2021 • 25 • 3

Print ISSN 2500-0462
Online ISSN 2500-3259

Научный рецензируемый журнал

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.
Периодичность 8 выпусков в год
DOI 10.18699/VJ21.027

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

2021 • 25 • 3

Print ISSN 2500-0462
Online ISSN 2500-3259

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/VJ21.027

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

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

СОДЕРЖАНИЕ • 2021 • 25 • 3

VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2021 • 25 • 3

Вавиловский журнал генетики и селекции. 2021;25(3):251-259
DOI 10.18699/VJ21.028

Рибосомное профилирование как инструмент исследования трансляции у растений: основные итоги, проблемы и перспективы

Д.А. Афонников[1, 2] ✉, О.И. Синицына[1, 2], Т.С. Голубева[1, 2], Н.А. Шмаков[1, 2], А.В. Кочетов[1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ ada@bionet.nsc.ru

Аннотация. Экспрессию эукариотических генов можно регулировать на нескольких этапах, включая трансляцию мРНК. Известно, что структура мРНК способна влиять как на эффективность взаимодействия с аппаратом трансляции в целом, так и на выбор сайтов инициации трансляции. Для исследования транслируемой фракции транскриптома были разработаны экспериментальные методы анализа, наиболее информативным из которых является рибосомное профилирование (РП, Ribo-seq). Первоначально созданный для использования в дрожжевых системах, этот метод был адаптирован для трансляционных исследований на многих видах растений. Технология включает выделение полисомной фракции и высокопроизводительное секвенирование пула сегментов мРНК, связанных с рибосомами. Сравнение результатов покрытия транскриптома прочтениями, полученными по протоколу рибосомного профилирования, с аналогичным результатами по секвенированию транскриптома дает возможность оценить эффективность трансляции для каждого транскрипта. Точные положения рибосом, определенные на последовательностях мРНК, позволяют определять трансляцию открытых рамок считывания и переключение между трансляцией нескольких рамок считывания – феномен, при котором с одной матрицы РНК происходят считывание двух или более перекрывающихся рамок и биосинтез разных белков. Преимущество метода заключается в том, что он дает возможность получить количественные оценки покрытия рибосомами мРНК и может выявлять относительно редкие события трансляции. Использование этой технологии позволило классифицировать гены растений по типу регуляции их экспрессии на уровне транскрипции, трансляции или на обоих уровнях. Обнаружены особенности структуры мРНК, которые влияют на уровни трансляции: формирование квадруплексов G2 и наличие специфических мотивов в области 5′-UTR, GC-состав, наличие альтернативных стартов трансляции, влияние uORF на трансляцию нижестоящих mORF. Показано, что изменения регуляции экспрессии генов на уровне трансляции возникают в ответ на биотический и абиотический стрессы, а также в процессе развития растений. В обзоре кратко рассмотрены методология РП и перспективы ее применения для исследования структурно-функциональной организации и регуляции экспрессии генов растений.
Ключевые слова: рибосомное профилирование; Ribo-seq; RNA-seq; трансляция; растения; абиотический стресс; биотический стресс.

Для цитирования: Афонников Д.А., Синицына О.И., Голубева Т.С., Шмаков Н.А., Кочетов А.В. Рибосомное профилирование как инструмент исследования трансляции у растений: основные итоги, проблемы и перспективы. *Вавиловский журнал генетики и селекции*. 2021;25(3):251-259. DOI 10.18699/VJ21.028

Ribosomal profiling as a tool for studying translation in plants: main results, problems and future prospects

D.A. Afonnikov[1, 2] ✉, O.I. Sinitsyna[1, 2], T.S. Golubeva[1, 2], N.A. Shmakov[1, 2], A.V. Kochetov[1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
✉ ada@bionet.nsc.ru

