10.1371/journal.pgen.1004605.
- Mustafin R.N., Khusnutdinova E.K. The role of transposable elements in the ecological morphogenesis under the influence of stress. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2019;23(4):380-389. DOI 10.18699/VJ19.506. (in Russian)
- Pekovic V., Gibbs-Seymour I., Markiewicz E., Alzogaibi F., Benham A.M., Edwards R., Wehnert M., von Zglincki T., Hutchison C.J. Conserved cysteine residues in the mammalian lamin A tail are essential for cellular responses to ROS generation. *Aging Cell.* 2011;10(6):1067-1079. DOI 10.1111/j.1474-9726.2011.00750.x.
- Piriyapongsa J., Marino-Ramirez L., Jordan I.K. Origin and evolution of human microRNAs from transposable elements. *Genetics.* 2007;176(2):1323-1337. DOI 10.1534/genetics.107.072553.
- Raihan O., Brishti A., Molla M.R., Li W., Zhang Q., Xu P., Khan M.I., Zhang J., Liu Q. The age-dependent elevation of miR-335-3p leads to reduced cholesterol and impaired memory in brain. *Neuroscience.* 2018;390:160-173. DOI 10.1016/j.neuroscience.2018.08.003.
- Qin S., Jin P., Zhou X., Chen L., Ma F. The role of transposable elements in the origin and evolution of microRNAs in human. *PLoS One.* 2015;10(6):e0131365. DOI 10.1371/journal.pone.0131365.
- Santi S., Cenni V., Capanni C., Lattanzi G., Mattioli E. PCAF Involvement in Lamin A/C-HDAC2 interplay during the early phase of muscle differentiation. *Cells.* 2020;9(7):1735. DOI 10.3390/cells9071735.
- Sato A., Ogino Y., Shimotsu A., Hiramoto A., Kim H., Wataya Y. Direct interaction analysis of microRNA-351-5p and nuclear scaffold lamin B1 mRNA by the cell-free in vitro mRNA/miRNA binding evaluation system. *Nucleosides Nucleotides Nucleic Acids.* 2020;39(6):799-805. DOI 10.1080/15257770.2019.1702675.
- Seipel K., Yanze N., Schmid V. The germ line and somatic stem cell gene *Cniwi* in the jellyfish *Podocoryne carnea*. *Int. J. Dev. Biol.* 2004;48(1):1-7. DOI 10.1387/ijdb.15005568.
- Setijono S.R., Park M., Kim G., Kim Y., Cho K.W., Song S.J. miR-218 and miR-129 regulate breast cancer progression by targeting Lamins. *Biochem. Biophys. Res. Commun.* 2018;496:826-833. DOI 10.1016/j.bbrc.2018.01.146.
- Silva L., Cliffe A., Chang L., Knipe D.M. Role for A-type lamins in herpesviral DNA targeting and heterochromatin modulation. *PLoS Patholog.* 2008;4(5):e1000071. DOI 10.1371/journal.ppat.1000071.
- Speiseder T., Nevels M., Dobner T. Determination of the transforming activities of adenovirus oncogenes. *Methods Mol. Biol.* 2014;1089: 105-115. DOI 10.1007/978-1-62703-679-5\_8. PMID: 24132481.
- Sun W.W., Jiao S., Sun L., Zhou Z., Jin X., Wang J. SUN2 modulates HIV-1 infection and latency through association with lamin A/C to maintain the repressive chromatin. *mBio.* 2018;9:e02408-17. DOI 10.1128/mBio.02408-17.
- Sylvius N., Bonne G., Straatman K., Reddy T., Gant T.W., Shackleton S. MicroRNA expression profiling in patients with lamin A/C-associated muscular dystrophy. *FASEB J.* 2011;25:3966-3978. DOI 10.1096/fj.11-182915.
- Tempel S., Pollet N., Tahi F. NcRNAClassifier: a tool for detection and classification of transposable element sequences in RNA hairpins. *BMC Bioinformatics.* 2012;13:246-258. DOI 10.1186/1471-2105-13-246.
- Toribio-Fernandez R., Zorita V., Rocha-Perugini V., Iborra S., Hoyo G.M., Andres V., Gonzalez-Granado J. Lamin A/C augments Th1 differentiation and response against vaccinia virus and *Leishmania major*. *Cell Death Dis.* 2018;9:9. DOI 10.1038/s41419-017-0007-6.
- Toro R., Blasco-Turron S., Morales-Ponce F.J., Gonzalez P., Mangas A., Llorente-Cortes V., Gonzalo-Calvo D. Plasma microRNA as biomarkers for Lamin A/C-related dilated cardiomyopathy. *J. Mol. Med. (Berl.).* 2018;96(8):845-856. DOI 10.1007/s00109-018-1666-1.
- Turgay Y., Eibauer M., Goldman A.E., Shimi T., Khayat M., Ben-Harush K., Dubrovsky-Gaupp A., Sapra K.T., Goldman R.D., Medalia O. The molecular architecture of lamins in somatic cells. *Nature.* 2017;543(7644):261-264. DOI 10.1038/nature21382.
- Vazquez B.N., Thackray J.K., Simonet N.G., An W., Vaquero A., Tischfield J.A., Serrano L. SIRT7 mediates L1 elements transcriptional repression and their association with the nuclear lamina. *Nucleic Acids Res.* 2019;47(15):7870-7885. DOI 10.1093/nar/gkz519.
- Vu A., Poyzer C., Roller R. Extragenic suppression of a mutation in herpes simplex virus 1 UL34 that affects lamina disruption and nuclear egress. *J. Virol.* 2016;90(23):10738-10751. DOI 10.1128/JVI.01544-16.
- Wang Y., Lichter-Konecki U., Anyane-Yeboah K., Shaw J.E., Lu J.T., Ostlund C., Shin J., Clark L.N., Gundersen G.G., Nagy P.L., Worman H.J. A mutation abolishing the ZMPSTE24 cleavage site in prelamin A causes a progeroid disorder. *J. Cell. Sci.* 2016;129(10): 1975-1980. DOI 10.1242/jcs.187302.
- Wang Y., Wang J., Devaraj A., Izsvak Z. Suicidal autointegration of sleeping beauty and piggyBac transposons in eukaryotic cells. *PLoS Genet.* 2014;10(3):e1004103. DOI 10.1371/journal.pgen.1004103.
- Wei G., Qin S., Li W., Chen L., Ma F. MDTE DB: a database for microRNAs derived from transposable element. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2016;13(6):1155-1160. DOI 10.1109/TCBB.2015.2511767.
- Wong N.W., Chen Y., Chen S., Wang X. OncomiR: and online resource for exploring pan-cancer microRNA dysregulation. *Bioinformatics.* 2018;34:713-715.
- Wu S., Pan S., Zhang L., Baines J., Roller R., Zhang C., Cao Y., He B. Herpes simplex virus 1 induces phosphorylation and reorganization of lamin A/C through the y134.5 protein that facilitates nuclear egress. *J. Virol.* 2016;90(22):10414-10422. DOI 10.1128/JVI.01392-16.
- Xiong Y., Eickbush T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequence. *EMBO J.* 1990;9(10): 3353-3362.



- Yu K.R., Lee S., Jung J.W., Hong I., Kim H., Seo Y., Shin T., Kang K.S. MicroRNA-141-3p plays a role in human mesenchymal stem cell ageing by directly targeting ZMPSTE24. *J. Cell Sci.* 2013;126:5422-5431. DOI 10.1242/jcs.133314.
- Yu Z.Y., Jiang X.Y., Zhao R.R., Luo C.J., Ren Y.X., Ma Z.J., Ye H.L., Shi W.G., Wang C., Jiao Z.Y. Lamin B1 deficiency promotes malignancy and predicts poor prognosis in gastric cancer. *Neoplasma.* 2020;67(6):1303-1313. DOI 10.4149/neo\_2020\_200109N33.
- Yuan Z., Sun X., Liu H., Xie J. MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. *PLoS One.* 2011;6:e17666. DOI 10.1371/journal.pone.0017666.
- Yue S., Zheng X., Zheng Y. Cell-type-specific role of lamin-B1 in thymus development and its inflammation-driven reduction in thymus aging. *Aging Cell.* 2019;18:e12952. DOI 10.1111/accel.12952.
- Zhang M.Y., Han Y.C., Han Q., Liang Y., Luo Y., Wei L., Yan T., Yang Y., Liu S.L., Wang E.H. Lamin B2 promotes the malignant phenotype of non-small cell lung cancer cells by upregulating dimethylation of histone 3 lysine 9. *Exp. Cell Res.* 2020;393(2):112090. DOI 10.1016/j.yexcr.2020.112090.
- Zhang X., Xu K., Wei D., Wu W., Yang K., Yuan M. Baculovirus infection induces disruption of the nuclear lamina. *Sci. Rep.* 2017;7(1):7823. DOI 10.1038/s41598-017-08437-5.
- Zhu Y., Gong K., Denholtz M., Chandra V., Kamps M.P., Alber F., Murre C. Comprehensive characterization of neutrophil genome topology. *Genes Dev.* 2017;31(2):141-153. DOI 10.1101/gad.293910.116.
- Zhuo X., Feschotte C. Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PLoS Pathog.* 2015;11:e1005279. DOI 10.1371/journal.ppat.1005279.

---

**ORCID ID**

R.N. Mustafin orcid.org/0000-0002-4091-382X  
E.K. Khusnutdinova orcid.org/0000-0003-2987-3334

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

Received March 14, 2021. Revised May 23, 2021. Accepted September 28, 2021.

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

## Transcriptional activity of repair, apoptosis and cell cycle genes (*TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PADI4*, *MAPK8*, *NF-KB1*, *STAT3*, *GATA3*) in chronically exposed persons with different intensity of apoptosis of peripheral blood lymphocytes

V.S. Nikiforov<sup>1, 2</sup>✉, E.A. Blinova<sup>1, 2</sup>, A.I. Kotikova<sup>1, 2</sup>, A.V. Akleyev<sup>1, 2</sup>

<sup>1</sup> Urals Research Center for Radiation Medicine, Chelyabinsk, Russia

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

✉ nikiforovx@mail.ru

**Abstract.** Transcriptional activity of genes involved in maintaining genetic homeostasis (genes for repair, cell cycle and apoptosis: *TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PADI4*, *MAPK8*, *NF-KB1*, *STAT3*, *GATA3*) was studied in chronically exposed persons with an increased intensity of early and late stages of apoptosis and necrosis of peripheral blood lymphocytes. The object of this study was peripheral blood mononuclear cells obtained from 132 chronically exposed residents of the Techa riverside villages. The mean accumulated dose to red bone marrow was  $426.4 \pm 48.2$  mGy (1.3–2930.0 mGy), to thymus and peripheral immune organs,  $58.9 \pm 7.9$  mGy (0.1–489.0 mGy). The study was performed more than 60 years after the onset of exposure, the average age of exposed persons was  $68 \pm 0.6$  years (55–86 years). The study of apoptotic and necrotic death of peripheral blood lymphocytes was based on the presence of phosphatidylserine on the cell membrane surface, as well as on its permeability for DNA-intercalating dye. Evaluation of the relative content of mRNA genes for repair, cell cycle, and apoptosis was carried out using real-time PCR. An increased relative content of *PADI4* gene mRNA was registered in the group of chronically exposed persons with the increased intensity of early apoptosis ( $p = 0.006$ ). Modulation of the relative content of mRNA of the *TP53* ( $p = 0.013$ ) and *BCL-2* ( $p = 0.021$ ) genes was detected in the group of chronically exposed individuals with the increased intensity of the late stage of apoptosis. A statistically significant increase in the transcriptional activity of the *TP53* gene was observed in the group of chronically exposed persons with the increased intensity of peripheral blood lymphocyte necrosis in the long-term period ( $p = 0.015$ ). In the course of the study it was noted that exposed people with increased intensity of apoptosis, first of all, demonstrate changes in the transcriptional activity of apoptotic genes. These data are consistent with current views on the activation of programmed cell death.

Key words: mRNA; apoptosis; necrosis; lymphocytes; chronic exposure.

**For citation:** Nikiforov V.S., Blinova E.A., Kotikova A.I., Akleyev A.V. Transcriptional activity of repair, apoptosis and cell cycle genes (*TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PADI4*, *MAPK8*, *NF-KB1*, *STAT3*, *GATA3*) in chronically exposed persons with different intensity of apoptosis of peripheral blood lymphocytes. *Vavilovskii Zhurnal Genetiki i Selektii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):50-58. DOI 10.18699/VJGB-22-08

## Транскрипционная активность генов репарации, апоптоза и клеточного цикла (*TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PADI4*, *MAPK8*, *NF-KB1*, *STAT3*, *GATA3*) у хронически облученных людей с различной интенсивностью апоптоза лимфоцитов периферической крови

В.С. Никифоров<sup>1, 2</sup>✉, Е.А. Блинова<sup>1, 2</sup>, А.И. Котикова<sup>1, 2</sup>, А.В. Аклеев<sup>1, 2</sup>

<sup>1</sup> Уральский научно-практический центр радиационной медицины Федерального медико-биологического агентства России, Челябинск, Россия

<sup>2</sup> Челябинский государственный университет, Челябинск, Россия

✉ nikiforovx@mail.ru

**Аннотация.** Исследовали транскрипционную активность генов, вовлеченных в поддержание генетического гомеостаза клетки (репарации, клеточного цикла и апоптоза: *TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PADI4*, *MAPK8*, *NF-KB1*, *STAT3*, *GATA3*), у лиц, подвергшихся хроническому радиационному облучению, с повышенной интенсивностью раннего, позднего апоптоза и некроза лимфоцитов периферической крови. Объектом изучения служили мононуклеарные клетки периферической крови, полученные от 132 жителей

прибрежных сел реки Течи, подвергшихся хроническому облучению. Доза облучения красного костного мозга составляла  $426.4 \pm 48.2$  мГр (1.3–2930.0 мГр), доза облучения тимуса и периферических органов иммунной системы –  $58.9 \pm 7.9$  мГр (0.1–489.0 мГр). Исследование проводили в отдаленные сроки (более 60 лет с начала облучения), возраст людей на время проведения обследования был  $68 \pm 0.6$  года (55–86 лет). Анализ апоптотической и некротической гибели лимфоцитов периферической крови основывался на наличии на поверхности мембраны клеток фосфолипида фосфатидилсерина, а также ее проницаемости для интеркалирующего ДНК-красителя. Оценку относительного содержания мРНК генов репарации, клеточного цикла и апоптоза проводили с использованием полимеразной цепной реакции в реальном времени. В группе хронически облученных людей с повышенной интенсивностью раннего апоптоза отмечено увеличение относительного содержания мРНК гена *PADI4* ( $p = 0.006$ ). Для хронически облученных людей с повышенной интенсивностью позднего апоптоза зафиксирована модуляция относительного содержания мРНК генов *TP53* ( $p = 0.013$ ) и *BCL-2* ( $p = 0.021$ ). В отдаленные сроки у хронически облученных людей с повышенной интенсивностью некроза лимфоцитов периферической крови отмечен статистически значимый рост транскрипционной активности гена *TP53* ( $p = 0.015$ ). Установлено, что у облученных людей с повышенной интенсивностью апоптоза регистрируются в первую очередь изменения со стороны транскрипционной активности апоптотических генов, что согласуется с существующими представлениями об активации программированной гибели клеток.

Ключевые слова: мРНК; апоптоз; некроз; лимфоциты; хроническое облучение.

## Introduction

Ionizing radiation is the factor that could trigger changes in transcriptional activity of genes that have a key role in maintaining the stability of cellular homeostasis (Kabacik et al., 2011). Complex molecular responses to genotoxic stress set into action a lot of regulatory mechanisms including apoptosis (Zeegers et al., 2017).

Apoptosis plays an important part in the development of both early and late effects of ionizing radiation (Verheij, Bartelink, 2000). Its activation starts with changes in the expression of the genes regulating the processes of DNA damage repair, cell cycle control, cell proliferation and differentiation, etc (Verheij, Bartelink, 2000). With cell death, a genetic program regulating the balance of intracellular pro- and anti-apoptosis factors starts developing. At the early stage of apoptosis, the expression of phosphatidylserine begins on the external surface of the membrane. However, its presence is not the strict requirement of cell death. Of great importance are its concentration and formation of a complex with other proteins. It sends a signal to the phagocytes to recognize the apoptotic cells (Bever, Williamson, 2016).

Protein p53 that regulates apoptotic genes, coding cells of cellular membrane (CD95, DR5), proteins of cytoplasm and proteins located on the mitochondrial membrane (proteins of the BCL-2 family), plays an important role in the activation of signaling cascade that induces apoptotic cell death (Chipuk, 2006). Moreover, the BAX/BCL-2 protein ratio predetermines the implementation of apoptotic cell death. It was demonstrated that in case of ionizing radiation apoptosis is initiated against the early repression of the *BAX* gene and increase in the activity of *BCL-2* in human blood cells (Azimian et al., 2015).

In physiological conditions, a strict balance of pro- and anti-apoptotic proteins is maintained. However, following radiation exposure as well as in the presence of various pathological conditions, a shift of this balance occurs due to changes in the expression of genes involved in apoptosis. In this respect, the study of the transcriptional activity of genes controlling cell proliferation and death is an important task of radiation

biology as the disturbance of apoptosis promotes the development of pathological conditions accompanied by retention of cells with unlimited proliferative potential in the exposed body (Baryshnikov, Shishkin, 2002), or by the development of cytopenic conditions associated with increased cell death (Kvatcheva, 2000).

In studies conducted earlier in chronically exposed residents of the Techa riverside settlements, changes in the intensity of apoptotic death of the peripheral blood lymphocytes were registered in the long-term period (Blinova et al., 2020a). Moreover, changes in the transcriptional activity of apoptotic genes accompanied by a decrease in the relative mRNA content of the *BCL-2* gene and increase in the relative content of mRNA of the *BAX* gene were demonstrated in exposed people 60 years after the onset of chronic radiation exposure (Nikiforov et al., 2020).

The next step of the work is the study of the relative content of mRNA of genes involved in cellular homeostasis in residents of the Techa riverside settlements, which are noted for the disturbed mechanism of cell elimination, in particular by increased intensity of apoptotic and necrotic cell death.

In this regard, the objective of the current study is to perform quantitative assessment of the content of mRNA of *TP53*, *MDM2*, *ATM*, *BAX*, *BCL-2*, *CDKN1A*, *OGG1*, *XPC*, *PADI4*, *MAPK8*, *NF-KB1*, *STAT3* and *GATA3* genes in the long-term period in chronically exposed residents of the Techa riverside villages who had an increased frequency of peripheral blood lymphocytes (PBL) at different stages of apoptosis and necrosis.

## Materials and methods

The study objects were the PBL of the 132 residents of the Techa riverside settlements who had been chronically exposed in 1949–1950 (Akleyev, 2016). Mean accumulated dose to the red bone marrow of all the exposed individuals was  $426.4 \pm 48.2$  mGy (1.3–2930.0 mGy), mean accumulated dose to the thymus and organs of the peripheral immune system was  $58.9 \pm 7.9$  mGy (0.1–489.0 mGy). Mean age of exposed people at the time of examination was  $68 \pm 0.6$  years (55–86 years).

The control group consisted of 32 people who were not chronically exposed and lived in similar social and economic conditions. Mean age of the control group members at the time of examination was  $67 \pm 1.25$  years (57–81 years). Intensity of apoptosis and necrosis was calculated in the control group according to the formula (1). For the purpose of the study the exposed people were subdivided into the following groups: exposed people with the intensity of apoptosis/necrosis exceeding the critical value, and exposed people with the intensity of apoptosis/necrosis within the normal range.

Intensity of apoptosis =  $\frac{\sum x}{n} + 2\sqrt{\frac{\sum (x_1 - \bar{x})^2}{n-1}}$ , (1)

where  $\frac{\sum x}{n}$  – mean value of the rate/frequency/intensity of

PBL apoptosis/necrosis;  $\sqrt{\frac{\sum (x_1 - \bar{x})^2}{n-1}}$  – standard deviation.

In the control group, mean frequency of cells at the early stage of apoptosis was 3.04, standard deviation – 4.52; at the late stage of apoptosis – 0.03, standard deviation – 0.06; at the stage of necrosis – 0.02, standard deviation – 0.04. Thus, the critical value of the frequency of cells for the early apoptosis was 12.08, for the late apoptosis – 0.15, for necrosis – 0.1. Exposed people who had frequency of cells at various stages of apoptotic death exceeding the critical value were included into the group with increased frequency of PBL apoptosis/necrosis. The description of the studied groups is given in Table 1.

Blood for the study of PBL apoptotic/necrotic death was taken from the cubital vein in a volume of 6 ml into Vacuette tubes with Li-heparin (Improvacuter, China). The study was performed using flow cytometer with the Annexin V FITC stain kit (BD, France). Leukocyte fraction was isolated in Ficoll-Urografin density gradient from whole blood (density 1077–1078 g/cm<sup>3</sup>) in accordance with the standard method (Kheifets, Abalakin, 1973). Annexin-V (human) recombinant (FITC conjugate) and DNA binding fluorescent dye propidium

iodide (PI) were added to cell-suspension. The analysis was performed using flow cytometer Navios (Beckman Coulter, USA). In the course of the analysis, cell populations at the stages of early and late apoptosis, and necrosis, as well as live cells were isolated. Results were given as a percentage ratio of cells that entered this or that stage of apoptosis and necrosis (see Table 1).

Blood for measuring mRNA relative content was taken from the cubital vein in a volume of 3 ml in sterile Tempus Blood RNA Collection Tubes (Applied Biosystem, USA). RNA was isolated through a column-based method with the Tempus Spin RNA Isolation Kit (Applied Biosystem). Information on concentration and purity of the isolated RNA samples was obtained using a NanoDrop 2000C spectrophotometer (Thermo Scientific, USA). Purity of the samples was measured by the values of absorption at wavelength of 260 and 280 nm (A260/280). The ratio of absorbances measured at A260/280 for purified RNA extracted from all blood samples was  $2.1 \pm 0.02$ . The total yield of the RNA was from 50 to 90 µg/ml. Reaction of reverse transcription for the cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystem). Relative quantitative content of the mRNA was measured with RT-PCR using the CFX96 Touch amplifier (Bio-Rad Laboratories, USA).

The relative amount of mRNA in the studied samples was determined using 2<sup>–ΔΔCt</sup>-method. The data were evaluated with respect to the relative level of mRNA of the “housekeeping” *COMT* and *B2M* genes and control group averaged values. Amplification curves were analyzed with the Bio-Rad CFX Manager 2.1 (Bio-Rad Laboratories) using the threshold line method. The calculation was performed taking into account three replicates for each gene and the efficiency of amplification was obtained by constructing calibration curves. Oligonucleotide sequence of the primers, temperature conditions of the RT-PCR were taken from international published papers and adapted to our experiments. The characteristics of primers

Table 1. Description of the individuals under study

Studied groups of exposed people with various intensity of apoptosis and necrosis		Number of people	Frequency of apoptotic cells, % M ± SD (range)	Age at the time of examination, years (range)	Accumulated dose to RBM, mGy M ± SE (range)	Accumulated dose to the thymus and organs of the peripheral immune system, mGy
Early apoptosis	Normal intensity	104	5.69 ± 2.46 (0.1–11.75)	68.19 ± 0.6 (55–86)	427.6 ± 58.1 (1.3–2930.0)	62.7 ± 9.65 (0.15–489.0)
	Increased intensity	26	17.17 ± 4.25 (13.09–30.45)	67.69 ± 1.3 (58–85)	415.5 ± 75.4 (10.4–1226.3)	41.3 ± 10.5 (0.74–252.2)
Late apoptosis	Normal intensity	88	0.05 ± 0.04 (0–0.14)	68.00 ± 0.7 (56–86)	380.2 ± 47.3 (2.13–187.1)	52.9 ± 8.1 (0.15–456.2)
	Increased intensity	43	0.67 ± 0.75 (0.15–3.66)	68.61 ± 0.9 (55–82)	523.1 ± 109.6 (1.3–2930.0)	70.7 ± 17.7 (0.18–489.0)
Necrosis	Normal intensity	106	0.02 ± 0.02 (0–0.09)	67.91 ± 0.6 (56–86)	395.4 ± 50.5 (1.3–2870.5)	59.5 ± 9.26 (0.2 – 489.0)
	Increased intensity	26	0.31 ± 0.36 (0.1–1.86)	69.31 ± 1.4 (55–82)	548.5 ± 130.4 (2.1–2930.0)	56.9 ± 14.5 (0.2–300.2)



are described in detail in (Blinova et al., 2020b; Nikiforov et al., 2020).

Statistical processing of the obtained data was performed using Statistica 10.0 and SigmaPlot software packages. The Kolmogorov–Smirnov test was used to check if the data in the samples were normally distributed. Since many of the investigated parameters did not have normal distribution, the non-parametric Mann–Whitney U-test and the Kruskal–Wallis test were used to compare the groups. The results were given as mean values, error of mean and range of the data (M, min–max) (Tables 2–4).

Correlation-regression analysis performed without taking into account the outliers was used to reveal the dependences of changes in the relative mRNA content in the studied genes on radiation factors (dose to RBM, thymus and organs of

the peripheral immune system).  $p$ -value  $\leq 0.05$  corrected for the multiple comparisons was used to exclude errors in the hypothesis wording.

## Results

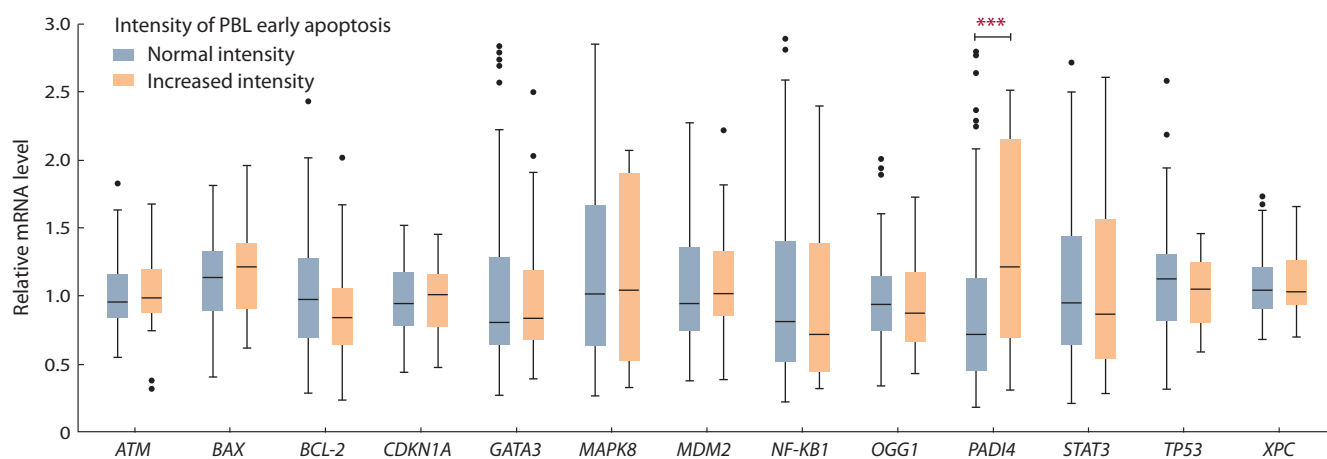
### Transcriptional activity of genes in chronically exposed people with increased intensity of early apoptosis

In the framework of the current study, a statistically significant increase (1.5 times) of the relative content of mRNA of the *PADI4* gene was registered in the group of exposed people with increased intensity of the PBL apoptosis relative to the exposed people with normal intensity of early apoptosis (see Table 2).

It can be seen from Fig. 1, that the changes in relative mRNA content of the *PADI4* gene are due to the shifts of median data

**Table 2.** Relative content of mRNA (rel. un.) of the genes in the groups of examined people with different intensity of PBL early apoptosis (M  $\pm$  SE; min–max)

Gene	Exposed people with normal intensity of PBL early apoptosis, N = 106	Exposed people with increased intensity of PBL early apoptosis, N = 26	$p$
<i>TP53</i>	1.15 $\pm$ 0.04 (0.31–2.55)	1.05 $\pm$ 0.05 (0.59–1.46)	0.523
<i>MDM2</i>	1.06 $\pm$ 0.05 (0.37–2.27)	1.08 $\pm$ 0.08 (0.39–2.22)	0.629
<i>BCL-2</i>	1.06 $\pm$ 0.05 (0.28–2.48)	0.90 $\pm$ 0.08 (0.23–2.01)	0.106
<i>OGG1</i>	0.98 $\pm$ 0.03 (0.34–2.00)	0.94 $\pm$ 0.07 (0.42–1.69)	0.460
<i>ATM</i>	0.99 $\pm$ 0.02 (0.55–1.83)	1.00 $\pm$ 0.06 (0.32–1.67)	0.441
<i>BAX</i>	1.12 $\pm$ 0.03 (0.40–1.81)	1.17 $\pm$ 0.07 (0.61–1.96)	0.438
<i>XPC</i>	1.07 $\pm$ 0.02 (0.68–1.73)	1.10 $\pm$ 0.05 (0.70–1.66)	0.693
<i>CDKN1A</i>	0.99 $\pm$ 0.05 (0.43–1.51)	0.98 $\pm$ 0.05 (0.48–1.45)	0.756
<i>STAT3</i>	1.05 $\pm$ 0.06 (0.21–2.71)	1.08 $\pm$ 0.13 (0.28–2.60)	0.919
<i>GATA3</i>	1.05 $\pm$ 0.07 (0.27–2.82)	1.03 $\pm$ 0.10 (0.39–2.50)	0.664
<i>MAPK8</i>	1.17 $\pm$ 0.07 (0.26–2.85)	1.14 $\pm$ 0.13 (0.32–2.07)	0.864
<i>NF-KB1</i>	0.99 $\pm$ 0.07 (0.22–2.88)	0.94 $\pm$ 0.12 (0.32–2.39)	0.610
<i>PADI4</i>	0.93 $\pm$ 0.07 (0.18–2.79)	1.37 $\pm$ 0.14 (0.30–2.51)	0.006



**Fig. 1.** Distribution of the relative content of mRNA of the studied genes in chronically exposed people with normal and increased intensity of PBL early apoptosis.

Here and in Figures 2 and 4 the data are presented as the median (25 and 75 percentile) and range (min–max); \*\*\* the differences between the groups are statistically significant ( $p < 0.05$ ).

to the area/region of high values in the group of chronically exposed people with increased intensity of early apoptosis of the PBL, and not due to the changes in the transcriptional activity of this gene in some exposed people.

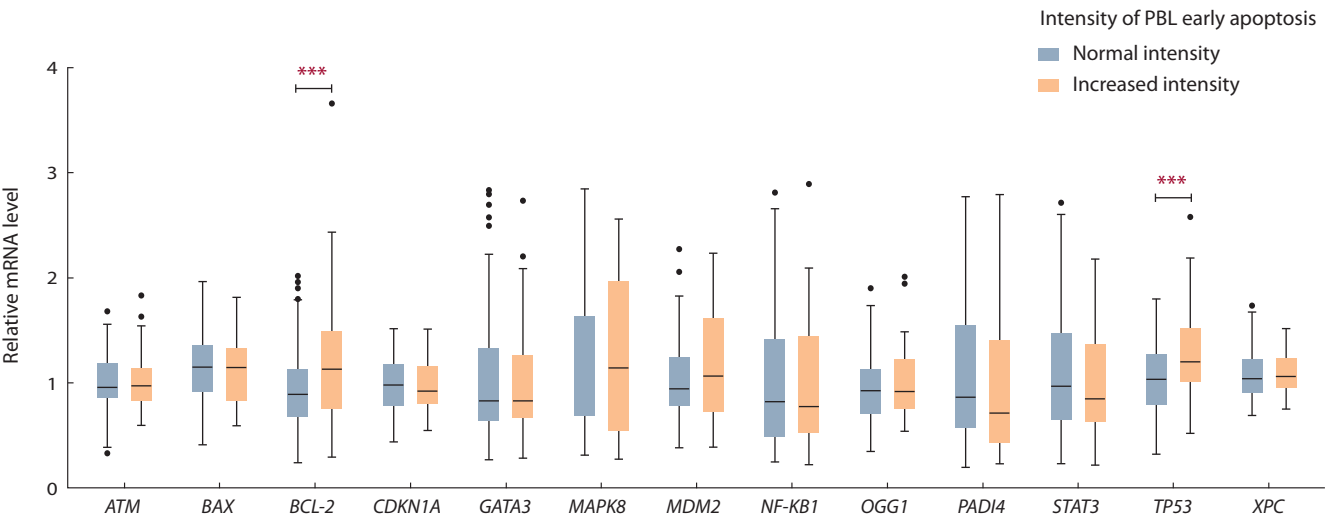
No statistically significant dependences were revealed when we checked the relationship between the mRNA content and dose characteristics (accumulated dose to RBM, thymus and organs of the peripheral immune system) in the group of exposed people with increased intensity of early apoptosis.

**Transcriptional activity of genes  
in chronically exposed people  
with increased intensity of late apoptosis**

In the study of the late stage of apoptosis, it was observed that chronically exposed people with increased intensity of late apoptosis have a statistically significant increase in the mRNA content of *TP53* and *BCL-2* genes relative to the exposed people with normal frequency of PBL at the late stage of apoptosis (see Table 3, Fig. 2).

**Table 3.** Relative mRNA content (rel. un.) of genes in the groups of examined people with different intensity of PBL late apoptosis (Me; Q1–Q3)

Gene	Exposed people with normal intensity of PBL late apoptosis, N = 89	Exposed people with increased intensity of PBL late apoptosis, N = 43	p
<i>TP53</i>	1.03±0.03 (0.30–1.79)	1.31±0.10 (0.50–2.65)	0.013
<i>MDM2</i>	1.01±0.04 (0.37–2.27)	1.18±0.09 (0.37–2.23)	0.140
<i>BCL-2</i>	0.94±0.04 (0.23–2.01)	1.19±0.09 (0.28–2.29)	0.021
<i>OGG1</i>	0.95±0.03 (0.33–1.90)	1.02±0.06 (0.53–2.00)	0.405
<i>ATM</i>	0.98±0.03 (0.32–1.67)	1.02±0.04 (0.58–1.82)	0.700
<i>BAX</i>	1.13±0.03 (0.40–1.96)	1.14±0.05 (0.59–1.81)	0.912
<i>XPC</i>	1.08±0.03 (0.68–1.73)	1.08±0.03 (0.74–1.51)	0.679
<i>CDKN1A</i>	0.95±0.02 (0.43–1.51)	1.07±0.12 (0.53–1.50)	0.940
<i>STAT3</i>	1.08±0.07 (0.22–2.71)	0.99±0.09 (0.21–2.17)	0.559
<i>GATA3</i>	1.06±0.07 (0.27–2.83)	1.04±0.11 (0.27–2.73)	0.902
<i>MAPK8</i>	1.13±0.07 (0.29–2.84)	1.22±0.13 (0.26–2.55)	0.699
<i>NF-KB1</i>	1.02±0.07 (0.24–2.80)	0.95±0.12 (0.22–2.89)	0.699
<i>PADI4</i>	1.06±0.08 (0.18–2.77)	0.94±0.12 (0.23–2.79)	0.345



**Fig. 2.** Distribution of the relative content of mRNA of the studied genes in chronically exposed people with normal and increased intensity of PBL late apoptosis.

As a result of the correlation analysis, negative correlation was revealed between relative content of the mRNA of the *BCL-2* ( $r = -0.6$ ;  $p = 0.001$ ) and *ATM* ( $r = -0.4$ ;  $p = 0.02$ ) genes and dose to RBM in chronically exposed people with increased intensity of PBL late apoptosis. In addition, a negative correlation between the content of mRNA and dose to the thymus and organs of the peripheral immune system was noted for the *BCL-2* ( $r = -0.4$ ;  $p = 0.002$ ) gene. The obtained dependences were studied using linear regression analysis (Fig. 3).

#### Transcriptional activity of genes in chronically exposed people with increased intensity of necrosis

Statistically significant differences between the groups of exposed people with different intensity of PBL necrosis were shown only for the *TP53* gene. An increase in the relative mRNA content of the *TP53* gene (almost by 1.5 times) was observed in chronically exposed persons with increased PBL intensity of necrosis compared to chronically exposed individuals with a normal frequency of PBL that entered the phase of necrosis (see Table 4, Fig. 4).

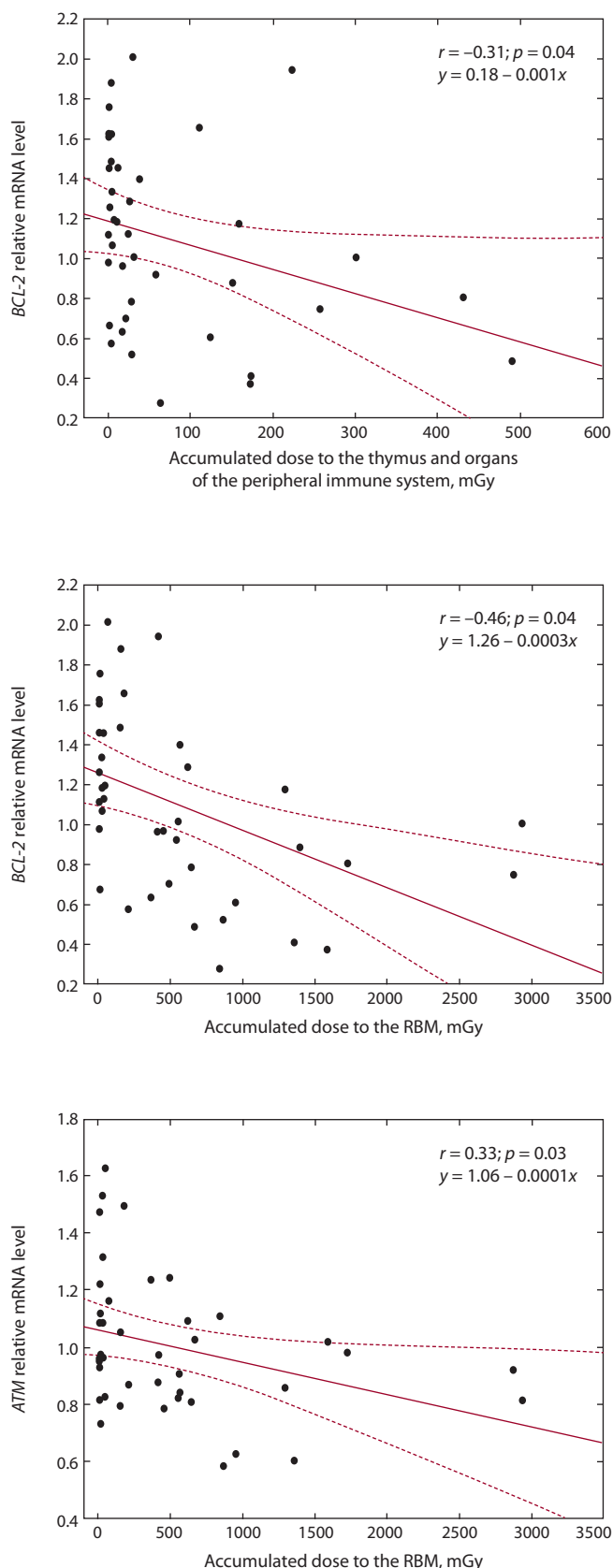
Negative correlation of the relative mRNA content of the *BCL-2* ( $r = -0.47$ ;  $p = 0.02$ ) and *ATM* ( $r = -0.6$ ;  $p = 0.001$ ) genes with RBM dose was registered in the group of exposed people with increased PBL frequency at the stage of necrosis. The results of linear regression analysis showed no significant dependence of changes of mRNA amount of the *BCL-2* gene on the accumulated RBM dose ( $p = 0.13$ ), while a statistically significant negative linear dependence of mRNA content on RBM dose was shown for the *ATM* gene in the group of chronically exposed persons with increased intensity of PBL necrosis (Fig. 5).

Verification of the relationship between relative mRNA content and intensity of necrotic cell death revealed a negative correlation for the *MAPK8* gene ( $r = -0.62$ ;  $p = 0.01$ ) in exposed people with increased frequency of PBL that entered the phase of necrosis.

#### Discussion

This study showed that chronically exposed people with increased frequency of PBL in the early stage of apoptosis have increased mRNA content of *PADI4* gene compared to exposed people with normal intensity of early apoptosis. *PADI4* is a  $\text{Ca}^{2+}$ -dependent enzyme that catalyzes protein citrullination in the presence of  $\text{Ca}^{2+}$  (Rogers et al., 1977). In particular, *PADI4* can mediate histone H3 citrullination on the promoters of p53 target genes such as *CDKN1A*, *BAX*, *BCL-2*, etc, and also bind to the p53 C-terminal regulatory domain, which causes repression of its activity (Tanikawa et al., 2012). In this regard, we can assume that *PADI4* protein is an important mediator of the p53 signaling pathway that can lead to apoptosis activation.

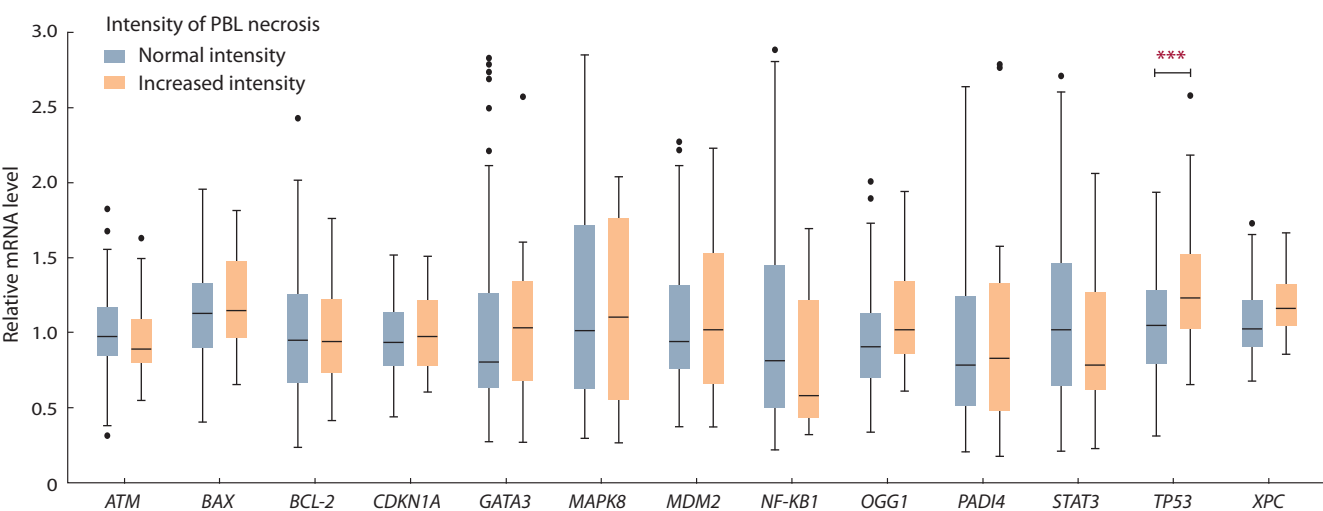
In the group of exposed people with increased intensity of PBL apoptotic death at a late stage, modification of transcriptional activity of *TP53* and *BCL-2* genes is observed. In particular, a statistically significant increase in the relative mRNA content of these genes has been shown.



**Fig. 3.** Linear dependence of the changes of the mRNA relative content of the *ATM* and *BCL-2* genes on the accumulated dose to the RBM, thymus and organs of the peripheral immune system in the group of chronically exposed people with increased intensity of PBL late apoptosis.

**Table 4.** Relative mRNA content (rel. unit) of genes in groups of exposed people with various intensity of PBL necrosis (M ± SE; min–max)

Gene	Exposed people with normal frequency of PBL necrosis, N = 106	Exposed people with increased frequency of PBL necrosis, N = 26	p
<i>TP53</i>	1.01 ± 0.03 (0.31–1.94)	1.40 ± 0.15 (0.64–2.22)	0.015
<i>MDM2</i>	1.05 ± 0.04 (0.37–2.27)	1.12 ± 0.13 (0.37–2.23)	0.741
<i>BCL-2</i>	1.00 ± 0.04 (0.23–2.43)	1.10 ± 0.12 (0.41–1.73)	0.750
<i>OGG1</i>	0.95 ± 0.04 (0.33–2.00)	1.08 ± 0.08 (0.61–1.94)	0.150
<i>ATM</i>	1.01 ± 0.02 (0.32–1.83)	0.96 ± 0.06 (0.55–1.63)	0.151
<i>BAX</i>	1.12 ± 0.03 (0.40–1.95)	1.19 ± 0.07 (0.65–1.81)	0.390
<i>XPC</i>	1.07 ± 0.02 (0.68–1.73)	1.17 ± 0.05 (0.85–1.67)	0.079
<i>CDKN1A</i>	0.95 ± 0.02 (0.43–1.51)	1.18 ± 0.19 (0.59–1.50)	0.411
<i>STAT3</i>	1.08 ± 0.06 (0.21–2.71)	0.89 ± 0.12 (0.23–2.06)	0.220
<i>GATA3</i>	1.05 ± 0.07 (0.27–2.83)	1.06 ± 0.14 (0.27–2.56)	0.592
<i>MAPK8</i>	1.16 ± 0.07 (0.29–2.84)	1.15 ± 0.15 (0.27–2.04)	0.960
<i>NF-KB1</i>	1.04 ± 0.07 (0.22–2.89)	0.78 ± 0.11 (0.32–1.69)	0.185
<i>PADI4</i>	1.03 ± 0.07 (0.20–2.64)	1.03 ± 0.19 (0.18–2.79)	0.893

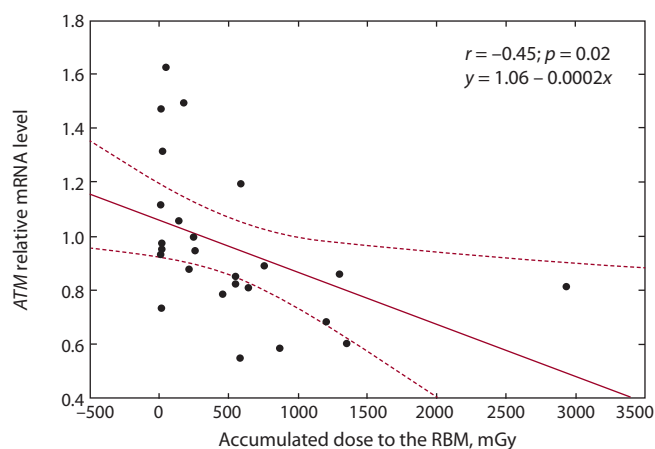


**Fig. 4.** Distribution of relative mRNA content of studied genes in chronically exposed people with normal and increased intensity of PBL necrosis.

One of the main functions of p53 is induction of signaling mechanisms aimed at the elimination of potentially harmful cells (Miyashita et al., 1994). However, against the background of increased transcriptional activity of *TP53*, there is an increase in the anti-apoptotic *BCL-2* gene in the group of exposed people with increased intensity of apoptosis. At the same time, the amount of mRNA of the *BCL-2* gene decreases with the increase of dose to RBM, thymus and organs of the peripheral immune system. At this stage of work, it is difficult to explain this phenomenon; there is probably a violation of the mechanism of cell elimination against the background of hyperexpression of anti-apoptotic factors in some exposed

people with increased intensity of apoptosis. This is also indicated by the fact that the transcriptional activity of the *TP53* gene is increased in an exposed person with increased intensity of necrosis. Moreover, negative correlation between the relative mRNA content of *ATM* genes and RBM dose was recorded in exposed people with increased intensity of late apoptosis. *ATM* gene dysfunction leads to a progression of genome instability, which is primarily accompanied by increased frequency of chromosome aberrations (telomere length shortening, increased level of paired and single chromosome fragments and frequency of translocations) (Hahn, Weinberg, 2002; Franco et al., 2006).





**Fig. 5.** Linear dependence of changes of relative mRNA content of ATM genes on accumulated RBM dose in the group of chronically exposed people with increased intensity of PBL necrosis.

It is possible that down-regulation of this gene transcription, which increases with an increasing dose to red bone marrow of the residents of the Techa Riverside villages, is associated with depletion of intracellular reserves for neutralizing the resulting DNA damage, and thus is the leading cause of the increased intensity of cell death.

In the group of chronically exposed people with increased intensity of necrosis, a decrease in relative mRNA content of *MAPK8* gene with increased intensity of PBL necrotic death is observed against the background of increased transcriptional activity of the *TP53* gene.

MAPK8 phosphorylates hundreds of substrates responsible for stress response control and apoptosis regulation, including p53 (Guimaraes, Hainaut, 2002). In addition, MAPK8 phosphorylates BMF (BCL-2 modulating factor) on specific serine residues located inside and immediately adjacent to the BMF binding domain. BMF released from actin enters mitochondria, physically interacts with the BCL-2 protein, which subsequently also initiates apoptosis (Puthalakath et al., 2001).

## Conclusion

Thus, in the framework of the study it has been noted that changes in the transcriptional activity of apoptotic genes are primarily registered in exposed people with increased intensity of apoptosis, which is consistent with current ideas about the activation of programmed cell death. It was shown that gene expression depends on the stage of PBL apoptosis.

The study should be continued with an expanded sample of examined people and studied targets, which will allow defining the significance of parameters of transcriptional activity of some genes as markers of cancer and non-cancer incidence risk associated with apoptosis registered in chronically exposed people in the long-term period.

## References

Akleyev A.V. (Ed.) Consequences of Radioactive Contamination of the Techa River. Chelyabinsk: Kniga Publ., 2016. (in Russian)

- Azimian H., Bahreyni-Toossi M.T., Rezaei A.R., Rafatpanah H., Hamzehloei T., Fardid H. Up-regulation of Bcl-2 expression in cultured human lymphocytes after exposure to low doses of gamma radiation. *J. Med. Phys.* 2015;40(1):38-44. DOI 10.4103/0971-6203.152249.
- Baryshnikov A.Yu., Shishkin Yu.V. Immunological Problems of Apoptosis. Moscow, 2002. (in Russian)
- Bevers E.M., Williamson P.L. Getting to the outer leaflet: physiology of phosphatidylserine exposure at the plasma membrane. *Physiol. Rev.* 2016;96(2):605-645. DOI 10.1152/physrev.00020.2015.
- Blinova E.A., Kotikova A.I., Yanishevskaya M.A., Akleyev A.V. Apoptosis of lymphocytes and polymorphisms of apoptosis regulation genes in individuals exposed to chronic radiation exposure. *Meditsinskaya Radiologiya i Radiatsionnaya Bezopasnost' = Medical Radiology and Radiation Safety.* 2020a;65(4):36-42. DOI 10.12737/1024-6177-2020-65-4-36-42. (in Russian)
- Blinova E.A., Nikiforov V.S., Yanishevskaya M.A., Akleyev A.A. Single nucleotide polymorphism and expression of genes for immune competent cell proliferation and differentiation in radiation-exposed individuals. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2020b;24(4):399-406. DOI 10.18699/VJ20.632. (in Russian)
- Chipuk J.E. Dissecting p53-dependent apoptosis. *Cell Death Differ.* 2006;13(6):994-1002. DOI 10.1038/sj.cdd.4401908.
- Franco S., Alt F.W., Manis J.P. Pathways that suppress programmed DNA breaks from progressing to chromosomal breaks and translocations. *DNA Repair. (Amst.).* 2006;5(9-10):1030-1041. DOI 10.1016/j.dnarep.2006.05.024.
- Guimaraes D.P., Hainaut P. TP53: a key gene in human cancer. *Biochimie.* 2002;84(1):83-93. DOI 10.1016/s0300-9084(01)01356-6.
- Hahn W.C., Weinberg R.A. Modelling the molecular circuitry of cancer. *Nat. Rev. Cancer.* 2002;2(5):331-341. DOI 10.1038/nrc795.
- Kabacik S., Mackay A., Tamber N., Manning G., Finnon P., Pailier F., Ashworth A., Bouffler S., Badie C. Gene expression following ionising radiation: identification of biomarkers for dose estimation and prediction of individual response. *Int. J. Radiat. Biol.* 2011;87(2):115-129. DOI 10.3109/09553002.2010.519424.
- Kheifets L.B., Abalakin V.A. Separation of human blood corpuscles in the verografin-ficoll density gradient. *Laboratornoye Delo = Laboratory Science.* 1973;10:579-581. (in Russian)
- Kvatcheva Yu.E. Human bone marrow repair processes and cell populations in acute radiation injury: morphological study. *Radiatsionnaya Biologiya. Radioekologiya = Radiological Biology. Radioecology.* 2000;40(1):5-9. (in Russian)
- Miyashita T., Krajewski S., Krajewska M., Wang H.G., Lin H.K., Liebermann D.A., Hoffman B., Reed J.C. Tumor suppressor p53 is a regulator of *bcl-2* and *bax* gene expression *in vitro* and *in vivo*. *Oncogene.* 1994;9(6):1799-1805.
- Nikiforov V.S., Blinova E.A., Akleyev A.V. The transcriptional activity of cell cycle and apoptosis genes in chronically exposed people with an increased frequency of TCR mutant lymphocytes. *Radiatsiya i Risk = Radiation and Risk.* 2020;29(2):89-100. DOI 10.21870/0131-3878-2020-29-2-89-100. (in Russian)
- Puthalakath H., Villunger A., O'Reilly L.A., Beaumont J.G., Coultas L., Cheney R.E., Huang D.C., Strasser A. BMF: a proapoptotic BH3-only protein regulated by interaction with the myosin V actin motor complex, activated by anoikis. *Science.* 2001;293(5536):1829-1832. DOI 10.1126/science.1062257.
- Rogers G.E., Harding H.W., Llewellyn-Smith I.J. The origin of citrulline-containing proteins in the hair follicle and the chemical nature

- of trichohyalin, an intracellular precursor. *Biochim. Biophys. Acta*. 1977;495(1):159-175. DOI 10.1016/0005-2795(77)90250-1.
- Tanikawa C., Espinosa M., Suzuki A., Masuda K., Yamamoto K., Tsuchiya E., Ueda K., Daigo Y., Nakamura Y., Matsuda K. Regulation of histone modification and chromatin structure by the p53-PAD14 pathway. *Nat. Commun.* 2012;3:676. DOI 10.1038/ncomms1676.
- Verheij M., Bartelink H. Radiation-induced apoptosis. *Cell Tissue Res.* 2000;301(1):133-142. DOI 10.1007/s004410000188.
- Zeegers D., Venkatesan S., Koh S.W., Low G.K., Srivastava P., Sundaram N., Sethu S., Banerjee B., Jayapal M., Belyakov O., Basakar R., Balajee A.S., Hande M.P. Biomarkers of ionizing radiation exposure: A multiparametric approach. *Genome Integr.* 2017;8:6. DOI 10.4103/2041-9414.198911.

---

#### ORCID ID

V.S. Nikiforov orcid.org/0000-0002-6685-1823  
E.A. Blinova orcid.org/0000-0002-2567-7945  
A.I. Kotikova orcid.org/0000-0002-1695-1340  
A.V. Akleyev orcid.org/0000-0003-2583-5808

**Acknowledgements.** The study was carried out within the framework of the State Contract No. 27.501.21.2 dated June 11, 2021 "Modernization of high-tech methods aimed at identifying the medical consequences of radiation effects on the personnel of the Mayak PA and the population of the Ural region".

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

Received June 29, 2021. Revised November 15, 2021. Accepted November 16, 2021.

# Association between *TP53*, *MDM2* and *NQO1* gene polymorphisms and viral load among women with human papillomavirus

A.H. AlBosale<sup>1, 2</sup>✉, E.V. Mashkina<sup>2</sup>

<sup>1</sup> Al-Dour Technical Institute, Northern Technical University, Department of Medical Laboratory Techniques, Mosul, Iraq

<sup>2</sup> Southern Federal University, Rostov-on-Don, Russia

✉ abbashammadi4@gmail.com

**Abstract.** The risk of cervical cancer is caused by persistent human papillomavirus (HPV) infection. Cervical cancer is the most frequent cancer among women. Our purpose was to investigate the association between *TP53* 215C>G (Pro72Arg), *MDM2* -410T>G, and *NQO1* 609C>T gene polymorphisms with a high HPV load and the influence of gene-gene interactions on prolonged HPV infection. Eighty-nine women with a high HPV viral load and 114 healthy women were involved in a case-control study. Genotyping for *TP53* 215C>G (Pro72Arg) and *MDM2* -410T>G SNPs was carried out by allele-specific PCR and genotyping for *NQO1* 609C>T was performed by a TaqMan assay. Quantitative analysis of HPV DNA was performed by AmpliSens® HPV HCR screen-titer-FRT test system. Gene-gene interactions were analyzed using the multifactor dimensionality reduction (MDR) method. The study of separate SNPs of *MDM2* -410T>G and *NQO1* 609C>T genes did not reveal any statistically significant difference in genotype and allele frequencies among women within the two groups. The frequency of the 215G (72Arg) allele and 215GG (72Arg/Arg) genotype of the *TP53* gene was significantly higher in the case group than in the control group (OR = 1.74, 95 % CI = 1.10–2.73;  $p = 0.02$  and OR = 1.97, 95 % CI = 1.13–3.46;  $p = 0.04$ , respectively). MDR analysis showed the significance of intergenic interactions of the three studied loci *TP53* (rs1042522) – *MDM2* (rs2279744) – *NQO1* (rs1800566) for the formation of a high HPV load (OR = 3.05, 95 % CI = 1.73–5.46;  $p = 0.0001$ ).

Key words: polymorphism; human papillomavirus; viral load; *TP53*; *MDM2*; *NQO1*.

**For citation:** AlBosale A.H., Mashkina E.V. Association between *TP53*, *MDM2* and *NQO1* gene polymorphisms and viral load among women with human papillomavirus. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):59-64. DOI 10.18699/VJGB-22-09

## Ассоциация полиморфизма генов *TP53*, *MDM2* и *NQO1* с вирусной нагрузкой среди женщин с вирусом папилломы человека

A.X. АлБосале<sup>1, 2</sup>✉, Е.В. Машкина<sup>2</sup>

<sup>1</sup> Аль-Доур технический институт, Северный технический университет, кафедра медицинской лабораторной техники, Мосул, Ирак

<sup>2</sup> Южный федеральный университет, Ростов-на-Дону, Россия

✉ abbashammadi4@gmail.com

**Аннотация.** Риск рака шейки матки вызван персистирующей инфекцией вируса папилломы человека (ВПЧ). Наша цель – исследовать связь между полиморфизмами генов *TP53* 215C>G (Pro72Arg), *MDM2* -410T>G и *NQO1* 609C>T с риском формирования высокой вирусной нагрузки при ВПЧ-инфекции. Восемьдесят девять женщин с высокой вирусной нагрузкой ВПЧ и 114 здоровых женщин были вовлечены в исследование случай-контроль. Генотипирование для SNP *TP53* Pro72Arg и *MDM2* -410T>G проводили методом аллель-специфичной ПЦР, а для *NQO1* 609C>T – путем анализа ПЦР в реальном времени с использованием TaqMan зондов. Количественный анализ ДНК ВПЧ выполняли с использованием тест-системы «АмплиСенс ВПЧ ВКР скрин-титр-FL». Анализ межгенных взаимодействий осуществляли с помощью алгоритма многофакторного снижения размерности (MDR). Исследование отдельных SNP генов, *MDM2* -410T>G и *NQO1* 609C>T, не выявило статистически значимой разницы в частотах генотипов и аллелей среди женщин в двух группах. Частота аллеля 72Arg и генотипа 72Arg/Arg гена *TP53* в группе ВПЧ-инфицированных женщин была значительно выше, чем в контрольной группе (OR = 1.74, 95 % CI = 1.10–2.73;  $p = 0.02$  и OR = 1.97, 95 % CI = 1.13–3.46;  $p = 0.04$  соответственно). MDR-анализ показал значимость межгенных взаимодействий исследуемых локусов *TP53* (rs1042522) – *MDM2* (rs2279744) – *NQO1* (rs1800566) для формирования высокой нагрузки ВПЧ (OR = 3.05, 95 % CI = 1.73–5.46;  $p = 0.0001$ ).

Ключевые слова: полиморфизм; вирус папилломы человека; вирусная нагрузка; *TP53*; *MDM2*; *NQO1*.

## Introduction

Human papillomavirus (HPV) is implicated in the development of cervical cancer. A key critical step in papillomavirus-related carcinogenesis is a persistent viral infection (Vonsky et al., 2019). There is heterogeneity in the development of human papillomavirus infection due to genetic variations, ethnicity, viral types involved in infection, viral load, and oncogenic expression, as well as environmental, and hormonal, physiological, and nutritional factors (Roura et al., 2016; Tasic et al., 2018). After HPV-infection, especially with high-risk HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59), HPV oncoproteins induce mutations in oncogenes, epigenetic modifications, and chromosomal rearrangements (Mittal, Banks, 2017). A disequilibrium in the relationship between virus and host results in a decrease in the effectiveness of the immune system, the imbalance between cellular and humoral immune processes, as well as alteration in pro- and anti-inflammatory cytokine levels, which increases the replicative ability of the virus (Fernandes et al., 2015; Bordignon et al., 2017). In addition, modifications that alter the stability of cell cycle proteins such as retinoblastoma protein (pRb), tumor suppressor p53, result in uncontrolled cell cycle progression and induce oncogenic transformation of cells (Sen et al., 2017; Balasubramaniam et al., 2019).

The *TP53* tumor suppressor gene plays a crucial role in regulating DNA repair, apoptosis, and cell cycle control. It has been observed that most human tumors contain mutated p53, with about 50 % of those mutations causing a reduction in DNA repair ability, irregular cell growth, and, eventually, progression to malignancy (Aubrey et al., 2016). Polymorphisms in the *TP53* gene change p53 protein conformation, which leads to p53 degradation through a process mediated by ubiquitin (Rampias et al., 2014). The most widely studied of the non-synonymous SNP *TP53* Pro72Arg (rs1042522) replaces proline (Pro) with arginine (Arg) in the p53 protein due to a substituted C to G base in the *TP53* gene. Both variants have the same binding affinity for DNA while their ability to bind components of the transcription factor is different. So, the two variants of the p53 protein are not functionally equivalent (Thomas et al., 1999). The p53 is ubiquitinated in the proteasome, which is regulated by MDM2 via a ubiquitin-dependent degradation pathway and NAD(P)H quinone oxidoreductase 1 via a ubiquitin-independent degradation pathway (Tsvetkov et al., 2010; Karni-Schmidt et al., 2016). As a result, the level of p53 is affected by MDM2 and NQO1 activity.

Oncoprotein MDM2 is a negative regulator of the p53 tumor protein (Saadatzaheh et al., 2017). A functional SNP in the *MDM2* gene promoter (-410T>G rs2279744) regulates MDM2 protein expression. When T is replaced with G, this increases the affinity of the transcriptional activator Sp1, resulting in higher MDM2 expression and subsequent suppression of the p53 pathway (Bond et al., 2004).

The NQO1 enzyme can catalyze the reduction of various quinones to hydroquinones by a two-electron reduction mechanism (NADH or NADPH) as a reducing cofactor. This two-electron reduction prevents the formation of free radicals (semiquinones) that protect the cells from oxidative stress (Atia, Abdullah, 2020). The SNP of *NQO1* at nucleotide position 609C>T in exon 6 (rs1800566) with the proline to serine amino acid substitution at codon 187 induces a change

of enzyme activity. The homozygotes (*TT*) genotype gives rise to an inactive enzyme NQO1, heterozygotes (*CT*) have the enzyme displays mild activity, while the wild homozygotes (*CC*) have the highest activity of the NQO1 (Ross, Siegel, 2004). Wild type NQO1 partially inhibits HPV E6-mediated p53 degradation, although this does not occur with the mutant type NQO1 (Niwa et al., 2005).

Thus, the efficiency of the cell cycle repair and control system depends not only on the p53 protein. Also, the levels of MDM2 and NQO1 proteins in the cell can affect the stability of the p53 protein and the activity of its degradation processes. However, human papillomavirus, as an exogenous factor, can be an additional cause affecting the work of the repair system. Most of the studies on the association of SNPs of genes with HPV infection and cervical cancer are devoted to the analysis of individual nucleotide substitutions. There is practically no data in the literature on the combined effect of polymorphic variants of these three genes in the presence of HPV load.

Our work aims to analyze the distribution of the polymorphisms of the *TP53* gene (rs1042522), *MDM2* gene (rs2279744), and *NQO1* gene (rs1800566) in patients with HPV load versus HPV-negative women.

## Materials and methods

Two hundred and three samples of epithelial cells scraped from the urogenital tract of women were used for molecular genetic studies. The study equipment has been provided by the clinical diagnostic laboratory, Nauka (Rostov-on-Don, Russia). The women were divided into two groups: women with a high HPV load (above 3 log of HPV genomes per 100 thousand human cells) ( $n = 89$ ), and HPV-negative women ( $n = 114$ ). All the women included in the study were over thirty years old. Criteria for women being included in the control group: a normal result of colposcopy, HPV-negative PCR-test. The comparative group of cases included women with symptoms such as vaginal discharge, bleeding menstrual abnormalities, and HPV-positive PCR-test with an HPV load of more than 3 log of HPV genomes per  $10^5$  human cells. The ethnic composition of the women involved in the study groups was as follows: Russians accounted for 86 %, Armenians accounted for 9 %, and other nationalities of the Caucasian race – 5 %.

All women have given formal written consent to take part in the study. The study was approved by the Bioethics Committee of the Academy of Biology and Biotechnology of the Southern Federal University (Protocol No. 2 of March 29, 2016). All the tests for clinical experimentation were carried out in line with the standards and ethical guidelines of the World Medical Association (Helsinki Declaration).

The total DNA was isolated from scraping epithelial cells of the cervical canal of women according to the DNA-sorb-AM (NextBio, Russia) reagent kit protocol. The quantification of DNA for high-risk HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) in biological material was analyzed according to the AmpliSens-HPV HCR screen-titre-FRT PCR kit (Interlabservice, Russia) protocol. According to the kit manufacturer's instructions and clinical reports, the viral load is interpreted as follows:  $\log \leq 3$  per  $10^5$  human cells – low clinical significance, 3–5 log per  $10^5$  human cells – clinically significance, risk of dysplasia; and  $> 5$  log per  $10^5$  human cells – clinically significance, strongly probable dysplasia.



Genotyping was performed for the SNP of *TP53* 215C>G (Pro72Arg) (rs1042522), *MDM2* -410T>G (rs2279744) genes by allele-specific PCR and the SNP-express reagent (Lytech, Russia) according to the kit protocol. *NQO1* 609C>T (rs1800566) was genotyped by a TaqMan genotyping assay. The amplification was carried out in a 25-ml reaction containing 2 µl 25 mM MgCl<sub>2</sub>, 1 µl 10 mmol/L of the forward primer (5'-CAG AGT GGC ATT CTG CAT TTC T-3') and reverse (5'-CTG GAG TGT GCC CAA TGC TA-3') primers and 0.5 µl mmol/L *NQO1* wild-type (5'-6FAM-CTT AGA ACC TCA ACT GA-MGBNFQ-3') and mutant (5'-VIC-CTT AGA ATC TCA ACT GAC A-MGBNFQ-3') probes, 0.5 µl Taq-polymerase (5 U/µl), 2.5 µl 2.5 mM of dNTP, 2 µl 10×PCR buffer B, 12 µl ddH<sub>2</sub>O (Syntol, Russia) and 3 µl DNA. Cycling conditions were as follows: initial denaturation at 95 °C for 10 min, followed by 40 cycles consisting of denaturation at 95 °C for 15 sec, then annealing at 54 °C for 60 sec. The PCR products for *NQO1* 609C>T (rs1800566) were analyzed in real-time using RotorGene thermocycler. PCR products for the *TP53* Pro72Arg and *MDM2* -410T>G genes were analyzed by 3 % agarose gel horizontal electrophoresis and visualized under the ultraviolet (UV) transilluminator GelDoc (Bio-Rad, USA).

To calculate the statistical data, the  $\chi^2$  test was used to compare the allele and genotype frequencies of the *TP53* 215C>G (Pro72Arg) (rs1042522), *MDM2* -410T>G (rs2279744), and *NQO1* 609C>T (rs1800566) genes in the case group and control group. The Hardy–Weinberg equilibrium test was performed to determine the goodness-of-fit of the  $\chi^2$  test with one degree of freedom by comparing the observed genotype frequencies with the expected genotype frequencies. The SNP

genetic association was assessed by the  $\chi^2$  test, odds ratio (OR), and its confidence interval (CI). A *p*-value < 0.05 was considered statistically significant. Statistical analyses were performed using GraphPad InStat 3.05 software.

The analysis of intergenic interactions was performed using the MDR software (<http://www.multifactorialdimensionalityreduction.org/>) and by using the exhaustive search algorithm. All potential combinations of genotypes were evaluated with respect to the risk of developing an HPV infection. Multilocus genotypes are summed up in the MDR program into groups of increased and reduced disease risk, which reduces the dimension of the number of calculated parameters. Using multiple cross-recalculations of the input primary data, the optimal model is selected for intergenic interaction, with the highest accuracy and, accordingly, with the least error, to predict the presence or absence of predisposition to the studied pathology.

## Results

In 89 HPV-positive women, the average age was 40.1 ± 7.3 years and 41.1 ± 7.6 years in 114 HPV-negative women. Among 89 HPV-infected women the minimum, middle, and maximum HPV DNA load were 3.2, 5.1, and 8.6 log of HPV genomes per 100 thousand human cells, respectively.

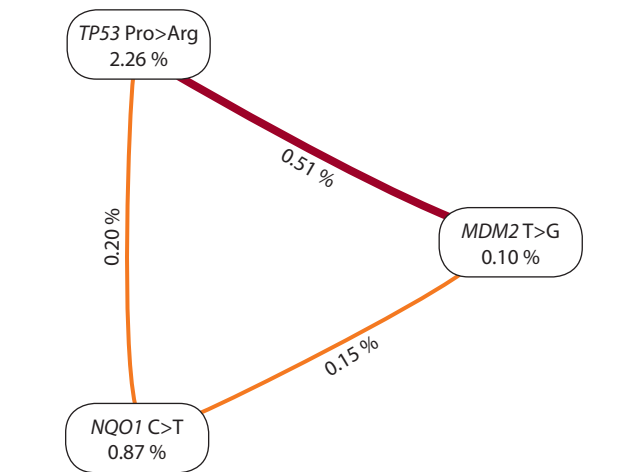
Frequency distributions for the three investigated *TP53*, *MDM2*, and *NQO1* gene polymorphisms are given in Table 1. The polymorphic variants of *MDM2* -410T>G and *NQO1* 609C>T were not associated with a high HPV load. At the same time females with a high HPV load had a significantly higher frequency of *TP53* 215G (72Arg) allele (OR = 1.74, 95 % CI = 1.10–2.73; *p* = 0.02) and 215GG (72ArgArg) genotype (OR = 1.97, 95 % CI = 1.13–3.46; *p* = 0.04) than

**Table 1.** Genotypes (abs., %) and alleles frequencies for *MDM2* -410T>G, *TP53* 215C>G (Pro72Arg), and *NQO1* 609C>T genes among women with high HPV load and without HPV

Gene, polymorphism	Group with high HPV load ( <i>n</i> = 89)	Control group without HPV ( <i>n</i> = 114)	<i>p</i>	OR (95 % CI)
<i>MDM2</i> -410T>G (rs2279744)				
<i>T</i>	0.651	0.671	0.68	0.92 (0.61–1.39)
<i>G</i>	0.348	0.328		1.09 (0.72–1.65)
<i>T/T</i>	35 (39.3)	47 (41.2)	0.86	0.92 (0.53–1.62)
<i>T/G</i>	46 (51.7)	59 (51.8)		1.00 (0.57–1.73)
<i>G/G</i>	8 (9)	8 (7)		1.31 (0.47–3.62)
<i>TP53</i> 215C>G (Pro72Arg) (rs1042522)				
<i>C</i>	0.213	0.321	0.02	0.58 (0.37–0.91)
<i>G</i>	0.787	0.679		1.74 (1.10–2.73)
<i>C/C</i>	3 (3.3)	9 (7.9)	0.04	0.41 (0.11–1.54)
<i>C/G</i>	32 (36)	55 (48.2)		0.60 (0.34–1.06)
<i>G/G</i>	54 (60.7)	50 (43.9)		1.97 (1.13–3.46)
<i>NQO1</i> 609C>T (rs1800566)				
<i>C</i>	0.516	0.587	0.18	0.75 (0.51–1.11)
<i>T</i>	0.483	0.412		1.33 (0.90–1.98)
<i>C/C</i>	20 (22.5)	36 (31.6)	0.29	0.63 (0.33–1.18)
<i>C/T</i>	52 (58.4)	62 (54.4)		1.18 (0.68–2.06)
<i>T/T</i>	17 (19.1)	16 (14)		1.45 (0.54–3.84)

**Table 2.** Analysis of intergenic interactions by the multifactor dimensional reduction algorithm (MDR)

Genes, polymorphisms in model	Testing balanced accuracy	Cross-validation consistency	$\chi^2$	$p$	OR (95 % CI)
<i>MDM2</i> (rs2279744) <i>TP53</i> (rs1042522) <i>NQO1</i> (rs1800566)	0.64	10/10	14.673	0.0001	3.05 (1.73–5.46)



Interaction analysis among three loci of *MDM2* -410T>G, *TP53* 215C>G (Pro72Arg), and *NQO1* 609C>T of women with high HPV viral load.

The informational value of each marker is presented on vertices; the informational value of the interaction for a pair of loci is presented on the edges. The nature of the interaction between genes is shown by the color of the line (red – pronounced synergism, orange – moderate synergism).

healthy controls. The existence of multiple allelic variations in genes that encode for protein molecules can lead to several related changes in the genome and proteome function. Therefore, an analysis of intergenic interactions of allelic variants was conducted.

An analysis of intergenic interactions showed that the three-locus model of gene interaction has a prediction accuracy of 64 % and cross-validation consistency (10/10) (Table 2). Interaction of *TP53* (rs1042522) – *MDM2* (rs2279744) – *NQO1* (rs1800566) genes is associated with the risk of high HPV load among women (OR = 3.05, 95 % CI = 1.73–5.46;  $p$  = 0.0001).

A radial diagram demonstrates the contribution of polymorphism of each gene, both individually and in combination with others for the three-loci. In the vertices of the diagram, the values of information for individual genes are indicated, on the edges – the information value of the interaction of a pair of genes. The studied SNPs affect the formation of the viral load to varying degrees. According to the model of loci interaction (see the Figure), the highest predictive potential is possessed by the SPN of the *TP53* gene (2.26 %). The *TP53* and *MDM2* loci have the greatest effect by intergenic interaction. A pronounced synergism was revealed between these loci – the total effect of the combination is 2.87 %. Its information value is higher than the sum of its individual effects.

**Discussion**

Cervical cancer is the most common gynecological cancer among women and the high-risk HPV genotypes play a major role in abnormal lesion development and cervical malignant

neoplasms (Malagón et al., 2019; So et al., 2019). The presence of a high viral load in HPV-positive women indicates that the virus has not been entirely removed and will likely continue to replicate in the body cells for a long time. Long-term virus persistence contributes to the incorporation of HPV DNA into the human genome, the expression of E6/E7 oncogenic proteins, and the development of cancer (McBride, Warburton, 2017; Gheit, 2019).

Human papillomatosis appears to be a polygenic disease, suggesting that recurrent, small-effect genetic variations can have consequences for disease susceptibility (Khoury et al., 2018). Tumor development is largely attributed to genetic variations in the host’s cell cycle control (Litwin et al., 2017). The relationships between the *TP53* gene (rs1042522), *MDM2* gene (rs2279744), *NQO1* gene (rs1800566), and the risk of high HPV load have been investigated in this study.

In our study, 43.8 % of women (89 out of 203) were positive for high-risk HPV types. Our analysis revealed an association of high viral load formation risk with 215G (72Arg) allele carriage (OR = 1.74, 95 % CI = 1.10–2.73;  $p$  = 0.02) and 215GG (72Arg/Arg) genotype of *TP53* gene (OR = 1.97, 95 % CI = 1.13–3.46;  $p$  = 0.04). On the contrary, 215C (Pro72) allele (OR = 0.58, 95 % CI = 0.37–0.91;  $p$  = 0.02) and 215CC (72ProPro) genotype (OR = 0.41, 95 % CI = 0.11–1.54;  $p$  = 0.002) showed protective effect compared to the control group. The polymorphic variants of the p53 protein with Pro or Arg in codon 72 have been shown to vary in the efficiency of interaction with the E6 oncoprotein of HPV (Storey et al., 1998). The Arg variant is degraded by the E6 oncoprotein more readily than the Pro variant. Therefore, the carriers of the Arg/Arg genotype have p53 protein more vulnerable to viral protein-induced degradation (Hu et al., 2010). Our results are consistent with several other studies suggesting that HPV-positive women are more vulnerable to cervical malignant neoplasms when having *TP53* 72Arg/Arg genotype and *TP53* 72Arg allele (Basyar et al., 2016; Moschonas et al., 2017).

*MDM2* promotes cell cycle progression through the activation of S-phase, via interaction with the retinoblastoma tumor suppressor protein and the transcriptional factor E2F (Oliner et al., 2016). *MDM2* is one of the central nodes in the p53 pathway regulation. It has been shown that even a small change in *MDM2* level may affect the p53 pathway and, subsequently, cancer development (Mendoza et al., 2014). Our analysis showed no statistically significant difference in the genotypes ( $p$  = 0.86) and allele frequencies ( $p$  = 0.68) distribution of *MDM2* -410T>G gene polymorphism in two women groups.

Our analysis showed no statistically significant difference in the genotypes ( $p$  = 0.29) and allele ( $p$  = 0.18) distribution of *NQO1* 609C>T gene polymorphism in the two groups of women. In agreement with our results, J. Chansaenroj and his coworkers showed no association of the *NQO1* 609C>T polymorphism with the risk of cervical cancer (Chansaenroj

et al., 2013). At the same time, several studies reported a relationship between *NQO1* 609TT genotypes and the risk of cervical cancer (Niwa et al., 2005; Yang et al., 2020). The *NQO1* gene (rs1800566) TT genotype is associated with null enzyme activity and could influence cancer progression by reducing cytotoxic agents containing the quinone moiety (Diao et al., 2017).

Favorable conditions for HPV persistence include multiple genetic substitutions which result in gene expression changes. In our work, the analysis of gene-gene interactions (MDR) showed significant interaction of the polymorphic loci (OR = 3.05, 95 % CI = 1.73–5.46;  $p = 0.0001$ ) for increased viral load (see Table 2). The interaction of the polymorphic variants for the three loci of the genes *TP53* 215C>G (Pro-72Arg), *MDM2* -410T>G, and *NQO1* 609C>T are associated with HPV viral load increase.

A synergistic effect was revealed between the studied loci. That is, the combined effect of these loci is more pronounced than individual effects. Thus, we revealed an increased risk of a high viral load in HPV infection in the case of a combination of polymorphic variants of the *TP53*, *MDM2*, and *NQO1* genes. The risk may be due to disturbances in the work of the checkpoints of the cell cycle due to the activation of the processes of degradation of the p53 protein.

The current study has several limitations. First, the small sample size: our results should be verified in larger populations as well as in other ethnic groups. Second, women with cervical cancer were not included in our research. Comparison of the different histological types of cervical cancer may also be warranted for future studies to determine whether the frequency of *TP53*, *MDM2*, and *NQO1* gene polymorphisms differ based on the histological types of cervical cancer. Third, the influence of epidemiologic risk factors such as smoking, alcohol intake, and sexual behavior or pathogenic factors like bacteria with the risk of HPV infection was not included. It would be interesting to analyze if *TP53*, *MDM2*, and *NQO1* production is associated with environmental or pathogenic factors.

## Conclusion

Our results demonstrate that the risk of high viral load formation is associated with *TP53* 215G (72Arg) allele and *TP53* 215GG (72ArgArg) genotype in HPV-positive women. Although the individual SNPs of *MDM2* -410T>G and *NQO1* 609C>T genes did not reveal a statistically significant frequency difference in our study, intergenic interactions analysis revealed significant interaction for all polymorphic variants. This demonstrated that the infection development depends on the synergistic effect of several polymorphisms that induce changes in gene expression and represent an allelic load for HPV-positive cells. However, the role of the genetic susceptibility to HPV infections and high HPV load with *TP53* rs1042522, *MDM2* rs2279744, *NQO1* rs1800566 polymorphisms requires further investigation.

## References

Atia A., Abdullah A. NQO1 enzyme and its role in cellular protection; an insight. *Iberoam. J. Med.* 2020;2(4):306-313. DOI 10.5281/zenodo.3877528.  
Aubrey B.J., Strasser A., Kelly G.L. Tumor-suppressor functions of the TP53 pathway. *Cold Spring Harb. Perspect. Med.* 2016;6(5):a026062. DOI 10.1101/cshperspect.a026062.

Balasubramaniam S.D., Balakrishnan V., Oon C.E., Kaur G. Key molecular events in cervical cancer development. *Medicina (Kaunas)*. 2019;55(7):384. DOI 10.3390/medicina55070384.  
Basyar R., Saleh A.Z., Sastradinata I., Yuwono Y. p53 gene codon 72 polymorphisms among cervical carcinoma patients. *Indones. J. Obstet. Gynecol.* 2016;3(3):165-169. DOI 10.32771/inajog.v3i3.48.  
Bond G.L., Hu W., Bond E.E., Robins H., Lutzker S.G., Arva N.C. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell*. 2004;119(5):591-602. DOI 10.1016/j.cell.2004.11.022.  
Bordignon V., Di Domenico E., Trento E., D'Agosto G., Cavallo I., Pontone M. How human papillomavirus replication and immune evasion strategies take advantage of the host DNA damage repair machinery. *Viruses*. 2017;9(12):390. DOI 10.3390/v9120390.  
Chansaenroj J., Theamboonlers A., Junyangdikul P., Swangvaree S., Karalak A., Chinchai T. Polymorphisms in TP53 (rs1042522), p16 (rs11515 and rs3088440) and NQO1 (rs1800566) genes in Thai cervical cancer patients with HPV 16 infection. *Asian Pac. J. Cancer Prev.* 2013;14(1):341-346. DOI 10.7314/APJCP.2013.14.1.341.  
Diao J., Bao J., Peng J., Mo J., Ye Q., He J. Correlation between NAD(P)H: quinone oxidoreductase 1 C609T polymorphism and increased risk of esophageal cancer: evidence from a meta-analysis. *Ther. Adv. Med. Oncol.* 2017;9(1):13-21. DOI 10.1177/1758834016668682.  
Fernandes J.V., De Medeiros T.A.A., De Azevedo J.C.V., Cobucci R.N.O., De Carvalho M.G.F., Andrade V.S. Link between chronic inflammation and human papillomavirus-induced carcinogenesis (Review). *Oncol. Lett.* 2015;9(3):1015-1026. DOI 10.3892/ol.2015.2884.  
Gheit T. Mucosal and cutaneous human papillomavirus infections and cancer biology. *Front. Oncol.* 2019;9:355. DOI 10.3389/fonc.2019.00355.  
Hu X., Zhang Z., Ma D., Huettner P.C., Massad L.S., Nguyen L. TP53, MDM2, NQO1, and susceptibility to cervical cancer. *Cancer Epidemiol. Biomarkers Prev.* 2010;19(3):755-761. DOI 10.1158/1055-9965.EPI-09-0886.  
Karni-Schmidt O., Lokshin M., Prives C. The roles of MDM2 and MDMX in cancer. *Annu. Rev. Pathol.* 2016;11:617-644. DOI 10.1146/annurev-pathol-012414-040349.  
Khouri R., Sauter S., Butsch Kovacic M., Nelson A., Myers K., Mehta P. Risk of human papillomavirus infection in cancer-prone individuals: What we know. *Viruses*. 2018;10(1):47. DOI 10.3390/v10010047.  
Litwin T., Clarke M., Dean M., Wentzensen N. Somatic host cell alterations in HPV carcinogenesis. *Viruses*. 2017;9(8):206. DOI 10.3390/v9080206.  
Malagón T., Louvanto K., Ramanakumar A.V., Koushik A., Coutlée F., Franco E.L. Viral load of human papillomavirus types 16/18/31/33/45 as a predictor of cervical intraepithelial neoplasia and cancer by age. *Gynecol. Oncol.* 2019;155(2):245-253. DOI 10.1016/j.jgyno.2019.09.010.  
McBride A.A., Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. *PLoS Pathog.* 2017;13(4):e1006211. DOI 10.1371/journal.ppat.1006211.  
Mendoza M., Mandani G., Momand J. The MDM2 gene family. *Biomol. Concepts*. 2014;5(1):9-19. DOI 10.1515/bmc-2013-0027.  
Mittal S., Banks L. Molecular mechanisms underlying human papillomavirus E6 and E7 oncoprotein-induced cell transformation. *Mutat. Res. Rev. Mutat. Res.* 2017;772:23-35. DOI 10.1016/j.mrrev.2016.08.001.  
Moschonas G.D., Tsakogiannis D., Lamprou K.A., Mastora E., Dimitriou T.G., Kyriakopoulou Z. Association of codon 72 polymorphism of p53 with the severity of cervical dysplasia, E6-T350G and HPV16 variant lineages in HPV16-infected women. *J. Med. Microbiol.* 2017;66(9):1358-1365. DOI 10.1099/jmm.0.000563.  
Niwa Y., Hirose K., Nakanishi T., Nawa A., Kuzuya K., Tajima K. Association of the NAD(P)H: quinone oxidoreductase C609T poly-

- morphism and the risk of cervical cancer in Japanese subjects. *Gynecol. Oncol.* 2005;96(2):423-429. DOI 10.1016/j.ygyno.2004.10.015.
- Oliner J.D., Saiki A.Y., Caenepeel S. The role of MDM2 amplification and overexpression in tumorigenesis. *Cold Spring Harb. Perspect. Med.* 2016;6(6):a026336. DOI 10.1101/cshperspect.a026336.
- Rampias T., Sasaki C., Psyrri A. Molecular mechanisms of HPV induced carcinogenesis in head and neck. *Oral. Oncol.* 2014;50(5):356-363. DOI 10.1016/j.oraloncology.2013.07.011.
- Ross D., Siegel D. NAD(P)H:quinone oxidoreductase 1 (NQO1, DT-diaphorase), functions and pharmacogenetics. *Methods Enzymol.* 2004;382:115-144. DOI 10.1016/S0076-6879(04)82008-1.
- Roura E., Travier N., Waterboer T., de Sanjosé S., Bosch F.X., Pawlita M. The influence of hormonal factors on the risk of developing cervical Cancer and pre-Cancer: results from the EPIC cohort. *PLoS One.* 2016;11(1):e0147029. DOI 10.1371/journal.pone.0147029.
- Saadatzadeh M., Elmi A., Pandya P., Bijangi-Vishehsaraei K., Ding J., Stamatkin C. The role of MDM2 in promoting genome stability versus instability. *Int. J. Mol. Sci.* 2017;18(10):2216. DOI 10.3390/ijms18102216.
- Sen P., Ganguly P., Ganguly N. Modulation of DNA methylation by human papillomavirus E6 and E7 oncoproteins in cervical cancer (Review). *Oncol. Lett.* 2017;15(1):11-22. DOI 10.3892/ol.2017.7292.
- So K.A., Lee I.H., Lee K.H., Hong S.R., Kim Y.J., Seo H.H. Human papillomavirus genotype-specific risk in cervical carcinogenesis. *J. Gynecol. Oncol.* 2019;30(4):e52. DOI 10.3802/jgo.2019.30.e52.
- Storey A., Thomas M., Kalita A., Harwood C., Gardiol D., Mantovani F. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. *Nature.* 1998;393(6682):229-234. DOI 10.1038/30400.
- Tasic D., Lazarevic I., Knezevic A., Tasic L., Pikula A., Perisic Z. The impact of environmental and behavioural cofactors on the development of cervical disorders in HR-HPV-infected women in Serbia. *Epidemiol. Infect.* 2018;146(13):1714-1723. DOI 10.1017/S0950268818001668.
- Thomas M., Kalita A., Labrecque S., Pim D., Banks L., Matlashewski G. Two polymorphic variants of wild-type p53 differ biochemically and biologically. *Mol. Cell Biol.* 1999;19(2):1092-1100. DOI 10.1128/MCB.19.2.1092.
- Tsvetkov P., Reuven N., Shaul Y. Ubiquitin-independent p53 proteasomal degradation. *Cell Death Differ.* 2010;17(1):103-108. DOI 10.1038/cdd.2009.67.
- Vonsky M., Shabaeva M., Runov A., Lebedeva N., Chowdhury S., Palefsky J.M. Carcinogenesis associated with human papillomavirus infection. Mechanisms and potential for immunotherapy. *Biochemistry.* 2019;84(7):782-799. DOI 10.1134/S0006297919070095.
- Yang S., Zhao J., Li L. NAD(P)H: quinone oxidoreductase 1 gene rs1800566 polymorphism increases the risk of cervical cancer in a Chinese Han sample. *Medicine (Baltimore).* 2020;99(20):e19941. DOI 10.1097/MD.00000000000019941.

#### ORCID ID

A.H. AlBosale orcid.org/0000-0003-1520-8245  
E.V. Mashkina orcid.org/0000-0002-4424-9508

**Acknowledgements.** This study was performed with the equipment of the Center of collective use "High technologies" (Southern Federal University, Russia).

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

Received May 11, 2021. Revised September 30, 2021. Accepted October 1, 2021.



# The functional insight into the genetics of cardiovascular disease: results from the post-GWAS study

L.O. Bryzgalov<sup>1</sup>, E.E. Korbolina<sup>1</sup>✉, I.S. Damarov<sup>1</sup>, T.I. Merkulova<sup>1, 2</sup>

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

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

✉ lungry@bionet.nsc.ru

**Abstract.** Cardiovascular diseases (CVDs), the leading cause of death worldwide, generally refer to a range of pathological conditions with the involvement of the heart and the blood vessels. A sizable fraction of the susceptibility loci is known, but the underlying mechanisms have been established only for a small proportion. Therefore, there is an increasing need to explore the functional relevance of trait-associated variants and, moreover, to search for novel risk genetic variation. We have reported the bioinformatic approach allowing effective identification of functional non-coding variants by integrated analysis of genome-wide data. Here, the analysis of 1361 previously identified regulatory SNPs (rSNPs) was performed to provide new insights into cardiovascular risk. We found 773,471 coding co-segregating markers for input rSNPs using the 1000 Genomes Project. The intersection of GWAS-derived SNPs with a relevance to cardiovascular traits with these markers was analyzed within a window of 10 Kbp. The effects on the transcription factor (TF) binding sites were explored by DeFine models. Functional pathway enrichment and protein-protein interaction (PPI) network analyses were performed on the targets and the extended genes by STRING and DAVID. Eighteen rSNPs were functionally linked to cardiovascular risk. A significant impact on binding sites of thirteen TFs including those involved in blood cells formation, hematopoiesis, macrophage function, inflammation, and vasoconstriction was found in K562 cells. 21 rSNP gene targets and 5 partners predicted by PPI were enriched for spliceosome and endocytosis KEGG pathways, endosome sorting complex and mRNA splicing REACTOME pathways. Related Gene Ontology terms included mRNA splicing and processing, endosome transport and protein catabolic processes. Together, the findings provide further insight into the biological basis of CVDs and highlight the importance of the precise regulation of splicing and alternative splicing.

Key words: non-coding genetic variation; rSNPs; cardio-vascular disease risk; GWAS association; 1000 Genomes Project; gene expression regulation; transcription factor binding.

**For citation:** Bryzgalov L.O., Korbolina E.E., Damarov I.S., Merkulova T.I. The functional insight into the genetics of cardiovascular disease: results from the post-GWAS study. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2022;26(1):65-73. DOI 10.18699/VJGB-22-10

## Функциональный подход к изучению генетики сердечно-сосудистых заболеваний: post-GWAS исследование

Л.О. Брызгалов<sup>1</sup>, Е.Е. Корболина<sup>1</sup>✉, И.С. Дамаров<sup>1</sup>, Т.И. Меркулова<sup>1, 2</sup>

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

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

✉ lungry@bionet.nsc.ru

**Аннотация.** Сердечно-сосудистые заболевания (ССЗ) занимают одно из ведущих мест по уровню смертности в развитых странах. На сегодняшний день значительное число геномных локусов продемонстрировали ассоциации с ССЗ, в основном в полногеномных исследованиях (GWAS), но лишь в немногих случаях установлены молекулярно-генетические механизмы, лежащие в основе ассоциации. Таким образом, задача определения функционально значимых для предрасположенности к ССЗ генетических вариантов остается актуальной. Ранее мы разработали биоинформатический подход, позволяющий эффективно идентифицировать регуляторные SNP (rSNP) – функциональные однонуклеотидные полиморфизмы, влияющие на экспрессию генов, путем анализа комплекса полногеномных данных. В данной работе с использованием данных проекта «1000 геномов» найден 773 471 кодирующий SNP, являющийся косегрегирующим маркером для 1361 идентифицированного нами rSNP. Проанализировано расположение этих маркеров в пределах геномного окна размером 10 Кб вокруг маркеров GWAS, ассоциированных с риском развития ССЗ или потенциально связанными с этим риском фенотипическими признаками. Влияние rSNP на сайты связывания транскрипционных факторов исследовано с помощью моделей DeFine. Анализ межбелковых взаимодействий и обогащения биологических путей для соответствующих генов-мишеней проведен с использованием баз STRING и DAVID. Найдено восемнадцать rSNP, функционально

связанных с риском развития ССЗ. В клеточной линии K562 выявлено значительное влияние этих rSNP на сайты связывания тринадцати транскрипционных факторов, в том числе участвующих в процессах кроветворения, функционирования макрофагов, воспаления и вазоконстрикции. Для двадцати одного соответствующего генамишени, а также пяти функционально взаимодействующих с ними генов показано обогащение биологическими категориями сплайсосомы и эндоцитоза (KEGG), комплекса сортировки эндосом и процессов сплайсинга (REACTOME), а также некоторыми терминами генетических онтологий, в том числе сплайсинга и процессинга мРНК. В целом полученные результаты расширяют имеющиеся представления о роли rSNP и связанных с ними изменений экспрессии генов в развитии ССЗ и подчеркивают важность точной регуляции процессов сплайсинга мРНК и альтернативного сплайсинга для формирования определенного фенотипа.

Ключевые слова: некодирующие полиморфизмы; регуляторные SNP; предрасположенность к сердечно-сосудистым заболеваниям; полногеномные исследования ассоциаций, проект «1000 геномов»; регуляция экспрессии генов; транскрипционные факторы.

## Introduction

While the running data indicate that the prevalence of cardiovascular disease may vary among regions of the world, they remain one of the leading causes of death and health loss and a large proportion of their forms are shown to have a familial aggregation and high heritability (Smith J.G., Newton-Cheh, 2015; Roth et al., 2017; Wang Y., Wang J.-G., 2018). The previous efforts led to the identification of candidate risk genes including the genes of renal homeostasis for Mendelian forms of abnormal blood pressure levels and several transcription factors (including NKX25, GATA4, TBX) for congenital septal (Kathiresan, Srivastava, 2012). However, the broad group of cardiovascular traits such as myocardial infarction/ischemia or coronary artery disease (CAD) show complex inheritance patterns, which suggest the collective and non-linear effects from multiple genetic and non-genetic factors. With the recent technological advances, the whole-exome sequencing (Li A.H. et al., 2017; Seidemann et al., 2017; Khera et al., 2019) and genome-wide association studies, GWASs (in particular (Erdmann et al., 2018; Schunkert et al., 2018)) have been shown to be a powerful tool for discovering the genetic variation associated with cardiovascular risk. The outcomes from multiple GWASs and their meta-analysis completed during the past decade have been deposited in the catalogs, such as the catalog of published GWASs from The National Human Genome Research Institute (Buniello et al., 2019), the Coronary ARtery Disease Genome-wide Replication And Meta-analysis, CARDIoGRAM (Preuss et al., 2010) plus The Coronary Artery Disease (C4D) Consortium and UK Biobank (Ge et al., 2017).