Abstract. The expression of eukaryotic genes can be regulated at several stages, including the translation of mRNA. It is known that the structure of mRNA can affect both the efficiency of interaction with the translation apparatus in general and the choice of translation initiation sites. To study the translated fraction of the transcriptome, experimental methods of analysis were developed, the most informative of which is ribosomal profiling (RP, Ribo-seq). Originally developed for use in yeast systems, this method has been adapted for research in translation mechanisms in many plant species. This technology includes the isolation of the polysomal fraction and high-performance sequencing of a pool of mRNA fragments associated with ribosomes. Comparing the results of transcript coverage with reads obtained using the ribosome profiling with the transcriptional efficiency of genes allows the translation efficiency to be evaluated for each transcript. The exact positions of ribosomes determined on mRNA sequences allow determining the translation of open reading frames and switching between the translation of several reading frames – a pheno-

menon in which two or more overlapping frames are read from one mRNA and different proteins are synthesized. The advantage of this method is that it provides quantitative estimates of ribosome coverage of mRNA and can detect relatively rare translation events. Using this technology, it was possible to identify and classify plant genes by the type of regulation of their expression at the transcription, translation, or both levels. Features of the mRNA structure that affect translation levels have been revealed: the formation of G2 quadruplexes and the presence of specific motifs in the 5′-UTR region, GC content, the presence of alternative translation starts, and the influence of uORFs on the translation of downstream mORFs. In this review, we briefly reviewed the RP methodology and the prospects for its application to study the structural and functional organization and regulation of plant gene expression.
Key words: ribosome profiling; Ribo-seq; RNA-seq; translation; plants; abiotic stress; biotic stress.

For citation: Afonnikov D.A., Sinitsyna O.I., Golubeva T.S., Shmakov N.A., Kochetov A.V. Ribosomal profiling as a tool for studying translation in plants: main results, problems and future prospects. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):251-259. DOI 10.18699/VJ21.028 (in Russian)

Введение

Разработка и использование современных масштабных и высокопроизводительных геномных технологий привели к радикальному изменению инструментария проведения молекулярно-биологических экспериментов за последние несколько десятилетий. Так, исследование транскриптомного профиля в масштабах генома (RNA-seq) использовано для идентификации генов, экспрессия которых изменяется в ответ на сигналы окружающей среды и развития у большинства, если не у всех, сельскохозяйственных растений, а также у патогенных микробов (Kazan, Gardiner, 2018; Lanver et al., 2018; Baek et al., 2019; Zumaquero et al., 2019; Kang et al., 2020). Такие исследования обычно выявляют тысячи генов, которые при определенных условиях экспрессируются по-разному. Группировка дифференциально экспрессирующихся генов по ключевым функциональным категориям или онтологиям генов (GO) обеспечивает идентификацию главных клеточных процессов, лежащих в основе развития растений и их ответа на стресс. Однако становится все более очевидным, что мРНК разных генов не одинаковы по эффективности и специфичности трансляции и связь между количеством транскрипта и количеством синтезируемого с него белка может носить сложный характер, поскольку на экспрессию продукта гена оказывает влияние множество факторов, существенная доля которых связана с процессами трансляции мРНК (Lei et al., 2015; Merchante et al., 2015). Таким образом, хотя анализ транскриптома представляет собой необходимый этап при изучении паттернов экспрессии генов и механизмов генетического контроля различных процессов, он не является достаточным. Поэтому первостепенное значение для понимания того, какие продукты гена и в каком количестве образуются в результате его экспрессии, имеет исследование других процессов, включающих стабильность мРНК, особенности ее трансляции, стабильность полипептида, возможность посттрансляционных модификаций и т. д.