The genes related to regulating blood pressure, the tone and elasticity of the vascular wall, the inflammation process, the proliferation of vascular smooth muscle cells and the levels of low density lipoprotein cholesterol (LDL-C) are 'traditionally' involved in cardiovascular risk. Moreover, the GWA studies resulted in numerous loci for cardiometabolic risk factors such as plasma biomarkers of lipid metabolism, thrombosis, inflammation and metabolic status playing a role in risk analysis. In the case of long QT syndrome, in particular, fifteen candidate genes have been reported to date, including several genes for ion channels (Refsgaard et al., 2012; Arking et al., 2014). Notably, three candidates (*KCNQ1*, *KCNH2*, and *SCN5A*) account for approximately 75 % of cases (Wallace et al., 2019). And there are more than 150 suggestive loci estimated for CAD although only 46 from them have reached the genome-wide significance threshold (den Hoed et al., 2015). Interestingly,

a considerable overlap has been shown between the risk genes for monogenic forms of CVDs and those generating an association signal in GWAS (Rau et al., 2015).

Together, these findings have developed a relatively comprehensible picture of the biology underlying the cardiovascular disease, but despite the advances, we are not able to analyze the functionality for the majority of the reported associated genetic loci. Among the reasons, there may be important limitation of GWAS for identifying risk genomic regions instead of risk genes and the non-coding localization of the majority of the susceptibility SNPs (Ward, Kellis, 2012). Current theories assume that so-called regulatory non-coding SNPs (rSNPs) seem to make the greatest contribution to the development of various multifactorial diseases including oncological and CAD as these are directly involved in the control of the gene expression levels.

One of the lessons learned is the growing need for the comprehensive post-GWAS analysis in order to translate the reported statistical association to uncover the causal variants amongst those in linkage disequilibrium (Mansur et al., 2018; Smith A.J.P. et al., 2018). Moreover, only some of GWAS-implicated loci (i. e. sixteen of 46 validated loci for CAD) are also associated with 'classic' genes of risk, therefore showing the potential involvement of 'non-traditional' biological pathways in the disease (Smith J.G., Newton-Cheh, 2015).

Since the advances in next-generation sequencing technologies have provided an expanding amount of large-scale *-omics* datasets, the research strategies have started to focus on the integration of various genome-wide information layers (Huang S. et al., 2017) using functional genomics assays. One way to find putative functional variants is to detect regions with allele-specific binding of transcription factors or histone modifications, suggesting their different regulatory downstream role. ChIP-seq data will provide the snapshots of protein-DNA interactions allowing the analysis of sites with significant difference in signal between the alleles or allelic differences. The employment of transcriptome (RNA-seq) data will provide the snapshots of gene expression levels depending on the allele. The epigenome (iTEA) analysis (Meng et al., 2018) in combination with the regulatory sequence annotations, i. e. DNase-seq and ChIP-seq datasets (Cavalli et al., 2019), is beginning to be used to screen for the causal variants changing gene expression, including within GWAS-derived loci. However, the researchers have not come close to solving the issue of identifying the rSNPs at the genome-wide scale. The limitations are imposed by the incomplete experimen-

tal data collected to date and some critical methodological problems. Notably, one of the major challenges has become the development of effective *in silico* (bioinformatic) approaches but these are relatively few in number to date and only the individual studies have been reported to elucidate the underlying mechanisms of CVDs (Gong et al., 2018; Roman, Mohlke, 2018).

To address the challenges, we have recently reported an effective bioinformatic approach that facilitates the systematic identification of functional non-coding variants from available genome-wide data (Korbolina et al., 2018). Our pipeline utilized multiple positional and functional criteria to reveal non-coding regulatory variants in the human genome and imputed curated GWAS association signals to select the potentially colorectal cancer-causal rSNPs within a 1 Kb window of genomic sequence centered at the GWAS-SNP. Initially, the regulatory properties of found rSNPs were shown on a number of human cell lines of different origins (HCT116, K562, MCF7). However, expression of tissue-specific transcription factors is suppressed in cell lines. For this reason, here we have adopted the list of 1361 regulatory SNPs from the said study. The data from 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015) were incorporated in the analysis to improve matching rSNPs with the phenotypic outcome that would be the risk of CVDs. Further, we tried to narrow the focus toward rSNPs that potentially result in a difference in predicted binding status of various transcription factors and performed the functional annotation of the targeted genes.

## Materials and methods

**Input data on non-coding regulatory variants in human genome.** Regulatory SNPs from our earlier study (Korbolina et al., 2018) were used for input including the data on identified targeted genes. All of these were associated with the allele-specific binding of various TFs and allele-specific expression from raw data.

**Genetics data.** Genetic variants and allele information were retrieved from dbSNP150 (Database resources..., 2016) and four 1000 Genomes Project super populations (AFR, AMR, ASN and EUR) (1000 Genomes Project Consortium et al., 2015). The GRCh37 annotation was used to map genetic variants to gene loci.

**Assessment of the SNP clustering with a distance measure.** To map significant GWAS associations to novel functional variants here we implemented the data of 1000 Genomes Project. First, we extracted the data on 2500 individual haplotype-resolved human genomes from various super populations (AFR, AMR, ASN and EUR). We found 1361 variants from the 1476 input rSNPs reported earlier (Korbolina et al., 2018) within this genomic data. Next, we extracted the data on all SNPs within the transcribed genomic regions. The lists of rSNPs defined from 1000 genomes ('population' set of 1361 variant) and coding SNPs defined from 1000 genomes were consolidated. Each individual from 2500 genome samples was genotyped separately by each SNP from the consolidated list. The genotype data were turned to binary, where "0" represented the most frequent allele and "1" – the minor allele within the individual genotype. The resulting genotype data were set to the 2D matrix containing 2500 individual genomes and genotypes for all SNPs from the consolidated

SNP1	11010011010011		
SNP2	11000011010011	XOR	Relative distance
Distance	00010000000000	=	1
SNP1	11010011010011		
SNPn	11010010010000	XOR	Relative distance
Distance	00000001000011	=	3

**Fig. 1.** Finding the relative SNP-to-SNP distance using XOR.

The binary code was used for genotypes (the symbols 0 and 1 were used to represent two different alleles in a heterozygous site).

list. The initial matrix was then transformed to the distance matrix using XOR logic gate. Figure 1 shows an example of SNP distance measurement by XOR.

All the relative distances were normalized on the number of times the minor allele was found there in the genomic data of 2500 genomes. Based on the analysis of the resulting distances, we quantified the likelihood of the repeated recognition of rSNPs and the coding SNP within one genotype in human populations. Here we found that in total 773 471 coding SNPs ( $p < 0.01$ ) may be merely co-segregating markers for our input rSNPs.

**Implementing GWAS data to interpret the input rSNPs functionality.** Next, we examined GWAS index SNPs available up to date (May 2019). We used the 'cardio' signature for querying the GWAS Catalog (including 'heart', 'coronary artery disease', 'CAD', 'platelet', 'blood', 'blood cells', 'pressure', 'count', 'vessel', 'caliber', 'pulse', 'artery'). Next, we evaluated whether the input rSNPs or any of the corresponding coding markers lying within a 10 Kbp window of GWAS SNPs (Brodie et al., 2016) could be related to heart and vascular disease association signals (Suppl. Table 1)<sup>1</sup>.

**A functional protein association network.** STRING v 11 (Szklarczyk et al., 2019) was selected as the PPI database with a subset of 21 genes targeted by 18 rSNPs that were found to be associated with cardiovascular risk, as input (see Suppl. Table 2 for details of enrichment analysis).

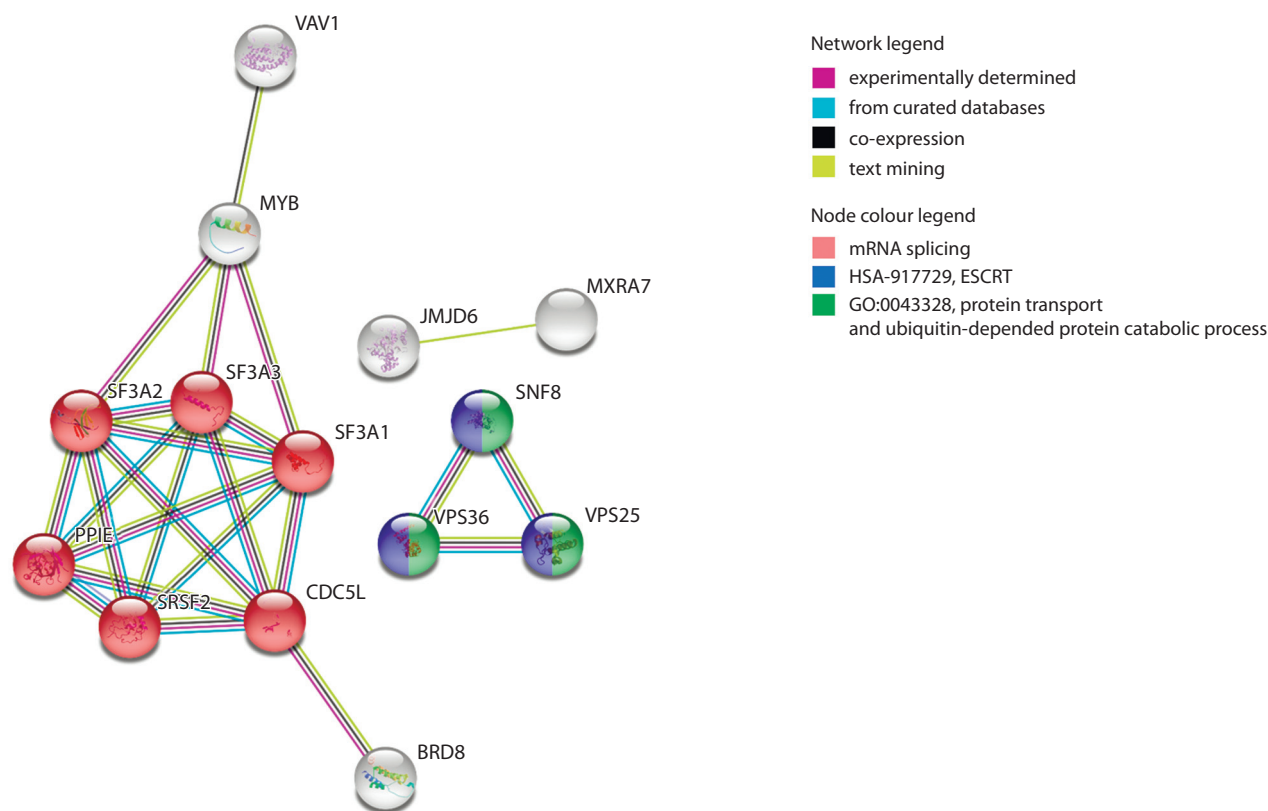
**Functional annotation by DAVID.** DAVID Database for Annotation, Visualization and Integrated Discovery (Database resources..., 2016) was used to further interpret the same targeted genes with the default values set for all parameters. The outputs of DAVID functional annotation and clustering tools are given in the Suppl. Table 3.

**In silico analysis of potentially affected TF binding sites.** The sequence-based DeFine deep learning models were employed to predict the effects on transcription factor binding in K562 cells (the data are accessible online via the DeFine tool) and to rank 18 rSNPs identified for cardiovascular risk. The DeFine functional scores predict the transcriptional factor-DNA binding intensities and are appointed in the view of the differences between the reference sequence and the altered sequence, as reviewed in (Wang M. et al., 2018). The outputs including the maximum TF functional scores, the most likely candidate TFs and top 10 contact genes for each rSNP position are given in the Suppl. Table 4.

**R code.** The R package, version 3.1.0, was used for data analysis. The custom-made Perl scripts employed are available upon request.

<sup>1</sup> Supplementary Tables 1–4 are available in the online version of the paper: <https://disk.icgbio.ru/s/zBiq4fm632zRywe>





**Fig. 2.** Protein association network for 21 rSNP-targeted genes and 5 gene partners with a functional relevance to cardiovascular risk in STRING.

The network for rSNP targets has been expanded by additional 5 proteins (via the 'More' button in the STRING interface and default confidence cut-off). The network contains 26 nodes with 24 edges (vs 10 expected edges, the disconnected nodes are hidden); enrichment  $p$ -value < 0.001. The legend inset at the right shows the various types of evidence for the predicted association and the enriched annotation term for the protein (by different colours of the nodes).

## Results

### Investigating rSNP functional relevance to cardiovascular risk via GWAS associations

We analyzed 438 GWAS-SNPs with relevance for cardiovascular traits and defined eighteen candidate rSNP variants at a 10 Kbp window size (see Suppl. Table 1) including within the associated loci for coronary heart and artery disease (CAD, 3 loci), HLD cholesterol, QT interval, red blood cells and platelet cells traits. One interesting result was that ten GWAS-derived SNPs including the ones for phenotypical associations with systolic and diastolic blood pressure, pulse pressure, retinal arteriolar microcirculation and one for CAD entered the list of founded regulatory SNPs in the study. We considered the input rSNP targets (Korbolina et al., 2018), and any gene targeted to these ten GWAS-SNPs (initially from GWAS catalog) to be a candidate for mediating the association. Actually, only one cardio-related SNP out of eighteen was linked to more than one target gene: rs3744061 (*MFSD11*, *JMJD6*).

### Functional annotation of the rSNP targeted genes

We further looked into the target genes to these eighteen cardio-vascular risk rSNPs as candidates for mediating the effects on CVDs. In our STRING enrichment analysis (Szkarczyk et al., 2019), 21 rSNP targeted genes had five predicted partners and these were shown significantly enriched in Kyoto Encyclopedia of Genes and Genomes (KEGG) spliceosome

pathway, two REACTOME pathways (endosomal sorting complex required for transport and mRNA splicing) and 40 Gene Ontology terms including gene expression, RNA splicing, regulation of mRNA splicing, regulation of alternative mRNA splicing via spliceosome, protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway; ubiquitin-dependent protein catabolic process (see Suppl. Table 2 for the details of STRING enrichment analysis). Figure 2 shows the corresponding protein association network with evidence for association between the targets. Such an enrichment indicates that the input proteins are at least partially biologically connected as a group.

The same enriched groups were found by DAVID functional annotation tool (Huang D.W. et al., 2008), including mRNA splicing and mRNA 3'-splice site recognition, regulation of transcription and endosomal transport GO terms (see Suppl. Table 3). However, these processes could all be linked to the pathological features of CVDs, but the most notable is the group of targets that are associated with splicing and alternative splicing regulation (given in red nodes).

### DeFine rSNP prioritization

The DeFine online tool revealed that eight rSNPs had positive functional scores meeting the pathogenic potential to enhance TF binding, and five – to weaken TF binding (see Suppl. Table 4) in the supported K562 cell line. The binding sites of



ten TFs were most strongly altered by 13 rSNPs according to DeFine models: TAL1, REST, NR1H2, SP2, RFX5, MXI1, PBX2, NFYB, ZNF274 and ZNF263. Nine out of ten of these TFs (with the exception of ZNF274) were the proteins with antibodies to which the immunoprecipitation was made by the original authors (as given in the Supplementary material, Korbolina et al., 2018). This would be expected in principle when examining certain regulatory regions.

In more detail, rs210962 (GWAS-derived) and rs7920217 functional variants shared the potential to weaken the binding of RE1 silencing transcription factor (REST) to the related genomic loci (the DeFine scores of  $-0.0778$  and  $-0.0888$ , respectively). rs2270574 and rs8106212 within the GWAS loci for CAD and platelet trait, respectively, shared the association with SP2 transcription factor, but had the opposite effects on TF binding according to DeFine scores ( $0.05020$  to enhance and  $-0.0973$  to weaken the binding, respectively). And again, two rSNPs shared the effects on TF binding sites for T-cell acute leukemia protein 1 (TAL1) in this study: rs140492 within the locus for HDL cholesterol (DeFine score  $0.0996$  to enhance the TF binding was counted) and rs10445033 – within GWAS locus for red blood cell levels with the DeFine score  $-0.0891$  to weaken the TF binding. Again, DeFine identified the same targeted genes (i. e. *SNF8* for rs2270574) with additional potential targets in each rSNP case (ten top candidate genes are listed in the DeFine output data, see Suppl. Table 4).

## Discussion

As has been mentioned in the Results, we have identified eighteen rSNPs with a functional relevance to CVDs that matched the GWAS loci for coronary heart and artery disease, HLD cholesterol, retinal arteriolar microcirculation, QT interval, red blood cells and platelet cells traits from this study (see Suppl. Table 1). The genome position of rSNP coincided with that for GWAS-derived SNP for ten identified variants out of eighteen. It was an interesting find as this was a relatively large part compared to our previously reported results for colorectal cancer (Korbolina et al., 2018) and cognitive disorders (Bryzgalov et al., 2018).

What we should like to mention is that a number of truly critical processes depend on the blood cells functionality and biological activities as has been widely demonstrated. Regarding erythrocytes, these include not only oxygen transport, but immune response (Astle et al., 2016), redox homeostasis (Kuhn et al., 2017) and regulation of vascular function (Helms et al., 2018; Rifkind et al., 2018). Moreover, several *ex vivo* studies on diabetes mellitus identified that red blood cells do act to mediate the development of endothelial dysfunction and cardiac injury (Yang et al., 2013; Zhou et al., 2018; Pernow et al., 2019). This means that any qualitative or quantitative deviations from the physiological ranges may be closely linked to the disease (Leal et al., 2018). The data suggest that the blood cell count and hematological parameters could be useful markers to improve the cardiovascular risk prediction; however, they have limited sensitivity (Mozos, 2015; Samman Tahhan et al., 2017; Lassale et al., 2018; Haybar et al., 2019). Interestingly, it was shown that the level of expression of some curated genes may independently aid in the prediction of heart failure prognosis when combined with neutrophil-to-lymphocyte ratio (Wan et al., 2018). The risk for CVDs

correlates well with platelet traits (Sloan et al., 2015; Vélez, García, 2015; Reinthaler et al., 2016; Gill et al., 2018), the initiation and progression of CAD in particular (Uysal et al., 2016). However, some have argued that shared genetic pathways linking blood cells with complex pathologies, including autoimmune diseases, schizophrenia, and CAD may be non-causal (Astle et al., 2016).

Still, our current knowledge of splicing regulation and alternative splicing in the heart is limited, but splicing analysis has emerged as an important line of research for the cardiovascular risk. The studies revealed that the regulation of splicing and alternative splicing events (reviewed in (van den Hoogenhof et al., 2016)) seem to play a causative role in heart development and cardiovascular disease. The promising therapeutic targets (Rexiati et al., 2018) have already been proposed. There is evidence that a significant number of alternate transcripts are increased in diseased hearts compared to controls, and can be involved in disease. Thus, abnormal splicing of apoptotic genes contributes to the pathogenesis of several CVDs including dilated and diabetic cardiomyopathy, atherosclerosis and heart failure as reviewed in (Dlamini et al., 2015). The dysregulation of cardiac splicing factors can also be sufficient to affect heart function and lead to disease. Thus, there is evidence that the decrease in RNA-binding motif protein 20 (RBM20) levels may be involved in dilated cardiomyopathy by providing input to splicing of at least several known target genes (Maatz et al., 2014). In the study, the target gene of the rs4360494 functional variant within the GWAS-derived locus for pulse pressure is the *SF3A3* gene that encodes subunit 3 of the splicing factor 3a protein heterotrimeric complex. As is known, the splicing factor 3a plays an important role in U2 snRNP biogenesis and thus, pre-mRNA splicing (Krämer et al., 2005; Huang C.-J. et al., 2011). With respect to other pathological states, targeting the components of the spliceosome has full potential as a strategy for cancer treatment and prognosis (Lin, 2017; El Marabti, Younis, 2018; Martinez-Montiel et al., 2018). Thus, the data suggest that SF3A3 is involved in the p53 activation, the induction of cell cycle arrest and cell death in non-small cell lung cancer (Siebring-van Olst et al., 2017). The *SRSF2* gene, encoding another splicing machinery component, could be used as a reliable prognostic factor in patients with hepatocellular carcinoma (Luo et al., 2017).

The role of endosomal system in the heart functioning and cardiovascular disease is described as critical, too (Yang et al., 2019), as endosomes contribute to control of cholesterol (LDL) plasma levels (Wijers et al., 2018),  $Ca^{2+}$  homeostasis and protein trafficking (Curran et al., 2015) and play an important part in atherosclerosis risk (Cai et al., 2018). However, surprisingly little is known regarding the regulation of endosome-based protein trafficking in the heart. Thus, our results could serve as useful background for further research. Protein quality and control and ubiquitin-proteasome system, UPS (Gilda et al., 2016; Barac et al., 2017; Gilda, Gomes, 2017; Dorsch et al., 2019; Shukla, Rafiq, 2019) have also played an essential role in the initiation and progression of CVDs. In short, the findings suggest that UPS contributes to structural remodeling of the myocardium, ischemia-reperfusion injury and myocardial cell loss, important components of progressive heart failure. There is evidence for the non-degradative role as well, as the ubiquitination was shown to affect the important

regulators of signaling pathways including those for the cell growth and apoptosis, DNA damage response, the innate immune response, endocytosis, and protein activity (reviewed partly in (Gupta et al., 2018)). Given that proteostasis is a dynamic multiple-step process involving complex molecular machinery, the deregulation at any stage could therefore be implicated in a wide variety of outcomes.

When addressing directly the list of targets for our 18 rSNPs that are likely relevant to cardiovascular function, a member of the VAV gene family – *VAV1* is the target of the rs8106212 variant for platelet distribution. Overall, the VAV proteins are known to interact with the receptors on cell surface to activate various downstream biological pathways thus leading to the alterations of transcription. *VAV1* is important in hematopoiesis, playing a role in T-cell and B-cell development and activation (Rodríguez-Fdez, Bustelo, 2019). Another candidate, the *PIEZO1* gene targeted by rs10445033 is widely associated with hereditary human diseases (Alper, 2017) and tissue homeostasis (Zhong et al., 2018). The homotetramer of the *PIEZO1* protein functions as a pore-forming subunit of a mechanically activated cation channel and contributes a lot to vascular biology and development (Li J. et al., 2014) including mechanistic signaling in endothelium (Albarrán-Juárez et al., 2018). Some studies suggest that the molecular events involved in the development of acute myocardial infarction may include MiR-103a microRNA expression in plasma and the subsequent regulation of the expression of *PIEZO1* protein (Huang L. et al., 2013). However, the up-regulation of *Piezo1* was demonstrated in rat model of heart failure (Liang et al., 2017).

Since the rSNPs do induce the variation in the gene expression, the significant rSNPs' effects on the binding affinity of a genomic locus for transcription factors can be a cause (Deplancke et al., 2016). There is relative evidence for the GWAS loci for complex diseases to be associated with not only ultimate changes in gene expression (Gallagher, Chen-Plotkin, 2018) but with the activity of various TFs (Harley et al., 2018). Here, we used the online classifier of variant pathogenicity, DeFine (Wang M. et al., 2018), to explore the functional effects of 18 rSNPs with the relevance to cardiovascular traits on the TF binding sites. DeFine classification approach employs the sequence-based deep learning models between the reference sequence and the altered sequence centered at the variant. The authors have shown that the given tool is capable to identify the causal non-coding variants within the reported GWAS loci for complex human diseases.

In this study, a significant functional impact on binding sites of thirteen TFs, including five genomic positions with potentially weakened and eight – with potentially enhanced TF binding, was found in K562 cells. Among ten identified TFs with maximum DeFine functional scores (see Suppl. Table 4), four TFs (NFYB, PBX2, SP2 and TAL1) were functionally reliable to CVDs when considering well-known roles. Thus, TAL1 is the erythroid differentiation factor that cooperates with various TFs to regulate hematopoiesis and normal differentiation of myeloid cells, and may also contribute to the process of malignant transformation (Vagapova et al., 2018). *PBX* genes encode homeodomain transcription factors, that were shown to determine the allele-specific phenotypic presentation of heart defects in mice and their loss

resulted in the insufficient expression of both genes controlling the blood vessel widening and narrowing and finally led to persistent vasoconstriction by multiple pathways (McCulley et al., 2017). NFYB (nuclear transcription factor Y subunit beta) is a subunit of a highly conserved trimeric TF that binds with high specificity to CCAAT motifs in the promoter regions in a variety of genes. Interestingly, the evidence for Nf-y, SP family factors (that bind to GC-boxes) and PBX1 to cooperate was identified (Suske, 2017; Völkel et al., 2018).

It is important that the regulatory elements of genome, the distribution of TF binding sites, and the effects of rSNPs on the gene expression can be highly tissue or cell line specific (Zhang et al., 2018). But since the tissue samples, in particular of the brain and heart, are very difficult to obtain from humans, it is not surprising that the approaches to genome-wide identifying of the functional variants were trained and available in cells and animal models first. It can be argued that genome-wide studying of the functional effects of non-coding variants on transcription very often relies on the modeling of the cell type-specific binding of transcription factors to regulatory elements of genome. The interest in the field of reliable *in silico* methods is increasing, but there are only a few that have been more or less broadly implemented according to PubMed analysis (Wang M. et al., 2018). Moreover, the evidence suggests that the performance of different functional prediction tools varies by disease phenotype (Anderson, Lassmann, 2018) and thus may give contradictory statements.

Overall, to date, using GWAS associations seems the most common way to explore the non-coding variants in the terms of functionality. Still, a survey from published studies showed that this approach helps to interpret just a minor part of thousands of identified rSNPs (Cavalli et al., 2016, 2019). Our results suggest that the reported analysis pipeline integrating the datasets from 1000 Genomes Project may serve as a general framework for future research and would eventually lead to investigation of novel functional variants within significant GWAS loci that confer human disease risk. Considering our previously obtained data for colorectal cancer (Korbolina et al., 2018) and a number of cognitive disorders (Bryzgalov et al., 2018), we have more evidence for the precise regulation of splicing mechanisms and alternative splicing to be among the key mechanisms underlying the effects of non-coding genetic variation on the phenotype including various pathological conditions.

## Conclusions

Overall using GWAS associations seems the most commonly used way to explore the non-coding variants in the terms of functionality to date. Still, a survey from published studies showed that this approach helps to interpret just a minor part of thousands of identified rSNPs (Cavalli et al., 2016, 2019). Our results suggest that the reported analysis pipeline integrating the datasets from 1000 Genomes Project may serve as a general framework for future research and would eventually lead to investigation of novel functional variants within significant GWAS loci that confer human disease risk. In consideration of our previously obtained data for colorectal cancer (Korbolina et al., 2018) and a number of cognitive disorders (Bryzgalov et al., 2018), we have got another evidence for the precise regulation of splicing mechanisms and alternative splicing to

be among the key mechanisms underlying the effects of non-coding genetic variation on the phenotype including various pathological conditions.

## References

- 1000 Genomes Project Consortium, Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A., Abecasis G.R. A global reference for human genetic variation. *Nature*. 2015;526:68-74. DOI 10.1038/nature15393.
- Albarrán-Juárez J., Iring A., Wang S., Joseph S., Grimm M., Strilic B., Wetschurack N., Althoff T.F., Offermanns S. Piezo1 and G<sub>q</sub>/G<sub>11</sub> promote endothelial inflammation depending on flow pattern and integrin activation. *J. Exp. Med.* 2018;215(10):2655-2672. DOI 10.1084/jem.20180483.
- Alper S.L. Genetic Diseases of PIEZO1 and PIEZO2 Dysfunction. *Curr. Top. Membr.* 2017;79:97-134. DOI 10.1016/bs.ctm.2017.01.001.
- Anderson D., Lassmann T. A phenotype centric benchmark of variant prioritisation tools. *NPJ Genom. Med.* 2018;3:5. DOI 10.1038/s41525-018-0044-9.
- Arking D.E., Pulit S.L., Crotti L., van der Harst P., Munroe P.B., Koopmann T.T., Sotoodehnia N., Rossin E.J., Morley M., Wang X., ... Schwartz P.J., Kääb S., Chakravarti A., Ackerman M.J., Pfeufer A., de Bakker P.I.W., Newton-Cheh C. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat. Genet.* 2014;46(8):826-836. DOI 10.1038/ng.3014.
- Astle W.J., Elding H., Jiang T., Allen D., Ruklisa D., Mann A.L., Mead D., Bouman H., Riveros-Mckay F., Kostadima M.A., ... Bourque G., Frontini M., Danesh J., Roberts D.J., Ouwehand W.H., Butterworth A.S., Soranzo N. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016; 167(5):1415-1429.e19. DOI 10.1016/j.cell.2016.10.042.
- Barac Y.D., Emrich F., Krutzwald-Josefson E., Schrepfer S., Sam-paio L.C., Willerson J.T., Robbins R.C., Ciechanover A., Mohr F.-W., Aravot D., Taylor D.A. The ubiquitin-proteasome system: a potential therapeutic target for heart failure. *J. Heart Lung Transplant.* 2017;36(7):708-714. DOI 10.1016/j.healun.2017.02.012.
- Brodie A., Azaria J.R., Ofra Y. How far from the SNP may the causative genes be? *Nucleic Acids Res.* 2016;44(13):6046-6054. DOI 10.1093/nar/gkw500.
- Bryzgalov L.O., Korbolina E.E., Brusentsov I.I., Leberfarb E.Y., Bondar N.P., Merkulova T.I. Novel functional variants at the GWAS-implicated loci might confer risk to major depressive disorder, bipolar affective disorder and schizophrenia. *BMC Neurosci.* 2018; 19(Suppl.1):22. DOI 10.1186/s12868-018-0414-3.
- Buniello A., MacArthur J.A.L., Cerezo M., Harris L.W., Hayhurst J., Malangone C., McMahon A., Morales J., Mountjoy E., Sollis E., Suveges D., Vrousitou O., Whetzel P.L., Amodio R., Guillen J.A., Riat H.S., Trevanion S.J., Hall P., Jenkins H., Flicek P., Burdett T., Hindorf L.A., Cunningham F., Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics. *Nucleic Acids Res.* 2019;47(1): 1005-1012. DOI 10.1093/nar/gky1120.
- Cai Y., Wang X.-L., Flores A.M., Lin T., Guzman R.J. Inhibition of endo-lysosomal function exacerbates vascular calcification. *Sci. Rep.* 2018;8(1):3377. DOI 10.1038/s41598-017-17540-6.
- Cavalli M., Baltzer N., Pan G., Bárcenas Walls J.R., Smolinska Garbulowska K., Kumar C., Skrtic S., Komorowski J., Wadelius C. Studies of liver tissue identify functional gene regulatory elements associated to gene expression, type 2 diabetes, and other metabolic diseases. *Hum. Genomics.* 2019;13(1):20. DOI 10.1186/s40246-019-0204-8.
- Cavalli M., Pan G., Nord H., Wallén Arzt E., Wallerman O., Wadelius C. Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals. *Genomics.* 2016; 107(6):248-254. DOI 10.1016/j.ygeno.2016.04.006.
- Curran J., Makara M.A., Mohler P.J. Endosome-based protein trafficking and Ca<sup>2+</sup> homeostasis in the heart. *Front. Physiol.* 2015;6:34. DOI 10.3389/fphys.2015.00034.
- Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2016;44(D1):D7-D19. DOI 10.1093/nar/gkv1290.
- den Hoed M., Strawbridge R.J., Almgren P., Gustafsson S., Axelson T., Engström G., de Faire U., Hedblad B., Humphries S.E., Lindgren C.M., Morris A.P., Östling G., Syvänen A.-C., Tremoli E., Hamsten A., Ingelsson E., Melander O., Lind L. GWAS-identified loci for coronary heart disease are associated with intima-media thickness and plaque presence at the carotid artery bulb. *Atherosclerosis.* 2015;239(2):304-310. DOI 10.1016/j.atherosclerosis.2015. 01.032.
- Deplancke B., Alpern D., Gardeux V. The genetics of transcription factor DNA binding variation. *Cell.* 2016;166(3):538-554. DOI 10.1016/j.cell.2016.07.012.
- Dlamin Z., Tshidino S., Hull R. Abnormalities in alternative splicing of apoptotic genes and cardiovascular diseases. *Int. J. Mol. Sci.* 2015;16(11):27171-27190. DOI 10.3390/ijms161126017.
- Dorsch L.M., Schuldt M., Knežević D., Wiersma M., Kuster D.W.D., van der Velden J., Brundel B.J.J.M. Untying the knot: protein quality control in inherited cardiomyopathies. *Pflug. Arch. Eur. J. Physiol.* 2019;471(5):795-806. DOI 10.1007/s00424-018-2194-0.
- El Marabti E., Younis I. The cancer spliceome: reprogramming of alternative splicing in cancer. *Front. Mol. Biosci.* 2018;5:80. DOI 10.3389/fmolb.2018.00080.
- Erdmann J., Kessler T., Munoz Venegas L., Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc. Res.* 2018;114(9):1241-1257. DOI 10.1093/cvr/cvy084.
- Gallagher M.D., Chen-Plotkin A.S. The post-GWAS era: from association to function. *Am. J. Hum. Genet.* 2018;102(5):717-730. DOI 10.1016/j.ajhg.2018.04.002.
- Ge T., Chen C.-Y., Neale B.M., Sabuncu M.R., Smoller J.W. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet.* 2017; 13(4):e1006711. DOI 10.1371/journal.pgen.1006711.
- Gilda J.E., Gomes A.V. Proteasome dysfunction in cardiomyopathies. *J. Physiol.* 2017;595(12):4051-4071. DOI 10.1113/JP273607.
- Gilda J.E., Lai X., Witzmann F.A., Gomes A.V. Delineation of molecular pathways involved in cardiomyopathies caused by troponin T mutations. *Mol. Cell. Proteom.* 2016;15(6):1962-1981. DOI 10.1074/mcp.M115.057380.
- Gill D., Monori G., Georgakis M.K., Tzoulaki I., Laffan M. Genetically determined platelet count and risk of cardiovascular disease. *Arterioscler. Thromb. Vasc. Biol.* 2018;38(12):2862-2869. DOI 10.1161/ATVBAHA.118.311804.
- Gong J., Qiu C., Huang D., Zhang Y., Yu S., Zeng C. Integrative functional analysis of super enhancer SNPs for coronary artery disease. *J. Hum. Genet.* 2018;63(5):627-638. DOI 10.1038/s10038-018-0422-2.
- Gupta I., Varshney N.K., Khan S. Emergence of members of TRAF and DUB of ubiquitin proteasome system in the regulation of hypertrophic cardiomyopathy. *Front. Genet.* 2018;9:336. DOI 10.3389/fgene.2018.00336.
- Harley J.B., Chen X., Pujato M., Miller D., Maddox A., Forney C., Magnusen A.F., Lynch A., Chetal K., Yukawa M., Barski A., Salomonis N., Kaufman K.M., Kottyan L.C., Weirauch M.T. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* 2018;50(5):699-707. DOI 10.1038/s41588-018-0102-3.
- Haybar H., Pezeshki S.M.S., Saki N. Evaluation of complete blood count parameters in cardiovascular diseases: an early indicator of prognosis? *Exp. Mol. Pathol.* 2019;110:104267. DOI 10.1016/j.yexmp.2019.104267.
- Helms C.C., Gladwin M.T., Kim-Shapiro D.B. Erythrocytes and vascular function: oxygen and nitric oxide. *Front. Physiol.* 2018;9:125. DOI 10.3389/fphys.2018.00125.



- Huang C.-J., Ferfoglía F., Raleff F., Krämer A. Interaction domains and nuclear targeting signals in subunits of the U2 small nuclear ribonucleoprotein particle-associated splicing factor SF3a. *J. Biol. Chem.* 2011;286(15):13106-13114. DOI 10.1074/jbc.M110.201491.
- Huang D.W., Sherman B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2008;4(1):44-57. DOI 10.1038/nprot.2008.211.
- Huang L., Li L., Chen X., Zhang H., Shi Z. MiR-103a targeting Piezo1 is involved in acute myocardial infarction through regulating endothelium function. *Cardiol. J.* 2013;23(5):556-562. DOI 10.5603/CJ.a2016.0056.
- Huang S., Chaudhary K., Garmire L.X. More is better: recent progress in multi-omics data integration methods. *Front. Genet.* 2017;8:84. DOI 10.3389/fgene.2017.00084.
- Kathiresan S., Srivastava D. Genetics of human cardiovascular disease. *Cell.* 2012;148(6):1242-1257. DOI 10.1016/j.cell.2012.03.001.
- Khera A.V., Chaffin M., Zekavat S.M., Collins R.L., Roselli C., Natarajan P., Lichtman J.H., D'Onofrio G., Mather J., Dreyer R., Speratus J.A., Taylor K.D., Psaty B.M., Rich S.S., Post W., Gupta N., Gabriel S., Lander E., Ida Chen Y.-D., Talkowski M.E., Rotter J.I., Krumholz H.M., Kathiresan S. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation.* 2019;139(13):1593-1602. DOI 10.1161/CIRCULATIONAHA.118.035658.
- Korbolina E.E., Brusentsov I.I., Bryzgalov L.O., Leberfarb E.Y., Degtyareva A.O., Merkulova T.I. Novel approach to functional SNPs discovery from genome-wide data reveals promising variants for colon cancer risk. *Hum. Mutat.* 2018;39(6):851-859. DOI 10.1002/humu.23425.
- Krämer A., Ferfoglía F., Huang C.-J., Mulhaupt F., Nesic D., Tanackovic G. Structure-function analysis of the U2 snRNP-associated splicing factor SF3a. *Biochem. Soc. Trans.* 2005;33(Pt.3):439-442. DOI 10.1042/BST0330439.
- Kuhn V., Diederich L., Keller T.C.S., Kramer C.M., Lückstädt W., Panknin C., Suvorova T., Isakson B.E., Kelm M., Cortese-Krott M.M. Red blood cell function and dysfunction: redox regulation, nitric oxide metabolism, anemia. *Antioxid. Redox Signal.* 2017;26(13):718-742. DOI 10.1089/ars.2016.6954.
- Lassale C., Curtis A., Abete I., van der Schouw Y.T., Verschuren W.M.M., Lu Y., Bueno-de-Mesquita H.B. Elements of the complete blood count associated with cardiovascular disease incidence: findings from the EPIC-NL cohort study. *Sci. Rep.* 2018;8(1):3290. DOI 10.1038/s41598-018-21661-x.
- Leal J.K.F., Adjubo-Hermans M.J.W., Bosman G.J.C.G.M. Red blood cell homeostasis: mechanisms and effects of microvesicle generation in health and disease. *Front. Physiol.* 2018;9:703. DOI 10.3389/fphys.2018.00703.
- Li A.H., Hanchard N.A., Furthner D., Fernbach S., Azamian M., Nicosia A., Rosenfeld J., Muzny D., D'Alessandro L.C.A., Morris S., Jhangiani S., Parekh D.R., Franklin W.J., Lewin M., Towbin J.A., Penny D.J., Fraser C.D., Martin J.F., Eng C., Lupski J.R., Gibbs R.A., Boerwinkle E., Belmont J.W. Whole exome sequencing in 342 congenital cardiac left sided lesion cases reveals extensive genetic heterogeneity and complex inheritance patterns. *Genome Med.* 2017;9(1):95. DOI 10.1186/s13073-017-0482-5.
- Li J., Hou B., Tumova S., Muraki K., Bruns A., Ludlow M.J., Sedo A., Hyman A.J., McKeown L., Young R.S., Yuldasheva N.Y., Majeed Y., Wilson L.A., Rode B., Bailey M.A., Kim H.R., Fu Z., Carter D.A.L., Bilton J., Imrie H., Ajuh P., Dear T.N., Cubbon R.M., Kearney M.T., Prasad K.R., Evans P.C., Ainscough J.F.X., Beech D.J. Piezo1 integration of vascular architecture with physiological force. *Nature.* 2014;515(7526):279-282. DOI 10.1038/nature13701.
- Liang J., Huang B., Yuan G., Chen Y., Liang F., Zeng H., Zheng S., Cao L., Geng D., Zhou S. Stretch-activated channel Piezo1 is up-regulated in failure heart and cardiomyocyte stimulated by AngII. *Am. J. Transl. Res.* 2017;9(6):2945-2955.
- Lin J.-C. Therapeutic applications of targeted alternative splicing to cancer treatment. *Int. J. Mol. Sci.* 2017;19(1):75. DOI 10.3390/ijms19010075.
- Luo C., Cheng Y., Liu Y., Chen L., Liu L., Wei N., Xie Z., Wu W., Feng Y. SRSF2 regulates alternative splicing to drive hepatocellular carcinoma development. *Cancer Res.* 2017;77(5):1168-1178. DOI 10.1158/0008-5472.CAN-16-1919.
- Maatz H., Jens M., Liss M., Schafer S., Heinig M., Kirchner M., Adami E., Rintisch C., Dauksaite V., Radke M.H., Selbach M., Barton P.J.R., Cook S.A., Rajewsky N., Gotthardt M., Landthaler M., Hubner N. RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J. Clin. Investig.* 2014;124(8):3419-3430. DOI 10.1172/JCI74523.
- Mansur Y.A., Rojano E., Ranea J.A.G., Perkins J.R. Analyzing the effects of genetic variation in noncoding genomic regions. In: Deigner H.-P., Kohl M. (Eds.). Precision Medicine. Tools and Quantitative Approaches. Acad. Press, 2018;119-144. DOI 10.1016/B978-0-12-805364-5.00007-X.
- Martinez-Montiel N., Rosas-Murrieta N., Anaya Ruiz M., Monjaraz-Guzman E., Martinez-Contreras R. Alternative splicing as a target for cancer treatment. *Int. J. Mol. Sci.* 2018;19:545. DOI 10.3390/ijms19020545.
- McCulley D.J., Wienhold M.D., Hines E.A., Hacker T.A., Rogers A., Pewowaruk R.J., Zewdu R., Chesler N.C., Selleri L., Sun X. PBX transcription factors drive pulmonary vascular adaptation to birth. *J. Clin. Investig.* 2017;128(2):655-667. DOI 10.1172/JCI93395.
- Meng F., Yuan G., Zhu X., Zhou Y., Wang D., Guo Y. Functional variants identified efficiently through an integrated transcriptome and epigenome analysis. *Sci. Rep.* 2018;8(1):2959. DOI 10.1038/s41598-018-21024-6.
- Mozos I. Mechanisms linking red blood cell disorders and cardiovascular diseases. *BioMed Res. Int.* 2015;2015:682054. DOI 10.1155/2015/682054.
- Pernow J., Mahdi A., Yang J., Zhou Z. Red blood cell dysfunction: a new player in cardiovascular disease. *Cardiovasc. Res.* 2019;115(11):1596-1605. DOI 10.1093/cvr/cvz156.
- Preuss M., König I.R., Thompson J.R., Erdmann J., Absher D., Assimes T.L., Blankenberg S., Boerwinkle E., Chen L., Cupples L.A., Hall A.S., Halperin E., Hengstenberg C., Holm H., Laaksonen R., Li M., März W., McPherson R., Musunuru K., Nelson C.P., Burnett M.S., Epstein S.E., O'Donnell C.J., Quertermous T., Rader D.J., Roberts R., Schillert A., Stefansson K., Stewart A.F.R., Thorleifsson G., Voight B.F., Wells G.A., Ziegler A., Kathiresan S., Reilly M.P., Samani N.J., Schunkert H., and on behalf of the CARDIOGRAM Consortium. Design of the Coronary ARtery Disease Genome-wide Replication And Meta-analysis (CARDIoGRAM) study: a genome-wide association meta-analysis involving more than 22000 cases and 60000 controls. *Circ. Cardiovasc. Genet.* 2010;3(5):475-483. DOI 10.1161/CIRCGENETICS.109.899443.
- Rau C.D., Lusi A.J., Wang Y. Genetics of common forms of heart failure. *Curr. Opin. Cardiol.* 2015;30(3):222-227. DOI 10.1097/HCO.0000000000000160.
- Refsgaard L., Holst A.G., Sadjadic G., Haunsø S., Nielsen J.B., Olesen M.S. High prevalence of genetic variants previously associated with LQT syndrome in new exome data. *Eur. J. Hum. Genet.* 2012;20(8):905-908. DOI 10.1038/ejhg.2012.23.
- Reinthal M., Braune S., Lendlein A., Landmesser U., Jung F. Platelets and coronary artery disease: interactions with the blood vessel wall and cardiovascular devices. *Biointerphases.* 2016;11(2):29702. DOI 10.1161/1.4953246.
- Rexiati M., Sun M., Guo W. Muscle-specific mis-splicing and heart disease exemplified by RBM20. *Genes.* 2018;9(1):18. DOI 10.3390/genes9010018.
- Rifkind J.M., Mohanty J.G., Nagababu E., Salgado M.T., Cao Z. Potential modulation of vascular function by nitric oxide and reactive oxygen species released from erythrocytes. *Front. Physiol.* 2018;9:690. DOI 10.3389/fphys.2018.00690.



- Rodríguez-Fdez S., Bustelo X.R. The Vav GEF family: an evolutionary and functional perspective. *Cells*. 2019;8(5):465. DOI 10.3390/cells8050465.
- Roman T.S., Mohlke K.L. Functional genomics and assays of regulatory activity detect mechanisms at loci for lipid traits and coronary artery disease. *Curr. Opin. Genet. Dev.* 2018;50:52-59. DOI 10.1016/j.gde.2018.02.004.
- Roth G.A., Johnson C., Abajobir A., Abd-Allah F., Abera S.F., Abyu G., Ahmed M., Aksut B., Alam T., Alam K., ... Yip P., Yonemoto N., Younis M., Yu C., Vos T., Naghavi M., Murray C. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J. Am. Coll. Cardiol.* 2017;70:1-25. DOI 10.1016/j.jacc.2017.04.052.
- Samman Tahhan A., Hammad M., Sandesara P.B., Hayek S.S., Kalogeropoulos A.P., Alkholder A., Mohamed Kelli H., Topel M., Ghasemzadeh N., Chivukula K., Ko Y.-A., Aida H., Hesaroie H., Mahar E., Kim J.H., Wilson P., Shaw L., Vaccarino V., Waller E.K., Quyyumi A.A. Progenitor cells and clinical outcomes in patients with heart failure. *Circ. Heart. Fail.* 2017;10(8):e004106. DOI 10.1161/CIRCHEARTFAILURE.117.004106.
- Schunkert H., von Scheidt M., Kessler T., Stiller B., Zeng L., Vilne B. Genetics of coronary artery disease in the light of genome-wide association studies. *Clin. Res. Cardiol.* 2018;107(Suppl.2):2-9. DOI 10.1007/s00392-018-1324-1.
- Seidemann S.B., Smith E., Subrahmanyam L., Dykas D., Abou Ziki M.D., Azari B., Hannah-Shmouni F., Jiang Y., Akar J.G., Marieb M., Jacoby D., Bale A.E., Lifton R.P., Mani A. Application of whole exome sequencing in the clinical diagnosis and management of inherited cardiovascular diseases in adults. *Circ. Cardiovasc. Genet.* 2017;10(1):e001573. DOI 10.1161/CIRCGENETICS.116.001573.
- Shukla S.K., Rafiq K. Proteasome biology and therapeutics in cardiac diseases. *Transl. Res.* 2019;205:64-76. DOI 10.1016/j.trsl.2018.09.003.
- Siebring-van Olst E., Blijlevens M., de Menezes R.X., van der Meulen-Muileman I.H., Smit E.F., van Beusechem V.W. A genome-wide siRNA screen for regulators of tumor suppressor p53 activity in human non-small cell lung cancer cells identifies components of the RNA splicing machinery as targets for anticancer treatment. *Mol. Oncol.* 2017;11(5):534-551. DOI 10.1002/1878-0261.12052.
- Sloan A., Gona P., Johnson A.D. Cardiovascular correlates of platelet count and volume in the Framingham Heart Study. *Ann. Epidemiol.* 2015;25(7):492-498. DOI 10.1016/j.annepidem.2015.01.010.
- Smith A.J.P., Deloukas P., Munroe P.B. Emerging applications of genome-editing technology to examine functionality of GWAS-associated variants for complex traits. *Physiol. Genomics.* 2018;50(7):510-522. DOI 10.1152/physiolgenomics.00028.2018.
- Smith J.G., Newton-Cheh C. Genome-wide association studies of late-onset cardiovascular disease. *J. Mol. Cell. Cardiol.* 2015;83:131-141. DOI 10.1016/j.yjmcc.2015.04.004.
- Suske G. NF-Y and SP transcription factors – new insights in a long-standing liaison. *Biochim. Biophys. Acta Gene Regul. Mech.* 2017;1860(5):590-597. DOI 10.1016/j.bbagr.2016.08.011.
- Szklarczyk D., Gable A.L., Lyon D., Junge A., Wyder S., Huerta-Cepas J., Simonovic M., Doncheva N.T., Morris J.H., Bork P., Jensen L.J., von Mering C. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-D613. DOI 10.1093/nar/gky1131.
- Uysal H.B., Dağlı B., Akgüllü C., Avcil M., Zencir C., Ayhan M., Sönmez H.M. Blood count parameters can predict the severity of coronary artery disease. *Korean J. Intern. Med.* 2016;31(6):1093-1100. DOI 10.3904/kjim.2015.199.
- Vagapova E.R., Spirin P.V., Lebedev T.D., Prassolov V.S. The role of TAL1 in hematopoiesis and leukemogenesis. *Acta Naturae.* 2018;10(1):15-23.
- van den Hoogenhof M.M.G., Pinto Y.M., Creemers E.E. RNA splicing. *Circ. Res.* 2016;118:454. DOI 10.1161/CIRCRESAHA.115.307872.
- Vélez P., García Á. Platelet proteomics in cardiovascular diseases. *Transl. Proteomics.* 2015;7:15-29. DOI 10.1016/j.trprot.2014.09.002.
- Völkel S., Stielow B., Finkernagel F., Berger D., Stiewe T., Nist A., Suske G. Transcription factor Sp2 potentiates binding of the TALE homeoproteins Pbx1:Prep1 and the histone-fold domain protein NF-y to composite genomic sites. *J. Biol. Chem.* 2018;293(50):19250-19262. DOI 10.1074/jbc.RA118.005341.
- Wallace E., Howard L., Liu M., O'Brien T., Ward D., Shen S., Prendiville T. Long QT syndrome: genetics and future perspective. *Pediatr. Cardiol.* 2019;40(7):1419-1430. DOI 10.1007/s00246-019-02151-x.
- Wan G., Ji L., Xia W., Cheng L., Zhang Y. Screening genes associated with elevated neutrophil-to-lymphocyte ratio in chronic heart failure. *Mol. Med. Rep.* 2018;18(2):1415-1422. DOI 10.3892/mmr.2018.9132.
- Wang M., Tai C., E W., Wei L. DeFine: deep convolutional neural networks accurately quantify intensities of transcription factor-DNA binding and facilitate evaluation of functional non-coding variants. *Nucleic Acids Res.* 2018;46(11):e69-e69. DOI 10.1093/nar/gky215.
- Wang Y., Wang J.-G. Genome-wide association studies of hypertension and several other cardiovascular diseases. *Pulse.* 2018;6(3-4):169-186. DOI 10.1159/000496150.
- Ward L.D., Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* 2012;30(11):1095-1106. DOI 10.1038/nbt.2422.
- Wijers M., Rimbart A., Dalila N., Fedoseienko A., Wolters K., Dekker D., Smit M., Levels H., Huijckman N., Kloosterhuis N., Hofker M., Billadeau D., van Deursen J., Horton J., Burstein E., Tybjaerg-Hansen A., Kuivenhoven J.A., van de Sluis B. A regulatory role of the endosomal sorting machinery in controlling plasma LDL cholesterol levels and atherosclerosis in mice and humans. *Atheroscler. Suppl.* 2018;32:18. DOI 10.1016/j.atherosclerosis.2018.04.053.
- Yang J., Gonon A.T., Sjoquist P.-O., Lundberg J.O., Pernow J. Arginase regulates red blood cell nitric oxide synthase and export of cardioprotective nitric oxide bioactivity. *Proc. Natl. Acad. Sci. USA.* 2013;110(37):15049-15054. DOI 10.1073/pnas.1307058110.
- Yang J., Villar V.A.M., Rozyayev S., Jose P.A., Zeng C. The emerging role of sorting nexins in cardiovascular diseases. *Clin. Sci.* 2019;133(5):723-737. DOI 10.1042/CS20190034.
- Zhang L., Xue G., Liu J., Li Q., Wang Y. Revealing transcription factor and histone modification co-localization and dynamics across cell lines by integrating ChIP-seq and RNA-seq data. *BMC Genom.* 2018;19(Suppl.10):914. DOI 10.1186/s12864-018-5278-5.
- Zhong M., Komarova Y., Rehman J., Malik A.B. Mechanosensing Piezo channels in tissue homeostasis including their role in lungs. *Pulm. Circ.* 2018;8(2):1-6. DOI 10.1177/2045894018767393.
- Zhou Z., Mahdi A., Tratsiakovich Y., Zahrán S., Kövamees O., Nordin F., Uribe Gonzalez A.E., Alvarsson M., Östenson C.-G., Andersson D.C., Hedin U., Hermesz E., Lundberg J.O., Yang J., Pernow J. Erythrocytes from patients with type 2 diabetes induce endothelial dysfunction via arginase I. *J. Am. Coll. Cardiol.* 2018;72(7):769-780. DOI 10.1016/j.jacc.2018.05.052.

**Acknowledgements.** We greatly thank the people who contributed to a valuable, worldwide reference for human genetic variation, 1000 Genomes Project, by contributing their samples and their efforts. The study was funded by the grant 18-29-09041 from the Russian Foundation for Basic Research and the State Budget Project FWNR-2022-0016.

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

Received September 7, 2021. Revised November 22, 2021. Accepted November 23, 2021.

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

## Species delimitation and microalgal cryptic diversity analysis of the genus *Micractinium* (*Chlorophyta*)

E.S. Krivina<sup>1</sup>✉, A.D. Temraleeva<sup>1</sup>, Yu.S. Bukin<sup>2</sup>

<sup>1</sup> Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Moscow region, Russia

<sup>2</sup> Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

✉ pepelisa@yandex.ru

**Abstract.** In this article, the system of the green microalgal genus *Micractinium*, based on morphological, physiological, ecological and molecular data, is considered. The main diagnostic species characteristics and the taxonomic placement of some taxa are also discussed. Phylogenetic analysis showed that the genus *Micractinium* is characterized by high cryptic diversity. The algorithms used for species delimitation had different results on the number of potentially species-level clusters allocated. The ABGD method was less "sensitive". The tree-based approaches GMYC and PTP showed a more feasible taxonomy of the genus *Micractinium*, being an effective additional tool for distinguishing species. The clustering obtained by the latter two methods is in good congruence with morphological (cell size and shape, ability to form colonies, production of bristles, chloroplast type), physiological (vitamin requirements, reaction to high and low temperatures), molecular (presence of introns, level of genetic differences, presence of CBCs or special features of the secondary structure in ITS1 and ITS2) and ecological characteristics (habitat). The polyphyly of the holotype of the genus *M. pusillum* as well as *M. belenophorum* is shown. The intron was effective as an additional tool for distinguishing species, and the results of the intron analysis should be taken into account together with other characteristics. The CBC approach, based on the search for compensatory base changes in conservative ITS2 regions, was successful only for distinguishing cryptic species from "true" members of *M. pusillum*. Therefore, to distinguish species, it is more effective to take into account all the CBC in ITS1 and ITS2 and analyze characteristic structural differences (molecular signatures) in the secondary structure of internal transcribed spacers. The genetic distances analysis of 18S–ITS1–5.8S–ITS2 nucleotide sequences showed that intraspecific differences in the genus ranged from 0 to 0.5 % and interspecific differences, from 0.6 to 4.7 %. Due to the polyphasic approach, it was possible to characterize 29 clusters and phylogenetic lines at the species level within the genus *Micractinium* and to make assumptions about the species.

Key words: green microalgae; ABGD; GMYC; PTP; species delimitation; morphology; ecology; phylogeny; 18S–ITS1–5.8S–ITS2 fragment.

**For citation:** Krivina E.S., Temraleeva A.D., Bukin Yu.S. Species delimitation and microalgal cryptic diversity analysis of the genus *Micractinium* (*Chlorophyta*). *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):74-85. DOI 10.18699/VJGB-22-11

## Разграничение видов и анализ криптического разнообразия микроводорослей рода *Micractinium* (*Chlorophyta*)

Е.С. Кривина<sup>1</sup>✉, А.Д. Темралеева<sup>1</sup>, Ю.С. Букин<sup>2</sup>

<sup>1</sup> Федеральный исследовательский центр «Пушинский научный центр биологических исследований Российской академии наук», Пушино, Московская область, Россия

<sup>2</sup> Лимнологический институт Сибирского отделения Российской академии наук, Иркутск, Россия

✉ pepelisa@yandex.ru

**Аннотация.** В статье рассматривается система зеленых микроводорослей рода *Micractinium*, построенная на основании морфологических, физиологических, экологических и молекулярно-генетических данных. Обсуждаются главные диагностические признаки видов, а также систематическое положение некоторых таксонов. Филогенетический анализ показал, что род *Micractinium* характеризуется достаточно высоким криптическим разнообразием. Используемые алгоритмы разграничения видов имели различные результаты по количеству выделенных кластеров потенциально видового уровня. Метод ABGD, основанный на дистанциях, является менее «чувствительным». Алгоритмы GMYC и PTP, анализирующие топологию филогенетического дерева, более реалистично отражают систематику рода *Micractinium* и служат эффективными вспомогательными инструментами для разграничения видов. Кластеризация, полученная двумя последними методами, хорошо согласуется с морфологическими (размеры и форма клеток, способность формировать колонии, продуцирование щетинок, тип хлоропласта), физиологическими (потребность в витаминах, реакция на воздействие высоких и низ-

ких температур), молекулярно-генетическими (наличие интронов и их длина, уровень генетических различий, наличие компенсаторных замен (CBC) или характерных особенностей вторичной структуры в ITS1 и ITS2) и экологическими признаками (среда обитания). Показана полифилетичность типового вида рода *M. pusillum*, а также *M. belenophorum*. Интрон был эффективен как вспомогательный инструмент для разграничения видов, однако результаты анализа интронов необходимо учитывать в совокупности с другими признаками. Применение CBC-подхода, базирующегося на поиске компенсаторных замен в консервативных регионах ITS2, было успешным только для отграничения криптических видов от «истинных» представителей *M. pusillum*. При разграничении видов эффективнее учитывать все CBC в ITS1 и ITS2 и анализировать характерные структурные различия (молекулярные подписи) во вторичной структуре внутренних транскрибируемых спейсеров. Анализ генетических дистанций нуклеотидных последовательностей 18S–ITS1–5.8S–ITS2 показал, что внутривидовые различия у представителей рода колебались в пределах 0–0.5 %, межвидовые – 0.6–4.7 %. Благодаря полифазному подходу удалось охарактеризовать 29 кластеров и филогенетических линий видового уровня в рамках рода *Micractinium* и выдвинуть предположения о видах внутри выделенных групп.

Ключевые слова: зеленые микроводоросли; ABGD; GMYC; PTP; морфология; экология; филогения; фрагмент 18S–ITS1–5.8S–ITS2.

## Introduction

The genus *Micractinium* was described by G. Fresenius in 1858 and was referred to the family Micractiniaceae. For a long time, it was thought that this genus includes only microalgae which unlike the genus *Chlorella* and other ‘small green balls’ form colonies and produce bristles consisting of protein, devoid of cellulose fibers and developing after the formation of a cell wall (Schnepf et al., 1980). The species differences were based on minor changes in the formation of colonies, as well as the length and number of bristles.

Based on the results of the phylogenetic analysis of the 18S rRNA gene, Wolf et al. (2003) concluded that strains of the genus *Micractinium* are members of the *Trebouxio-phyceae* class and are closely related to the genus *Chlorella* Beijerinck. Later, Luo et al. (2005, 2006) found that the formation of colonies and the production of bristles is often a reaction to the so-called algophages ‘grazing’ load from (primarily rotifers and ciliates), and in their studies, the authors also suggested that the type species of the genus, *M. pusillum*, is polyphyletic. Using molecular genetic analysis, Pröschold et al. (2010) proved that the genus *Diacanthos* with its type species *D. belenophorus* is a member of the genus *Micractinium*. Summarizing the results of molecular genetic, morphological, and ontogenetic analyses by Wolf et al. (2003), Krienitz et al. (2004), Fawley et al. (2005), Luo et al. (2010), Pröschold et al. (2010) proposed a new concept of the *Chlorella*-clade, according to which the genus *Micractinium* was transferred to the family Chlorellaceae.

Currently, there are 20 species of microalgae in this genus. However, the fragment 18S–ITS1–5.8S–ITS2 was sequenced only for 9 species, among which there are both microalgae with a classical *Micractinium*-like morphotype, i. e. forming colonies and producing bristles, and organisms with a typical *Chlorella*-like morphology (for example, *M. singularis*, *M. variabile*, *M. simplicissimum*, *M. inermum*, *M. tetrahymenae*, which have single cells and lack bristles under standard conditions) (Hoshina, Fujiwara, 2013; Chae et al., 2019; Pröschold et al., 2020).

Members of the genus *Micractinium* are widely distributed in various biotopes, including freshwater and brackish water reservoirs, hot springs, and cold waters of Antarctica,

at temperatures from zero to above 70 °C (Hoshina, Fujiwara, 2013; Onay et al., 2014; Adar et al., 2016; Chae et al., 2019). They play an important role in the life of ecosystems, actively participating in the processes of photosynthesis of organic substances and photosynthetic aeration, as well as natural self-purification of the reservoir through the accumulation, transformation, and mineralization of pollutants (Vaishlya, Kulyatov, 2011; Mehrabadi et al., 2017). These microalgae are also actively used for the production of animal feed, food additives, and wastewater treatment (Lipstein, Hurwitz, 1983; Onay et al., 2014; Mehrabadi et al., 2017). In addition, some species of the genus *Micractinium* are recognized as suitable raw materials for biofuels due to a high growth rate combined with a high lipid content (Onay et al., 2014; Adar et al., 2016). Currently, thermophilic and cryo-tolerant representatives of the genus *Micractinium*, which are able to accumulate lipids or other valuable substances, are of considerable interest for biotechnology (Onay et al., 2014; Adar et al., 2016; Chae et al., 2021). Accurate species identification, in this case, becomes a priority task, since the ecological plasticity of species to abiotic environmental factors can vary significantly not only within the entire genus but also between closely related species (Onay et al., 2014; Chae et al., 2021).

The goal of this research was a comprehensive study of representatives of the genus *Micractinium*, including new strains of the Algal Collection of Soil Science Institute (ACSSI), for reliable differentiation of closely related taxa at the species level. For the first time, morphological, physiological, and ecological characteristics were generalized for all the described members of *Micractinium*, the results of phylogenetic analysis of the 18S–ITS1–5.8S–ITS2 fragment were considered, including the presence of introns and their characteristics, the values of genetic distances, differences in the secondary structures of spacers ITS1 and ITS2, among them the presence of compensatory substitutions (CBC) and structural differences, and species boundaries were determined using GMYC, PTP and ABGD methods. Based on the polyphasic approach, assumptions were made about the criteria for species distinguishing within the genus.



## Materials and methods

**Objects of research.** The objects of this study were the genetic sequences of strains belonging to the genus *Micractinium* described and deposited in GenBank, as well as six new strains of microalgae from the ACSSI collection. Strains ACSSI 198, ACSSI 287 and ACSSI 345 were isolated from water from the surface horizon of the pelagic zone of the lake Prudovikov (53°31'44.4" N, 49°30'58.0" E, Tolyatti, Samara region, Russia), ACSSI 343 and ACSSI 344 – from water from the surface horizon of the pelagic zone and macrophyte thickets of the lake Bolshoe Vasilyevskoe, respectively (53°32'45.2" N, 49°32'02.0" E, Tolyatti, Samara region, Russia). The strain ACSSI 332 (= IPPAS C-16) is a subculture of the IPPAS Collection of microalgae and cyanobacteria, and was isolated from hot springs on the Chukchi Peninsula.

**Isolation and cultivation of new strains.** A drop of lake water without prefiltration was applied to a solid medium BG-11 with nitrogen (1 % agar, pH = 7.2) and then individual colonies were repeatedly replanted. The obtained isolates were cultured in a climatostat under standard conditions (temperature +23...+25 °C, light 60–75  $\mu\text{M}$  of photons/( $\text{m}^2 \cdot \text{s}$ ), photoperiod of 12 hours).

**Microscopy.** The morphology and life cycle of these strains were studied by light microscopy (light field and interference contrast) using Leica DM750 and Carl Zeiss Axio Scope A1 microscopes (Germany) at the Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences". The results of the observations were documented by working drawings and photographs taken with the help of color digital cameras Videosavr (Russia) and Carl Zeiss MRc 5 (Germany). The follow-up period ranged from 2 weeks to 12 months. To determine the limits of variation of morphological features, the characteristics of 200 vegetative cells of each strain were analyzed.

**Isolation, amplification, purification, and sequencing of DNA.** The total DNA from the strains was isolated using a DNeasy Plant Mini Kit (Qiagen, USA), following the manufacturer's protocol. For amplification, Screen Mix-HS mixture was used (Eurogen, Russia). Primers for PCR of the 18S and 5.8S rRNA genes and ITS1, ITS2 spacers, and amplification conditions are given in the work of Krivina and Temraleeva (2020). The detection of the target PCR products was carried out electrophoretically in a 1 % agarose gel. For further purification of amplicons from the gel, a Cleanup Standard kit (Eurogen, Russia) was used. The sequencing of the nucleotide sequences was carried out based on CJSC "Syntol" (Russia).

**Molecular phylogenetic analysis.** To analyze the phylogeny and clarify the taxonomic position of the studied strains, the homology of the nucleotide sequences 18S–ITS1–5.8S–ITS2 was searched using the BLASTn algorithm in GenBank (<https://blast.ncbi.nlm.nih.gov>). The selection of sequences was carried out based on the criteria of maximum identity (similarity  $\geq 95$  %), reading quality, reading length (at least 2300 bp) and belonging to type species and authentic strains.

The sample for phylogenetic analysis included 59 strains. The names of taxa are given according to the International Electronic Database AlgaeBase (Guiry M.D., Guiry G.M., 2021). In the BioEdit program, multiple alignment was performed using the ClustalW algorithm. The phylogenetic tree reconstructed by the maximum likelihood (ML) method in the IQ-TREE program (with an assessment of the reliability of the topology by ultra-fast bootstrap analysis and testing of the evolutionary model using the AIC criterion) was used to distinguish species using the Poisson tree processes (PTP) algorithm on an online server <https://species.h-its.org/>.

To distinguish species in the data array, the method of automatic search for interspecific gap in genetic distances (automatic barcode gap discovery, ABGD) (Puillandre et al., 2012) was used on an online server <https://bioinfo.mnhn.fr/abi/public/abgd/>. To analyze ABGD, a matrix of genetic distances calculated using the maximum likelihood method in the IQ-TREE program was used. When using the ABGD method, the results were analyzed both in the initial partition mode and in the recursive partition mode. The third method was a generalized mixed Yule model taking into account the integrity of species (general mixed Yule coalescent model, GMYC) (Fujisawa, Barraclough, 2013), implemented in the 'splits' package for the R programming language v. 3.4.4 (<https://www.R-project.org/>). For GMYC analysis, an ultrametric tree reconstructed in the BEAST v. 1.10.4 program was used.

The reconstruction of the tree in BEAST was carried out using four speciation models: the Yule speciation model (Aldous, 2001) with a strict molecular clock; the Yule speciation model with a relaxed molecular clock with evolution rates distributed according to a lognormal distribution; the birth–death speciation model (Lambert, Stadler, 2013) with a strict molecular clock; a model of speciation of the birth–death of species with a relaxed molecular clock with the rates of evolution distributed according to the lognormal distribution. The selection of the best speciation model was carried out by comparing the marginal likelihood values calculated by the method of sequential sampling (Lartillot, Philippe, 2006) in the BEAST v. 1.10.4 program. During the reconstruction of the tree, the BEAST program set 50,000,000 generations for Markov chains and 250,000 generations of Markov chains and 200 steps for calculating the marginal likelihood. With these parameters of the number of generations, all the values of the ESS statistics (the convergence indicator of the BEAST analysis) were more than 200. An ultrametric Bayesian phylogenetic tree was used to visualize the analysis results. A sample of ultra-fast bootstrap analysis trees obtained in the IQ-TREE program was combined with the topology of a Bayesian ultrametric tree to calculate bootstrap supports by the maximum likelihood method (ultra-fast bootstrap analysis). Thus, the support of the ultrametric Bayesian tree topology was evaluated by the Bayesian inference (BI) using a posterior probabilities and bootstrap analysis. To calculate bootstrap supports, we used an algorithm previously developed by us (Temraleeva et al., 2018), implemented using the functions



of the APE package (Paradis et al., 2004) for the R statistical software environment v. 3.4.4. A representative of the sister genus *Chlorella* (*Trebouxiophyceae*, *Chlorophyta*), *C. vulgaris*, was chosen as an outgroup during phylogenetic reconstructions. The distribution of genetic distances was visualized as a histogram in the R statistical software environment v. 3.4.4.

Genetic differences between nucleotide sequences were characterized using genetic distances (K2P distances), which were calculated in the MEGA 6.0 program. The boxplot of genetic distances was built in the R statistical software environment v. 3.4.4 (<https://www.R-project.org/>). To compare the topology of trees, we used data from articles (Krienitz et al., 2004; Luo et al., 2006; Hoshina et al., 2010, 2017; Pröschold et al., 2010, 2011, 2020; Bock et al., 2011; Hoshina, Nakada, 2018).

Folding of ITS1 and ITS2 was performed using the RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) in accordance with the principle of minimum energy. When assessing the correctness of the prediction of the secondary structure, ITS1 and ITS2 were guided by A. Coleman (2015) and Caisová et al. (2013), respectively. The comparison of the secondary structure of spacers between strains, the search for conservative motives and compensatory substitutions (CBCs) was carried out in the 4SALE program (Seibel et al., 2008). In the analysis of ITS2 for the species distinguishing, special attention is paid to the approach of *sensu* A. Coleman (2000, 2015), according to which the presence of even one CBC in conservative regions of ITS2 (5 bp of I helix, 10 bp of II helix, all III helix) in two microalgae correlates with their sexual incompatibility. The secondary structures of spacers are visualized in the PseudoViewer3 program.

**Statistical analysis of various characteristics of representatives of the genus *Micractinium*.** For comparative analysis, the characteristics of the strains were encoded in the form of binary vectors. The length of the binary vector of the analyzed feature was equal to the number of its possible states, while each element corresponded to a certain state. For the analyzed strains, 1 was recorded in the position corresponding to the state of the characteristic, the remaining elements had the value 0. All binary vectors determining the states of each of the traits for all the strains studied in the analysis were summarized in a single table. The analysis used strains for which the states of 80 % or more of the considered traits were known, the remaining strains were excluded from the analysis.

On the basis of a binary table of feature states, the similarity and difference of strains were visualized using multidimensional scaling, for which a matrix of Jacquard distances was used (one minus the share of common non-zero states in the total number of non-zero states in the two strains being compared), during the calculation of which for each pair of strains, features that were indeterminate for one of the strains were excluded. In order to determine the significance of a trait in the overall distribution of distances between strains, the Mantel test (Mantel, 1967) was used based on the Pearson correlation coefficient, the reliability of the correlation was determined by a permutation test (10,000 permutations). During the Mantel test, the general matrix of Jacquard distances was compared with the distance matrices calculated for each feature separately. The higher the value of the Pearson correlation coefficient for the trait under consideration, the greater the contribution it makes to the separation of strains. All calculations were performed using the functions of the ‘vegan’ package (Dixon, 2003) for the R statistical software environment.

## Results

**Morphology of ACSSI strains.** All studied strains had a *Chlorella*-like morphotype: the cells were single, spherical in shape, without bristles. The vegetative cell sizes of ACSSI 343, ACSSI 344, and ACSSI 345 were more than ACSSI 198, ACSSI 287, and ACSSI 332 (Table 1).

The chloroplast is parietal, mainly cup-shaped. However, in the strains ACSSI 343, ACSSI 344 and ACSSI 345, a saucer-shaped chloroplast (~20 %) and a hollow spherical with a hole (~20 %) are found in adult cells. The pyrenoid is single, spherical or broadly oval with a starch sheath. All strains reproduce by autospores. The number of autospores in the strains ACSSI 343, ACSSI 344, and ACSSI 345 varies from 2 to 8, in the strains ACSSI 198, ACSSI 287, as a rule, 2–4 autospores were noted (8 autospores are rare), while in ACSSI 332, there were no more than 4. Based on the morphological characteristics, the studied strains were initially assigned to the genus *Chlorella*.

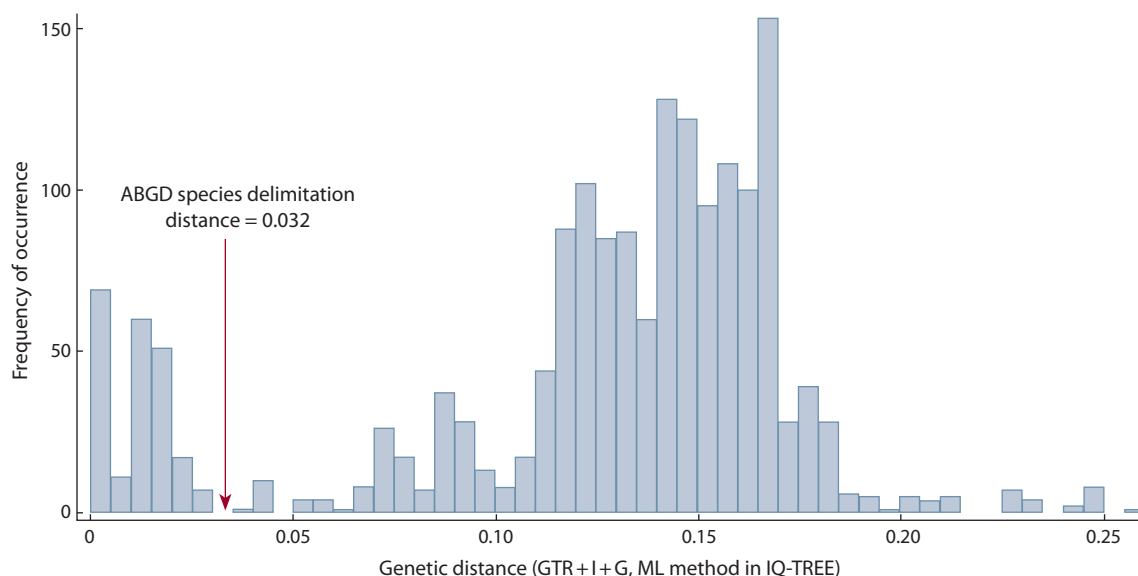
**Phylogenetic analysis.** The best model of DNA evolution for the studied dataset of nucleotide sequences (18S–ITS1–5.8S–ITS2) is GTR+I+G (AIC = 38198.0101), which was used for all further calculations. The results of the “path sampling” analysis showed that the best model for the reconstruction of the phylogenetic tree by the Bayesian method in the BEAST program is a model of speciation of the birth–death of species with a relaxed molecular clock

**Table 1.** The vegetative cell sizes of the studied strains

Strains	Average size, $\mu\text{m} \pm \text{sd}^*$	Minimum size, $\mu\text{m}$	Maximum size, $\mu\text{m}$
ACSSI 198, ACSSI 287	4.3 $\pm$ 1.08	2.2	6.2
ACSSI 343, ACSSI 344, ACSSI 345	5.7 $\pm$ 1.61	3.2	10.5
ACSSI 332	4.4 $\pm$ 1.04	2.5	6.1

\* sd – standard deviation. 200 measurements for each strain.





**Fig. 2.** Histogram of the distribution of genetic distances between representatives of the genus *Micractinium*. The species delimitation boundary determined by the ABGD method is shown.

ACSSI 332 with maximum statistical support was clustered with *Chlorella*-like strains TvB, SH, CCAP 211/92, ehime, IC-80. The sister phylogenetic lines to this cluster are the incorrectly identified *Pseudochlorella pringsheimii* and *M. tetrahymenae* (PP = 0.99–1.00, BP = 98–100 %). The strain ACSSI 332 does not have an intron of the 18S rRNA gene, unlike TvB, SH, CCAP 211/92. The genetic distances between it and the strains TvB, SH, CCAP 211/92, ehime, IC-80 varied in the range of 0.1–0.5 %, with *P. pringsheimii* and *M. tetrahymenae* – 1.1–1.2 %. The sister strains to ACSSI 198, 287 are *M. conductrix* and the KNUA032 strain with a *Chlorella*-like morphotype (statistical support is maximum). The level of genetic differences ranged from 0.7 to 1.3 %.

**The secondary structure of ITS1 and ITS2.** The length of ITS1 of the studied strains was 238–267 nt, ITS2 – 242–243 nt. The ITS1 and ITS2 secondary structures generally corresponded to the models proposed by Coleman (2000, 2015) for eukaryotic organisms. The strains ACSSI 343, ACSSI 344, and ACSSI 345 had 1 CBC in the III helix of ITS1 compared to *M. inermum*, 1 CBC in the conservative region of II helix and 1 CBC in the variable IV helix of ITS2 compared to *M. simplicissimum*. Strains ACSSI 332, 198, 287 did not have CBC compared to similar species. However, the strains ACSSI 198, 287 differed in structure of ITS2 helix II from *M. conductrix*. The mismatch in its upper part of strains ACSSI 198, 287 consisted of 4 unpaired nucleotides, and of *M. conductrix* – of 10.

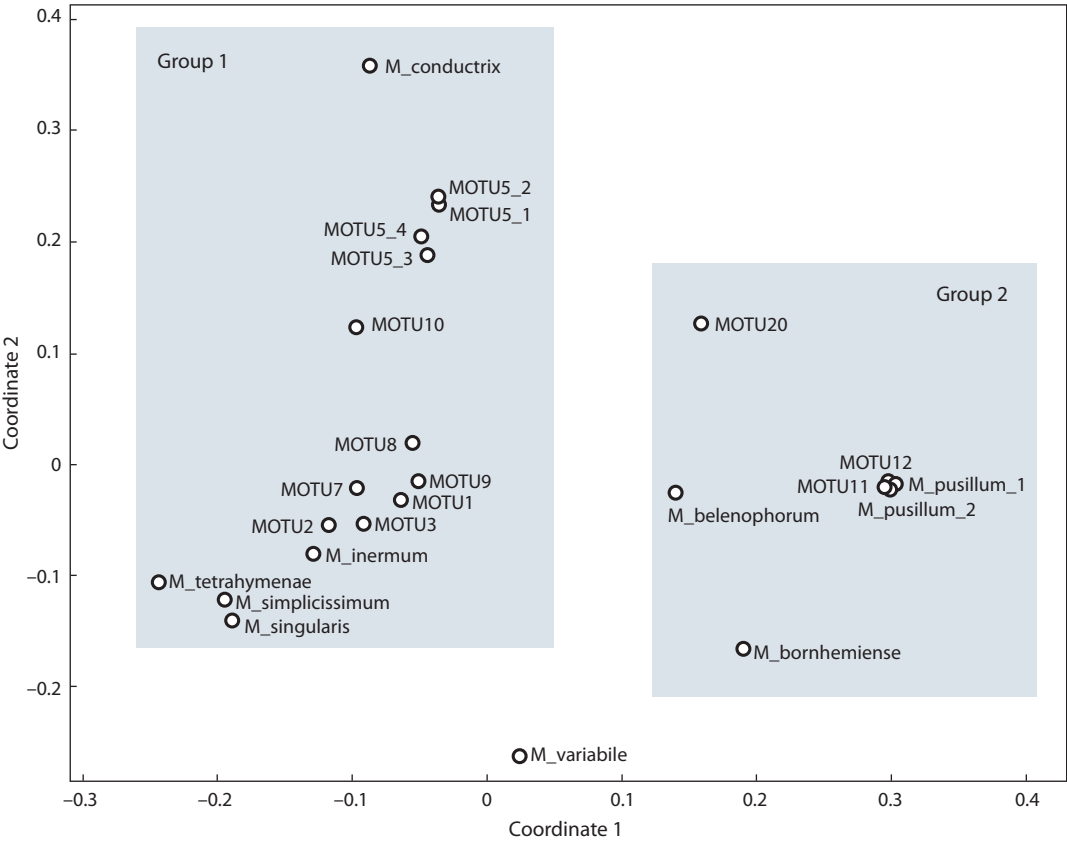
**Delimitation of species.** The ABGD method of species delimitation identified 18 MOTUs (molecular operational taxonomic units) of the species level in the genus *Micractinium*, not counting the external group. The ABGD distance of species differentiation in the pairwise comparison of sequences was 0.032 (Fig. 2). The results of ABGD analysis in the range of species distinction distances according to

the variants of the algorithm of initial delimitation (initial partition) and recursive delimitation (recursive partition) coincided with each other.

Using the GMYC method, the largest number of clusters of the species level was identified – 33 (the delimitation distance of species is 0.0015). Statistical support for the results of differentiation  $P = 1.07493e-07 < 0.05$ , therefore, there is enough data in the array to obtain reliable results. Using the PTP method, 30 species were identified, which is close to the results of the GMYC method. The results of species differentiation by ABGD, GMYC, and PTP methods are shown on the phylogenetic tree (see Fig. 1). All clusters of the species level identified by these methods have high statistical support (PP = 0.95–1.00, BP = 90–100 %).

**Multidimensional scaling.** To clarify the taxonomic status, we correlated the MOTUs isolated by the GMYC algorithm with their morphological, physiological, ecological, and molecular genetic characteristics (Suppl. Material 2). It should be noted that during the multidimensional scaling only the presence of an intron was taken into account from the genetic characteristics, while the remaining parameters are discussed separately. According to the results of the analysis, the studied MOTUs were divided into two groups (Fig. 3).

Group 1 included strains with single cells that do not produce bristles. Within it, only representatives of MOTU5\_1, MOTU5\_2 were united into one subgroup. Members of the other species/MOTU had a unique position. It should be noted separately that the studied strains ACSSI 343, ACSSI 344, ACSSI 345 (MOTU1), and strains ACSSI 198, ACSSI 287 (MOTU8) did not form a single complex with related species according to the results of phylogenetic analysis. The ACSSI strain 332 and IC-80 (MOTU5\_4) are localized next to the ehime strain (MOTU5\_3), while representatives of MOTU5\_1 and MOTU5\_2 are somewhat removed.



**Fig. 3.** The dot – MOTUs scattering diagram in the space of two coordinates of multidimensional scaling, constructed on the basis of similarity and difference of strains by a set of features.

**Table 2.** Mantel test results

Features	Code	Pearson correlation coefficient	<i>p</i> -value e
Ability to produce bristles	<b>B</b>	<b>0.56</b>	<b>0.0001</b>
Form (single cells or colonies)	<b>A</b>	<b>0.50</b>	<b>0.0001</b>
Chloroplast type	<b>E</b>	<b>0.45</b>	<b>0.0001</b>
Intron number	<b>J</b>	<b>0.38</b>	<b>0.0001</b>
Reproduction type	<b>F</b>	<b>0.37</b>	<b>0.0002</b>
Maximum cell size	<b>D</b>	<b>0.36</b>	<b>0.0002</b>
Cell shape	<b>C</b>	<b>0.35</b>	<b>0.0004</b>
Lifestyle	<b>H</b>	<b>0.32</b>	<b>0.0003</b>
Requirement for B vitamins	G	0.19	0.087
Relation to temperature	I	0.12	0.13

Note. Statistically significant features are highlighted in bold.

All representatives of Group 2, on the contrary, have bristles and, as a rule, form colonies. The strains initially identified as *M. pusillum* formed a single group. All the other species were quite distant from each other. An intermediate position between the groups is occupied by *M. variable*, in which only a part of the population is able to produce bristles

and form colonies in the presence of algophages. According to the results of the Mantel test, in addition to the ability to produce bristles and form colonies, chloroplast type, intron number, reproduction type, cell maximum size and shape, and lifestyle were considered significant features when distinguishing MOTUs (Table 2).



## Discussion

The ABGD method, based on the analysis of genetic distance matrices, compared with other methods, identified the smallest number of MOTUs of the species level with the greatest distance of their differentiation. This is consistent with the research of Zou et al. (2016a, b), who also noted a lower sensitivity of this method. The GMYC and PTP algorithms, using phylogenetic trees as initial data, are able to capture the features of genetic divergence between strains, identify a larger number of putative species, and are more consistent with the modern concept of the genus *Micractinium*. To clarify the taxonomic status of MOTU, the results of the GMYC algorithm, which identified the largest number of potential species, were correlated with their morphological, physiological, ecological, and molecular genetic characteristics.

**Morphological characteristics.** All representatives of the genus *Micractinium*, for which morphological characteristics are known, had a number of common features: a coccoid thallome, one parietal chloroplast, one pyrenoid with fragmented starch sheath, asexual reproduction by autospores. Important morphological criteria are the ability to produce bristles and form colonies (Krienitz et al., 2004; Luo et al., 2006). According to these characteristics, 2 morphotypes can be distinguished within this genus: *Chlorella*-like (under standard conditions, single cells do not produce bristles) and *Micractinium*-like (single cells or colonies producing bristles), which is confirmed by the results of multidimensional scaling (see Fig. 3). Other significant morphological characteristics for representatives of the genus *Micractinium* are chloroplast type, maximum size and shape of cells. However, *Micractinium* morphology is rather poor. The study did not reveal a single feature that could be considered as a universal tool for species distinguishing. For example, the phenotypes of *M. simplicissimum* and *M. singularis* are extremely similar, and it is quite problematic to separate them morphologically. It should be noted that microalgae of genus *Micractinium* have high phenotypic plasticity, and their morphotype can vary depending on the “grazing” load from algophages (Krienitz et al., 2004; Luo et al., 2006). Thus, in some cells of *M. variable*, which usually exhibits a *Chlorella*-like morphotype, the formation of colonies and the production of bristles is noted at high trophic pressure of algophages (Chae et al., 2019). At the same time, representatives of *M. pusillum*, *M. bornheimense*, *M. belenophorum* can stop producing bristles and form colonies during prolonged cultivation, especially on solid agar (Krienitz et al., 2004; Luo et al., 2006).

**Reproduction.** The main reproduction type of members of the genus *Micractinium* is asexual with the help of autospores. At the moment, an exception is *M. pusillum* strains that reproduce using oogamy (sexual process) (Krienitz et al., 2004; Luo et al., 2006). However, according to the whole genome analysis, meiotic genes, the presence of which suggests a sexual process, were found in many representatives of the *Trebouxiophyceae* class, for which only asexual reproduction was observed (Fučíková et al., 2015). This

question is still open and needs to be studied for members of *Micractinium*.

**The vitamins requirement and lifestyle.** Most of the species are free-living organisms and vitamins do not need to be added when culturing them under laboratory conditions. At the same time, a specific feature of *M. conductrix* is the requirement for vitamins B<sub>1</sub> and B<sub>12</sub> for normal implementation of vital processes. This species is an obligate endosymbiont and naturally receives vitamins from the host organism (Vorobyev et al., 2009; Hoshina et al., 2010; Pröschold et al., 2011). It is noteworthy that other obligate endosymbionts of the clade *Chlorella* (*C. variabilis*, *Caro-librandtia ciliaticola*) also grow only on media enriched with vitamins (Pröschold et al., 2011; Hoshina et al., 2017; Hoshina, Nakada, 2018). The strain CCAP 211/11F isolated from lichen and the facultative endosymbiont *M. tetrahymenae* are also cultivated on media containing B<sub>1</sub> and B<sub>12</sub>. However, there is no information that vital activity of the strains is not possible without them. According to the results of multidimensional scaling, lifestyle is one of the significant characteristics when distinguishing species, while the need for B vitamins is a highly specific property characteristic only of *M. conductrix*. However, it is a unique feature of this species and helps to separate the representatives of this species from the “sister” ones at the cultivation stage.

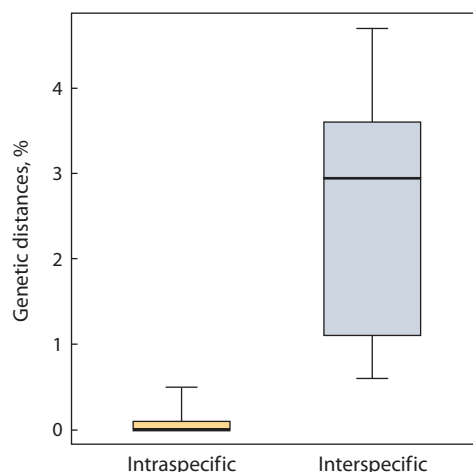
**Temperature.** In relation to temperature, members of the genus *Micractinium*, for the most part, show mesophilic characteristics (Hong et al., 2015). However, *M. simplicissimum*, *M. variable*, *M. singularis* and the strain of *Micractinium* sp. KNUA032 withstand the effects of low temperatures. They are able to survive and reproduce at temperatures up to +5 °C, showing their cryotolerant properties. One of the main adaptation strategies of these microalgae species is to maintain vital activity through the accumulation of unsaturated fatty acids (Hong et al., 2015; Chae et al., 2019). The strains TvB, SH, CCAP 211/92 are thermophiles that can withstand high temperatures (Adar et al., 2016). The ACSSI 332 strain is presumably resistant to high temperatures, since it was isolated from a hot source. The question of the thermophilicity of related strains chime and IC-76 remains unexplored. Multidimensional scaling has shown that resistance to the effects of extremely low or high temperatures are specific properties characteristic of only a small number of species. However, such species have a great biotechnological potential, and therefore they need to be carefully studied (Onay et al., 2014; Adar et al., 2016; Chae et al., 2021).

**Intron.** As an auxiliary tool for distinguishing species, it was effective in proving the species status of strain KNUA032, ACSSI 198 and ACSSI 287 cluster, and representatives of *M. conductrix*, all strains of which have an intron 324 nt long in the 18S rRNA gene. The composition of this intron and its specific position in 18S rRNA have been repeatedly considered by researchers as a characteristic feature of this species (Vorobyev et al., 2009; Hoshina et al., 2010; Spanner et al., 2020). The intron was also useful in distinguishing between the strains of *M. belenophorum*: CCAP

strain 271/1, in contrast to the authentic strain SAG 42.98, has an intron 315 nt long in the 18S rRNA gene. At the same time, the presence of introns in some species may indicate that speciation processes began but are still occurring at the population level (Goankar et al., 2018). For example, within the clade *Chlorella* Hoshina et al. (2021) found in some populations of *C. variabilis*, geographically isolated from each other, the length of introns can vary. Within the genus *Micractinium*, a similar situation can be observed among MOTU5 members with similar morphology, genetic distances at the intraspecific level and without CBC (see Suppl. Material 2). In the strains TvB, SH (MOTU5\_1), CCAP 211/92 (MOTU5\_2), unlike the related ehome (MOTU5\_3), ACSSI 332, and IC-76 (MOTU5\_4), an intron with a length of 351 nt is present in the 18S rRNA gene. In other words, although the intron is a statistically significant feature in the differentiation of MOTUs, it cannot be used as the main criterion for the division of species, but only as an auxiliary one.

**Comparative analysis of the secondary structure of internal transcribed spacers.** The application of the *sensu* Coleman (2000, 2015) CBC approach, based on the search for CBC exclusively in conservative ITS2 regions, was successful in distinguishing the strains SAG 48.93 and SAG 72.80 from the “true” representatives of *M. pusillum*, *M. conductrix* from the KNUA032 strain, as well as the authentic strain *M. belenophorum* SAG 42.98 compared to strain CCAP 271/1. The low efficiency of the *sensu* Coleman CBC approach for distinguishing green microalgae species with low genetic divergence was also noted in (Hoshina, Fujiwara, 2013; Song et al., 2018). Therefore, at present, when distinguishing species of the genus *Micractinium*, all CBCs in ITS1 and ITS2 are often taken into account (Hoshina, Fujiwara, 2013; Chae et al., 2019; Pröschold et al., 2020). However, for example, between the species *M. singularis* and *M. variabile*, there are no CBCs in both ITS1 and ITS2. At the same time, Chae et al. (2019) noted that these species differ in the structure of the ITS2 helix I.

The use of characteristic structural differences in the secondary structure of internal transcribed spacers as an analogue of CBC among members of the genus *Micractinium* was first proposed by Hoshina et al. (2010), who found a specific feature in the *M. conductrix* ITS2 secondary structure. In all representatives of the clade *Chlorella* in general and the genus *Micractinium* in particular, the II helix of ITS2 consists of two double-stranded regions articulated by an “elbow-like bulge”. Compared to other species, *M. conductrix* has a large “elbow” of 10 unpaired nucleotides (bachelor nucleotides), although other species have from three to six unpaired nucleotides. We believe that this feature can be considered a “molecular signature” of *M. conductrix*. For comparison, the sister strains KNUA032, ACSSI 198, and ACSSI 287 had only four unpaired nucleotides in this region. Thus, the CBC approach is not a universal tool for distinguishing species of the genus *Micractinium*. In addition, when analyzing internal transcribed spacers, one should not limit oneself only to searching for CBC in conservative



**Fig. 4.** Genetic distances within the genus *Micractinium*.

The borders of the box show the first and third quartiles, the bold horizontal line – the median value, the “whiskers” – the span.

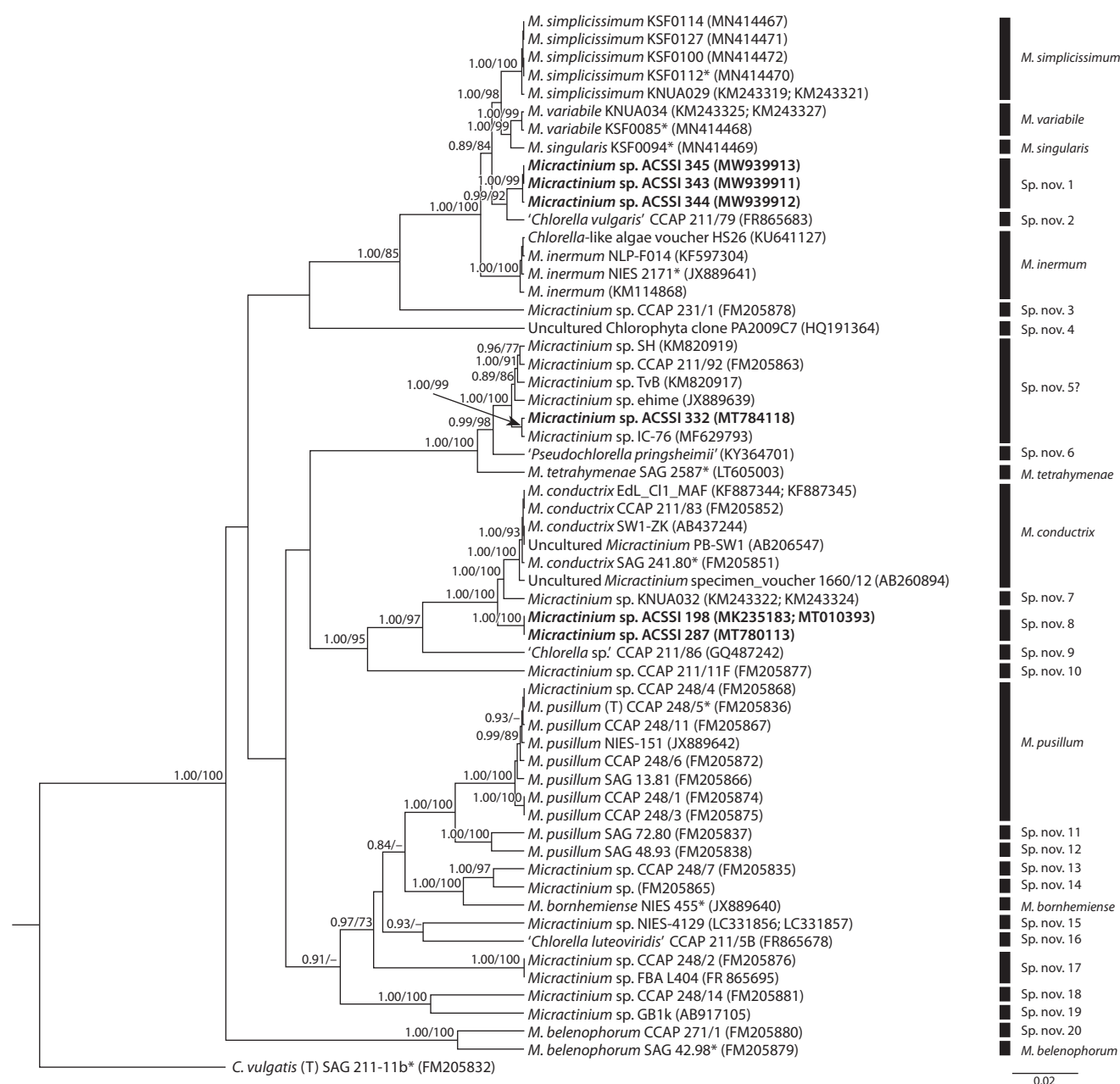
areas, it is also important to take into account the structural features of their secondary structures.