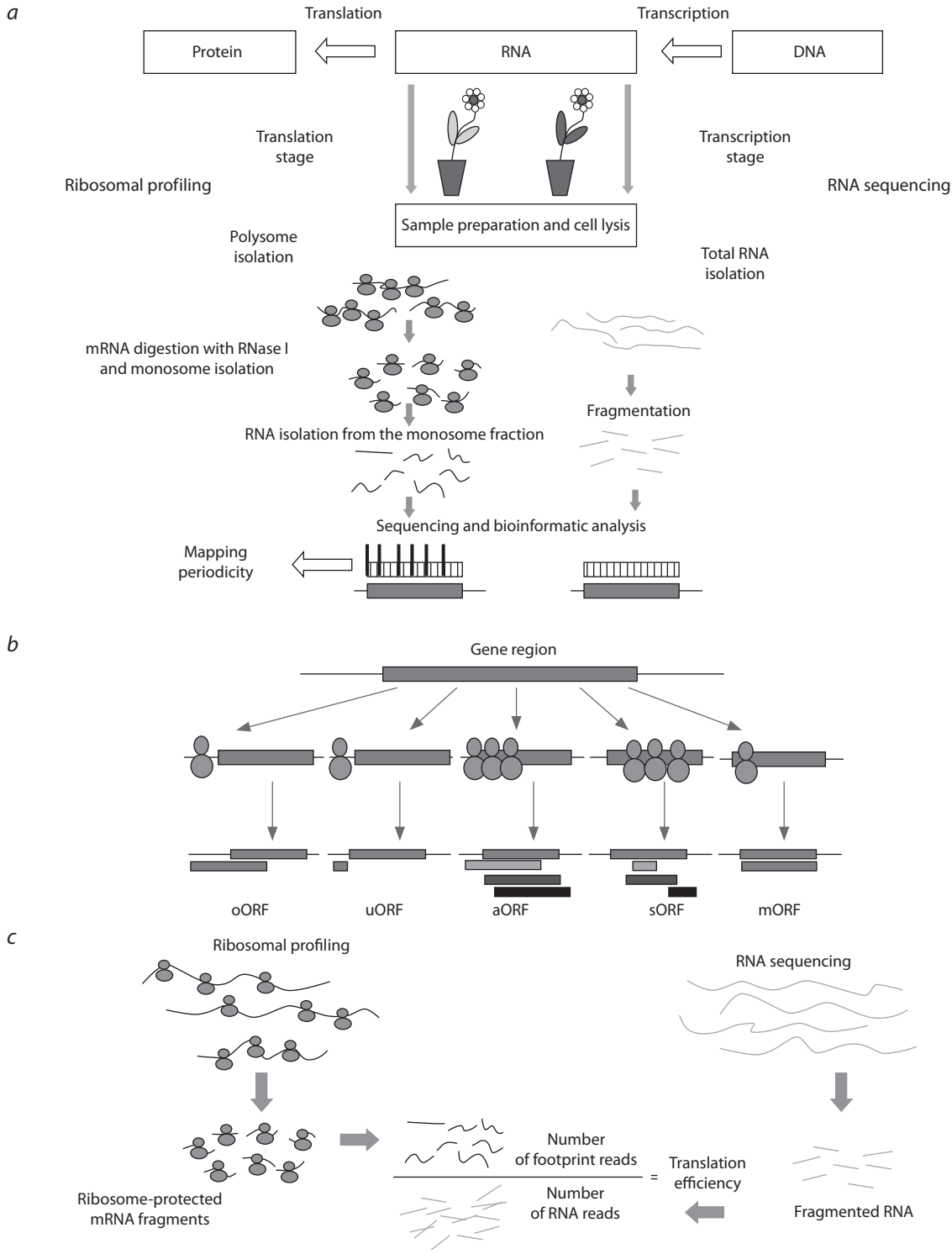
Для экспериментального изучения этого процесса в последние годы было разработано несколько высокопроизводительных подходов, например профилирование рибосом или рибо-секвенирование (далее РП, в литературе – Ribo-seq). Они дают возможность в масштабе всего транскриптома определять различные особенности процесса трансляции и оценивать их интенсивность на количественном уровне. В этой статье мы привели краткий обзор технологии РП, несколько примеров применения метода для исследования генов растений, которые позволили расширить представления о процессах, влияющих на развитие растений, а также на устойчивость к биотическим и абиотическим стрессам.

Рибосомное профилирование как инструмент исследования трансляции

Точный мониторинг процесса трансляции был технически невозможен до тех пор, когда о методологии Ribo-seq впервые сообщили в 2009 г. (Ingolia et al., 2009). Этот метод позволяет получить «отпечаток» всех транслируемых мРНК (транслатом) с помощью транскриптомной идентификации коротких (~30 нт) фрагментов мРНК, физически связанных с рибосомами (см. рисунок, *а*). Метод достаточно прост и основан на выделении полисом (т. е. транслируемой фракции мРНК), обработке РНКазой 80S рибосом с находящимся внутри сегментом мРНК, выделении пула этих сегментов и их секвенировании. Далее биоинформатический анализ позволяет определить количество таких сегментов для различных мРНК в транскриптоме, что дает важную информацию об эффективности трансляции матриц – если таких сегментов мало, то уровень синтеза белка будет низким даже при большом количестве мРНК. Не менее важно то, что позиционирование защищенных рибосомами сегментов при их выравнивании на нуклеотидную последовательность мРНК дает возможность определить сайты инициации трансляции, т. е. транслируемые открытые рамки считывания.

Первоначально разработанный для использования в дрожжевых системах (Ingolia et al., 2009), этот метод был адаптирован для трансляционных исследований на многих видах растений (Liu et al., 2013; Zoschke et al., 2013; Juntawong et al., 2014; Lei et al., 2015; Merchante et al., 2015; Hsu et al., 2016; Lukoszek et al., 2016). Протоколы Ribo-seq также были расширены для изучения процессов функционирования хлоропластных (Zoschke et al., 2013; Gawroński et al., 2018) и митохондриальных (Rooijers et al., 2013) рибосом.

В методе рибосомального профилирования ткани растений экстрагируют в буфере и обрабатывают РНКазой I. При гидролитическом расщеплении одноцепочечной мРНК остаются ее фрагменты, связанные с рибосомами. Фрагменты мРНК с рибосомами далее выделяют с помощью хроматографии или методом центрифугирования в ступенчатом градиенте сахарозы («сахарозная подушка»). Получается набор олигонуклеотидов характерной длины,



Schematic presentation of the ribosomal profiling technology in plants.

(*a*) Plant tissue is homogenized, the cells are lysed, and each sample is divided into two portions. The polysomal fraction is extracted from the first portion of the sample, the RNA is digested with RNase I to form mRNA fragments associated with individual ribosomes (monosomes) for the ribosome fingerprint method ("ribosomal footprint"). RNA is isolated from the monosomal fractions and used for library preparation and subsequent high-coverage sequencing of ribosome-bound mRNA fragments to determine the rate of instantaneous protein synthesis. The other portion of the sample is used to isolate total RNA to obtain a transcriptional profile in addition to ribosomal profiling.

(*b*) Ribosomal profiling provides experimental identification of intensely translated transcript regions and identification of new translated open reading frames (ORFs), such as upstream ORFs (uORF), overlapping ORFs (oORF), alternative ORFs (aORF), small ORFs (sORF), in addition to main ORFs (mORFs).

(*c*) Ribosomal profiling data not only provides information about the rate of instantaneous protein synthesis, but also allows determination of the relative translation efficiency of individual genes by comparing the reading intensity during sequencing of mRNA fragments of the "ribosomal footprint" with the number of specific transcripts.

~30 нт, которая приблизительно соответствует размеру сайта посадки рибосомы. Последовательности произведенных таким образом олигонуклеотидов определяют на секвенаторах нового поколения с высоким покрытием.

При сравнении результатов покрытия транскриптома прочтениями, полученными по протоколу рибосомного профилирования, с аналогичными данными, полученными для секвенирования транскриптома, можно оценить эффективность трансляции для каждого транскрипта. Такой анализ в масштабе всего транскриптома позволяет установить, как трансляция отдельных транскриптов соотносится между образцами. В основе этого метода лежит предположение, что все мРНК, связанные с рибосомой, подвергаются трансляции и, если скорость удлинения эквивалентна по всему транслатому, то средняя занятость рибосом является хорошим показателем скорости трансляции.

Рибосомное профилирование имеет много преимуществ перед некоторыми другими методами профилирования трансляции (Jackson, Standart, 2015). Одно из них заключается в том, что локализация рибосом вдоль последовательности мРНК позволяет производить дополнительный контроль собранных данных. Например, участки, связанные с рибосомами, должны быть сконцентрированы в кодирующей части мРНК, отсутствовать в области 3′-UTR и располагаться с периодичностью трех нуклеотидов, как следствие, кодонной структуры мРНК. Если в результате покрытия прочтениями библиотек Ribo-seq транскриптома такие особенности отсутствуют, то это может свидетельствовать о низком качестве полученных библиотек.

Возможность получить расположение пиков покрытия рибосомами последовательностей мРНК на основе выравнивания прочтений Ribo-seq с разрешением на уровне кодонов сделала этот метод полезным для изучения механизмов и динамики трансляции. Точные положения рибосом, определенные на последовательностях мРНК, отображают периодичность кодонов. Это свойство позволяет точно определять трансляцию открытых рамок считывания (ORF) и переключение между трансляцией нескольких рамок считывания – феномен, при котором с одной матрицы РНК происходят считывание двух или более перекрывающихся ORF и биосинтез разных белков. Это, в свою очередь, помогает идентифицировать новые механизмы контроля трансляции, такие как события инициации трансляции в кодонах, отличных от кодона AUG, трансляции uORF, трансляции малых (sORF) и альтернативных ORF (aORF), которые ранее считались некодирующими или псевдогенами (Ingolia et al., 2009, 2011; Brar et al., 2012; Stern-Ginossar et al., 2012; Hsu et al., 2016).