**Genetic distances.** A comparative analysis of the level of genetic differences of the fragment 18S–ITS1–5.8S–ITS2 of the studied strains with such diacritical features as the cell shape and size, the ability to produce bristles, the chloroplast type, the intron presence in the 18S rRNA gene, CBC in ITS1 and ITS2, molecular signatures, the ratio to temperature, vitamin requirement, lifestyle, clustering by ABGD, GMYC, PTP, allowed us to clarify intraspecific and interspecific levels genetic differences (Fig. 4). Within the species, the genetic distances varied in the range of 0–0.5 %, between species – 0.6–4.7 %. Minimal genetic distances were observed between single-celled and non-bristle-producing cryotolerant Antarctic species *M. singularis* and *M. variabile*, which, under the influence of “grazing” load, is able to form colonies and release bristles. The maximum genetic distances are between the *Chlorella*-like cryotolerant *M. simplicissimum* and *M. bornhemiense*, which under standard conditions has a classical *Micractinium*-like morphotype.

Based on the results of a comprehensive analysis of the above parameters, 29 species were identified within the genus *Micractinium* (Fig. 5), including candidates for three new species from the ACSSI Algological Collection, whose validation is yet to be performed.

## Conclusion

At present, only 9 species were described in the genus *Micractinium* using a combination of morphological and molecular genetic methods, but according to the analysis results, its true species richness turned out to be significantly higher – at least 29 species. The delimitation method ABGD, which is based on a matrix of genetic distances, is less “sensitive” and identified only 18 MOTUs of the species level, while the more advanced topological algorithms GMYC and PTP found 33 and 30, respectively. In our opinion, GMYC and PTP reflect the taxonomy of the genus *Micractinium*



**Fig. 5.** The proposed separation of species within the genus *Micractinium* based on a comprehensive analysis of features. The ACSSI strains studied in this work are highlighted in bold; \* – authentic strains; (T) – type species; ? – taxonomic status needs to be clarified.

more realistically, being an effective auxiliary tool for distinguishing species.

Multidimensional scaling of qualitative characteristics of the strains under consideration showed that the most significant for representatives of the genus *Micractinium* is the ability to produce bristles and form colonies, the chloroplast type, the intron presence, the reproduction type, the cell maximum size and shape, and lifestyle. However, not a single trait has been identified that could be considered as a universal species criterion. The requirements for B vitamins and resistance to extremely low or high temperatures are highly specific properties that are characteristic of only a small number of species and help in distinguishing them

from “sister” species. The application of the CBC approach based on the search for CBC in conservative ITS2 regions was successful only for the separation of “true” representatives of cryptic species (SAG 48.93, SAG 72.80) from *M. pusillum*, *M. conductrix* from strain KNUA032 and *M. belenophorum* from strain CCAP 271/1. When analyzing ITS1 and ITS2, in addition to searching for CBC, the structural features of their secondary structures should be taken into account. Based on the results of the analysis of the genetic distances of the 18S–ITS1–5.8S–ITS2 nucleotide sequences, it can be assumed that intraspecific differences are in the range of 0–0.5 %, interspecific differences are in the range of 0.6–4.7 %.



Thus, based on the joint use of morphological, physiological, ecological, and genetic characteristics (the polyphasic approach), it was possible to characterize 29 species within the genus *Micractinium* and propose additional criteria for their separation. Among the strains of the Algological Collection ACSSI, candidates for three new species of the genus *Micractinium*, the validation of which is yet to be performed, were found.

## References

- Adar O., Kaplan-Levy R.N., Banet G. High temperature Chlorellaceae (Chlorophyta) strains from the Syrian-African Rift Valley: the effect of salinity and temperature on growth, morphology and sporulation mode. *Eur. J. Phycol.* 2016;51(4):387-400. DOI 10.1080/09670262.2016.1193772.
- Aldous D.J. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat. Sci.* 2001;16:23-24. DOI 10.1214/ss/998929474.
- Bock C., Krienitz L., Pröschold T. Taxonomic reassessment of the genus *Chlorella* (Trebouxiophyceae) using molecular signatures (barcodes), including description of seven new species. *Fottea*. 2011;11(2):293-312. DOI 10.5507/fot.2011.028.
- Caisová L., Marin B., Melkonian M. A consensus secondary structure of ITS2 in the Chlorophyta identified by phylogenetic reconstruction. *Protist*. 2013;164:482-496. DOI 10.1016/j.protis.2013.04.005.
- Chae H., Lim S., Kim H., Choi H.-G., Kim J.H. Morphology and phylogenetic relationships of *Micractinium* (Chlorellaceae, Trebouxiophyceae) taxa, including three new species from Antarctica. *Algae*. 2019;34(4):267-275. DOI 10.4490/algae.2019.34.10.15.
- Chae H., Seo J.B., Kim S., Youn E.J. Antarctic freshwater microalga, *Micractinium simplicissimum*, suppresses inflammation. *J. Nanosci. Nanotechnol.* 2021;21(7):4098-4103(6). DOI 10.1166/jnn.2021.19158.
- Coleman A.W. The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist*. 2000;151:1-9. DOI 10.1078/1434-4610-00002.
- Coleman A.W. Nuclear rRNA transcript processing versus internal transcribed spacer secondary structure. *Trends Genet.* 2015;31(3):157-163. DOI 10.1016/j.tig.2015.01.002.
- Dixon P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 2003;14(6):927-930.
- Fawley M.W., Fawley K.P., Owen H.A. Diversity and ecology of small coccoid green algae from Lake Itasca, Minnesota, USA, including *Meyerella planktonica*, gen. et sp. nov. *Phycologia*. 2005;44:35-48.
- Fučíková K., Pažoutová M., Rindi F. Meiotic genes and sexual reproduction in the green algal class Trebouxiophyceae (Chlorophyta). *J. Phycol.* 2015;51(3):419-430. DOI 10.1111/jpy.12293.
- Fujisawa T., Barraclough T.G. Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* 2013;62(5):707-724. DOI 10.1093/sysbio/syt033.
- Gaonkar C.C., Piredda R., Minucci C., Mann D.G., Montresor M., Sarno D., Kooistra W.H.C.F. Annotated 18S and 28S rDNA reference sequences of taxa in the planktonic diatom family Chaetocerotaceae. *PLoS One*. 2018;13(12):e0208929. DOI 10.1371/journal.pone.0208929.
- Guiry M.D., Guiry G.M. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. (Cited on April 16, 2021). Available from: <http://www.algaebase.org>
- Hong J.W., Jo S.-W., Cho H.-W., Nam S.W., Shin W., Park K.M., Lee K.I., Yoon H.-S. Phylogeny, morphology, and physiology of *Micractinium* strains isolated from shallow ephemeral freshwater in Antarctica. *Phycol. Res.* 2015;63:212-218. DOI 10.1111/pre.12097.
- Hoshina R., Fujiwara Y. Molecular characterization of *Chlorella* cultures of the National Institute for Environment Studies culture collection with description of *Micractinium inermum* sp. nov., *Didymogenes sphaerica* sp. nov. and *Didymogenes soliella* sp. nov. (Chlorellaceae, Trebouxiophyceae). *Phycol. Res.* 2013;61(2):124-132. DOI 10.1111/pre.12010.
- Hoshina R., Iwataki M., Imamura N. *Chlorella variabilis* and *Micractinium reisseri* sp. nov. (Chlorellaceae, Trebouxiophyceae): redescription of the endosymbiotic green algae of *Paramecium bursaria* (Peniculia, Oligohymenophorea) in the 120th year. *Phycol. Res.* 2010;58(3):188-210. DOI 10.1111/j.1440-1835.2010.00579.x.
- Hoshina R., Kobayashi M., Suzaki T., Kusuok Y. *Brandtia ciliaticola* gen. et sp. nov. (Chlorellaceae, Trebouxiophyceae) a common symbiotic green coccoid of various ciliate species. *Phycol. Res.* 2017;66(1):76-81. DOI 10.1111/pre.12194.
- Hoshina R., Nakada T. *Carolibrandtia* nom. nov. as a replacement name for *Brandtia* Hoshina (Chlorellaceae, Trebouxiophyceae). *Phycol. Res.* 2018;66(1):82-83. DOI 10.1111/pre.12208.
- Hoshina R., Tsukii Y., Harumoto T., Suzaki T. Characterization of a green *Stentor* with symbiotic algae growing in an extremely oligotrophic environment and storing large amounts of starch granules in its cytoplasm. *Sci. Rep.* 2021;11:2865. DOI 10.1038/s41598-021-82416-9.
- Krienitz L., Hegewald E.H., Hepperle D., Huss V.A.R., Rohr T., Wolf M. Phylogenetic relationship of *Chlorella* and *Parachlorella* gen. nov. (Chlorophyta, Trebouxiophyceae). *Phycologia*. 2004;43:529-542. DOI 10.2216/i0031-8884-43-5-529.1.
- Krivina E., Temraleeva A. Identification problems and cryptic diversity of *Chlorella*-clade microalgae (Chlorophyta). *Microbiology*. 2020;89(6):720-732. DOI 10.1134/S0026261720060107.
- Lambert A., Stadler T. Birth-death models and coalescent point processes: the shape and probability of reconstructed phylogenies. *Theor. Popul. Biol.* 2013;90:113-128. DOI 10.1016/j.tpb.2019.07.002.
- Lartillot N., Philippe H. Computing Bayes factors using thermodynamic integration. *Syst. Biol.* 2006;55:195-207.
- Lipstein B., Hurwitz S. The nutritional value of sewage-grown samples of *Chlorella* and *Micractinium* in broiler diets. *Poultry Sci.* 1983;62(7):1254-1260. DOI 10.3382/ps.0621254.
- Luo W., Krienitz L., Pflugmacher S., Walz N. Genus and species concept in *Chlorella* and *Micractinium* (Chlorophyta, Chlorellaceae): genotype versus phenotypic variability under ecosystem conditions. *Verh. Internat. Verein. Limnol.* 2005;29:170-173. DOI 10.1080/03680770.2005.11901988.
- Luo W., Pflugmacher S., Pröschold T., Walz N., Krienitz L. Genotype versus phenotype variability in *Chlorella* and *Micractinium* (Chlorophyta, Trebouxiophyceae). *Protist*. 2006;157:315-333. DOI 10.1016/j.protis.2006.05.006.
- Luo W., Pröschold T., Bock C., Krienitz L. Generic concept in *Chlorella*-related coccoid green algae (Chlorophyta, Trebouxiophyceae). *Plant Biol.* 2010;12:545-553. DOI 10.1111/j.1438-8677.2009.00221.x.
- Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 1967;27(2):209-220.
- Mehrabadi A., Farid M.M., Craggs R. Potential of five different isolated colonial algal species for wastewater treatment and biomass energy production. *Algal Res.* 2017;21:1-8. DOI 10.1016/j.algal.2016.11.002.
- Onay M., Sonmez C.A., Oktem H., Yücel M. Thermo-resistant green microalgae for effective biodiesel production: isolation and characterization of unialgal species from geothermal flora of Central Anatolia. *Bioresour. Technol.* 2014;169:62-71. DOI 10.1016/j.biortech.2014.06.078.
- Paradis E., Claude J., Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20(2):289-290. DOI 10.1093/bioinformatics/btg412.
- Pröschold T., Bock C., Luo W., Krienitz L. Polyphyletic distribution of bristle formation in Chlorellaceae: *Micractinium*, *Diacanthos*, *Didymogenes* and *Hegewaldia* gen. nov. (Trebouxiophyceae, Chlorophyta). *Phycol. Res.* 2010;58:1-8. DOI 10.1111/j.1440-1835.2009.00552.x.



- Pröschold T., Darienko T., Silva P.C., Reisser W., Krienitz L. The systematics of *Zoochlorella* revisited employing an integrative approach. *Environ. Microbiol.* 2011;13:350-364. DOI 10.1111/j.1462-2920.2010.02333.x.
- Pröschold T., Pitsch G., Darienko T. *Micractinium tetrahymenae* (Trebouxiophyceae, Chlorophyta), a new endosymbiont isolated from ciliates. *Diversity.* 2020;12:200. DOI 10.3390/d12050200.
- Puillandre N., Lambert A., Brouillet S., Achaz G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* 2012;21:1864-1877. DOI 10.1111/j.1365-294X.2011.05239.x.
- Schnepp E., Deichgräber G., Glaab M., Hegewald E. Bristles and spikes in chlorococcales: ultrastructural studies in *Acanthosphaera*, *Micractinium*, *Pediastrum*, *Polyedriopsis*, *Scenedesmus*, and *Siderocystopsis*. *J. Ultrastr. Res.* 1980;72:367-379. DOI 10.1016/S0022-5320(80)90071-4.
- Seibel P.N., Müller T., Dandekar T., Wolf M. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Res. Notes.* 2008;1:1-91.
- Song H., Qinghua W.Q., Liu X., Hu Y., Long J., Guoxiang L.G., Hu Z. Phylogenetic diversity and taxonomic problems of the *Dictyosphaerium* morphotype within the *Parachlorella* clade (Chlorellaceae, Trebouxiophyceae). *J. Eukaryot. Microbiol.* 2018;65:382-391. DOI 10.1111/jeu.12482.
- Spanner C., Darienko T., Biehler T., Sonntag B., Pröschold T. Endosymbiotic green algae in *Paramecium bursaria*: a new isolation method and a simple diagnostic PCR approach for the identification. *Diversity.* 2020;12(6):240. DOI 10.3390/d12060240.
- Temraleeva A., Moskalenko S., Mincheva E., Bukin Y., Sinetova M. *Spongiosarcinopsis terrestris* gen. et sp. nov. (Chlorophyta, Chlorophyceae): a new genus of green algae from gray forest soil, Russia. *Phytotaxa.* 2018;376(6):291-300. DOI 10.11646/phytotaxa.376.6.4.
- Vayshlya O.B., Kulyatov D.V. Microscopic alga species promising as biodegraders of water ecosystem pollutants in southern West Siberia. *Izvestia Samarskogo Nauchnogo Tsentra RAN = Proceedings of the Samara Research Center of the Russian Academy of Sciences.* 2011;1(4):787-789. (in Russian)
- Vorobyev K., Andronov E., Rautian M., Skoblo I., Migunova A., Kvitko K. An atypical *Chlorella* symbiont from *Paramecium bursaria*. *Protistology.* 2009;6(1):39-44.
- Wolf M., Hegewald E., Hepperle D., Krienitz L. Phylogenetic position of the Golenkiniaceae (Chlorophyta) as inferred from 18S rDNA sequence data. *Biologia.* 2003;58:433-6.
- Zou S., Fei C., Song J., Bao Y., He M., Wang C. Combining and comparing coalescent, distance and character-based approaches for barcoding microalgae: a test with *Chlorella*-like species (Chlorophyta). *PLoS One.* 2016a;11(4):e0153833. DOI 10.1371/journal.pone.0153833.
- Zou S., Fei C., Wang C., Gao Z., Bao Y., He M., Wang C. How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in *Scenedesmus* (Chlorophyceae). *Sci. Rep.* 2016b;6:36822. DOI 10.1038/srep36822.

#### ORCID ID

E.S. Krivina orcid.org/0000-0002-0849-5832  
A.D. Temraleeva orcid.org/0000-0002-3445-0507  
Yu.S. Bukin orcid.org/0000-0002-4534-3846

**Acknowledgements.** The ACSSI strains cultivation, microscopy and literature analysis were carried out with the financial support of the Russian Foundation for Basic Research (RFBR), project number 19-34-60002, molecular phylogenetic work was carried out with the support of the Russian Science Foundation, project number 19-74-00030.

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

Received April 23, 2021. Revised July 28, 2021. Accepted August 6, 2021.

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

# DNA metabarcoding of benthic algae and associated eukaryotes from Lake Baikal in the face of rapid environmental changes

Yu.S. Bukin<sup>1, 2</sup>✉, L.S. Kravtsova<sup>1</sup>, T.E. Peretolchina<sup>1</sup>, A.P. Fedotov<sup>1</sup>, A.E. Tupikin<sup>3</sup>, M.R. Kabilov<sup>3</sup>,  
D.Yu. Sherbakov<sup>1, 4</sup>, E.V. Mincheva<sup>1</sup>

<sup>1</sup> Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

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

<sup>3</sup> Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>4</sup> Novosibirsk State University, Novosibirsk, Russia

✉ bukinyura@mail.ru

**Abstract.** Here we report new data describing the biodiversity of phytobenthic communities based on DNA-metabarcoding using the 18S rDNA marker and the Illumina MiSeq system. The study was initiated due to the blooming of filamentous algae (mainly of the genus *Spirogyra*) and cyanobacteria in the coastal zone of Lake Baikal under climate change and anthropogenic impact. The composition and taxonomic diversity of algae and other organisms associated with them on different sites of Lake Baikal (near Bolshoi Ushkaniy Island, in Listvennichny Bay) and in the Kaya (within the city of Irkutsk, located in the same drainage basin as Lake Baikal) were determined using DNA-metabarcoding. About 15 thousand reads of the 18S rRNA marker were obtained by applying NGS (next-generation sequencing). The species of algae dominating in the number of reads, as well as the difficult-to-identify taxa (Stramenopiles, Alveolata, Euglenozoa, Chromista, Rhizaria, Amoebozoa, etc.), which play an important role in the functioning and formation of the structure of algal communities, were revealed. The Shannon index of the communities studied ranges from 1.56 to 2.72. The advantages and weaknesses of using DNA-metabarcoding based on the 18S rRNA gene fragment for studying the structure of algal communities are shown. The advantage of this method is the possibility to more fully determine the diversity of eukaryotes taxa, which are difficult to identify by morphology, without involving a large number of specialists, while the disadvantage of the method is the distortion that may occur during the PCR. Here, ways of solving this problem are proposed. The results of the study show that the analysis of the minor component of the eukaryotic community in samples (organisms with low biomass) consisting of a mixture of multicellular and unicellular organisms requires a read-depths of at least 100,000 sequences per sample. In general, the DNA-metabarcoding method is recommended for studying the structure of algal communities and eukaryotes associated with them.

Key words: algal communities; metabarcoding; 18S rDNA; Illumina MiSeq; Lake Baikal; green algae; *Spirogyra*.

**For citation:** Bukin Yu.S., Kravtsova L.S., Peretolchina T.E., Fedotov A.P., Tupikin A.E., Kabilov M.R., Sherbakov D.Yu., Mincheva E.V. DNA metabarcoding of benthic algae and associated eukaryotes from Lake Baikal in the face of rapid environmental changes. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1): 86-95. DOI 10.18699/VJGB-22-12

## ДНК-метабаркодирование бентосных водорослей и ассоциированных с ними эукариот оз. Байкал в условиях быстрых экологических изменений

Ю.С. Букин<sup>1, 2</sup>✉, Л.С. Кравцова<sup>1</sup>, Т.Е. Перетолчина<sup>1</sup>, А.П. Федотов<sup>1</sup>, А.Е. Тупикин<sup>3</sup>, М.Р. Кабилов<sup>3</sup>,  
Д.Ю. Щербakov<sup>1, 4</sup>, Е.В. Минчева<sup>1</sup>

<sup>1</sup> Лимнологический институт Сибирского отделения Российской академии наук, Иркутск, Россия

<sup>2</sup> Иркутский государственный университет, Иркутск, Россия

<sup>3</sup> Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>4</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ bukinyura@mail.ru

**Аннотация.** Впервые приводится оценка разнообразия фитобентосных сообществ на основе ДНК-метабаркодирования с использованием ампликонов фрагмента гена 18S рРНК и технологии Illumina MiSeq. Исследование проведено в связи с цветением нитчатых водорослей (преимущественно рода *Spirogyra*) и цианобактерий в прибрежной зоне озера Байкал в условиях изменения климата и антропогенного воздействия. С помощью ДНК-метабаркодирования определен видовой состав водорослей, а также таксономическое разнообразие ассоциированных с ними эукариот в разных районах Байкала (у острова Большой Ушканий, в заливе Лиственнич-

ный) и в р. Кая (в черте г. Иркутска), находящейся в одном водосборном бассейне с оз. Байкал. С помощью NGS (next generation sequencing) получено более 15 тыс. прочтений 18S рРНК маркера. Выявлены виды водорослей, доминирующие по количеству прочтений, а также трудно идентифицируемые таксоны Stramenopiles, Alveolata, Euglenozoa, Chromista, Rhizaria, Amoebozoa и др., играющие важную роль в функционировании и формировании структуры водорослевых сообществ. Охарактеризовано разнообразие грибов и грибоподобных организмов в изучаемых сообществах. Индекс Шеннона рассмотренных сообществ колеблется от 1.56 до 2.72. Показаны преимущества и слабые стороны использования ДНК-метабаркодинга на основе фрагмента гена 18S рРНК для изучения структуры сообществ водорослей. Метод позволяет более полно учесть разнообразие таксонов эукариот, трудно идентифицируемых по морфологии, без привлечения большого числа специалистов, что характеризует его преимущество. Недостатком метода являются искажения, возникающие при проведении ПЦР. Предложены пути решения для устранения этого недостатка. Результаты исследования показывают, что для анализа минорной компоненты сообщества эукариот в образцах (организмы с малой биомассой), состоящих из смеси многоклеточных и одноклеточных организмов, требуется глубина прочтения не менее чем 100 000 последовательностей на пробу. В целом метод ДНК-метабаркодинга рекомендован для исследования структуры сообществ водорослей и ассоциированных с ними эукариот.

Ключевые слова: водорослевые сообщества; метабаркодинг; 18S рРНК; Illumina MiSeq; озеро Байкал; зеленые водоросли; *Spirogyra*.

## Introduction

Recently, a number of catastrophic and rapidly developing ecological phenomena, including the expansion of filamentous Chlorophyta and Cyanobacteria, took place in some areas of the coastal zone of Lake Baikal (Timoshkin et al., 2016). The first reports of such changes appeared in 2011 (Kravtsova et al., 2012, 2014). Previously, no changes in the coastal zone of the lake had been observed, with the invasion of *Elodea canadensis* in the 1970s being the only exception (Izhboldina, 1990).

It is known that the algae of Lake Baikal are characterized by zoning in spatial distribution and seasonal dynamics, which persisted for a long time (Meyer, 1930; Izhboldina, 1990, 2007; Izhboldina et al., 2017). However, since 2011, researchers have begun to note the overgrowth of the bottom with filamentous algae in Listvennichny Bay (Kravtsova et al., 2012, 2014). The overgrowth of the bottom with filamentous algae is also recorded in other areas of the lake near the settlements Kultuk, Baikalsk, Severobaikalsk (Timoshkin et al., 2018; Kravtsova et al., 2020). Among the filamentous algae blooms in the littoral zone near the Listvyanka, representatives of the genus *Spirogyra* dominate. For the first time (in almost 100 years of research), benthic filamentous algae *Spirogyra* were found in the plankton communities of the coastal zone (Bondarenko, Logacheva, 2016). In some areas of Lake Baikal, large accumulations of algae washed up on the shore were recorded (Suturin et al., 2016; Timoshkin et al., 2016, 2018). Forming algal mats along the coastline, filamentous algae impede the penetration of light, concentrate suspension and thus negatively affect filter feeders organisms; in particular, Baikal sponges (Khanaev et al., 2018). It should be noted that algae of the genus *Spirogyra* were encountered in Lake Baikal earlier. Researchers have occasionally found single *spirogyra* filaments in bottom phytocenoses in the bays of Lake Baikal. Among them, 4 species of the genus *Spirogyra* and 3 forms were registered: *S. calospora*, *S. decimina* (*S. decimina* f. *jurgensis*, *S. decimina* f. *longata*), *S. weberi* (*S. weberi*

f. *weberi*), *S. hassallii* (Izhboldina, 2007). Later, the species *S. fluviatilis* was discovered. This species dominated in the littoral zone of Listvennichny Bay in accumulations of filamentous algae in 2012 (Timoshkin et al., 2014). It is possible that endemic species of the genus *Spirogyra*, which have adapted to the specific conditions of the lake's ecosystem, can also inhabit Baikal. Currently, the question about the number of species occurred in accumulations of filamentous algae remains open.

A change in the composition of algae communities (and the ratio of algae biomass) entails a change in the composition of eukaryotic organisms associated with them (unicellular algae, protozoa and fungus-like organisms). The role of parasitic forms of eukaryotes (fungus-like organisms), negatively affecting the development of algae typical of the littoral zone, is practically not studied.

Algae of the littoral zone of open Lake Baikal are a rather complex object for taxonomic identification. Therefore, the analysis of the species diversity of algae and their abundance in benthic communities using classical morphological and hydrobiological methods is a laborious and time-consuming process. An even more difficult task is to study the taxonomic composition and quantitative ratio of various groups of eukaryotic organisms (including microeukaryotes) associated with algae. Identification of such taxa may require long-term cultivation on selective media and laborious microscopic analysis.

The application of modern molecular genetic approaches, such as DNA metabarcoding, can simplify and speed up these studies. This method includes the amplification of universal genetic markers in a mixture of DNA extracted from a sample followed by next generation sequencing (NGS) and analysis of the obtained sequence data set. Metabarcoding allows the detection of all species present in the bulk of DNA extracts and determines the species composition and quantitative ratio of taxa.

The metagenomic analysis of amplicons was widely used in the study of bacterial community composition using the

universal marker 16S ribosomal DNA (Petrosino et al., 2009). Similar studies were carried out for different bacterial communities of Lake Baikal (Kurilkina et al., 2016). The number of studies where metabarcoding based on 18S rDNA fragments and Folmer fragment of COI of mtDNA is used for analysis of eukaryotic communities is continuously increasing (Leray et al., 2013; Taylor, Cunliffe, 2014; Hawkins et al., 2015; Smith et al., 2017). These markers were implemented for studying communities from Lake Baikal. In particular, 18S rDNA was used for metabarcoding micro-eukaryotic communities (Yi et al., 2017), and COI was used for studying invertebrates' communities (Metazoa) (Kravtsova et al., 2021). Data on the DNA-metabarcoding of benthic algae in Lake Baikal are extremely limited (Mincheva et al., 2017), and data on algae communities and eukaryotes associated with them are currently lacking.

18S rRNA is most suitable for studying algal communities and associated organisms. To amplify various regions of this gene, universal primers that cover a wide range of species belonging to different distant taxa have been developed. Interpretation of sequencing results is facilitated by the availability of databases containing templates that allow alignment of large arrays of sequences, taking into account the secondary structure.

The aim of the study was to test the DNA-metabarcoding approach using the 18S rRNA marker to assess the diversity of benthic algal communities and associated eukaryotic organisms.

## Materials and methods

The sampling of algae (meio-, macrophytes  $\geq 2$  mm in size) was carried out in July–August 2015 on a stony littoral near Bolshoi Ushkaniy Island in Northern Baikal (background area, stony littoral), Listvennichny Bay opposite Listvyanka village in South Baikal (area overgrowing of the bottom with filamentous algae, stony littoral). For comparison, sample was collected in the Kaya River, flowing within the city of Irkutsk, located in the same drainage basin as Lake Baikal (Table 1).

In Northern and Southern Baikal, algae were collected by divers from three depths: 0–2, 2–5 and 6–10 m, and in the Kaya River – from a depth of 0.05–0.10 m. In each sampling site, algae collected from different depths were combined into one integral sample. The identification of algae was carried out according to L.A. Izhboldina (2007).

For molecular genetic analysis, the collected algae samples were fixed with 80 % ethyl ethanol, and then refixed with 70 % ethanol a day later.

Total DNA was isolated according to the modified method of Doyle and Dickson (Doyle, Dickson, 1987). A fragment of the 18S rRNA gene was used as a molecular genetic marker (Katana et al., 2001). Amplification was carried out with a set of PCR reagents with HS-Taq (Biolabmix, Novosibirsk, www.biolabmix.ru) in 25  $\mu$ L of the reaction mixture in a Bio-Rad-T100 thermal cycler (Bio-Rad, USA). The genetic marker (about 400 base pairs in length) encod-

ing the V1–V2 variable region of 18S rRNA was amplified using the 18SF universal primers: 5'-AACCTGGTTGATC CTGCCAGT-3' and 416-37R: 5'-ATTTGCGCGCCTGCT GCCTTCC-3' (Katana et al., 2001). The amplification conditions were as follows: predenaturation at 95 °C for 5 minutes, then 25 cycles: denaturation at 95 °C for 1 minute, annealing of primers at 55 °C for 1 minute, elongation at 72 °C for 2 minutes (5 minutes on the last cycle).

The reaction products were analyzed by electrophoresis in 1 % agarose gel. The band of the expected size was excised and purified using an agarose gel DNA elution kit (Biosilica, Novosibirsk).

The paired DNA sequencing of amplification products was performed using the Illumina MiSeq technology at the SB RAS Genomics Core Facility of the Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia).

All stages of the analysis of Illumina MiSeq DNA reads were carried out using the MOTHUR program (Schloss et al., 2009) and the SILVA 18S rRNA sequence database (Quast et al., 2012) according to the MiSeq standard operating procedure (MiSeq SOP) (Kozich et al., 2013). The analysis consisted of the following procedures: (1) merging of paired MiSeq reads of amplification products into consensus sequences; (2) trimming of consensus sequences by reading quality (deleting sequences with an average quality below 20 units); (3) removal of chimeric sequences from the data set; (4) deletion of sequences that do not correspond to the amplified 18S rRNA fragment in the SILVA database; (5) aligning sequences according to the SILVA database template; (6) calculation of the matrix of genetic distances (the proportion of mismatched nucleotides in pairwise comparison of sequences was used as a metric of distances); (7) clustering of sequences based on genetic distances; (8) identification of OTUs (operational taxonomic units) at the level of cluster distance (0.01) corresponding to interspecific differences (1 %); (9) drawing up of a table indicating the number of sequences per OTU in the sample; (10) secretion of representative sequences for each OTU; (11) taxonomic identification of representative sequences using the online BLAST application.

The statistical convergence of the results of assessing taxonomic diversity was characterized using saturation curves and the Chao1 index (Chao, 1987). The Chao1 index gives an estimate of the expected  $\alpha$  diversity in the studied community based on the observed number of taxa at the current number of reads per sample. In other words, Chao1's calculations allow the researcher to understand how many more taxa (species) can potentially be found in a sample if the number of reads increases from the existing value to infinity. A significant excess of the expected  $\alpha$  diversity calculated using the Chao1 index over the observed one indicates an insufficient number of reads in the sample and a loss of taxa.

Data on the taxonomic composition of communities (representation of OTU species rank) were compared using clus-



**Table 1.** General characteristics of the sample studied

Sampling locality	Coordinates	Dominant species identified by morphology	Reads number	Species richness (abundance)	Chao1	Shannon index
Bolshoi Ushkaniy Island (sample UI) (stony littoral)	N 53.848626° E 108.616931°	<i>Draparnaldioides baicalensis</i>	6610	19	22	2.72
Listvennichny Bay (sample LB) (stony littoral)	N 51.867102° E 104.832101°	<i>Spirogyra</i> sp.	5054	15	16	2.45
Kaya River (sample KR)	N 52.265051° E 104.235322°	<i>Cladophora glomerata</i> , <i>Draparnaldia plumosa</i>	3400	7	10	1.56

ter analysis by the UPGMA method, where the Bray–Curtis dissimilarity coefficient was used as a measure of distance. Before clustering, the data on the number of sequences per each presented OTU in the samples were normalized to the average number of reads per sample. The structure of the identified communities was visualized on a heat map.

The diversity of communities was assessed by the Shannon index, as well as by the species rank abundance curve, taking into account that the slower the abundance curve goes to zero, the more diverse the community is.

Statistical calculations (diversity index, clustering, data visualization) were carried out using the vegan package (Dixon, 2003) for the R programming language.

## Results

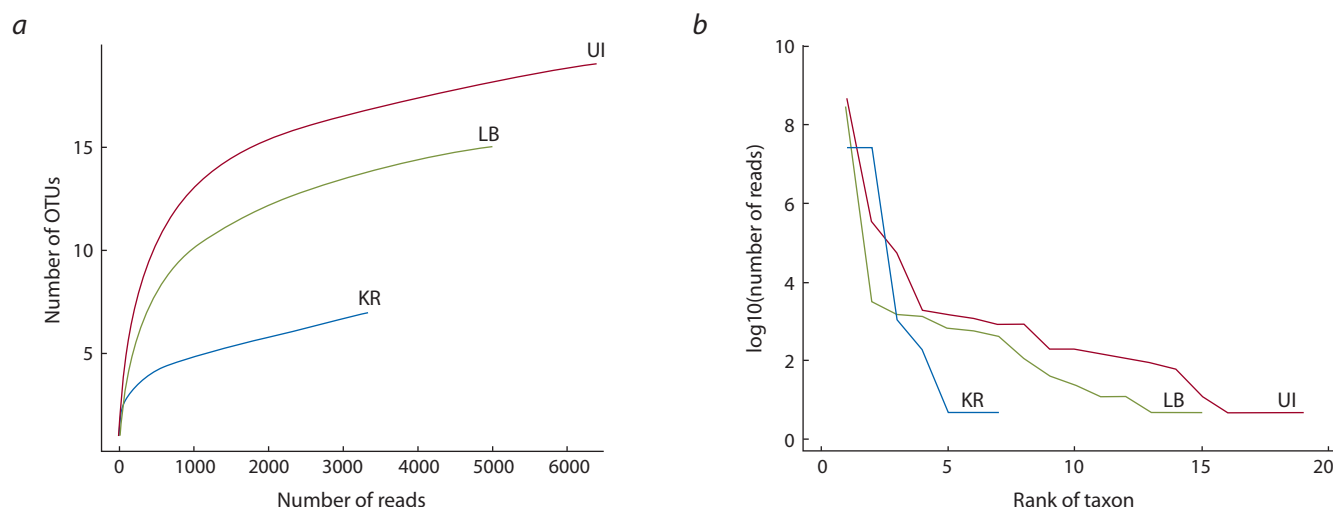
After applying initial data filtering, the dataset included 6610 reads from the Bolshoi Ushkaniy Island, 5054 reads from the Listvennichny Bay and 3400 reads from the Kaya River (see Table 1).

OTUs grouped at genetic distances of 1 % (0.01) had different numbers of sequences: 88 OTUs contained more than one sequence (2 – 4000), and 378 OTUs were presented by

a single sequence. The OTUs presented by a single sequence accounted for 2.57 % of the entire dataset, which was lower than the permissible 5 % threshold and indicated the absence of errors in the amplification, sequencing and initial data filtering steps. According to the practice of metabarcoding research (Kozich et al., 2013), only those OTUs that included 4 or more sequences were used for analysis (Table 2).

The convergence curves of species abundance in samples from different localities showed the absence of saturation (Fig. 1, a). The same result is provided by the values of the Chao1 index (see Table 1), according to which the number of species in the communities was also underestimated. Moreover, the most underestimated was the species composition of the minor component represented by unicellular eukaryotic organisms.

We identified 27 OTUs characterizing different taxa of algae and associated organisms (Fig. 2 and 3, see Table 2). Most taxa identified by BLAST belonged to the Charophyta and Chlorophyta algae: *Spirogyra*, *Draparnaldioides*, *Cladophora*, and *Draparnaldia* (see Table 2). Dominant taxa identified by DNA-metabarcoding are fully corroborated with those revealed by the morphologic analysis of samples



**Fig. 1.** Curves of saturation of the number of taxa of species rank in samples with various sample sizes of reads (a) and curves of the species abundance (b).

Here and in Fig. 2 and 3: UI – Lake Baikal, stony littoral, Bolshoi Ushkaniy Island; LB – Lake Baikal, stony littoral, Listvennichny Bay; KR – Irkutsk, Kaya River.

**Table 2.** Results of the taxonomic identification of OTU based on homology with sequences from the NCBI database

Taxon of species rank	Taxon of high rank	The homology with the reference sequences from NCBI	Proportion of the total reads, %		
			Lake Baikal, Bolshoi Ushkaniy Island (stony littoral)	Lake Baikal, Listvennichny Bay (stony littoral)	Kaya River (Irkutsk City)
<i>Draparnaldioides baicalensis</i>	Plantae	100	91.88	0.04	–
<i>Spirogyra</i> sp. 1		100	0.078	97.157	–
<i>Cladophora glomerata</i>		100	–	0.02	49.836
<i>Draparnaldia plumosa</i>		99	–	–	49.21
<i>Navicula radiosa</i>	Stramenopiles	100	3.928	–	–
Eukaryote sp.	Unclassified eukaryota	99	1.761	–	–
<i>Vorticella</i> sp.	Alveolata	100	0.358	0.26	0.03
<i>Gomphonema</i> cf. <i>angustatum</i>	Stramenopiles	99	–	0.641	0.03
<i>Didymosphenia geminata</i>		99	0.405	0.02	–
<i>Spumella</i> sp.		100	0.016	0.45	–
<i>Prokryptobia sorokinii</i>		100	–	0.44	–
<i>Nitzschia aequorea</i>	Stramenopiles	99	0.3	–	–
<i>Pythium myophilum</i>		100	0.15	0.03	0.596
<i>Adriamonas</i> sp.		93	0.25	–	–
<i>Pseudovorticella coscinodisci</i>		98	0.281	–	–
<i>Oikomonas</i> sp.	Stramenopiles	99	0.016	0.3	–
No identification Stramenopiles		100	–	0.3	–
<i>Ulnaria ulna</i>	Chromista	100	0.140	0.02	–
<i>Aspidisca</i> sp.	Alveolata	93	0.140	–	–
<i>Cladophora</i> sp.	Plantae	93	–	–	0.268
<i>Ichthyobodo</i> sp.	Euglenozoa	90	0.125	–	–
<i>Bodomorpha</i> sp.	Rhizaria	99	–	0.14	–
<i>Chaetophora</i> sp.	Plantae	97	0.109	–	–
<i>Hartmannellidae</i> sp.	Amoebozoa	93	0.094	–	–
<i>Achnanthidium</i> sp.	Stramenopiles	98	0.031	0.06	–
<i>Amphileptus</i> sp.	Alveolata	99	0.016	0.08	–
<i>Cocconeis</i> sp.	Stramenopiles	99	0.016	0.04	0.03

(see Table 1). In addition, a significant proportion of the sequences belonged to high-level taxa: Stramenopiles, Alveolata, Euglenozoa, Chromista, Rhizaria, and Amoebozoa (see Fig. 2). All samples differed by spectrum of dominant species (see Fig. 3, Table 2). Endemic *Draparnaldioides baicalensis* dominated in the background area near Bolshoi Ushkaniy Island of Northern Baikal. *Spirogyra* sp. dominated in List-

vennichny Bay. Communities of Bolshoi Ushkaniy Island and Listvennichny Bay include 10 common taxa in their composition and form one cluster on the dendrogram (see Fig. 3). *Cladophora glomerata* and *Draparnaldia plumosa* dominated in the Kaya River. The community from the Kaya River differed significantly from two other communities in the taxonomic composition and shared two taxa with the

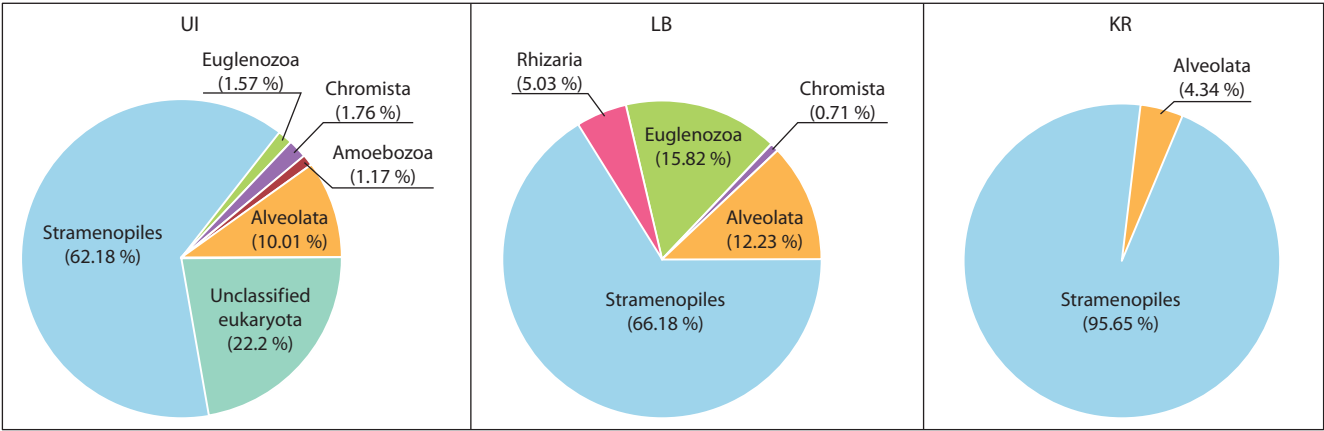


Fig. 2. Distribution of high rank taxa associated with algae (percentage ratios from their number of reads).

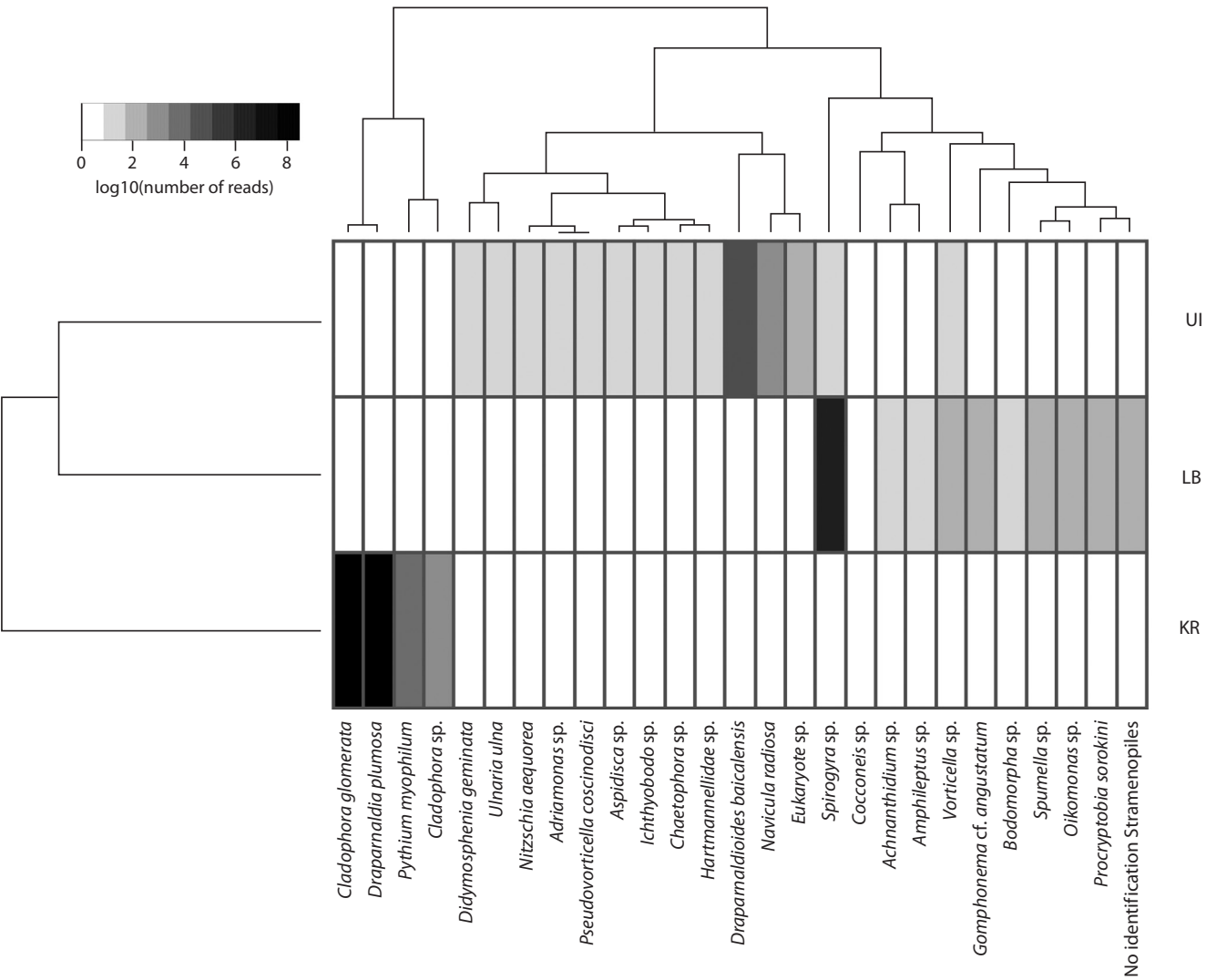


Fig. 3. Heat map of the structure of the benthic algae communities with eukaryotes associated with them and their clustering according to the degree of similarity based on the Bray–Curtis distances. Gray gradient shows the normalized number of reads (in a logarithmic scale) per taxon.

community of Bolshoi Ushkaniy Island, and four taxa with the community from Listvennichny Bay.

In general, according to the Shannon index, the communities from the Bolshoi Ushkaniy Island and Listvennichny Bay are more diverse compared to Kaya River (see Table 1). The abundance curves also confirm this result (see Fig. 1, b).

## Discussion

In the coastal zone near the Bolshoi Ushkaniy Island, located in the central conservation area of Lake Baikal, where there is practically no anthropogenic impact, typical representatives of the algal flora of the stony littoral of the lake are endemic species *Draparnaldioides baicalensis*, *D. arnoldii*, *D. arenaria* and *Cladophora floccosa* f. *floccosa* (Izhboldina, 1990). According to the results of the study, no structural changes were observed in the community near Bolshoi Ushkaniy Island (see Table 2, Fig. 3), although it contained trace amounts of *Spirogyra* sp. 1. The findings of *Spirogyra* in this area are due to the circulation currents that exist in Lake Baikal (Kravtsova et al., 2020).

During the summer period, in the littoral zone of Listvennichny Bay (before 2000s), at a depth of more than 1.5 metres, *Dermatochrysis reticulata*, *Didymosphenia geminata*, and *Nitella flexilis* species dominated. However, currently, they have lost the leading role in phytocenoses due to the expansion of green filamentous algae *Spirogyra* atypical for the open littoral zone of Lake Baikal (Kravtsova et al., 2014, 2020), which is confirmed by our results (see Table 2, Fig. 3). Sequences of 18S rDNA fragment obtained for filamentous algae *Spirogyra* from Listvennichny Bay were 100 % identical to a previously published sequence *Spirogyra* sp. 1 found in the littoral of this bay in 2013 (Romanova et al., 2013). Thus, we can assume that during the period from 2011 to 2015, the same species, *Spirogyra* sp. 1, developed in Listvennichny Bay. The spread of filamentous algae creates stressful conditions for the habitation of algal communities. It is known that there is a decrease in diversity according to Shannon index due to the expansion of species in ecosystems (Ling, 2008; Powell et al., 2013). The same regularities are observed in Listvennichny Bay, where algae typical for this period are suppressed; moreover, the Shannon index is lower here than in the community near Bolshoi Ushkaniy Island (see Table 1). Even less diversity by Shannon index was observed in the Kaya River (see Table 1). But we cannot unambiguously conclude that this community is in stressful conditions, because in the literature, we did not find information on the value of the Shannon diversity index characteristic of communities of bottom algae and associated eukaryotic organisms for river ecosystems. It is possible that the value of the Shannon index obtained for Kaya River is, in principle, typical for such an ecosystem.