Другие примеры полезного использования точного позиционирования рибосом на мРНК – картирование сайтов начала трансляции с ингибиторами элонгации, наличие вышележащих открытых рамок считывания (uORF) и неканонических стартовых кодонов (Ingolia et al., 2011), более точное определение механизмов сканирования и инициации путем картирования малых рибосомных субъединиц 40S (Archer et al., 2016), а также действие специфических стрессоров или динамика рибосом после терминации (Andreev et al., 2017).

Дополнительная особенность метода – он позволяет получить количественные оценки покрытия рибосомами мРНК и может определять относительно редкие события трансляции (см. рисунок, *б*). Среднее нормализованное (для длины ORF и глубины секвенирования) количество защищенных фрагментов рибосом (ribosome protected fragments, RPF), обнаруженное с помощью профилирования, обеспечивает оценку интенсивности синтеза белка (Ingolia et al., 2009). Кроме того, профилирование рибосом в сочетании с секвенированием РНК тех же образцов дает информацию об эффективности трансляции *in vivo* (см. рисунок, *в*), определяемой как скорость трансляции мРНК. Ее можно рассчитать, разделив среднюю плотность участков рибосомы данного гена на уровень экспрессии его мРНК, оцененный на основе анализа RNA-seq (Ingolia et al., 2009).

Важной особенностью рибосомного профилирования также является то, что этот метод может быть адаптирован для изучения самых разных типов клеток или тканей любых организмов с очень небольшими модификациями из-за консервативности молекулярных и биофизических свойств рибосом, хотя для применения этой технологии у различных организмов может потребоваться техническая оптимизация (Brar, Weissman, 2015).

Основные итоги применения технологии рибосомного профилирования у растений

Экспрессию генов во время адаптации к биотическим и абиотическим факторам стресса, а также в процессе развития интенсивно изучали на уровне транскриптомов растений и различных генов-кандидатов, которыми можно манипулировать для повышения устойчивости к стрессу. С появлением технологий рибосомного профилирования эти данные могут быть сопоставлены с результатами анализа профилей трансляции для оценки вклада различных молекулярных механизмов трансляции в регуляцию экспрессии генов.

Трансляционная регуляция экспрессии генов в условиях абиотического стресса

Абиотический стресс, такой как повышенные или пониженные температуры, или повышенная или пониженная интенсивность освещения, засуха, засоление или избыточное увлажнение почвы, оказывает значительное влияние на физиологическое состояние растений и, следовательно, на экспрессию их генов. Поэтому среди первых приложений метода РП были эксперименты по анализу реакции растений на абиотические стрессы. Метод РП был особенно полезен для понимания потенциальной роли, которую играют uORF в регуляции трансляции во время быстрых реакций на внешние стимулы, такие как кислородное голодание (гипоксия), тепловой стресс и водная депривация.

Ранее предполагалось, что uORF участвуют в регуляции генов в ответ на различные изменения окружающей среды (Hanfrey et al., 2002; Imai et al., 2006; Alatorre-Cobos et al., 2012). Однако из-за отсутствия экспериментальных данных оставалось неясным, транслируются ли uORF в белки. Профилирование рибосом позволяет идентифицировать uORF и другие регуляторные области, такие как

квадруплексы G2 (третичные структуры ДНК, состоящие из тетрад гуанина). Анализ адаптивных ответов на тепловой стресс у *Arabidopsis thaliana* с помощью метода РП в сочетании с секвенированием РНК выявил, что экспрессия генов перепрограммируется во время длительного теплового стресса за счет преимущественной трансляции генов, содержащих квадруплексы G2 в их 5′-UTR. Это было очевидно из корреляции между плотностью считывания RPF по структурам квадруплексов G2 и повышенными уровнями экспрессии нижележащих основных ORF (mORF) (Lukoszek et al., 2016).