The composition of eukaryotes associated with algae is quite diverse: Stramenopiles, Alveolata, Euglenozoa, Chromista, Rhizaria, and Amoebozoa. Of particular interest are oomycetes of the genus *Pythium* found in samples from the littoral zone of Lake Baikal (see Fig. 3, Table 2). It should be

noted that at present the mycofauna of Lake Baikal has not been practically studied, although the DNA of representatives of the genus *Pythium* was found during the study of microeukaryotes of the surface layer of bottom sediments of Lake Baikal (Yi et al., 2017). However, it is known that most of freshwater fungi of the genus *Pythium* are parasites of green algae (Raghukumar, 1987; Li et al., 2010; Carney, Lane, 2014). An increase in the concentration of *Pythium* in phytocenoses occurs at the end of the growing season, as it contributes to a more rapid destruction of the primary organic matter of plant origin. It is possible that the algae in the community near Bolshoi Ushkaniy Island ended their growing season, and therefore, they were more affected by parasitic fungi, as evidenced by the greater amount of DNA of *P. myophilum* in samples from this area compared to samples from Listvennichny Bay. On the one hand, *P. myophilum*, apparently, has not yet had time to spread in the community of the littoral zone of Listvennichny Bay, since during the study *Spirogyra* sp. 1 was in a good phenological state. On the other hand, *Spirogyra* sp. 1 can potentially be resistant to the *P. myophilum*, in contrast to other typical species (for example, *Ulothrix*), which are widespread in the littoral of Lake Baikal. Then, in this situation, *Spirogyra* sp. 1 gains a competitive advantage over other species and becomes a community-forming taxon. The distribution of *P. myophilum* probably makes a certain contribution to changes in the structure of benthic communities of algae in the littoral of Lake Baikal.

Despite the originality of the results obtained, we would like to draw attention to the features of the use of the DNA-metabarcoding approach in the study of algal communities and associated eukaryotes. Analyzing the 18S rRNA metabarcoding data, we faced the problem of determining the threshold of the genetic distance between species within a genus. For 16S rRNA (a marker for bacterial communities), this distance was chosen as 3 % (0.03) of mismatched nucleotides between the compared sequences (Petrosino et al., 2009; Kurilkina et al., 2016). Some researchers use the same distance to separate the OTUs of the species level for analysis of eukaryotic communities based on 18S rRNA (Yi et al., 2017). If this is justified for microeukaryotes to some extent (due to the high rate of evolution and rapid change of generations), then for multicellular organisms such a measure cannot be used to separate species, since it is proved that 18S rRNA is one of the slowly evolving markers for them (Anne, 2006). Our analysis of the literature data showed that for multicellular algae, the distance threshold corresponding to 3 % of mismatched nucleotides separates not species, but different genera and families (Chen et al., 2012; Romanova et al., 2013; Sherwood et al., 2014; Taylor, Cunliffe, 2014). Thus, if we used a threshold of 3 % for isolating OTUs, then we could not identify species, and the study of species diversity in this group of organisms would become impossible. A more detailed analysis of the published information (Chen et al., 2012; Romanova et al., 2013; Sherwood et al., 2014; Taylor, Cunliffe, 2014) allowed us to choose a threshold



of 1 % (0.01) substitutions for separating different species within one genus and use it in our study.

In the course of the study, it was found that with 3400–6610 reads of the 18S rRNA fragment for each sample, the species diversity of eukaryotic organisms associated with benthic algae is clearly underestimated. In the DNA mixture consisting of macro and microorganisms, the main pool of reads was presented by multicellular species (algae), and accounted for from 91 to 99 % of all 18S rRNA reads in each sample. In this case, we adequately evaluate the species diversity of macroorganisms (algae), but lose most of the diversity of eukaryotic organisms associated with them. In the methodological works on the study of bacterial communities based on 16S rRNA metabarcoding (Bukin et al., 2019), it is given that the final number of filtered in quality reads in the sample for an adequate assessment of the species diversity of microorganisms should be 10,000. Considering this, in order to evaluate the species diversity of the minor component of the community of eukaryotic organisms associated with algae, it is necessary to increase the number of reads per sample 50 or more times. Then the minor component will account for approximately 10,000 DNA sequences. In total, several hundred thousand reads will be required. With the current level of development of NGS technologies, this is a completely accessible task. Another way to obtain 10,000 readings on the minor component of the community is a preliminary mechanical separation of the sample into two parts: one of them will include algae, and another will contain organisms associated with algae. In this case, the DNA of the two selected parts should be sequenced separately.

It should also be noticed that the specificity of the universal primers varies for different taxa in samples, and a distortion of DNA-metabarcoding results may arise. As a result, the initial DNA concentration of different taxa in the sample is changed after PCR. It is possible to decrease this effect by reducing the number of PCR cycles during the preparation of a sample for sequencing, using the same sets of reagents and standardizing the sampling method. Another important note refers to the methodology of statistical analysis, which must contain the data range stage. To do this, you can convert the number of reads to the proportion of reads on a taxon in a sample, or normalize the entire dataset (as done in our study) to the average number of reads on the sample. Such data rationing, despite the distortions associated with PCR, will allow determining the trend of changes in the concentration of DNA of any taxon in different compared natural samples using multidimensional statistics (clustering methods, etc.).

## Conclusion

Due to the growth of filamentous green algae, one of the *Spirogyra* species began to dominate in the community of the littoral in Listvennichny Bay. DNA of this species was also found in the samples of the background area near Bolshoi Ushkaniy Island of Lake Baikal. DNA-metabarcoding based on a fragment of the 18S rRNA gene is a perspective method

for studying the structure of algae communities and makes it possible to obtain statistically representative results. This method is also effective for accurate taxonomic identification of a morphologically complex group of organisms, such as filamentous algae. At the same time, DNA-metabarcoding allows determining the representation in the samples of difficult-to-study taxa associated with algae, which play an important role in the formation of the diversity and the functioning of the communities. For a representative assessment of the minor component of the community (eukaryotic organisms associated with algae), a significant increase in the sample sizes of DNA sequences is necessary.

## References

- Anne C. Choosing the right molecular genetic markers for studying biodiversity: from molecular evolution to practical aspects. *Genetica*. 2006;127(1-3):101-120. DOI 10.1007/s10709-006-9118-1.
- Bondarenko N.A., Logacheva N.F. Structural changes in phytoplankton of the littoral zone of Lake Baikal. *Gidrobiologicheskii Zhurnal = Hydrobiological Journal*. 2017;53(2):16-24. DOI 10.1615/HydrobJ.v53.i2.20.
- Bukin Y.S., Galachyants Y.P., Morozov I.V., Bukin S.V., Zakharenko A.S., Zemskaia T.I. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci. Data*. 2019; 6(1):1-14. DOI 10.1038/sdata.2019.7.
- Bukin Yu.S., Bondarenko N.A., Rusanov I.I., Pimenov N.V., Bukin S.V., Pogodaeva T.V., Chernitsyna S.M., Shubenkova O.V., Ivanov V.G., Zakharenko A.S., Zemskaia T.I. Interconnection of bacterial and phytoplanktonic communities with hydrochemical parameters from ice and under-ice water in coastal zone of Lake Baikal. *Sci. Rep.* 2020;10(11087):1-12. DOI 10.1038/s41598-020-66519-3.
- Carney L.T., Lane T.W. Parasites in algae mass culture. *Front. Microbiol.* 2014;5:278. DOI 10.3389/fmicb.2014.00278.
- Chao A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*. 1987;43:783-791. DOI 10.2307/2531532.
- Chen C., Barfuss M.H., Pröschold T., Schagerl M. Hidden genetic diversity in the green alga *Spirogyra* (Zygnematophyceae, Streptophyta). *BMC Evol. Biol.* 2012;12:77. DOI 10.1186/1471-2148-12-77.
- Dixon P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 2003;14(6):927-930. DOI 10.1111/j.1654-1103.2003.tb02228.x.
- Doyle J.J., Dickson E.E. Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*. 1987;36(4):715-722. DOI 10.2307/1221122.
- Hawkins J., de Vere N., Griffith A., Ford C., Allainguillaume J., Hegarty M., Baillie L., Adams-Groom B. Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. *PLoS One*. 2015; 10(8):e0134735. DOI 10.1371/journal.pone.0134735.
- Izhboldina L.A. Meio- and Macrophytobenthos of Lake Baikal (algae). Irkutsk: Irkutsk State Univ. Publ. House, 1990. (in Russian)
- Izhboldina L.A. Atlas and Keys to Algae of Benthos and Periphyton of Lake Baikal (meio and macrophytes) with Brief Essays on Their Ecology. Novosibirsk: Nauka-Tsentr Publ., 2007. (in Russian)

- Izhboldina L.A., Chepinoga V.V., Mincheva E.V. Meio- and macrophytobenthos distribution in the littoral zone along the open coasts of Lake Baikal according to profiling data from 1963–1988. Part 2. Eastern coast. *Izvestiya Irkutskogo Gosudarstvennogo Universiteta. Ser. Biologiya, Ekologiya = Irkutsk State University Bulletin. Ser. "Biology. Ecology"*. 2017;19:36–57. (in Russian)
- Katana A., Kwiatowski J., Spalik K., Zakryś B., Szalacha E., Szymańska H. Phylogenetic position of *Koliella* (Chlorophyta) as inferred from nuclear and chloroplast small subunit rDNA. *J. Phycology*. 2001;37(3):443–451. DOI 10.1046/j.1529-8817.2001.037003443.x.
- Khanaev I.V., Kravtsova L.S., Maikova O.O., Bukshuk N.A., Sakirko M.V., Kulakova N.V., Butina T.V., Nebesnykh I.A., Belikov S.I. Current state of the sponge fauna (Porifera: Lubomirskiidae) of Lake Baikal: sponge disease and the problem of conservation of diversity. *J. Great Lakes Res.* 2018;44(1):77–85. DOI 10.1016/j.jglr.2017.10.004.
- Kozich J.J., Westcott S.L., Baxter N.T., Highlander S.K., Schloss P.D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 2013;79(17):5112–5120. DOI 10.1128/AEM.01043-13.
- Kravtsova L.S., Izhboldina L.A., Khanaev I.V., Pomazkina G.V., Domyшева V.M., Kravchenko O.S., Grachev M.A. Disturbances of the vertical zoning of green algae in the coastal part of the Listvennichnyy gulf of Lake Baikal. *Doklady Biological Sciences*. 2012;447(1):350–352. DOI 10.1134/S0012496612060026.
- Kravtsova L.S., Izhboldina L.A., Khanaev I.V., Pomazkina G.V., Rodionova E.V., Domyшева V.M., Sakirko M.V., Tomberg I.V., Kostornova T.Ya., Kravchenko O.S., Kupchinsky A.B. Nearshore benthic blooms of filamentous green algae in Lake Baikal. *J. Great Lakes Res.* 2014;40(2):441–448. DOI 10.1016/j.jglr.2014.02.019.
- Kravtsova L.S., Mizandrontsev I.B., Vorobyova S.S., Izhboldina L.A., Mincheva E.V., Potyomkina T.G., Golobokova L.P., Sakirko M.V., Triboy T.I., Khanaev I.V., Sherbakov D.Yu., Fedotov A.P. Influence of water motion on the spatial distribution of *Spirogyra* in Lake Baikal. *J. Great Lakes Res.* 2020;46(1):29–40. DOI 10.1016/j.jglr.2019.09.004.
- Kravtsova L.S., Peretolchina T.E., Triboy T.I., Nebesnykh I.A., Kupchinskiy A.B., Tupikin A.E., Kabilov M.R. The study of the diversity of hydrobionts from Listvennichny Bay of Lake Baikal by DNA metabarcoding. *Russ. J. Genet.* 2021;57(4):460–467. DOI 10.1134/S1022795421040050.
- Kurilkina M.I., Zakharova Yu.R., Galachyants Yu.P., Petrova D.P., Bukin Yu.S., Domyшева V.M., Blinov V.V., Likhoshvay Ye.V. Bacterial community composition in the water column of the deepest freshwater Lake Baikal as determined by next-generation sequencing. *FEMS Microbiol. Ecol.* 2016;92(7):1–19. DOI 10.1093/femsec/fiw094.
- Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Machida R.J. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front. Zool.* 2013;10(1):1–14. DOI 10.1186/1742-9994-10-34.
- Li W., Zhang T., Tang X., Wang B. Oomycetes and fungi: important parasites on marine algae. *Acta Oceanologica Sinica*. 2010;29(5):74–81. DOI 10.1007/s13131-010-0065-4.
- Ling S.D. Range expansion of a habitat-modifying species leads to loss of taxonomic diversity: a new and impoverished reef state. *Oecologia*. 2008;156(4):883–894. DOI 10.1007/s00442-008-1043-9.
- Meyer K.I. Introduction to the algal flora of Lake Baikal. *Bulleten MOIP. Otdelenie Biol. = Bulletin of Moscow Society of Naturalists. Biological Series*. 1930;38:179–396. (in Russian)
- Mincheva E.V., Bukin Yu.S., Kravtsova L.S., Koval V.V., Kabilov M.R., Tupikin A.E., Shcherbakov D.Yu. Study of algal-fungal communities in Listvennichny Bay and Bol'shoy Ushkaniy island of Lake Baikal. In: Bychkov I.V., Kazakov A.L. (Eds.) Topical Problems in Baikal Region Studies. Iss. 2. Irkutsk, 2017;138–143. (in Russian)
- Petrosino J.F., Highlander S., Luna R.A., Gibbs R.A., Versalovic J. Metagenomic pyrosequencing and microbial identification. *Clin. Chem.* 2009;55(5):856–866. DOI 10.1373/clinchem.2008.107565.
- Powell K.I., Chase J.M., Knight T.M. Invasive plants have scale-dependent effects on diversity by altering species-area relationships. *Science*. 2013;339(6117):316–318. DOI 10.1126/science.1226817.
- Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Glöckner F.O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2012;41(D1):D590–D596. DOI 10.1093/nar/gks1219.
- Raghukumar C. Fungal parasites of the green alga *Chaetomorpha* media. *Infection*. 1987;50:100. DOI 10.3354/dao00314.7.
- Romanova E.V., Kravtsova L.S., Izhboldina L.A., Khanayev I.V., Shcherbakov D.Yu. Identification of filamentous green algae from an area of local biogenic pollution of Lake Baikal (Listvennichnyy bay) using SSU 18S rDNA molecular marker. *Ekologicheskaya Genetika = Ecological Genetics*. 2013;11(4):23–33. DOI 10.17816/ecogen11423-33. (in Russian)
- Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Sahl J.W. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 2009;75(23):7537–7541. DOI 10.1128/AEM.01541-09.
- Sherwood A.R., Carlile A.L., Neumann J.M., Kociolek J.P., Johansen J.R., Lowe R.L., Presting G.G. The Hawaiian freshwater algae biodiversity survey (2009–2014): systematic and biogeographic trends with an emphasis on the macroalgae. *BMC Ecol.* 2014;14(1):28. DOI 10.1186/s12898-014-0028-2.
- Smith K.F., Kohli G.S., Murray S.A., Rhodes L.L. Assessment of the metabarcoding approach for community analysis of benthic-epiphytic dinoflagellates using mock communities. *N. Z. J. Mar. Freshwater Res.* 2017;51(1):555–576. DOI 10.1186/s12898-014-0028-2.
- Suturin A.N., Chebykin E.P., Malnik V.V., Khanaev I.V., Minaev A.V., Minaev V.V. The role of anthropogenic factors in the development of ecological stress in Lake Baikal littoral (the Listvyanka settlement lakescape). *Geografiya i Prirodnye Resursy = Geography and Natural Resources*. 2016;6:43–54. DOI 10.21782/GIPR0206-1619-2016-6(43-54). (in Russian)
- Taylor J.D., Cunliffe M. High-throughput sequencing reveals neustonic and planktonic microbial eukaryote diversity in coastal waters. *J. Phycol.* 2014;50(5):960–965. DOI 10.1111/jpy.12228.
- Timoshkin O.A., Bondarenko N.A., Volkova Y.A., Tomberg I.V., Vishnyakov V.S., Malnik V.V. Mass development of green filamentous algae of the genera *Spirogyra* and *Stigeoclonium* (Chlorophyta) in the littoral zone of the southern part of Lake Baikal. *Gidrobiologicheskii Zhurnal = Hydrobiological Journal*. 2015;51(1):13–23. DOI 10.1615/Hydrobiol.v51.i1.20. (in Russian)
- Timoshkin O.A., Moore M.V., Kulikova N.N., Tomberg I.V., Malnik V.V., Shimaraev M.N., Troitskaya E.S., Shirokaya A.A., Sinyukovich V.N., Zaitseva E.P., Domyшева V.M., Yamamuro M., Po-

- berezhnaya A.E., Timoshkina E.M. Groundwater contamination by sewage causes benthic algal outbreaks in the littoral zone of Lake Baikal (East Siberia). *J. Great Lakes Res.* 2018;44(2):230-244. DOI 10.1016/j.jglr.2018.01.008.
- Timoshkin O.A., Samsonov D.P., Yamamuro M., Moore M.V., Belykh O.I., Malnik V.V., Sakirko M.V., Shirokaya A.A., Bondarenko N.A., Domysheva V.M., Fedorova G.A., Kochetkov A.I., Kuzmin A.V., Lukhnev A.G., Medvezhonkova O.V., Nepokrytykh A.V., Pasynkova E.M., Poberezhnaya A.E., Potapskaya N.V., Rozhkova N.A., Sheveleva N.G., Tikhonova I.V., Timoshkina E.M., Tomberg I.V., Volkova E.A., Zaitseva E.P., Zvereva Yu.M., Kupchinsky A.B., Bukshuk N.A. Rapid ecological change in the coastal zone of Lake Baikal (East Siberia): is the site of the world's greatest freshwater biodiversity in danger? *J. Great Lakes Res.* 2016;42(3):487-497. DOI 10.1016/j.jglr.2016.02.011.
- Yi Z., Berney C., Hartikainen H., Mahamdallie S., Gardner M., Boenigk J., Bass D. High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiol. Ecol.* 2017;93(8):fix073. DOI 10.1093/femsec/fix073.

---

#### ORCID ID

Yu.S. Bukin orcid.org/0000-0002-4534-3846  
L.S. Kravtsova orcid.org/0000-0003-0862-4726  
T.E. Peretolchina orcid.org/0000-0002-2950-9762  
A.P. Fedotov orcid.org/0000-0003-3020-9895

A.E. Tupikin orcid.org/0000-0002-8194-0322  
M.R. Kabilov orcid.org/0000-0003-2777-0833  
D.Yu. Sherbakov orcid.org/0000-0002-1410-392X  
E.V. Mincheva orcid.org/0000-0003-4447-6345

**Acknowledgements.** This study was supported by the governmentally funded project of the Limnological Institute of the SB RAS No. 121032300196-8, Russian Foundation for Basic Research, projects Nos. 17-44-388071\_r, and 19-05-00398\_a.

We are grateful to the Irkutsk Supercomputer Center of the SB RAS for providing access to the high-performance cluster "Academician V.M. Matrosov". We thank the administrator of the Irkutsk Supercomputer Center of the SB RAS Ivan Sidorov for help in conducting computing.


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

Received July 30, 2021. Revised October 5, 2021. Accepted October 21, 2021.

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

# A bioinformatic search for correspondence between differentially expressed genes of domestic versus wild animals and orthologous human genes altering reproductive potential


M.P. Ponomarenko , I.V. Chadaeva, P.M. Ponomarenko, A.G. Bogomolov, D.Yu. Oshchepkov, E.B. Sharypova, V.V. Suslov, A.V. Osadchuk, L.V. Osadchuk, Yu.G. Matushkin

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia  
 [pon@bionet.nsc.ru](mailto:pon@bionet.nsc.ru)

**Abstract.** One of the greatest achievements of genetics in the 20th century is D.K. Belyaev's discovery of destabilizing selection during the domestication of animals and that this selection affects only gene expression regulation (not gene structure) and influences systems of neuroendocrine control of ontogenesis in a stressful environment. Among the experimental data generalized by Belyaev's discovery, there are also findings about accelerated extinction of testes' hormonal function and disrupted seasonality of reproduction of domesticated foxes in comparison with their wild congeners. To date, Belyaev's discovery has already been repeatedly confirmed, for example, by independent observations during deer domestication, during the use of rats as laboratory animals, after the reintroduction of endangered species such as Przewalski's horse, and during the creation of a Siberian reserve population of the Siberian grouse when it had reached an endangered status in natural habitats. A genome-wide comparison among humans, several domestic animals, and some of their wild congeners has given rise to the concept of self-domestication syndrome, which includes autism spectrum disorders. In our previous study, we created a bioinformatic model of human self-domestication syndrome using differentially expressed genes (DEGs; of domestic animals versus their wild congeners) orthologous to the human genes (mainly, nervous-system genes) whose changes in expression affect reproductive potential, i.e., growth of the number of humans in the absence of restrictions caused by limiting factors. Here, we applied this model to 68 human genes whose changes in expression alter the reproductive health of women and men and to 3080 DEGs of domestic versus wild animals. As a result, in domestic animals, we identified 16 and 4 DEGs, the expression changes of which are codirected with changes in the expression of the human orthologous genes decreasing and increasing human reproductive potential, respectively. The wild animals had 9 and 11 such DEGs, respectively. This difference between domestic and wild animals was significant according to Pearson's  $\chi^2$  test ( $p < 0.05$ ) and Fisher's exact test ( $p < 0.05$ ). We discuss the results from the standpoint of restoration of endangered animal species whose natural habitats are subject to an anthropogenic impact.  
Key words: human; reproductive potential; animal model of human disease; domestication; RNA-Seq; most recent common ancestor.

**For citation:** Ponomarenko M.P., Chadaeva I.V., Ponomarenko P.M., Bogomolov A.G., Oshchepkov D.Yu., Sharypova E.B., Suslov V.V., Osadchuk A.V., Osadchuk L.V., Matushkin Yu.G. A bioinformatic search for correspondence between differentially expressed genes of domestic versus wild animals and orthologous human genes altering reproductive potential. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):96-108. DOI 10.18699/VJGB-22-13

## Биоинформатический поиск соответствия дифференциально экспрессируемых генов домашних и диких животных с ортологичными генами, изменяющими репродуктивный потенциал человека

М.П. Пономаренко , И.В. Чадаева, П.М. Пономаренко, А.Г. Богомолов, Д.Ю. Ощепков, Е.Б. Шарыпова, В.В. Суслов, А.В. Осадчук, Л.В. Осадчук, Ю.Г. Матушкин

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия  
 [pon@bionet.nsc.ru](mailto:pon@bionet.nsc.ru)

**Аннотация.** Одним из крупнейших достижений генетики XX в. стало открытие Д.К. Беляевым дестабилизирующего отбора при одомашнивании животных, который затрагивает регуляцию экспрессии генов, но не их структуру, и влияет на системы нейроэндокринного контроля онтогенеза при стрессовом воздействии окружающей среды. Среди экспериментов, результаты которых были обобщены этим открытием, были также наблюдения ускоренного угасания гормональной функции семенников и нарушения сезонности размножения



одомашненных лисиц по сравнению с дикими сородичами. На сегодняшний день открытие Д.К. Беляева уже многократно подтверждено независимыми наблюдениями при одомашнивании, например, оленя, использовании крыс как лабораторных животных, восстановлении таких исчезающих видов, как лошадь Пржевальского, а также при создании резервной популяции дикуши, оказавшейся на грани исчезновения в естественных местах обитания. В результате сравнения геномов человека, ряда домашних животных и некоторых их диких сородичей был введен термин «синдром самодоместикации», в симптомы которого включены расстройства аутистического спектра. Ранее мы создали биоинформатическую модель синдрома самодоместикации человека с использованием дифференциально экспрессируемых генов (ДЭГ) домашних животных по сравнению с дикими сородичами, ортологичных генам преимущественно нервной системы человека, чьи изменения экспрессии могут влиять на репродуктивный потенциал, т.е. рост численности людей при отсутствии ограничений со стороны лимитирующих факторов. В настоящей работе мы применили эту модель к 68 генам человека, изменения экспрессии которых влияют на репродуктивное здоровье женщин и мужчин, и к 3080 ДЭГ животных. Обнаружено 16 и 4 ДЭГ домашних животных, изменения экспрессии которых являются сонаправленными изменениям экспрессии генов-ортологов человека, соответственно понижающим и повышающим его репродуктивный потенциал, тогда как у диких животных было 9 и 11 таких ДЭГ. Это различие было достоверно по критерию  $\chi^2$  Пирсона ( $p < 0.05$ ) и точному критерию Фишера ( $p < 0.05$ ). Полученный результат обсуждается в связи с восстановлением видов животных, исчезающих под антропогенной нагрузкой. Ключевые слова: человек; репродуктивный потенциал; модель болезни человека с использованием животных; доместикация; RNA-Seq; ближайший общий предок.

## Introduction

One of the greatest achievements of genetics in the 20th century was D.K. Belyaev's discovery of destabilizing selection during the domestication of animals and his finding that this selection affects the regulation of gene expression (i. e., specificity and level of expression) but not gene structure. In this context, destabilizing selection directly or indirectly affects systems of neuroendocrine control of ontogenesis when pre-existing stress factors strengthen or new ones emerge in the environment: "In a genetic and biochemical sense, what may be selected for are changes in the regulation of genes – that is, in the timing and the amount of gene expression rather than changes in individual structural genes. Selection having such an effect is called by me destabilizing selection. The selection becomes destabilizing when it affects, directly or indirectly, the systems of neuroendocrine control of ontogenesis. This seems always to be the case when some new stressful factors appear in the environment, or when stresses usual for the species increase in strength." (Belyaev, 1979, p. 307).

This discovery is the result of many years of unique experiments on the domestication of the mink (Belyaev et al., 1972) and fox (Belyaev et al., 1975) as well as the mouse as a laboratory model of human cancer (Belyaev, Gruntenko, 1972). In these experiments, there were findings about accelerated extinction of testes' hormonal function (Osadchuk et al., 1978a) and disturbances in reproduction seasonality (Osadchuk et al., 1978b) of domesticated foxes versus wild foxes; these experiments were conducted with the participation of a coauthor of the present study.

Subsequent comparative analysis of reproductive indices of domesticated foxes versus wild ones (taken as the norm) revealed decreases in the activity indicators of the female endocrine system (Osadchuk, 1992a), in sexual activity of first-year males (Osadchuk, 1992b, 2006), in embryonic gonad mass, and in developmental heterochrony of their pituitary-testicular axis (Osadchuk, 1998) as evidence of destabilizing selection during the domestication of animals (Belyaev, 1979). Additionally, in a laboratory model of animal domestication involving outbred rat strains, a delay in puberty was independently documented in males of a tame strain compared

to an aggressive strain (Prasolova et al., 2014). The results of a comparison among the genomes of humans, numerous domestic animals, and some of their wild congeners have been generalized by the term "self-domestication syndrome", the symptoms of which include autism spectrum disorders (Theofanopoulou et al., 2017), although the idea of human self-domestication is still subject to debate (Del Savio, Marneli, 2020) to this day.

Following a trend in the postgenomic era of life sciences (Qian et al., 2021), we have created a bioinformatic model of self-domestication syndrome using differentially expressed genes (DEGs) – of domestic animals versus their wild congeners – that are orthologous to human genes associated with rheumatoid arthritis (Klimova et al., 2021) and with reproductive potential (Vasiliev et al., 2021), i. e., with an increase in the number of humans when there are no restrictions caused by limiting factors (Chapman, 1931; Pianka, 1976).

In the present work, we analyzed 68 human genes whose expression changes affect the reproductive health of women (Chadaeva et al., 2018) and men (Ponomarenko et al., 2020). The results are discussed in terms of restoration of animal species that are disappearing under anthropogenic pressure (Esmacili et al., 2019).

## Materials and methods

**The analyzed human genes.** We analyzed 68 human genes in the promoters of which we have previously evaluated candidate SNP markers of changes in the reproductive health of women (Chadaeva et al., 2018) and men (Ponomarenko et al., 2020); the examples are presented in Table 1, and complete descriptions – in Supplementary Material<sup>1</sup>.

For instance, in the promoter of the human *ACKR1* gene (atypical chemokine receptor 1), we previously found SNP rs2814778, which lowers the affinity of TATA-binding protein (TBP) for this promoter (Chadaeva et al., 2018), thereby (Mogno et al., 2010) lowering the expression of this gene (see Table 1, column iii,  $N_{\text{SNP}} = 1$ ). This finding is consistent with independent clinical data on patients carrying rs2814778

<sup>1</sup> Supplementary Material is available in the online version of the paper: [http://vavilov.elpub.ru/jour/manager/files/Suppl\\_Ponomarenko\\_Engl.pdf](http://vavilov.elpub.ru/jour/manager/files/Suppl_Ponomarenko_Engl.pdf)

**Table 1.** Examples of the 68 studied human genes for which a significant effect of an SNP(s) in the binding site for TATA-binding protein (TBP) on the affinity of TBP for the promoter of these genes has been previously documented, as have the effects on the levels of their expression and corresponding changes in the reproductive system of women (Chadaeva et al., 2018) and men (Ponomarenko et al., 2020). The complete list is provided in Supplemental Material

No.	Human gene	Lowered expression (–)		Elevated expression (+)			
		<i>N</i> <sub>SNP</sub> (references)	impact on reproductive system and health (references)	♂♀	<i>N</i> <sub>SNP</sub> (references)	impact on reproductive system and health (references)	♂♀
i	ii	iii	iv	v	vi	vii	viii
1	<i>ACKR1</i>	1 (Chadaeva et al., 2018)	Increased risk of preeclampsia as one of the most pressing problems of modern obstetrics (Velzing-Aarts et al., 2002)	↓	–	According to pathology reports, mortality from atherosclerosis and coronary artery diseases is higher (Hernandez-Aguilera et al., 2020)	↓
18	<i>DNMT1</i>	2 (Chadaeva et al., 2018)	Decitabine as an anticancer drug lowers DNMT1 levels (Awada et al., 2020)	↑	7 (Chadaeva et al., 2018)	In a mouse model of a human disease: impaired fetal brain development under stress (Matrisciano et al., 2013)	↓
49	<i>PLCXD1</i>	15 (Ponomarenko et al., 2020)	Gender-specific increased risk in men of middle (reproductive) age of ischemic stroke (Tian et al., 2012)	↓	35 (Ponomarenko et al., 2020)	Transfection of a vector carrying the <i>PLCXD1</i> gene into human cultured melanoma cells inhibits their growth (Mithani et al., 2011)	↑
68	<i>ZFY</i>	–	In a bovine model of a human disease: asthenozoospermia (Xi et al., 2019)	↓	2 (Ponomarenko et al., 2020)	Higher risk of meiosis stoppage in spermatocytes, leading to their apoptosis and to infertility (Jan et al., 2018)	↓

Note. No. is the ID number of a gene in the full list, sorted alphabetically in Supplementary Material. *N*<sub>SNP</sub> is the number of candidate SNP markers that significantly reduce or increase the affinity of TBP for a promoter of a gene (Chadaeva et al., 2018; Ponomarenko et al., 2020), thus decreasing (–) or increasing (+) its expression (Mogno et al., 2010; Ponomarenko et al., 2010); impact on reproductive system and health: deterioration (↓) or improvement (↑). Genes: *ACKR1* – atypical chemokine receptor 1; *DNMT1* – DNA methyltransferase 1; *PLCXD1* – phosphatidylinositol-specific phospholipase CX domain – containing 1; *ZFY* – Y-linked zinc finger protein.

(Michon et al., 2001; Nalls et al., 2008), and therefore we proposed rs2814778 as a candidate SNP marker of preeclampsia as one of the most pressing problems of modern obstetrics (Velzing-Aarts et al., 2002), which worsens the reproductive health of women (Chadaeva et al., 2018), as indicated by the down arrow (↓) in column v of Table 1. On the other hand, according to pathology reports (Hernandez-Aguilera et al., 2020), an excess of the *ACKR1* protein contributes to increased human mortality from atherosclerosis and other coronary artery diseases (see Table 1, column vii), thus reducing human reproductive potential (see Table 1, column viii).

Another example of a gene studied by us earlier (Ponomarenko et al., 2020), the decrease and increase in expression of which impair the reproductive system of humans, is *ZFY* (located on the Y chromosome) encoding a protein with a zinc finger (see Table 1).

In addition, we previously found two candidate SNP markers, rs758026532 and rs772821225, in the promoter of *DNMT1* encoding human DNA methyltransferase 1 – that reduce *DNMT1* expression (Chadaeva et al., 2018), as does anticancer drug decitabine (Awada et al., 2020), thereby increasing the reproductive potential of people (see Table 1, column v, symbol “↑”). Besides, in the promoter of this gene, we previously found seven candidate SNP markers of *DNMT1* overexpression (Chadaeva et al., 2018), which, according to a mouse model of a human disease (Matrisciano et al., 2013), can cause epigenetic aberrations of fetal brain development

under the influence of stressors, thus impairing the human reproductive system (see Table 1, column viii, symbol “↓”).

Finally, in Table 1, readers can see that the previously studied (Ponomarenko et al., 2020) *PLCXD1* gene (phosphatidylinositol-specific phospholipase CX domain-containing 1) represents a diametrically opposite situation (see Table 1, symbols “↓” and “↑” in columns v and viii, respectively). Indeed, underexpression of this gene is a risk factor for stroke in men of reproductive age (Tian et al., 2012), whereas its overexpression improves human reproductive potential by suppressing the progression of melanomas: some of the deadliest human malignant tumors (Mithani et al., 2011).

As done for the genes *ACKR1*, *DNMT1*, *PLCXD1*, and *ZFY* above, Supplementary Material describes all 68 human genes analyzed in the present study.

**The studied DEGs of domestic versus wild animals.** A total of 3080 DEGs of domestic versus wild animals were analyzed here, which are freely available in the PubMed database (Lu, 2011), as described in Table 2 and characterized by examples in Table 3. At the same time, according to (Klimova et al., 2021; Vasiliev et al., 2021), here, RNA-Seq data were examined in accordance with one of the oldest (Samet, 1985), widely used (Sun et al., 2008; Morozova et al., 2020; Hakizimana et al., 2021), and fundamental (Zhang et al., 2021) concepts of phylogenetic analysis – “most recent common ancestor” (Samet, 1985). In this regard, domestic animals and their wild relatives were studied by means of oppositely

**Table 2.** The analyzed RNA-Seq data on DEGs of domestic vs wild animals available in the PubMed database (Lu, 2011)

No.	Domestic animals	Wild animals	Organ/tissue	No. of DEGs	References
i	ii	iii	iv	v	vi
1	Dog ( <i>Canis familiaris</i> )	Wolf ( <i>C. lupus</i> )	Blood	450	Yang X. et al., 2018
2	Dog ( <i>C. familiaris</i> )	Wolf ( <i>C. lupus</i> )	Frontal cortex	19	Albert et al., 2012
3	Tame fox ( <i>Vulpes vulpes</i> )	Aggressive fox ( <i>V. vulpes</i> )	Pituitary gland	327	Hekman et al., 2018
4	Pig ( <i>Sus scrofa</i> )	Boar ( <i>S. scrofa</i> )	Frontal cortex	61	Albert et al., 2012
5	Pig ( <i>S. scrofa</i> )	Boar ( <i>S. scrofa</i> )	Frontal cortex	34	Long et al., 2018
6	Pig ( <i>S. scrofa</i> )	Boar ( <i>S. scrofa</i> )	Pituitary gland	22	Yang Y. et al., 2018
7	Domestic guinea pig ( <i>Cavia porcellus</i> )	Wild guinea pig ( <i>C. aperea</i> )	Frontal cortex	1174	Albert et al., 2012
8	Domestic rabbit ( <i>Oryctolagus cuniculus domesticus</i> )	Wild rabbit ( <i>O. cuniculus</i> )	Frontal cortex	19	Albert et al., 2012
9	Domestic rabbit ( <i>O. cuniculus domesticus</i> )	Wild rabbit ( <i>O. cuniculus</i> )	Parietal and temporal cortex	216	Sato et al., 2020
10	Domestic rabbit ( <i>O. cuniculus domesticus</i> )	Wild rabbit ( <i>O. cuniculus</i> )	Amygdala	118	Sato et al., 2020
11	Domestic rabbit ( <i>O. cuniculus domesticus</i> )	Wild rabbit ( <i>O. cuniculus</i> )	Hypothalamus	43	Sato et al., 2020
12	Domestic rabbit ( <i>O. cuniculus domesticus</i> )	Wild rabbit ( <i>O. cuniculus</i> )	Hypothalamus	100	Sato et al., 2020
13	Tame rat ( <i>Rattus norvegicus</i> )	Aggressive rat ( <i>R. norvegicus</i> )	Frontal cortex	23	Albert et al., 2012
14	Domestic chickens ( <i>Gallus gallus</i> )	Wild chickens ( <i>G. gallus</i> )	Pituitary gland	474	Fallahshahroudi et al., 2019
Total				3080	

directed equivalent changes in gene expression in comparison with their unknown most recent common ancestor.

For example, the *Ckbl* gene (creatine kinase B-like protein) was characterized in column v of Table 1 by a positive score of 4.33  $\log_2$  units of relative expression in the blood of dogs (*Canis familiaris*) versus wolves (*C. lupus*), as reported by (Yang X. et al., 2018). Therefore, dogs and wolves respectively show increased and decreased expression of this gene as compared to their most recent common ancestor (see Table 3, columns vii and viii). Likewise, a negative score of (–1.55) on the relative expression of *Adm* (adrenomedullin) in the dog's frontal cortex as compared to the wolf (see Table 3, column v) corresponds to decreased and increased expression of this gene in this part of the brain during divergence from their most recent common ancestor (see Table 3, columns vii and viii). A total of 450 DEGs in the blood (Yang X. et al., 2018) and 19 DEGs in the frontal cortex (Albert et al., 2012) of dogs and wolves (see Table 2, column v) were characterized in this way.

The score of (–0.47) on the differential expression of the *Hpd* gene, which encodes 4-hydroxyphenylpyruvate dioxygenase, in the pituitary gland of tame *versus* aggressive foxes *Vulpes vulpes* (Hekman et al., 2018) denotes respectively decreased and increased expression of this gene during divergence from their most recent ancestor (see Table 3).

In addition, positive scores on relative expression of genes *Mdk* (Albert et al., 2012) and *C7* (Long et al., 2018) (respectively encoding midkine and component 7 of the complement system of innate immunity) in the frontal cortex of the pig (*Sus scrofa*) as compared to the boar (*S. scrofa*) indicates their higher expression in the pig than in the boar when these species diverged from their most recent common ancestor (see Table 3). On the contrary, the negative score of (–1.32) for the *Ano3* gene in the pituitary gland of the pig compared to the boar (Yang Y. et al., 2018) denotes respectively a deficiency and an excess of anoctamin 3 (encoded by this gene) in this part of the brain when these species diverged from their most recent common ancestor (see Table 3).

Accordingly, a negative score on the differential expression of the *Agt* gene (angiotensinogen) in the frontal cortex of domestic guinea pigs *Cavia porcellus* relative to wild guinea pigs *C. aperea* (Albert et al., 2012) corresponds to decreased and increased expression of this gene as these animals diverged from their most recent common ancestor (see Table 3, columns v, vii, and viii). Table 3 provides similar examples of description for some of the 3080 DEGs of domestic animals versus their wild congeners, as investigated in this work (groups of all genes are described in Table 2).

**A search for orthologous genes of humans and animals.** For each analyzed DEG of domestic animals versus their wild

**Table 3.** Examples of the studied DEGs of domestic vs wild animals. These DEGs are collectively characterized in Table 2

Animals		RNA-Seq				Change in expression upon divergence of animals from most recent common ancestor		References
domestic	wild	Organ/tissue	DEG	$log_2$	$p$	domestic animals	wild animals	
i	ii	iii	iv	v	vi	vii	viii	ix
Dog	Wolf	Blood	<i>Ckbl</i>	4.33	10 <sup>-3</sup>	Elevated (+)	Lowered (–)	Yang X. et al., 2018
Dog	Wolf	Frontal cortex	<i>Adm</i>	–1.55	10 <sup>-4</sup>	Lowered (–)	Elevated (+)	Albert et al., 2012
Tame fox	Aggressive fox	Pituitary gland	<i>Hpd</i>	–0.47	10 <sup>-5</sup>	Lowered (–)	Elevated (+)	Hekman et al., 2018
Pig	Boar	Fontal cortex	<i>Mdk</i>	1.10	10 <sup>-8</sup>	Elevated (+)	Lowered (–)	Albert et al., 2012
Pig	Boar	Frontal cortex	<i>C7</i>	1.64	10 <sup>-4</sup>	Elevated (+)	Lowered (–)	Long et al., 2018
Pig	Boar	Pituitary gland	<i>Ano3</i>	–1.32	0.05	Lowered (–)	Elevated (+)	Yang Y. et al., 2018
Domestic guinea pig	Wild guinea pig	Frontal cortex	<i>Agt</i>	–1.80	10 <sup>-16</sup>	Lowered (–)	Elevated (+)	Albert et al., 2012
Domestic rabbit	Wild rabbit	Frontal cortex	<i>Gp2</i>	6.36	10 <sup>-3</sup>	Elevated (+)	Lowered (–)	Albert et al., 2012
Domestic rabbit	Wild rabbit	Parietal and temporal cortex	<i>ApoD</i>	1.05	0.05	Elevated (+)	Lowered (–)	Sato et al., 2020
Domestic rabbit	Wild rabbit	Amygdala	<i>Pgk1</i>	–1.22	0.05	Lowered (–)	Elevated (+)	Sato et al., 2020
Domestic rabbit	Wild rabbit	Hypothalamus	<i>Aqp1</i>	–3.46	0.05	Lowered (–)	Elevated (+)	Sato et al., 2020
Domestic rabbit	Wild rabbit	Hippocampus	<i>Irf6</i>	1.66	0.05	Elevated (+)	Lowered (–)	Sato et al., 2020
Tame rat	Aggressive rat	Frontal cortex	<i>Alb</i>	1.64	10 <sup>-6</sup>	Elevated (+)	Lowered (–)	Albert et al., 2012
Domestic chickens	Wild chickens	Pituitary gland	<i>Fst</i>	–1.21	10 <sup>-3</sup>	Lowered (–)	Elevated (+)	Fallahshahroudi et al., 2019

Note.  $log_2$  – expression in domesticated relative to wild animals (in  $log_2$  units);  $p$  – statistical significance as determined by the authors cited in column ix. Genes: *Ckbl* – creatine kinase B-like protein; *Adm* – adrenomedullin; *Hpd* – 4-hydroxyphenylpyruvate dioxygenase; *Mdk* – midkine; *C7* – component 7 of the complement system of innate immunity; *Ano3* – anoctamin 3; *Agt* – angiotensinogen; *Gp2* – glycoprotein 2; *ApoD* – apolipoprotein D; *Pgk1* – phosphoglycerate kinase 1; *Aqp1* – aquaporin 1; *Irf6* – interferon regulatory factor 6; *Alb* – albumin; *Fst* – follistatin.

congeners (see Tables 2 and 3), an orthologous gene was sought among all the 68 studied human genes (see Table 1 and Supplemental Material). If no such orthologous human gene was found, then the animal DEG in question was excluded from further analysis. Otherwise, we collated the effects of codirected changes in the expression of the found orthologous genes on the reproductive potential of humans (see Table 1 and Supplemental Material, columns v and viii) with expression changes during the emergence of a domesticated species or during preservation of the wild species of the respective animal in the microevolution of their most recent common ancestor (see Table 3, columns vii and viii). For example, the *Apoa1* gene (apolipoprotein A1) is characterized by a negative score of (–3.2) on differential expression in domestic versus wild guinea pigs (Albert et al., 2012), indicating decreased and increased expression of this gene, respectively, in the process of their divergence from their most recent common ancestor (Table 4, columns ii, iv, and vi). Accordingly, underexpression of a human orthologous gene, *APOA1*, was clinically associated with a predisposition to cognitive disorders (Peng et al., 2017), whereas its overexpression correlates with infertility in women (Manohar et al., 2014), as illustrated in columns vii

and ix of Table 4. Thus, a deficiency and excess of *APOA1* in humans impair the reproductive system of humans (see Table 4, columns viii and x). In the present study, within the framework of the previously proposed bioinformatic model of human diseases involving DEGs of domestic versus wild animals (Klimova et al., 2021; Vasiliev et al., 2021), all of the above means that the expression changes of *Apoa1* during the divergence of domestic and wild guinea pigs from their most recent common ancestor correspond to a negative impact of expression changes of the human orthologous gene *APOA1* on human reproductive potential. Similarly, the *CETP* gene encoding cholesteryl ester transfer protein is overexpressed in hypercholesterolemia of pregnancy (Silliman et al., 1993), thereby impairing the reproductive health of women (see Table 4, columns ix and x). The excess of *CETP* in humans is consistent with an excess of *Cetp* in the domestic guinea pig during its divergence from the most recent common ancestor with the wild guinea pig (Albert et al., 2012), as shown in Table 4 (columns ii and iv). By contrast, a *CETP* deficiency in humans is a clinically proven marker of slowing atherogenesis as well as lower risks of stroke and myocardial infarction (Plengpanich et al., 2011); these correlations can be



**Table 4.** A comparison between the effects of expression changes of human orthologous genes on reproductive potential and expression changes during the divergence of domestic and wild animals from their most recent common ancestor

Animals					Humans				
RNA-Seq		Change in expression upon divergence of animals from most recent common ancestor			Gene	Effect of changed gene expression on reproductive potential (♂♀)			
DEG	$\log_2$	$p$	domestic	wild		low expression (-)	♂♀	high expression (+)	♂♀
i	ii	iii	iv	v	vi	vii	viii	ix	x
Domestic and wild guinea pigs (Albert et al., 2012)									
<i>Apoa1</i>	-3.2	$10^{-2}$	Lowered (-)	Elevated (+)	<i>APOA1</i>	Higher risk of cognitive impairment (Peng et al., 2017)	↓	Infertility in women (Manohar et al., 2014)	↓
<i>Cetp</i>	2.1	$10^{-4}$	Elevated (+)	Lowered (-)	<i>CETP</i>	Slowing of atherogenesis thus preventing stroke and heart attack (Plengpanich et al., 2011)	↑	In pregnancy, the risk of hypercholesterolemia is higher (Silliman et al., 1993)	↓
<i>Cyp17a1</i>	-1.1	$10^{-2}$	Lowered (-)	Elevated (+)	<i>CYP17A1</i>	Higher risk of impaired fertility (Marsh, Auchus, 2014)	↓	Malaysian propolis upregulates CYP17A1 thereby correcting subfertility (Nna et al., 2020)	↑
<i>Gcg</i>	3.0	$10^{-2}$	Elevated (+)	Lowered (-)	<i>GCG</i>	Lower pregnancy rates (Sugiyama et al., 2012)	↓	Lower pregnancy rates (Sun et al., 2019)	↓
<i>Il1b</i>	2.3	$10^{-4}$	Elevated (+)	Lowered (-)	<i>IL1B</i>	No bone deformity during bacterial invasion (Sasaki et al., 2020)	↑	Higher circadian hypersensitivity to pain (Oikkonen et al., 2015)	↓
<i>Nr5a1</i>	-2.2	$10^{-3}$	Lowered (-)	Elevated (+)	<i>NR5A1</i>	Gonadal dysgenesis (Nagaraja et al., 2019)	↓	Better Sertoli cell differentiation and sperm quality (Wood et al., 2011)	↑
<i>Proc</i>	1.8	$10^{-2}$	Elevated (+)	Lowered (-)	<i>PROC</i>	Higher risk of deadly purpura fulminans in newborns (Dinarvand, Moser, 2019)	↓	Higher risk of miscarriage (Lay et al., 2005)	↓
Dog and wolf (Yang X. et al., 2018)									
<i>Gh1</i>	2.8	$10^{-4}$	Elevated (+)	Lowered (-)	<i>GH1</i>	Higher mortality from cardiovascular pathologies (Jorgensen, Juul, 2018)	↓	Growth hormone prolongs reproductive age of women (Regan et al., 2018)	↑
<i>Hba1</i>	-4.1	$10^{-7}$	Lowered (-)	Elevated (+)	<i>HBB</i>	Thalassemia impairs women's reproductive health (Takhviji et al., 2020)	↓	In Chinese medicine, Jian-Pi-Yi-Shen decoction increases hemoglobin levels, relieving chronic anemia (Wang et al., 2020)	↑
<i>Hbb1</i>	-5.9	$10^{-11}$	Lowered (-)	Elevated (+)					
<i>Hbm</i>	-6.5	$10^{-9}$	Lowered (-)	Elevated (+)					
<i>Hbz1</i>	-7.1	$10^{-4}$	Lowered (-)	Elevated (+)					
Tame and aggressive foxes (Hekman et al., 2018)									
<i>Esr2</i>	-0.3	$10^{-2}$	Lowered (-)	Elevated (+)	<i>ESR2</i>	ESR2 deficiency in adolescents reduces sperm quality in adults (Ivanski et al., 2020)	↓	Excess ESR2 in adolescents reduces sperm quality in adults (Ivanski et al., 2020)	↓
<i>Il9r</i>	0.4	$10^{-5}$	Elevated (+)	Lowered (-)	<i>IL9R</i>	Impaired trophoblast Implantation (Sun et al., 2020)	↓	Higher risk of fatal anaphylactic shock (Osterfeld et al., 2010)	↓
Domestic and wild rabbits (Albert et al., 2012)									
<i>F7</i>	-0.3	$10^{-2}$	Lowered (-)	Elevated (+)	<i>F7</i>	Spontaneous difficult-to-stop life-threatening hemorrhages (Senol, Zulfikar, 2020)	↓	Exogenous F7 is a life-saving drug for obstetric bleeding (Burad et al., 2012)	↑

Table 4 (end)

Animals					Humans				
RNA-Seq		Change in expression upon divergence of animals from most recent common ancestor			Gene	Effect of changed gene expression on reproductive potential (♂♀)			
DEG	<i>log</i> <sub>2</sub>	<i>p</i>	domestic	wild		low expression (–)	♂♀	high expression (+)	♂♀
i	ii	iii	iv	v	vi	vii	viii	ix	x
Domestic and wild chickens (Fallahshahroudi et al., 2019)									
<i>Cyp17a1</i>	0.6	10 <sup>–9</sup>	Elevated (+)	Lowered (–)	<i>CYP17A1</i>	Higher risk of impaired fertility (Marsh, Auchus, 2014)	↓	Malaysian propolis upregulates CYP17A1 thereby correcting subfertility (Nna et al., 2020)	↑
<i>F3</i>	0.8	10 <sup>–4</sup>	Elevated (+)	Lowered (–)	<i>F3</i>	Ozone therapy downregulates F3 for the treatment of thromboischemic bowel injury (Yu et al., 2020)	↑	Higher risk of stroke and myocardial infarction (Arnaud et al., 2000)	↓
<i>Hbad</i>	–1.1	10 <sup>–2</sup>	Lowered (–)	Elevated (+)	<i>HBD</i>	Thalassemia impairs women’s reproductive health (Takhviji et al., 2020)	↓	In Chinese medicine, Jian-Pi-Yi-Shen decoction upregulates hemoglobin alleviating chronic anemia (Wang et al., 2020)	↑
<i>Pgr</i>	1.3	10 <sup>–6</sup>	Elevated (+)	Lowered (–)	<i>PGR</i>	Infertility due to sexual-behavior aberrations (Kubota et al., 2016)	↓	Increased fertility (Yao et al., 2020)	↑
<i>Slc25a6</i>	0.5	10 <sup>–5</sup>	Elevated (+)	Lowered (–)	<i>SLC25A6</i>	Higher risk of muscle dystrophy (Clemencon et al., 2013)	↓	Higher resistance to the herpes virus (Guo et al., 2015)	↑

Note. See the footnote of Table 3. Genes: *Apoa1* – apolipoprotein A1; *Cetp* – cholesteryl ester transfer protein; *Cyp17a1* – steroid 17α-monooxygenase; *Gcg* – glucagon; *Il1b* – interleukin 1β; *Nr5a1* – steroidogenic factor 1; *F3*, *F7*, and *Proc* – blood coagulation factors III, VII, and XIV, respectively; *Gh1* – growth hormone; *HBD*, *Hba1*, *Hbad*, *Hbbl*, *Hbm*, and *Hbz1* are hemoglobin subunits δ, α1, αD, β-like, μ, and ζ1, respectively; *Esr2* – estrogen receptor 2; *Il9r* – interleukin 9 receptor; *Pgr* – progesterone receptor; *Slc25a6* – mitochondrial solute transporter.

regarded as factors increasing human reproductive potential (see Table 4, columns vii and viii). *CETP* downregulation in humans is consistent with *Cetp* downregulation in the wild guinea pig when it diverged with the domestic guinea pig from the most recent common ancestor (Albert et al., 2012) (see Table 4, column v).

Finally, the human *CYP17A1* gene produces steroid 17α-monooxygenase, underexpression of which impairs fertility in humans (Marsh, Auchus, 2014), thereby reducing their reproductive potential, as displayed in Table 4. The deficiency of CYP17A1 in humans is consistent with the deficiency of *Cyp17a1* in the domestic guinea pig (Albert et al., 2012) and in wild chickens *Gallus gallus* (Fallahshahroudi et al., 2019) when domestic and wild forms of these animals diverged from their respective most recent common ancestors (see Table 4, columns ii and iv). On the contrary, a CYP17A1 excess in humans overcomes subfertility (Nna et al., 2020), thus increasing human reproductive potential (see Table 4, columns ix and x). This influence is consistent with higher expression of the orthologous *Cyp17a1* gene in the wild guinea pig and domestic chicken as compared with this gene’s expression during their microevolution from the corresponding most recent common (see Table 4).

In Table 4, the reader can find similar descriptions for all the human and animal orthologous genes that we identified among the 68 human genes under study (see Table 1 and Supplemental Material) and among the 3080 DEGs of domestic animals versus their wild congeners (see Tables 2 and 3). In this context, it is noteworthy that because of the concept of “divergence from the most recent common ancestor,” it was possible to compare phenotypic manifestations of increased and decreased expression of human genes (see Table 1, columns v and viii; Table 4, columns viii and x) with changes in the expression of respective orthologous genes in domestic and wild animals as they diverged from their most recent common ancestor (see Table 3, columns v and vi; Table 4, columns iv and v).

**Knowledge base PetDEGsDB on human diseases as candidate symptoms of self-domestication syndrome.** Identified here as the main finding, the matches – between the effects of changed expression of human genes on human reproductive potential and expression changes of orthologous animal genes during the divergence of domestic and wild animals from their most recent common ancestors – were compiled into a flat text Excel-compatible file and were finally transformed in the MariaDB 10.2.12 Web environment (MariaDB Corp AB,

Espoo, Finland) into a knowledge base, named PetDEGsDB, on human diseases that are candidates for self-domestication syndrome (Vasiliev et al., 2021). This knowledge base is freely available at <https://www.sysbio.ru/domestic-wild>.

**Statistical analysis.** The correspondences (see Table 4) between the phenotypic manifestations of codirected changes in the expression of orthologous genes of humans and animals were summarized in a standard Fisher 2×2 table represented by intersections of the rows “domestic animals” and “wild animals” (Table 5, columns iii and iv). This Fisher 2×2 table was analyzed using the Statistica package (Statsoft™, Tulsa, USA); its operating mode was chosen via the sequence of commands Statistics → Nonparametrics → 2×2 Table”, which enabled us to perform a binomial distribution analysis, Fisher’s exact test, and Pearson’s  $\chi^2$  test (see Table 5, columns v, vi, vii, and viii).

## Results and discussion

In this work, we examined 68 human genes (see Table 1 and Supplemental Material) and 3080 DEGs of domestic animals versus wild congeners (see Tables 2 and 3), which are described in the “Materials and methods” section. As a result of the technique described in the subsection “A search for orthologous genes of humans and animals” (Materials and methods), 20 animal DEGs were found that turned out to be orthologous to the studied human genes, as presented in Table 4 and described in the “Materials and methods,” with human genes *APOA1*, *CETP*, and *CYP17A1* as examples. Let us review the identified orthologous genes of humans and animals.

The human *CGC* gene codes for glucagon; both a deficiency (Sugiyama et al., 2012) and an excess (Sun et al., 2019) of this protein are clinically proven markers of a reduced pregnancy rate and hence impairment of the reproductive system in humans (see Table 4). Upregulation and downregulation of glucagon in humans are consistent with increased and decreased expression of *Gcg* in domestic and wild guinea pigs (Albert et al., 2012) during their divergence from their most recent common ancestor.

The *IL1B* gene codes for interleukin 1 $\beta$ . An excess of this interleukin increases circadian sensitivity to pain (Olkkonen et al., 2015), thereby reducing human reproductive potential (see Table 4). By contrast, IL1B deficiency prevents bone deformation during bacterial invasion (Sasaki et al., 2020), thus expectedly increasing human reproductive potential (see Table 4). The excess and deficiency of IL1B in humans are the expression changes codirected with the upregulation and downregulation of *Il1b* in the wild guinea pig during its divergence with the domestic guinea pig from a common ancestor (Albert et al., 2012).

The *NR5A1* gene encoding human steroidogenic factor 1 is characterized by underexpression in gonadal dysgenesis (Nagaraja et al., 2019), which reduces human reproductive potential (see Table 4), whereas overexpression of the NR5A1 protein improves sperm quality (Wood et al., 2011). Both the NR5A1 deficiency and excess in humans are consistent with the decreased and increased expression of *Nr5a1* in the domestic guinea pig in the process of divergence with the wild guinea pig from a common ancestor (Albert et al., 2012).

The *PROC* gene represents human coagulation factor XIV, a deficiency of which in neonates can cause deadly purpura fulminans (Dinarvand, Moser, 2019), whereas its overexpression increases miscarriage risk (Lay et al., 2005). These alterations of *PROC* expression are in agreement with the decreased and increased expression of *Proc* in wild and domestic guinea pigs (Albert et al., 2012) during their microevolution (see Table 4).

The *GHI* gene codes for growth hormone, which increases the reproductive potential of women (Regan et al., 2018). The excess of GH1 in humans is similar to the excess of Gh1 in dogs (*C. familiaris*) when compared to the most recent common ancestor of dogs and wolves (*C. lupus*) (Yang X. et al., 2018). GH1 deficiency increases human mortality from cardiovascular disease (Jorgensen, Juul, 2018) in line with Gh1 deficiency in wolves during their microevolution.

Genes *HBB* and *HBD* encode hemoglobin subunits  $\beta$  and  $\delta$ . Their deficiency is associated with thalassemia, a contributing factor of poor reproductive potential in women (Takhviji et al., 2020). Human hemoglobin deficiency is consistent with hemoglobin underexpression in dogs (Yang X. et al., 2018) and domestic chickens (Fallahshahroudi et al., 2019) when compared with the most recent common ancestors for their wild counterparts (see Table 4). Conversely, an excess of hemoglobin in humans is in agreement with overexpression of hemoglobin in wolves and wild chickens (see Table 4).

The human *ESR2* gene (estrogen receptor 2) – both in the case of underexpression in adolescents and in the case of its overexpression in this segment of the population – was associated with decreased sperm quality in adults (Ivanski et al., 2020). These alterations of its expression in humans are consistent with those of an orthologous gene, *Esr2*, in tame and aggressive foxes (Hekman et al., 2018) during their microevolution (see Table 4).

The *IL9R* gene encodes human interleukin 9 receptor, the deficiency of which disrupts trophoblast implantation (Sun et al., 2020), whereas its excess contributes to deadly anaphylactic shock (Osterfeld et al., 2010). The upregulation and downregulation of this receptor in humans are consistent with increased and decreased expression of the *Il9r* gene in tame and aggressive foxes (Hekman et al., 2018) as they diverged from their most recent common ancestor (see Table 4).

The *F7* gene encodes proconvertin. Its recombinant activated form is used as an emergency life-saving modality against obstetric bleeding (Buraud et al., 2012). Upregulation of F7 in humans is consistent with that of its ortholog in wild rabbits in the process of divergence with domestic rabbits from a common ancestor (Albert et al., 2012). A proconvertin deficiency accompanies spontaneous life-threatening bleeding (Senol, Zulfikar, 2020) and is consistent with F7 deficiency in domestic rabbits (see Table 2).

The *F3* gene (thromboplastin) is overexpressed in stroke and myocardial infarction (Arnaud et al., 2000) and thus may reduce human reproductive potential (see Table 4). An excess of F3 in humans is consistent with an excess of F3 in domestic chickens (Fallahshahroudi et al., 2019). On the other hand, thromboplastin deficiency contributes to an increase in human reproductive potential (Yu et al., 2020), in agreement with F3 deficiency in wild chickens during their divergence with domestic chickens from the most recent common ancestor.

**Table 5.** Significant matches between the effects of codirected changes in the expression of human orthologous genes on reproductive potential and gene expression changes during the divergence of wild and domestic animals from their most recent common ancestor

Animals		Humans		Statistics			
		Changes in gene expression that cause changes in reproductive potential (♂♀)		Binomial distribution analysis	Pearson $\chi^2$ -test		Fisher's exact test
		worse (↓)	better (↑)		$p$	$\chi^2$	$p$
i	ii	iii	iv	v	vi	vii	viii
Change in DEG expression during divergence from most recent common ancestor	Domestic	16	4	0.01	5.2	0.05	0.05
	Wild	9	11	> 0.4			

The *PGR* gene codes for progesterone receptor. A human disease model based on *Pgr* knockout rats features infertility due to impaired sexual behavior (Kubota et al., 2016). PGR deficiency in humans is codirected with Pgr deficiency in wild chickens during their divergence from a common ancestor with domestic chickens (Fallahshahroudi et al., 2019). A human fertility model based on ewes revealed a positive correlation between Pgr and fertility (Yao et al., 2020). Upregulation of PGR in humans is consistent with Pgr overexpression in domestic chickens as a consequence of their selection by humans for egg production (see Table 4).

The *SLC25A6* gene encodes human steroidogenic factor 1. Its overexpression correlates with resistance to the herpes virus (Guo et al., 2015), in line with *Slc25a6* overexpression in domestic chickens compared to their most recent common ancestor with wild chickens (Fallahshahroudi et al., 2019). An *SLC25A6* deficiency is accompanied by an increased risk of muscle dystrophy (Clemencon et al., 2013) in agreement with the *Slc25a6* underexpression in wild chickens as compared to their most recent common ancestor with domestic chickens selected for muscle growth by humans.

All the results of this study are summarized in Table 5, where we present domestic animals' 16 and 4 DEGs the changes in expression of which are consistently codirected with changes in the expression of the orthologous genes in humans that respectively decrease and increase human reproductive potential. By contrast, in the wild animals, there were 9 and 11 such DEGs, respectively (almost equal numbers of oppositely acting DEGs). This difference between wild and domestic animals is statistically significant according to Pearson's  $\chi^2$  test ( $p < 0.05$ ) and Fisher's exact test ( $p < 0.05$ ). Finally, the binomial distribution analysis ( $p < 0.01$ ) indicates that the anthropogenic living conditions of animals during their domestication usually alter gene expression in a direction corresponding to the expression changes of human orthologous genes that decrease reproductive potential.

On the contrary, microevolution of wild animals in a natural habitat has changed the expression of genes equally often in the directions that either decrease or increase reproductive potential, judging from expression changes of respective human orthologous genes (binomial distribution:  $p > 0.4$ ). This finding is in agreement with the generally accepted choice of the wild type as the norm.

While discussing this result, we should note, first of all, that in laboratory animal models of human diseases, DEGs

are usually detected in inbred strains having symptoms of a disease in comparison with outbred strains as the norm (Fedoseeva et al., 2019).

Nevertheless, in the literature, we were unable to find unequivocal evidence that codirected changes in the expression of orthologous genes cause similar pathologies in humans and animals, probably owing to different genetic contexts of these changes in different species.

Among parameters of the harmful anthropogenic impact on animal populations, a decrease in their effective size is often mentioned, which promotes their inbreeding, which in turn negatively correlates with sperm quality, for example, in the domestic cat *Felis catus* (Pukazhenthil et al., 2006), deer *Cervus elaphus* (Gomendio et al., 2007), and finch *Taeniopygia guttata* (Forstmeier et al., 2017) as well as in Mexican wolves (*Canis lupus baileyi*), which disappeared from the wild in the 20th century and exist only as part of a program for their restoration and reintroduction into their former habitats (Asa et al., 2007).

When endangered cranes *Grus americana* are reintroduced, a high degree of inbreeding of their *ex situ* population (~400 individuals) delays the onset of reproduction, and as a consequence, decreases egg production; this problem is expected to be overcome by sperm cryopreservation and artificial insemination (Songsasen et al., 2019).

For the feline family Felidae, sperm cryopreservation and artificial insemination have already been successfully implemented for the reintroduction of the endangered wild cat *Prionailurus bengalensis euptilurus* (Amstislavsky et al., 2018). The creation of protected areas for natural habitats of the Amur tiger *Panthera tigris altaica* has contributed to the restoration of its population (Xiao et al., 2016). Due to an anthropogenic reduction in the geographic range of the Florida cougar *Puma concolor coryi*, only ~20 individuals are left. On the basis of theoretical populational calculations (Hedrick, 1995), individuals of the closely related Texas cougar *P. concolor cougar* were transported to restore this species, thereby ensuring the success of the reintroduction (Hedrick, 2010).

Crossing of subspecies has facilitated the reintroduction of Przewalski's horses *Equus caballus przewalskii*, which disappeared from the wild half a century ago (Der Sarkissian et al., 2015).

As a continuation of these successes, we can cite examples of the comparison of genomic diversity of inbred with out-



bred populations of the bull *Bos taurus*, comparisons of F<sub>1</sub> descendants (from crosses between them) and descendants of F<sub>1</sub> backcrosses with parental populations, as well as similar comparisons for the bison (*Bison bison*). The results of these studies independently confirm the finding of a decrease in the inbreeding degree when inbred strains of animals are crossed with their outbred relatives (Cronin, Leesburg, 2016). Finally, through the deciphering of the genome in the Austrian Fleckvieh bull *Bos (primigenius) taurus*, geographic locations influencing sperm quality were identified, and interbreeding options were found that improve this quality (Ferencakovic et al., 2017).

An increase in mortality from infections, as, for example, at the beginning of the reintroduction of Przewalski's horses, is a much less studied parameter of the negative anthropogenic impact on animal populations (Robert et al., 2005).

Besides, during the creation of a reserve population of the Siberian grouse *Falcipennis falcipennis*, which had been on the verge of extinction in natural habitats, the intestinal microbiota of these birds changed, acting as a stressor of the immune system (Konyaev et al., 2013). An analysis of phylogenetic inertia of the infection–host network revealed an increase in the number of common infections of humans and domestic animals with the growing number of new tamed animals; this increase may be an epidemiological bridge connecting the anthropogenic environment with wildlife (Morand et al., 2014).

Finally, a possible counterargument to the above notion of a decrease in the reproductive potential of animals under the influence of humans is the domestic pig, which surpasses the wild boar in sperm quality (Almeida et al., 2006). The reason is selection for fertility for the sake of meat. Another counterargument is an increased proportion of females among domestic chickens in comparison with wild chickens as a consequence of selection for egg production (Zhang et al., 2020).

All of the above means that the decrease in reproductive potential during the domestication of new economically valuable species of animals (for example, the Asiatic wild ass *Equus hemionus hemionus* (Soilemetzidou et al., 2020)) can be compensated either by artificial selection for fertility in addition to the main desired trait or through interbreed crosses. When natural habitats of wild animals are included into economic land rotation by humans, an inbreeding-related diminution of their reproductive potential takes place (up to extinction), which can be compensated by subspecies crossings of these animals and by methods of assisted reproductive technology.

## Conclusion

We examined 68 human genes (see Table 1 and Supplemental Material) and 3080 DEGs of domestic animals versus their wild congeners. We found that the anthropogenic impact during the domestication of animals usually changes the expression of their genes in the same direction as seen in the expression alterations of orthologous human genes that worsen reproductive potential. By contrast, the natural habitat of wild animals maintains the intraspecific variation of expression of their genes in a way that equally corresponds to decreases and increases of reproductive potential in people, according to the expression alterations of the orthologous human genes.

## References

- Albert F.W., Somel M., Carneiro M., Aximu-Petri A., Halbwax M., Thalmann O., Blanco-Aguilar J.A., Plyusnina I.Z., Trut L., Villafuerte R., Ferrand N., Kaiser S., Jensen P., Paabo S. A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet.* 2012;8(9):e1002962. DOI 10.1371/journal.pgen.1002962.
- Almeida F.F., Leal M.C., França L.R. Testis morphometry, duration of spermatogenesis, and spermatogenic efficiency in the wild boar (*Sus scrofa scrofa*). *Biol. Reprod.* 2006;75(5):792-799. DOI 10.1095/biolreprod.106.053835.
- Amstislavsky S., Brusentsev E., Kizilova E., Mokrousova V., Kozhevnikova V., Abramova T., Rozhkova I., Naidenko S. Sperm cryopreservation in the Far-Eastern wildcat (*Prionailurus bengalensis euptilurus*). *Reprod. Domest. Anim.* 2018;53(5):1219-1226. DOI 10.1111/rda.13230.
- Arnaud E., Barbalat V., Nicaud V., Cambien F., Evans A., Morrison C., Arveiler D., Luc G., Ruidavets J.B., Emmerich J., Fiessinger J.N., Aiach M. Polymorphisms in the 5' regulatory region of the tissue factor gene and the risk of myocardial infarction and venous thromboembolism: the ECTIM and PATHROS studies. *Arterioscler. Thromb. Vasc. Biol.* 2000;20(3):892-898. DOI 10.1161/01.atv.20.3.892.
- Asa C., Miller P., Agnew M., Rebollo J.A.R., Lindsey S.L., Callahan M., Bauman K. Relationship of inbreeding with sperm quality and reproductive success in Mexican gray wolves. *Anim. Conserv.* 2007;10(3):326-331. DOI 10.1111/j.1469-1795.2007.00116.x.
- Awada H., Mahfouz R.Z., Kishtagari A., Kuzmanovic T., Durrani J., Kerr C.M., Patel B.J., Visconte V., Radivoyevitch T., Lichtin A., Carraway H.E., Maciejewski J.P., Saunthararajah Y. Extended experience with a non-cytotoxic DNMT1-targeting regimen of decitabine to treat myeloid malignancies. *Br. J. Haematol.* 2020;188(6):924-929. DOI 10.1111/bjh.16281.
- Belyaev D.K. The Wilhelmine E. Key 1978 invitational lecture. Destabilizing selection as a factor in domestication. *J. Hered.* 1979;70(5):301-308. DOI 10.1093/oxfordjournals.jhered.a109263.
- Belyaev D.K., Evsikov V.I., Matysko E.K. Genetics of animal fertility: 3. Effect of monohybrid heterosis on the fertility and viability of minks and prospects of its use in breeding. *Sov. Genet.* 1972;8(1):46-51.
- Belyaev D.K., Gruntenko E.V. Strain differences in thymus weight in mice with different predispositions to spontaneous mammary cancer. *Nature.* 1972;237(5355):401-402. DOI 10.1038/237401a0.
- Belyaev D.K., Trut L.N., Ruvinsky A.O. Genetics of the *W* locus in foxes and expression of its lethal effects. *J. Hered.* 1975;66(6):331-338. DOI 10.1093/oxfordjournals.jhered.a108643.
- Burad J., Bhakta P., Sharma J. Timely 'off-label' use of recombinant activated factor VII (NovoSeven®) can help in avoiding hysterectomy in intractable obstetric bleeding complicated with disseminated intravascular coagulation: A case report and review of the literature. *Indian J. Anaesth.* 2012;56(1):69-71. DOI 10.4103/0019-5049.93349.
- Chadaeva I.V., Ponomarenko P.M., Rasskazov D.A., Sharypova E.B., Kashina E.V., Zhechev D.A., Drachkova I.A., Arkova O.V., Savinkova L.K., Ponomarenko M.P., Kolchanov N.A., Osadchuk L.V., Osadchuk A.V. Candidate SNP markers of reproductive potential are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics.* 2018;19(Suppl. 3). DOI 10.1186/s12864-018-4478-3.
- Chapman R.N. Animal Ecology with Special Reference to Insects. New York: McGraw-Hill, 1931.
- Clemençon B., Babot M., Trezeguet V. The mitochondrial ADP/ATP carrier (SLC25 family): pathological implications of its dysfunction. *Mol. Aspects Med.* 2013;34(2-3):485-493. DOI 10.1016/j.mam.2012.05.006.
- Cronin M.A., Leesburg V.L. Genetic variation and differentiation in parent-descendant cattle and bison populations. *J. Anim. Sci.* 2016;94(11):4491-4497. DOI 10.2527/jas.2016-0476.

- Del Savio L., Mameli M. Human domestication and the roles of human agency in human evolution. *Hist. Philos. Life Sci.* 2020;42(2):21. DOI 10.1007/s40656-020-00315-0.
- Der Sarkissian C., Ermini L., Schubert M., Yang M.A., Librado P., Fumagalli M., Jonsson H., Bar-Gal G.K., Albrechtsen A., Vieira F.G., Petersen B., Ginolhac A., Seguin-Orlando A., Magnussen K., Fages A., Gamba C., Lorente-Galdos B., Polani S., Steiner C., Neuditschko M., Jagannathan V., Feh C., Greenblatt C.L., Ludwig A., Abramson N.I., Zimmermann W., Schafberg R., Tikhonov A., Sicheritz-Ponten T., Willerslev E., Marques-Bonet T., Ryder O.A., McCue M., Rieder S., Leeb T., Slatkin M., Orlando L. Evolutionary genomics and conservation of the endangered Przewalski's Horse. *Curr. Biol.* 2015;25(19):2577-2583. DOI 10.1016/j.cub.2015.08.032.
- Dinarvand P., Moser K.A. Protein C deficiency. *Arch. Pathol. Lab. Med.* 2019;143(10):1281-1285. DOI 10.5858/arpa.2017-0403-rs.
- Esmacili S., Hemami M.R., Goheen J.R. Human dimensions of wild-life conservation in Iran: Assessment of human-wildlife conflict in restoring a wide-ranging endangered species. *PLoS One.* 2019;14(8):e0220702. DOI 10.1371/journal.pone.0220702.
- Fallahshahroudi A., Lotvedt P., Belteky J., Altimiras J., Jensen P. Changes in pituitary gene expression may underlie multiple domesticated traits in chickens. *Heredity (Edinb.)*. 2019;122(2):195-204. DOI 10.1038/s41437-018-0092-z.
- Fedoseeva L.A., Klimov L.O., Ershov N.I., Efimov V.M., Markel A.L., Orlov Y.L., Redina O.E. The differences in brain stem transcriptional profiling in hypertensive ISIAH and normotensive WAG rats. *BMC Genomics.* 2019;20(Suppl. 3):297. DOI 10.1186/s12864-019-5540-5.
- Ferencakovic M., Solkner J., Kaps M., Curik I. Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. *J. Dairy Sci.* 2017;100(6):4721-4730. DOI 10.3168/jds.2016-12164.
- Forstmeier W., Ihle M., Opatova P., Martin K., Knief U., Albrechtova J., Albrecht T., Kempenaers B. Testing the phenotype-linked fertility hypothesis in the presence and absence of inbreeding. *J. Evol. Biol.* 2017;30(5):968-976. DOI 10.1111/jeb.13062.
- Gomendio M., Malo A.F., Garde J., Roldan E.R. Sperm traits and male fertility in natural populations. *Reproduction.* 2007;134(1):19-29. DOI 10.1530/rep-07-0143.
- Guo X., Huang Y., Qi Y., Liu Z., Ma Y., Shao Y., Jiang S., Sun Z., Ruan Q. Human cytomegalovirus miR-UL36-5p inhibits apoptosis via downregulation of adenine nucleotide translocator 3 in cultured cells. *Arch. Virol.* 2015;160(10):2483-2490. DOI 10.1007/s00705-015-2498-8.
- Hakizimana J.N., Yona C., Kamana O., Nauwynck H., Misinzo G. African swine fever virus circulation between Tanzania and neighboring countries: a systematic review and meta-analysis. *Viruses.* 2021;13(2):306. DOI 10.3390/v13020306.
- Hedrick P. Genetic future for Florida panthers. *Science.* 2010;330(6012):1744. DOI 10.1126/science.330.6012.1744-a.
- Hedrick P.W. Gene flow and genetic restoration: the Florida panther as a case study. *Conserv. Biol.* 1995;9(5):996-1007. DOI 10.1046/j.1523-1739.1995.9050988.x-11.
- Hekman J.P., Johnson J.L., Edwards W., Vladimirova A.V., Gulevich R.G., Ford A.L., Kharlamova A.V., Herbeck Y., Acland G.M., Raetzman L.T., Trut L.N., Kukekova A.V. Anterior pituitary transcriptome suggests differences in ACTH release in tame and aggressive foxes. *G3 (Bethesda).* 2018;8(3):859-873. DOI 10.1534/g3.117.300508.
- Hernandez-Aguilera A., Fibla M., Cabre N., Luciano-Mateo F., Camps J., Fernandez-Arroyo S., Martin-Paredero V., Menendez J.A., Sirvent J.J., Joven J. Chemokine (C-C motif) ligand 2 and coronary artery disease: tissue expression of functional and atypical receptors. *Cytokine.* 2020;126:154923. DOI 10.1016/j.cyto.2019.154923.
- Ivanski F., de Oliveira V.M., de Oliveira I.M., de Araújo Ramos A.T., de Oliveira Tonete S.T., de Oliveira Hykavei G., Bargi-Souza P., Schiessel D.L., Martino-Andrade A.J., Romano M.A., Marino Romano R. Prepubertal acrylamide exposure causes dose-response decreases in spermatid production and functionality with modulation of genes involved in the spermatogenesis in rats. *Toxicology.* 2020;436:152428. DOI 10.1016/j.tox.2020.152428.
- Jan S.Z., Jongejan A., Korver C.M., van Daalen S.K.M., van Pelt A.M.M., Repping S., Hamer G. Distinct prophase arrest mechanisms in human male meiosis. *Development.* 2018;145(16):dev160614. DOI 10.1242/dev.160614.
- Jorgensen J.O.L., Juul A. Therapy of endocrine disease: growth hormone replacement therapy in adults: 30 years of personal clinical experience. *Eur. J. Endocrinol.* 2018;179(1):R47-R56. DOI 10.1530/EJE-18-0306.
- Klimova N.V., Oshchepkova E., Chadaeva I., Sharypova E., Ponomarenko P., Drachkova I., Rasskazov D., Oshchepkov D., Ponomarenko M., Savinkova L., Kolchanov N.A., Kozlov V. Disruptive selection of human immunostimulatory and immunosuppressive genes both provokes and prevents rheumatoid arthritis, respectively, as a self-domestication syndrome. *Front. Genet.* 2021;12:610774. DOI 10.3389/fgene.2021.610774.
- Konyaev S.V., Klimova S.N., Shilo V.A. Invasions of wild birds from the order Galliformes bred in captivity. *Rossiyskiy Veterinarnyy Zhurnal. Melkiye Domashniye i Dikiye Zhivotnye = Russian Veterinary Journal. Small Pets and Wild Animals.* 2013;5:19-22. (in Russian)
- Kubota K., Cui W., Dhakal P., Wolfe M.W., Rumi M.A., Vivian J.L., Roby K.F., Soares M.J. Rethinking progesterone regulation of female reproductive cyclicity. *Proc. Natl. Acad. Sci. USA.* 2016;113(15):4212-4217. DOI 10.1073/pnas.1601825113.
- Lay A.J., Liang Z., Rosen E.D., Castellino F.J. Mice with a severe deficiency in protein C display prothrombotic and proinflammatory phenotypes and compromised maternal reproductive capabilities. *J. Clin. Invest.* 2005;115(6):1552-1561. DOI 10.1172/jci24030.
- Long K., Mao K., Che T., Zhang J., Qiu W., Wang Y., Tang Q., Ma J., Li M., Li X. Transcriptome differences in frontal cortex between wild boar and domesticated pig. *Anim. Sci. J.* 2018;89(6):848-857. DOI 10.1111/asj.12999.
- Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford).* 2011;2011:baq036. DOI 10.1093/database/baq036.
- Manohar M., Khan H., Sirohi V.K., Das V., Agarwal A., Pandey A., Siddiqui W.A., Dwivedi A. Alteration in endometrial proteins during early- and mid-secretory phases of the cycle in women with unexplained infertility. *PLoS One.* 2014;9(11):e111687. DOI 10.1371/journal.pone.0111687.
- Marsh C.A., Auchus R.J. Fertility in patients with genetic deficiencies of cytochrome P450c17 (CYP17A1): combined 17-hydroxylase/17,20-lyase deficiency and isolated 17,20-lyase deficiency. *Fertil. Steril.* 2014;101(2):317-322. DOI 10.1016/j.fertnstert.2013.11.011.
- Matrisciano F., Tueting P., Dalal I., Kadriu B., Grayson D.R., Davis J.M., Nicoletti F., Guidotti A. Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by prenatal stress in mice. *Neuropharmacology.* 2013;68:184-194. DOI 10.1016/j.neuropharm.2012.04.013.
- Michon P., Woolley I., Wood E.M., Kastens W., Zimmerman P.A., Adams J.H. Duffy-null promoter heterozygosity reduces DARC expression and abrogates adhesion of the *P. vivax* ligand required for blood-stage infection. *FEBS Lett.* 2001;495(1-2):111-114. DOI 10.1016/S0014-5793(01)02370-5.
- Mithani S.K., Smith I.M., Califano J.A. Use of integrative epigenetic and cytogenetic analyses to identify novel tumor-suppressor genes in malignant melanoma. *Melanoma Res.* 2011;21(4):298-307. DOI 10.1097/CMR.0b013e328344a003.
- Mogno I., Vallania F., Mitra R.D., Cohen B.A. TATA is a modular component of synthetic promoters. *Genome Res.* 2010;20(10):1391-1397. DOI 10.1101/gr.106732.110.
- Morand S., McIntyre K.M., Baylis M. Domesticated animals and human infectious diseases of zoonotic origins: domestication time matters. *Infect. Genet. Evol.* 2014;24:76-81. DOI 10.1016/j.meegid.2014.02.013.

- Morozova O.V., Alekseeva A.E., Sashina T.A., Brusnigina N.F., Epifanova N.V., Kashnikov A.U., Zverev V.V., Novikova N.A. Phylogenetics of G4P[8] and G2P[4] strains of rotavirus A isolated in Russia in 2017 based on full-genome analyses. *Virus Genes*. 2020; 56(5):537-545. DOI 10.1007/s11262-020-01771-3.
- Nagaraja M.R., Gubbala S.P., Delphine Silvia C.R.W., Amanchy R. Molecular diagnostics of disorders of sexual development: an Indian survey and systems biology perspective. *Syst. Biol. Reprod. Med.* 2019;65(2):105-120. DOI 10.1080/19396368.2018.1549619.
- Nalls M.A., Wilson J.G., Patterson N.J., Tandon A., Zmuda J.M., Huntsman S., Garcia M., Hu D., Li R., Beamer B.A., Patel K.V., Akyzbekova E.L., Files J.C., Hardy C.L., Buxbaum S.G., Taylor H.A., Reich D., Harris T.B., Ziv E. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am. J. Hum. Genet.* 2008;82(1):81-87. DOI 10.1016/j.ajhg.2007.09.003.
- Nna V.U., Bakar A.B.A., Ahmad A., Umar U.Z., Suleiman J.B., Zakaria Z., Othman Z.A., Mohamed M. Malaysian propolis and metformin mitigate subfertility in streptozotocin-induced diabetic male rats by targeting steroidogenesis, testicular lactate transport, spermatogenesis and mating behaviour. *Andrology*. 2020;8(3):731-746. DOI 10.1111/andr.12739.
- Olkkonen J., Kouri V.P., Hynninen J., Kontinen Y.T., Mandelin J. Differentially expressed in chondrocytes 2 (DEC2) increases the expression of IL-1 $\beta$  and is abundantly present in synovial membrane in rheumatoid arthritis. *PLoS One*. 2015;10(12):e0145279. DOI 10.1371/journal.pone.0145279.
- Osadchuk L.V. Endocrine gonadal function in silver fox under domestication. *Scientific (Denmark)*. 1992a;16(2):116-121.
- Osadchuk L.V. Some peculiarities in reproduction in silver fox males under domestication. *Scientific (Denmark)*. 1992b;16(4):285-288.
- Osadchuk L.V. Biosynthesis of testosterone in the fetal gonads of the silver fox after long-term domestication. *Russ. J. Genet.* 1998;34(7):780-784.
- Osadchuk L.V. Reproductive potential of male silver foxes *Vulpes vulpes* after long selection for the domesticated behavior type. *J. Evol. Biochem. Physiol.* 2006;42(2):182-189.
- Osadchuk L.V., Krass P.M., Trut L.N., Belyaev D.K. Effects of selection for behavior on the endocrine function of the gonads in male silver-black foxes. *Doklady Akademii Nauk SSSR = Proceedings of the USSR Academy of Sciences*. 1978a;240(5):1255-1258. (in Russian)
- Osadchuk L.V., Krass P.M., Trut L.N., Ivanova L.N. Gonadal endocrine function in male silver foxes with different hereditary determined forms of defensive behavior. *Izvestiya Sibirskogo Otdeleniya Akademii Nauk SSSR = Proceedings of the Siberian Branch of the USSR Academy of Sciences*. 1978b;10: 79-86. (in Russian)
- Osterfeld H., Ahrens R., Strait R., Finkelman F.D., Renauld J.C., Hogan S.P. Differential roles for the IL-9/IL-9 receptor alpha-chain pathway in systemic and oral antigen-induced anaphylaxis. *J. Allergy Clin. Immunol.* 2010;125(2):469-476.e2. DOI 10.1016/j.jaci.2009.09.054.
- Peng Y., Zhou L., Cao Y., Chen P., Chen Y., Zong D., Ouyang R. Relation between serum leptin levels, lipid profiles and neurocognitive deficits in Chinese OSAHS patients. *Int. J. Neurosci.* 2017;127(11): 981-987. DOI 10.1080/00207454.2017.1286654.
- Pianka E.R. Natural selection of optimal reproductive tactics. *Amer. Zool.* 1976;16(4):775-784.
- Plengpanich W., Le Goff W., Poolsuk S., Julia Z., Guerin M., Khovidhunkit W. CETP deficiency due to a novel mutation in the CETP gene promoter and its effect on cholesterol efflux and selective uptake into hepatocytes. *Atherosclerosis*. 2011;216:370-373. DOI 10.1016/j.atherosclerosis.2011.01.051.
- Ponomarenko M., Kleshchev M., Ponomarenko P., Chadaeva I., Sharypova E., Rasskazov D., Kolmykov S., Drachkova I., Vasiliev G., Gutorova N., Ignatieva E., Savinkova L., Bogomolov A., Osadchuk L., Osadchuk A., Oshchepkov D. Disruptive natural selection by male reproductive potential prevents underexpression of protein-coding genes on the human Y chromosome as a self-domestication syndrome. *BMC Genomics*. 2020;21(Suppl. 1):89. DOI 10.1186/s12863-020-00896-6.
- Ponomarenko P.M., Suslov V.V., Savinkova L.K., Ponomarenko M.P., Kolchanov N.A. A precise equilibrium equation for four steps of binding between TBP and TATA-box allows for the prediction of phenotypical expression upon mutation. *Biophysics (Mosk.)*. 2010; 55(3):358-369. <https://pubmed.ncbi.nlm.nih.gov/20586319/>.
- Prasolova L.A., Gerbek Y.E., Gulevich R.G., Shikhevich S.G., Konoshenko M.Yu., Kozhemyakina R.V., Oskina I.N., Plyusina I.Z. The effects of prolonged selection for behavior on the stress response and activity of the reproductive system of male grey mice (*Rattus norvegicus*). *Russ. J. Genet.* 2014;50(8):846-852. DOI 10.1134/S1022795414080031.
- Pukazhenthi B.S., Neubauer K., Jewgenow K., Howard J., Wildt D.E. The impact and potential etiology of teratospermia in the domestic cat and its wild relatives. *Theriogenology*. 2006;66(1):112-121. DOI 10.1016/j.theriogenology.2006.03.020.
- Qian Y., Li L., Sun Z., Liu J., Yuan W., Wang Z. A multi-omics view of the complex mechanism of vascular calcification. *Biomed. Pharmacother.* 2021;135:111192. DOI 10.1016/j.biopha.2020.111192.
- Regan S.L.P., Knight P.G., Yovich J.L., Arfuso F., Dharmarajan A. Growth hormone during *in vitro* fertilization in older women modulates the density of receptors in granulosa cells, with improved pregnancy outcomes. *Fertil. Steril.* 2018;110(7):1298-1310. DOI 10.1016/j.fertnstert.2018.08.018.
- Robert N., Walzer C., Ruegg S.R., Kaczynsky P., Ganbaatar O., Stauffer C. Pathologic findings in reintroduced Przewalski's horses (*Equus caballus przewalskii*) in southwestern Mongolia. *J. Zoo Wildl. Med.* 2005;36(2):273-285. DOI 10.1638/03-035.1.
- Samet H. A top-down quadtree traversal algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* 1985;7(1):94-98. DOI 10.1109/tpami.1985.4767622.
- Sasaki Y., Otsuka K., Arimochi H., Tsukumo S.I., Yasutomo K. Distinct roles of IL-1 $\beta$  and IL-18 in NLRP4-induced autoinflammation. *Front. Immunol.* 2020;11:591713. DOI 10.3389/fimmu.2020.591713.
- Sato D.X., Rafati N., Ring H., Younis S., Feng C., Blanco-Aguilar J.A., Rubin C.J., Villafuerte R., Hallbook F., Carneiro M., Andersson L. Brain transcriptomics of wild and domestic rabbits suggests that changes in dopamine signaling and ciliary function contributed to evolution of tameness. *Genome Biol. Evol.* 2020;12(10):1918-1928. DOI 10.1093/gbe/evaa158.
- Senol B.K., Zulfikar B. Clinical problems and surgical interventions in inherited factor VII deficiency. *Turk. Pediatri Ars.* 2020;55(2):184-190. DOI 10.14744/TurkPediatriArs.2020.79069.
- Silliman K., Tall A.R., Kretschmer N., Forte T.M. Unusual high-density lipoprotein subclass distribution during late pregnancy. *Metabolism*. 1993;42(12):1592-1599. DOI 10.1016/0026-0495(93)90156-i.
- Soilemetzidou E.S., de Bruin E., Eschke K., Azab W., Osterrieder N., Czirjak G.A., Buuveibaatar B., Kaczynsky P., Koopmans M., Walzer C., Greenwood A.D. Bearing the brunt: Mongolian khulan (*Equus hemionus hemionus*) are exposed to multiple influenza A strains. *Vet. Microbiol.* 2020;242:108605. DOI 10.1016/j.vetmic.2020.108605.
- Songsasen N., Converse S.J., Brown M. Reproduction and reproductive strategies relevant to management of whooping cranes *ex situ*. In: Biodiversity of the World: Conservation from Genes to Landscapes. San Diego, California: Academic Press, 2019;373-387. DOI 10.1016/B978-0-12-803555-9.00017-7.
- Sugiyama C., Yamamoto M., Kotani T., Kikkawa F., Murata Y., Hayashi Y. Fertility and pregnancy-associated  $\beta$ -cell proliferation in mice deficient in proglucagon-derived peptides. *PLoS One*. 2012;7(8): e43745. DOI 10.1371/journal.pone.0043745.
- Sun G.L., Shen W., Wen J.F. Triosephosphate isomerase genes in two trophic modes of euglenoids (euglenophyceae) and their phylogenetic analysis. *J. Eukaryot. Microbiol.* 2008;55(3):170-177. DOI 10.1111/j.1550-7408.2008.00324.x.
- Sun S., Liu S., Luo J., Chen Z., Li C., Loo J.J., Cao Y. Repeated pregnant mare serum gonadotropin-mediated oestrous synchronization alters gene expression in the ovaries and reduces reproductive per-



- formance in dairy goats. *Reprod. Domest. Anim.* 2019;54(6):873-881. DOI 10.1111/rda.13439.
- Sun Y., Liu S., Hu R., Zhou Q., Li X. Decreased placental IL9 and IL9R in preeclampsia impair trophoblast cell proliferation, invasion, and angiogenesis. *Hypertens. Pregnancy.* 2020;39(3):228-235. DOI 10.1080/10641955.2020.1754852.
- Takhviji V., Zibara K., Azarkeivan A., Mehrvar N., Mehrvar N., Mezghinejad F., Khosravi A. Fertility and pregnancy in Iranian thalassemia patients: An update on transfusion complications. *Transfus. Med.* 2020;30(5):352-360. DOI 10.1111/tme.12707.
- Theofanopoulou C., Gastaldon S., O'Rourke T., Samuels B.D., Martins P.T., Delogu F., Alamri S., Boeckx C. Self-domestication in *Homo sapiens*: Insights from comparative genomics. *PLoS One.* 2017;12(10):e0185306. DOI 10.1371/journal.pone.0185306.
- Tian Y., Stamova B., Jickling G.C., Xu H., Liu D., Ander B.P., Bushnell C., Zhan X., Turner R.J., Davis R.R., Verro P., Pevec W.C., Hedayati N., Dawson D.L., Khoury J., Jauch E.C., Pancioli A., Broderick J.P., Sharp F.R. Y chromosome gene expression in the blood of male patients with ischemic stroke compared with male controls. *Genet. Med.* 2012;9(2):68-75.e3. DOI 10.1016/j.genm.2012.01.005.
- Vasiliev G., Chadaeva I., Rasskazov D., Ponomarenko P., Sharypova E., Drachkova I., Bogomolov A., Savinkova L., Ponomarenko M., Kolchanov N., Osadchuk A., Oshchepkov D., Osadchuk L. A bioinformatics model of human diseases on the basis of differentially expressed genes (of domestic versus wild animals) that are orthologs of human genes associated with reproductive-potential changes. *Int. J. Mol. Sci.* 2021;22(5):2346. DOI 10.3390/ijms22052346.
- Velzing-Aarts F.V., van der Dijks F.P., Muskiet F.A., Duits A.J. The association of pre-eclampsia with the Duffy negative phenotype in women of West African descent. *BJOG.* 2002;109(4):453-455. DOI 10.1111/j.1471-0528.2002.01181.x.
- Wang F., Yu H., Huang S., Zheng L., Zheng P., Zhang S., Li S., Chen J. Jian-Pi-Yi-Shen regulates EPO and iron recycling protein expressions in anemic rats with chronic kidney disease: accumulation of hypoxia inducible factor-2 $\alpha$  via ERK signaling. *Evid. Based Complement. Altern. Med.* 2020;2020:8894257. DOI 10.1155/2020/8894257.
- Wood M.A., Mukherjee P., Toocheck C.A., Walker W.H. Upstream stimulatory factor induces *Nr5a1* and *Shbg* gene expression during the onset of rat Sertoli cell differentiation. *Biol. Reprod.* 2011;85(5):965-976. DOI 10.1095/biolreprod.111.093013.
- Xi J.F., Wang X.Z., Zhang Y.S., Jia B., Li C.C., Wang X.H., Ying R.W. Sex control by Zfy siRNA in the dairy cattle. *Anim. Reprod. Sci.* 2019;200:1-6. DOI 10.1016/j.anireprosci.2018.05.015.
- Xiao W., Feng L., Mou P., Miquelle D.G., Hebblewhite M., Goldberg J.F., Robinson H.S., Zhao X., Zhou B., Wang T., Ge J. Estimating abundance and density of Amur tigers along the Sino-Russian border. *Integr. Zool.* 2016;11(4):322-332. DOI 10.1111/1749-4877.12210.
- Yang X., Zhang H., Shang J., Liu G., Xia T., Zhao C., Sun G., Dou H. Comparative analysis of the blood transcriptomes between wolves and dogs. *Anim. Genet.* 2018;49(4):291-302. DOI 10.1111/age.12675.
- Yang Y., Adeola A.C., Xie H.B., Zhang Y.P. Genomic and transcriptomic analyses reveal selection of genes for puberty in Bama Xiang pigs. *Zool. Res.* 2018;39(6):424-430. DOI 10.24272/j.issn.2095-8137.2018.068.
- Yao X., Wang Z., Gao X., Li X., Yang H., Ei-Samahy M.A., Bao Y., Xiao S., Meng F., Wang F. Unconservative\_15\_2570409 suppresses progesterone receptor expression in the granulosa cells of Hu sheep. *Theriogenology.* 2020;157:303-313. DOI 10.1016/j.theriogenology.2020.08.011.
- Yu Q., Yang X., Zhang C., Zhang X., Wang C., Chen L., Liu X., Gu Y., He X., Hu L., Liu W.T., Li Y. AMPK activation by ozone therapy inhibits tissue factor-triggered intestinal ischemia and ameliorates chemotherapeutic enteritis. *FASEB J.* 2020;34(9):13005-13021. DOI 10.1096/fj.201902717r.
- Zhang J., Nie C., Li X., Ning Z., Chen Y., Jia Y., Han J., Wang L., Lv X., Yang W., Qu L. Genome-wide population genetic analysis of commercial, indigenous, game, and wild chickens using 600K SNP microarray data. *Front. Genet.* 2020;11:543294. DOI 10.3389/fgene.2020.543294.
- Zhang Y., Katoh T.K., Finet C., Izumitani H.F., Toda M.J., Watabe H.A., Katoh T. Phylogeny and evolution of mycophagy in the *Zygothrica* genus group (Diptera: Drosophilidae). *Mol. Phylogenet. Evol.* 2021;163:107257.

#### ORCID ID

M.P. Ponomarenko orcid.org/0000-0003-1663-318X  
I.V. Chadaeva orcid.org/0000-0002-2724-5441  
P.M. Ponomarenko orcid.org/0000-0003-2715-9612  
A.G. Bogomolov orcid.org/0000-0003-4359-6089

D.Yu. Oshchepkov orcid.org/0000-0002-6097-5155  
E.B. Sharypova orcid.org/0000-0002-5517-920X  
V.V. Suslov orcid.org/0000-0002-1940-9389  
A.V. Osadchuk orcid.org/0000-0002-4210-7354  
L.V. Osadchuk orcid.org/0000-0002-7597-9204

**Acknowledgements.** The authors are thankful to Shevchuk Editing (Brooklyn, NY, United States; URL: <http://www.shevchuk-editing.com>) for translation from Russian into English. The idea (LVO, AVO, MPP, IVC) was supported by Russian Science Foundation grant No. 19-15-00075. The data analysis (DYUO, PMP, EBS, and AGB) was carried out with the help of the computing resources of the Multi-Access Center "Bioinformatics" with the support of publicly funded project No. FWNR-2022-0020. The knowledge base (VVS) and study coordination (YuGM) were supported by the Federal Scientific and Technical Program for the Development of Genetic Technologies in Russia.

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

Received October 30, 2020. Revised August 20, 2021. Accepted August 24, 2021.



## De novo assembly and analysis of the transcriptome of the Siberian wood frog *Rana amurensis*

D.N. Smirnov<sup>1, 2</sup>, S.V. Shekhovtsov<sup>3, 4</sup>✉, A.A. Shipova<sup>4</sup>, G.R. Gazizova<sup>5</sup>, E.I. Shagimardanova<sup>5</sup>, N.A. Bulakhova<sup>3</sup>, E.N. Meshcheryakova<sup>3</sup>, T.V. Poluboyarova<sup>4</sup>, E.E. Khrameeva<sup>1</sup>, S.E. Peltek<sup>4</sup>, D.I. Berman<sup>3</sup>

<sup>1</sup> Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>2</sup> Ben-Gurion University of the Negev, Department of Life Sciences, Beer Sheva, Israel

<sup>3</sup> Institute of the Biological Problems of the North of the Far-Eastern Branch of the Russian Academy of Sciences, Magadan, Russia

<sup>4</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>5</sup> Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia

✉ shekhovtsov@bionet.nsc.ru

**Abstract.** The Siberian wood frog *Rana amurensis* Boulenger, 1886 is the most hypoxia-tolerant amphibian. It can survive for several months in an almost complete absence of oxygen. Little is known about the mechanisms of this remarkable resilience, in part because studies of amphibian genomes are impeded by their large size. To make the Siberian wood frog more amenable for genetic analysis, we performed transcriptome sequencing and *de novo* assembly for the *R. amurensis* brain under hypoxia and normoxia, as well as for the normoxic heart. In order to build a *de novo* transcriptome assembly of *R. amurensis*, we utilized 125-bp paired-end reads obtained from the brain under normoxia and hypoxia conditions, and from the heart under normoxia. In the transcriptome assembled from about 100,000,000 reads, 81.5 % of transcripts were annotated as complete, 5.3 % as fragmented, and 13.2 % as missing. We detected 59,078 known transcripts that clustered into 22,251 genes; 11,482 of them were assigned to specific GO categories. Among them, we found 6696 genes involved in protein binding, 3531 genes involved in catalytic activity, and 576 genes associated with transporter activity. A search for genes encoding receptors of the most important neurotransmitters, which may participate in the response to hypoxia, resulted in a set of expressed receptors of dopamine, serotonin, GABA, glutamate, acetylcholine, and norepinephrine. Unexpectedly, no transcripts for histamine receptors were found. The data obtained in this study create a valuable resource for studying the mechanisms of hypoxia tolerance in the Siberian wood frog, as well as for amphibian studies in general.

Key words: Siberian wood frog; *Rana amurensis*; transcriptome; *de novo* assembly; neurotransmitters.

**For citation:** Smirnov D.N., Shekhovtsov S.V., Shipova A.A., Gazizova G.R., Shagimardanova E.I., Bulakhova N.A., Meshcheryakova E.N., Poluboyarova T.V., Khrameeva E.E., Peltek S.E., Berman D.I. *De novo* assembly and analysis of the transcriptome of the Siberian wood frog *Rana amurensis*. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2022;26(1):109-116. DOI 10.18699/VJGB-22-07

## De novo сборка и анализ транскриптома сибирской лягушки *Rana amurensis*

Д.Н. Смирнов<sup>1, 2</sup>, С.В. Шеховцов<sup>3, 4</sup>✉, А.А. Шипова<sup>4</sup>, Г.Р. Газизова<sup>5</sup>, Е.И. Шагимарданова<sup>5</sup>, Н.А. Булахова<sup>3</sup>, Е.Н. Мещерякова<sup>3</sup>, Т.В. Полубоярова<sup>4</sup>, Е.Е. Храмева<sup>1</sup>, С.Е. Пельтек<sup>4</sup>, Д.И. Берман<sup>3</sup>

<sup>1</sup> Центр наук о жизни Сколковского института науки и технологий, Москва, Россия

<sup>2</sup> Университет им. Давида Бен-Гуриона в Негеве, Отделение наук о жизни, Беэр-Шева, Израиль

<sup>3</sup> Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия

<sup>4</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>5</sup> Институт фундаментальной медицины и биологии Казанского федерального университета, Казань, Россия

✉ shekhovtsov@bionet.nsc.ru

**Аннотация.** Сибирская лягушка *Rana amurensis* Boulenger, 1886 – наиболее устойчивый к гипоксии вид амфибий. Она может прожить несколько месяцев при почти полном отсутствии кислорода. О механизмах этой замечательной устойчивости мало что известно, отчасти потому, что исследования геномов амфибий затруднены из-за их большого размера. Чтобы сделать сибирскую лягушку более доступной для генетического анализа, мы провели секвенирование и сборку *de novo* транскриптома мозга *R. amurensis* в условиях гипоксии и нормоксии, а также для сердца – в нормоксии. Для сборки транскриптома *de novo* использовали парные прочтения длиной 125 п.н., полученные для мозга сибирской лягушки в нормоксии и гипоксии, а также для сердца контрольных особей. В транскриптом, собранном из примерно 100 млн ридов, 81.5 % транскриптов были аннотированы как полные, 5.3 – как фрагментированные и 13.2 % – как отсутствующие. Мы обнаружили 59 078 известных транскриптов, которые были сгруппированы в 22 251 ген, 11 482 из них были отнесены к определенным категориям

Gene Ontology. Среди них – 6696 генов, участвующих в белок-белковом взаимодействии, 3531 ген, кодирующий белок с каталитической функцией, и 576 генов, связанных с транспортной активностью. Для большинства транскриптов были установлены тканеспецифичные различия в экспрессии. Известно, что нейротрансмиттеры играют важную роль в ответе на гипоксию различных организмов, устойчивых к недостатку кислорода. Поиск генов, кодирующих рецепторы важнейших нейромедиаторов, которые могут участвовать в реакции на гипоксию, выявил набор экспрессируемых рецепторов дофамина, серотонина, гамма-аминомасляной кислоты (ГАМК), глутамата, ацетилхолина и норадреналина. При этом не обнаружено транскриптов для рецепторов гистамина. Данные, полученные в нашей работе, представляют собой ценный ресурс для исследования механизмов толерантности к гипоксии у сибирской лягушки, а также для изучения амфибий в целом.

Ключевые слова: сибирская лягушка; *Rana amurensis*; транскриптом; сборка *de novo*; нейромедиаторы.

## Introduction

Next-generation sequencing revolutionized the studies in the field of molecular genetics. In contrast to early whole-genome projects, this technology presents a quick and relatively cheap way to obtain genome-wide information for non-model organisms. However, for organisms with large genome sizes, such as amphibians, this is still a challenge due to many repeat sequences, frequent cases of polyploidy and high costs associated with the sequences of large genomes (Schatz et al., 2010). Among the family Ranidae, there are currently only three genome assemblies: *Rana temporaria*, *Glandirana rugosa*, and *Lithobates catesbeianus* (Hammond et al., 2017; Katsura et al., 2021; Streicher et al., 2021). Available transcriptomes are more numerous; however, they are still provided only for a limited number of members of the family Ranidae and do not always meet the high-quality standards of modern transcriptome assemblies.

Assembled transcriptomes would be useful resources for studies on the emergent model species. Among these species are the northern amphibians that adapted to extreme conditions of the northern Palearctic. These include highly freeze-tolerant *Rana sylvatica* LeConte, 1825 (Storey, 1984), *R. arvalis* Nilsson, 1842 (Berman et al., 2020), *Hyla japonica* Günther, 1859 (Berman et al., 2016a), and the urodela *Salamandrella keyserlingii* Dybowski, 1870 and *S. schrenkii* (Strauch, 1870) (Berman et al., 1984, 2010, 2016b), as well as the hypoxia-tolerant Siberian wood frog *Rana amurensis* Boulenger, 1886 (Berman et al., 2019). These species are intensely studied because they represent one of the most remarkable adaptations of vertebrates to extreme conditions and could give insights into ischemia treatment and organ transplantation. Earlier studies of freeze- and hypoxia tolerant amphibians were mostly aimed at their physiology and biochemistry, but studying genetic systems becomes more important (Bickler, Buck, 2007; Storey K.B., Storey J.M., 2017).

Amphibians in general are believed to be not particularly tolerant to hypoxia: adults of the different studied species can survive for a few hours to a few days even at low (near-zero) temperatures in water with low oxygen content (Bickler, Buck, 2007). However, the Siberian wood frog *R. amurensis* Boulenger, 1886 is unique among amphibians in its ability to survive almost complete anoxia for several months (Berman et al., 2019). This makes it a promising model object for studying hypoxia tolerance. Metabolomic patterns in its organs indicate dramatic changes in biochemical pathways under hypoxia (Shekhovtsov et al., 2020). However, these patterns are not easy to interpret, and this could be facilitated by the analysis of gene expression and gene networks. In order to create a resource for studying gene expression in the Siberian

wood frog, we performed sequencing, *de novo* assembly, and annotation of the transcriptome of this species.

Brain and heart are the most sensitive to hypoxia (Nilsson et al., 2015; Swenson, 2016), so we used transcripts from these organs for transcriptome construction. To test the assembled transcriptome, we also performed a search for neurotransmitter receptor genes: it was demonstrated (Nilsson et al., 1990, 1991) that neurotransmitters mediate hypoxia response in turtles, so we hypothesized that this might also be true for the Siberian wood frog.

## Materials and methods

**RNA extraction and sequencing.** Specimens of the Siberian wood frog were collected in September 2019 near the Lesopilnoye village, Khabarovsk Krai (46° N, 134° E). We followed approved methods under appropriate permits issued by cognizant governmental agencies (No. 001/04-19). Frog handling, hypoxia exposure, and organ extraction were performed as described in S.V. Shekhovtsov et al. (2020): briefly, the frogs were distributed by 5–7 individuals into 10 L containers filled with water (oxygen level 7–8 mg/L) and acclimated to low temperatures: 2 days at 14–15 °C, for 4 days at 8, 4, and 2–3 °C. Acclimation was performed in a TSO-1/80 SPU thermostat (SKTB SPU, Russia) and in a WT-64/75 climatic test chamber (Weiss Umwelttechnik GmbH, Germany). Control animals were kept in open containers; those exposed to hypoxia, in closed airtight bottles. The dissolved oxygen content was measured daily by a HACH HQ30D Flexi digital single-channel device with a luminescent LDO101 sensor until it reached 0.2 mg/L. After 17 days in hypoxia, animals were slaughtered as quickly as possible, and the organs were extracted and immediately submerged in liquid nitrogen. RNA was extracted using commercial kits (Biolabmix, Russia) following the manufacturer's protocol.

The purity of total RNA was estimated on a NanoPhotometer (Implen, Germany). The quantity of total RNA was measured by fluorimeter Qubit 4.0 (ThermoFisher Scientific, USA). The quality of total RNA was evaluated using a Bioanalyzer 2100 (Agilent Technologies, USA). Then, from 800–1000 ng of pure and good quality total RNA (RIN ≥ 7), polyA mRNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, USA).

cDNA libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer's protocol. The concentration of amplified libraries was estimated by fluorimeter Qubit 3.0 (ThermoFisher Scientific). Size selection of pooled libraries was performed on the BluePippin system (SAGE Science, USA) using 1.5 % agarose gel cassettes with

300–400 bp target size. The quality of libraries was verified on a Bioanalyzer 2100 (Agilent Technologies) using DNA High Sensitivity Kit. Range of library fragment size was 200–1000 bp. The concentration of libraries was validated by qPCR using 2.5× EVA Green Mix (Synthol, Russia) and primers for Illumina adapters (Evrogen, Russia). Libraries were then sequenced on a HiSeq 2500 (Illumina, USA) with paired-end 125 bp reads.

**Transcriptome assembly.** RNA reads from *R. amurensis* brain and heart samples (brain+normoxia, brain+hypoxia, heart+normoxia, 3 samples in total) were used for *de novo* transcriptome assembly. The quality of raw reads was estimated with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adaptor trimming and read filtering were performed using fastp (Chen et al., 2018) with default parameters. The rCorrector tool (Song, Florea, 2015) was used for correcting the non-solid k-mers within reads and removing unfixable ones. Transcriptome assembly was performed on all samples via Trinity (Grabherr et al., 2011) with `--SS_lib_type FR` parameter for a stranded library. The basic assembly metrics were calculated using the ‘*TrinityStats.pl*’ script incorporated in Trinity. Redundant transcripts were identified and removed from the assembly via *CD-HIT* (Fu et al., 2012) with the following parameters: `-c 0.98 -p 1 -d 0 -b 3 -T 5 -M 2000`. Completeness of the assembled transcriptome was estimated using BUSCO (Simão et al., 2015) with the mode `-m transcriptome` and lineage ‘*tetrapoda\_odb10*’ parameters.

**Transcript quantification.** The transcript abundance was estimated using the ‘*align\_and\_estimate\_abundance.pl*’ script (`--est_method salmon`) included in Trinity. Both gene- and isoform-level abundance matrices for all samples were constructed with the ‘*abundance\_estimates\_to\_matrix.pl*’ script. The comparison of samples based on their expression level and the subsequent visualization procedures were performed using the ‘*PtR*’ script as well as custom scripts.

**Assembly annotation and candidate coding regions identification.** We used a collection of scripts from TransDecoder (Grabherr et al., 2011) to identify the candidate coding regions. First, open reading frames (ORF) were retrieved from the assembly file. A set of the longest obtained ORFs were then queried against Swiss-Prot (Bairoch, Apweiler, 1999; The Uniprot Consortium, 2021) and Pfam (Mistry et al., 2021) databases to search for sequence similarity with known proteins and Pfam protein domains. To achieve better computational efficiency, we used *hmmsearch* v3.3.2 (Eddy, 2011) scripts instead of *hmmscan* for domain identification, and the homology search was done with *Blast+* (Camacho et al., 2009). The output generated from the database searching step was then used for the prediction of coding regions using the *TransDecoder.Predict* script from TransDecoder and for transcriptome assembly annotation via the Trinotate (Bryant et al., 2017) pipeline.

**Gene Ontology (GO) analysis.** The Trinotate report obtained in the assembly annotation step was used to characterize the annotated genes according to their biological role and the occupied cell compartments. We counted the number of annotated genes per GO category for the cellular component (CC), biological process (BP), and molecular function (MF) sub-ontologies at level 2. Graphical representation was done

using an in-house R script. Annotated genes without assigned GO categories were classified according to the PFAM protein families they associated with.

**Searching for neurotransmitter receptors.** We extracted a set of genes encoding receptors of the main neurotransmitters (dopamine, serotonin, GABA, glutamate, acetylcholine, histamine, and norepinephrine) from the *Xenopus* genome database (Xenbase; <http://www.xenbase.org>). Xenbase was chosen over more closely related species due to its longer history and better annotation. For each annotated gene, the transcripts were taken for *Xenopus tropicalis*, or, if absent, for *X. laevis*. We performed a *blastn* search for this *Xenopus* transcript dataset in the assembled transcriptome (*Trinity\_filtered.fasta*) with *e*-value <  $1e^{-5}$ . Transcripts with > 70 % sequence similarity were included in the final dataset.

## Results and discussion

### De novo transcriptome assembly

In order to build a *de novo* transcriptome assembly of *R. amurensis*, we utilized 125 bp paired-end reads obtained from the brain under normoxia and hypoxia conditions (RABN and RABH samples, respectively) and from the heart under normoxia (RAHN sample).

After filtering out low-quality reads, a total of 98,948,825 reads from all three samples were used for the subsequent transcriptome assembling procedure. An initial assembly consisted of 610,890 Trinity ‘genes’ composed of 839,939 transcripts with an average contig length of 639 bp (or 481 bp based on the longest isoform per Trinity ‘gene’). In addition, we filtered out 56,044 redundant transcripts using *CD-HIT*. Once filtering was done, the final assembly was generated (Table 1).

Estimates using BUSCO demonstrated that 81.5 % of the transcripts were annotated as complete, 5.3 % as fragmented, and 13.2 % as missing.

### Abundance quantification

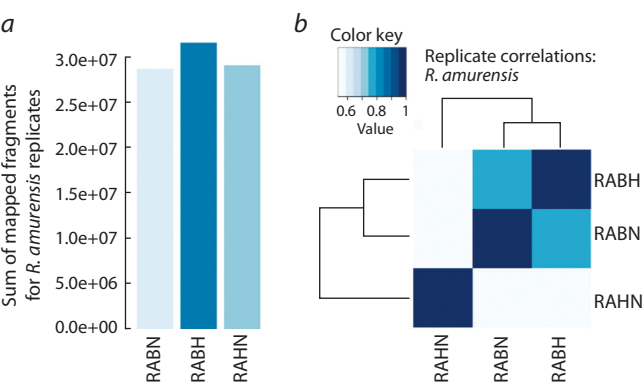
For each RNA sample, we calculated the levels of transcript abundances using the resulting assembled transcriptome. The Salmon alignment rate varied from 87.9 to 91 % across all samples, which is an additional indicator of good quality of the final assembly. The majority of Trinity ‘genes’ turned out to be low-expressed, and only 20,251 out of 588,475 ‘genes’ had expression levels  $\geq 10$  TPM (transcripts per million) in at least one sample. The sum of gene expression counts per sample

**Table 1.** The statistics of the final Trinity assembly

Number of assembled transcripts	783,895
Number of ‘genes’	588,475
Median contig length	347/323*
Average contig	616,66/487,50*
Contig N50	857/533*
Total assembled bases	483,398,726
Percent GC	43,37

\* Indicates values for statistics based on the longest isoform per Trinity ‘gene’.





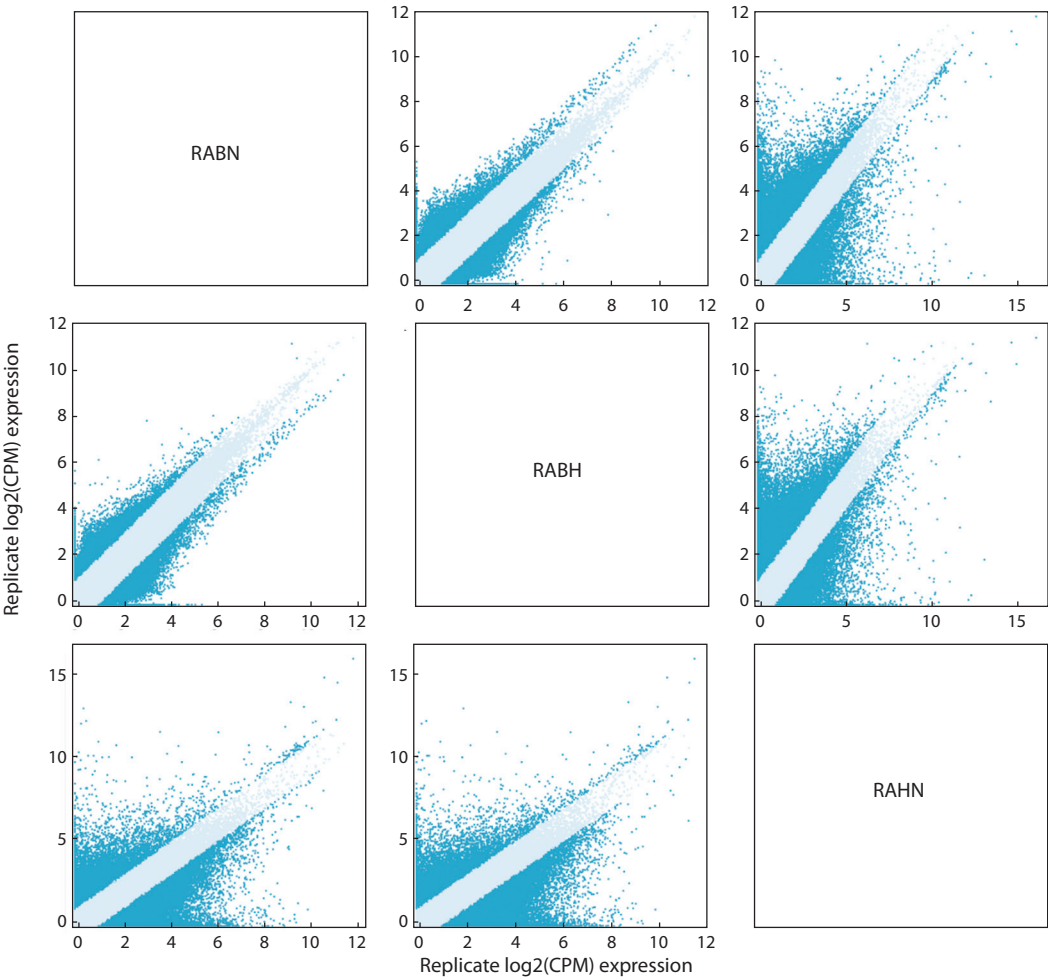
**Fig. 1.** Comparative analysis of expression of *R. amurensis* samples.  
*a* – barplot representing the distribution of the total number of mapped fragments across all samples; *b* – heatmap showing hierarchical clustering of samples based on their expression levels. In both panels, RABH corresponds to *R. amurensis* brain sample under hypoxia; RABN – *R. amurensis* brain sample under normoxia, and RAHN is a *R. amurensis* heart sample under normoxia.

varied from 28,587,718 to 31,477,299 with the largest value for the *R. amurensis* brain sample under hypoxia (Fig. 1, *a*).  
Moreover, we observed tissue-specific differences in gene expression levels between brain and heart samples (see Fig. 1, *b*).

As was expected, we found that the brain transcriptomes correlated better with each other (Pearson’s  $r = 0.770$ ) than with the heart transcriptome (Pearson’s  $r = 0.538$  and  $0.535$  for RABH correlation with RABN and RABH, respectively). In addition, we counted the number of genes with more than 2-fold expression change between each pair of samples (Fig. 2). For brain-brain transcriptome comparisons, the number of such genes was equal to 34,488, while for brain-heart comparisons, this number almost doubled (70,005 genes for RABN versus RAHN and 68,497 genes for RABH versus RAHN). Taken together, all these findings on gene and transcript quantification indicate the correctness of the transcriptome assembly.

**Transcriptome assembly and annotation**

Once the expression quantification step was done, we annotated the obtained transcripts to evaluate the number of biologically relevant ones. We first identified a total of 141,950 candidate coding regions using TransDecoder and obtained information about known protein homologs and protein domains. Using the Trinotate pipeline for functional annotation of transcripts, we then detected 59,078 known transcripts in our assembly that clustered into 22,251 genes. Finally, we explored the fraction of annotated genes with TPM > 0 that are common between replicates. We retrieved a total of



**Fig. 2.** Sample-to-sample scatter plots showing gene expression differences between *R. amurensis* samples. Circles corresponding to 2-fold changes are marked in blue.

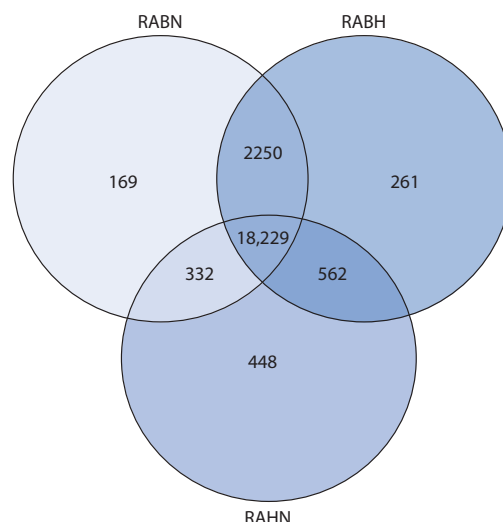


18,229 genes with expression in all three replicates as well as 2680 and 448 genes expressed only in brain and heart samples, respectively (Fig. 3).

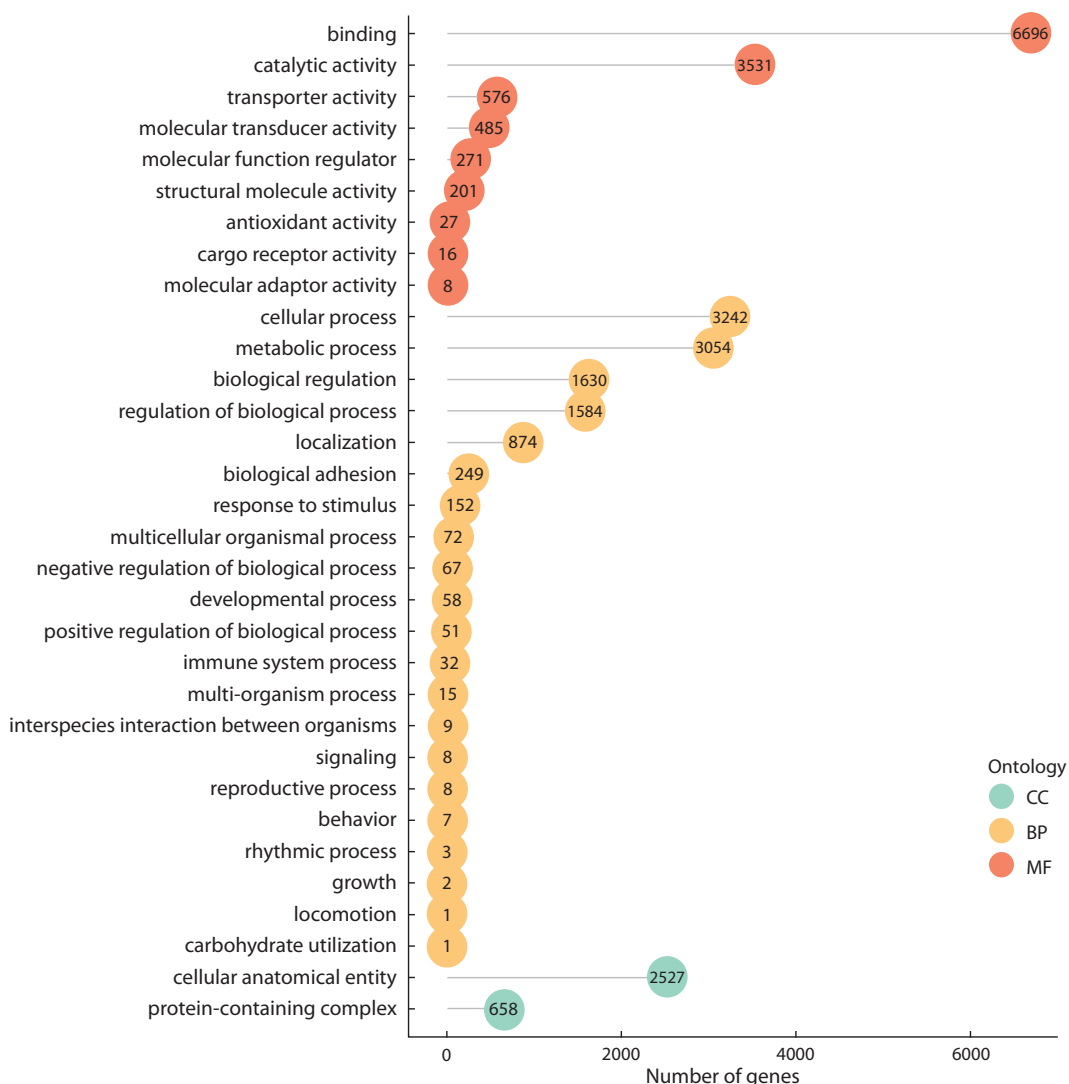
### Gene Ontology analysis

During the annotation step, we identified a total of 11,482 genes for which corresponding GO categories were described. In particular, we found 6696 genes involved in binding (including 2988 genes associated with protein binding and 878 genes responsible for DNA binding), 3531 genes involved in catalytic activity, and 576 genes associated with transporter activity (Fig. 4).

For 10,769 annotated genes without assigned GO categories, we performed an additional analysis of their functional roles. Among them, we found several large functional gene groups, including 1969 genes associated with the RVT\_1 (Reverse transcriptase) family, 1365 genes encoding zf-C2H2 (zinc finger) protein domains, and 455 genes belonging to the endonuclease/exonuclease/phosphatase family.



**Fig. 3.** Venn diagram showing the number of common and sample-specific expressed genes between *R. amurensis* samples.



**Fig. 4.** Distribution of the gene number per GO category in the transcriptome of *R. amurensis*.

Red circles represent molecular function (MF) terms, yellow circles represent biological process (BP) terms, and light green circles show terms related to the cellular component (CC) sub-ontology.

**Table 2.** Transcripts of neurotransmitter receptors detected in *R. amurensis* transcriptome

Receptors	% id.	Receptors	% id.	Receptors	% id.
Dopamine receptors		GABA receptors		Glutamate receptors	
Dopamine receptor d1	83.9	GABAA receptor α1	84.4	NMDA receptor 2B	84.8
Dopamine receptor d1c	78.5	GABAA receptor α2	85.5	NMDA receptor 2C	82.6
Dopamine receptor d2	83.2	GABAA receptor α3	79.3	NMDA receptor 2D	83.2
Dopamine receptor d4	80.9	GABAA receptor α5	76.1	NMDA receptor 3A	78.8
Dopamine receptor d5	82.2	GABAA receptor β1	82.1	Delta receptor GRID1	82.3
Acetylcholine receptors		GABAA receptor γ3	85.3	Delta receptor GRID2	82.6
Muscarinic receptor 5	81.8	GABAA receptor rho1	81.9	Metabotropic receptor 1	80.4
Nicotinic receptor α5	91.7	GABAA receptor rho3	83.1	Metabotropic receptor 3	81.1
Nicotinic receptor α6	85.6	Glutamate receptors		Metabotropic receptor 4	81.4
Nicotinic receptor β2	82.4	AMPA receptor 1/2	82.9	Metabotropic receptor 5	81.0
Nicotinic receptor β3	83.2	AMPA receptor 3	86.0	Metabotropic receptor 7	80.0
Serotonin receptors		AMPA receptor 4	84.7	Metabotropic receptor 8	85.1
Serotonin receptor 1A	80.1	Kainate receptor 1	83.7	Norepinephrine receptors	
Serotonin receptor 1B	82.7	Kainate receptor 2	86.5	Adrenoreceptor α2a	78.35
Serotonin receptor 4	82.5	Kainate receptor 4	80.8	Adrenoreceptor α2b	80.00
Serotonin receptor 6	80.6	Kainate receptor 5	82.0	Adrenoreceptor α2c	78.46
		NMDA receptor 1	81.0	Adrenoreceptor α2d	83.97
		NMDA receptor 2A	81.5	Adrenoreceptor β2	77.00

Note. % id. shows the percentage of identity to the respective *X. tropicalis* transcript along the alignable part.

Neurotransmitter receptor genes

A search for neurotransmitter receptors recovered a total of 47 transcripts belonging to six classes (Table 2). All detected transcripts could be unambiguously attributed to particular classes of receptors. Unexpectedly, we failed to detect any transcripts of histamine receptors. Our blastn and blastx search for these genes in the available ranid genome and transcriptome data resulted in no expressed histamine receptors in any sequenced cDNA data from any tissue. However, the genome of *R. temporaria* was found to contain the full gene set of histamine receptor genes. This may indicate that the histamine pathway has very limited expression in the family Ranidae.

The information on neurotransmitters is of special interest because they are known to be involved in hypoxia response in various organisms. G.E. Nilsson et al. (1990, 1991) found that levels of different neurotransmitters in the brain and other organs changed significantly upon exposure to hypoxia: the concentrations of GABA increased, and those of glutamate decreased in the crucian carp and the red-eared slider turtle, but not in the hypoxia-intolerant species. The authors also found that the levels of serotonin, dopamine, and norepinephrine remained unchanged, although their synthesis is oxygen-dependent. This response probably involves not just an upregulation of neurotransmitter synthesis, but the rearrangement of the whole pathway, and thus the obtained transcriptome data will be of particular use to elucidate this issue.

Conclusion

In recent years, transcriptome analysis is increasingly used for amphibians, e. g., to study the effects of pathogens (Price et al., 2015; Xu et al., 2017), insecticides (Ma et al., 2018), or the changes occurring during metamorphosis (Birol et al., 2015; Zhao et al., 2016). Many of those studies combine data from different tissues to obtain a more or less comprehensive set of transcripts expressed in the most important organs (Yang et al., 2012; Qiao et al., 2013; Robertson, Cornman, 2014; Christenson et al., 2014). In this study, we sequenced and assembled the transcriptome of the Siberian frog *R. amurensis*. We also provided a quality assessment of the obtained assembly and characterized the functional roles of annotated transcripts. The available information on amphibian transcriptomes is still limited; therefore, our dataset contributes to the understanding of genome functioning and evolution of amphibians. Moreover, the majority of the previously published transcriptome assemblies for other species of the genus *Rana*, e. g., in I. Birol et al. (2015) and S.J. Price et al. (2015), are probably not the best option for studying the mechanisms of the hypoxia tolerance in these species. Because these studies do not focus on hypoxia and are not based on hypoxia samples, the assembled transcriptomes might miss or under-represent some transcripts specific for hypoxia. In contrast, our work creates a useful resource for studying the mechanisms of the tolerance of *R. amurensis* to hypoxia.

## References

- Bairoch A., Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* 1999; 27(1):49-54. DOI 10.1093/nar/27.1.49.
- Berman D.I., Bulakhova N.A., Meshcheryakova E.N. The Siberian wood frog survives for months underwater without oxygen. *Sci. Rep.* 2019;9(1):13594. DOI 10.1038/s41598-018-31974-6.
- Berman D.I., Bulakhova N.A., Meshcheryakova E.N., Shekhovtsov S.V. Overwintering and cold tolerance in the moor frog (*Rana arvalis*) across its range. *Can. J. Zool.* 2020;98(11):705-714. DOI 10.1139/cjz-2019-0179.
- Berman D.I., Leirikh A.N., Meshcheryakova E.N. The Schrenck newt (*Salamandrella schrenckii*, Amphibia, Caudata, Hynobiidae) is the second amphibian that withstands extremely low temperatures. *Dokl. Biol. Sci.* 2010;431(1):131-134. DOI 10.1134/S0012496610020171.
- Berman D.I., Leirikh A.N., Mikhailova E.I. Winter hibernation of the Siberian salamander *Hynobius keyserlingi*. *J. Evol. Biochem. Physiol.* 1984;3(1-2):323-327. (in Russian)
- Berman D.I., Meshcheryakova E.N., Bulakhova N.A. The Japanese tree frog (*Hyla japonica*), one of the most cold-resistant species of amphibians. *Dokl. Biol. Sci.* 2016a;471(1):276-279. DOI 10.1134/S0012496616060065.
- Berman D.I., Meshcheryakova E.N., Bulakhova N.A. Extreme negative temperatures and body mass loss in the Siberian salamander (*Salamandrella keyserlingii*, Amphibia, Hynobiidae). *Dokl. Biol. Sci.* 2016b;468(1):137-141. DOI 10.1134/S001249661603011X.
- Bickler P.E., Buck L.T. Hypoxia tolerance in reptiles, amphibians, and fishes: Life with variable oxygen availability. *Annu. Rev. Physiol.* 2007;69(1):145-170. DOI 10.1146/annurev.physiol.69.031905.162529.
- Birol I., Behsaz B., Hammond S.A., Kucuk E., Veldhoen N., Helbing C.C. De novo transcriptome assemblies of *Rana* (*Lithobates*) *catesbeiana* and *Xenopus laevis* tadpole livers for comparative genomics without reference genomes. *PLoS One.* 2015;10(6):e0130720. DOI 10.1371/journal.pone.0130720.
- Bryant D.M., Johnson K., DiTommaso T., Tickle T., Couger M.B., Payzin-Dogru D., Lee T.J., Leigh N.D., Kuo T.-H., Davis F.G., Bateman J. A tissue-mapped axolotl *de novo* transcriptome enables identification of limb regeneration factors. *Cell Rep.* 2017;18(3):762-776. DOI 10.1016/j.celrep.2016.12.063.
- Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10(1):421. DOI 10.1186/1471-2105-10-421.
- Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884-i890. DOI 10.1093/bioinformatics/bty560.
- Christenson M.K., Trease A.J., Potluri L.P., Jezewski A.J., Davis V.M., Knight L.A., Kolok A.S., Davis P.H. De novo assembly and analysis of the northern leopard frog *Rana pipiens* transcriptome. *J. Genomics.* 2014;2:141-149. DOI 10.7150/jgen.9760.
- Eddy S.R. Accelerated profile HMM searches. *PLoS Comput. Biol.* 2011;7(10):e1002195. DOI 10.1371/journal.pcbi.1002195.
- Fu L., Niu B., Zhu Z., Wu S., Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28(23):3150-3152. DOI 10.1093/bioinformatics/bts565.
- Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 2011;29(7):644. DOI 10.1038/nbt.1883.
- Hammond S.A., Warren R.L., Vandervalk B.P., Kucuk E., Khan H., Gibb E.A., Pandoh P., Kirk H., Zhao Y., Jones M., Mungall A.J., Coope R., Pleasance S., Moore R.A., Holt R.A., Round J.M., Ohora S., Walle B.V., Veldhoen N., Helbing C.C., Birol I. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. *Nat. Commun.* 2017;8:1433. DOI 10.1038/s41467-017-01316-7.
- Katsura Y., Ikemura T., Kajitani R., Toyoda A., Itoh T., Ogata M., Miura I., Wada K., Wada Y., Satta Y. Comparative genomics of *Glandirana rugosa* using unsupervised AI reveals a high CG frequency. *Life Sci. Alliance.* 2021;4(5):e202000905. DOI 10.26508/lsa.202000905.
- Ma Y., Li B., Ke Y., Zhang Y., Zhang Y. Transcriptome analysis of *Rana chensinensis* liver under trichlorfon stress. *Ecotoxicol. Environ. Saf.* 2018;147:487-493. DOI 10.1016/j.ecoenv.2017.09.016.
- Mistry J., Chuguransky S., Williams L., Qureshi M., Salazar G.A., Sonnhammer E.L.L., Tosatto S.C.E., Paladin L., Raj S., Richardson L.J., Finn R.D., Bateman A. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419. DOI 10.1093/nar/gkaa913.
- Nilsson G.E. Long-term anoxia in crucian carp: changes in the levels of amino acid and monoamine neurotransmitters in the brain, catecholamines in chromaffin tissue, and liver glycogen. *J. Exp. Biol.* 1990;150(1):295-320. DOI 10.1242/jeb.150.1.295.
- Nilsson G.E., Lutz P.L., Jackson T.L. Neurotransmitters and anoxic survival of the brain: a comparison of anoxia-tolerant and anoxia-intolerant vertebrates. *Physiol. Zool.* 1991;64(3):638-652. DOI 10.1086/physzool.64.3.30158198.
- Nilsson G.E., Vaage J., Stenslökken K.O. Oxygen- and temperature-dependent expression of survival protein kinases in crucian carp (*Carassius carassius*) heart and brain. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 2015;308(1):R50-R61. DOI 10.1152/ajpregu.00094.2014.
- Price S.J., Garner T.W.J., Balloux F., Ruis C., Paszkiewicz K.H., Moore K., Griffiths A.G. A *de novo* assembly of the common frog (*Rana temporaria*) transcriptome and comparison of transcription following exposure to *Ranavirus* and *Batrachochytrium dendrobatidis*. *PLoS One.* 2015;10(6):e0130500. DOI 10.1371/journal.pone.0130500.
- Qiao L., Yang W., Fu J., Song Z. Transcriptome profile of the green odorless frog (*Odorrana margaretae*). *PLoS One.* 2013;8(9):e75211. DOI 10.1371/journal.pone.0075211.
- Robertson L.S., Corman R.S. Transcriptome resources for the frogs *Lithobates clamitans* and *Pseudacris regilla*, emphasizing antimicrobial peptides and conserved loci for phylogenetics. *Mol. Ecol. Resour.* 2014;14(1):178-183. DOI 10.1111/1755-0998.12164.
- Schatz M.C., Delcher A.L., Salzberg S.L. Assembly of large genomes using second-generation sequencing. *Genome Res.* 2010;20(9):1165-1173. DOI 10.1101/gr.101360.109.
- Shekhovtsov S.V., Bulakhova N.A., Tsentalovich Y.P., Zelentsova E.A., Yanshole L.V., Meshcheryakova E.N., Berman D.I. Metabolic response of the Siberian wood frog *Rana amurensis* to extreme hypoxia. *Sci. Rep.* 2020;10(1):14604. DOI 10.1038/s41598-020-71616-4.
- Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210-3212. DOI 10.1093/bioinformatics/btv351.
- Song L., Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience.* 2015;4(1):48. DOI 10.1186/s13742-015-0089-y.
- Storey K.B. Freeze tolerance in the frog, *Rana sylvatica*. *Experientia.* 1984;40(11):1261-1262. DOI 10.1007/BF01946664.
- Storey K.B., Storey J.M. Molecular physiology of freeze tolerance in vertebrates. *Physiol. Rev.* 2017;97(2):623-665. DOI 10.1152/physrev.00016.2016.
- Streicher J.W. Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations (Mead D., Saccheri I., Yung C.J., Lohse K., Lohse C., Ashmole P., Smith M., Corton C., Oliver K., Skelton J., Betteridge E., Quail M.A., Dolucan J., McCarthy S.A., Howe K., Wood J., Torrance J., Tracey A., Whiteford S., Challis R., Durbin R., Blaxter M.). The genome sequence of

- the common frog, *Rana temporaria* Linnaeus 1758. *Wellcome Open Res.* 2021;6:286. DOI 10.12688/wellcomeopenres.17296.1.
- Swenson E.R. Hypoxia and its acid-base consequences: from mountains to malignancy. *Adv. Exp. Med. Biol.* 2016;903:301-323. DOI 10.1007/978-1-4899-7678-9\_21. PMID: 27343105.
- The Uniprot Consortium. UniProt: the universal protein knowledge-base in 2021. *Nucleic Acids Res.* 2021;49(D1):D480-D489. DOI 10.1093/nar/gkaa1100.
- Xu Y.G., Chai L.H., Shi W., Wang D.D., Zhang J.Y., Xiao X.H. Transcriptome profiling and digital gene expression analysis of the skin of Dybowskii's frog (*Rana dybowskii*) exposed to *Aeromonas hydrophila*. *Appl. Microbiol. Biotechnol.* 2017;101(14):5799-5808. DOI 10.1007/s00253-017-8385-3.
- Yang W., Qi Y., Bi K., Fu J. Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: a comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R. kukunoris*. *BMC Genomics.* 2012;13:588. DOI 10.1186/1471-2164-13-588.
- Zhao L., Liu L., Wang S., Wang H., Jiang J. Transcriptome profiles of metamorphosis in the ornamented pygmy frog *Microhyla fissipes* clarify the functions of thyroid hormone receptors in metamorphosis. *Sci. Rep.* 2016;6:27310. DOI 10.1038/srep27310.

---

#### ORCID ID

D.N. Smirnov [orcid.org/0000-0002-9842-8181](https://orcid.org/0000-0002-9842-8181)  
S.V. Shekhovtsov [orcid.org/0000-0001-5604-5601](https://orcid.org/0000-0001-5604-5601)  
A.A. Shipova [orcid.org/0000-0001-7167-4481](https://orcid.org/0000-0001-7167-4481)  
G.R. Gazizova [orcid.org/0000-0003-4978-8495](https://orcid.org/0000-0003-4978-8495)  
E.I. Shagimardanova [orcid.org/0000-0003-2339-261X](https://orcid.org/0000-0003-2339-261X)

D.I. Berman [orcid.org/0000-0002-2137-1324](https://orcid.org/0000-0002-2137-1324)  
N.A. Bulakhova [orcid.org/0000-0002-3000-6476](https://orcid.org/0000-0002-3000-6476)  
E.N. Meshcheryakova [orcid.org/0000-0001-8089-4734](https://orcid.org/0000-0001-8089-4734)  
T.V. Poluboyarova [orcid.org/0000-0002-5652-0553](https://orcid.org/0000-0002-5652-0553)  
E.E. Khrameeva [orcid.org/0000-0001-6188-9139](https://orcid.org/0000-0001-6188-9139)  
S.E. Peltek [orcid.org/0000-0002-3524-0456](https://orcid.org/0000-0002-3524-0456)

**Data availability.** Raw sequencing reads and transcriptome assembly are available through NCBI BioSample (accession number PRJNA774040). The datasets described in this study are provided at GitHub ([https://github.com/Khrameeva-Lab/Rana\\_amurensis\\_transcriptome\\_2021](https://github.com/Khrameeva-Lab/Rana_amurensis_transcriptome_2021)), as well as at <http://arcuda.skoltech.ru/~d.smirnov/Frogs/>

**Acknowledgements.** This study was supported by the Russian Science Foundation grant No. 21-74-20050.

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

Received August 10, 2021. Revised November 2, 2021. Accepted November 2, 2021.