Показано, что гипоксия оказывает крайне негативное влияние на эффективность трансляции у *A. thaliana* с примерно 100-кратным ее снижением для некоторых мРНК (Juntawong et al., 2014). Это в основном объясняется снижением инициации трансляции из-за уменьшения занятости рибосом в стартовых кодонах транскриптов генов, не чувствительных к гипоксии (Juntawong et al., 2014). Эти данные расширили понимание механизмов трансляции, подтвердив, что в нормальных условиях роста uORF снижают трансляцию многих mORF. Напротив, для небольшого числа проанализированных генов uORF не влияют на трансляцию mORF в условиях низкого содержания кислорода. P. Juntawong с коллегами (2014) установили, что «отпечатки рибосом» на мРНК после ее расщепления нуклеазой можно использовать для профилирования рибосом у растений.

Ограничение количества влаги также приводит к глобальному изменению уровней экспрессии генов, регулируемых как на транскрипционном, так и на трансляционном уровне. L. Lei с коллегами (2015) использовали профилирование рибосом для выяснения регуляции трансляции экспрессии генов в ответ на засуху у кукурузы. Они установили, что кратные изменения транскрипции, вызванные засухой, умеренно коррелировали с трансляционными изменениями. В их работе показано, что 41 % всех генов, чувствительных к засухе (т. е. генов, экспрессия которых регулируется на уровне транскрипции, трансляции или на обоих уровнях), регулируется на уровне транскриптции/трансляции несогласованно. Это указывает на то, что в условиях засухи регуляция экспрессии происходит независимыми путями: для одних – на уровне транскрипции, для других – на уровне трансляции. Авторы также сообщили, что на эффективность трансляции влияют такие характеристики последовательности, как содержание нуклеотидов G, C, длина кодирующих последовательностей и нормализованная минимальная свободная энергия, определяющая стабильность последовательности вторичных структур (Lei et al., 2015).

Один из ключевых гормонов стрессового ответа растений – этилен. Путь передачи его сигналов пересекается с путями передачи сигналов других фитогормонов в процессе опосредованного ответа как на биотические (Schenk et al., 2000), так и на абиотические стрессы (Abeles et al., 2012). Исследование, в котором проведено сравнение этилен-индуцированных состояний транскриптома и трансляторма у *A. thaliana* с использованием комбинированных подходов РНК-seq и Ribo-seq, показало, что зависимость между оценками количества транскриптов и количества их фрагментов, связанных с рибосомами, является слабой (коэффициент детерминации $R^2 = 0.22$) (Merchante et al., 2015). В этой работе также продемонстрировано, что после воздействия этилена изменяются активация транскрипции и эффективность трансляции регуляторов передачи сигналов этилена, таких как EBF1 и EBF2, что указывает на ключевую роль транскрипционной и трансляционной регуляции экспрессии генов ответа на этилен. Изменение эффективности трансляции EBF1 в ответ на обработку этиленом опосредуется 3′-UTR-областью гена *EBF1* в присутствии функционального *EIN2*, что не требует присутствия комплекса EIN3/EIL1 или других ключевых транскрипционных факторов ответа на этилен, как при регуляции на уровне транскрипции (Merchante et al., 2015). В данном случае профилирование рибосом выявило ключевой компонент в регуляции передачи сигналов этилена, который был упущен из виду при использовании только транскриптомных подходов.

Трансляционная регуляция экспрессии генов в условиях биотического стресса

В процессе ответа растений на биотический стресс происходит транскрипционное репрограммирование большого количества генов (Schenk et al., 2000). Однако о трансляционном репрограммировании во время иммунного ответа растений было известно очень мало. G. Xu с коллегами (2017) выполнили глобальное рибосомное профилирование растений *A. thaliana*, обработанных полипептидом elf18, который содержит первые 18 аминокислот белка бактериального фактора элонгации Tu. Пептид elf18 – молекулярный паттерн, ассоциированный с патогенами (pathogen associated molecular pattern, PAMP), он идентифицируется при помощи паттерн-распознающих рецепторов растений. В результате этого запускается PAMP-активируемый иммунитет (pattern triggered immunity, PTI). Показано, что при таком ответе экспрессия ряда генов регулируется на трансляционном уровне. При этом для мРНК этих генов uORF могут оказывать как положительный, так и отрицательный эффект на трансляцию нижестоящих mORF. Эти исследователи обнаружили также богатый пуринами вышестоящий элемент, называемый R-мотивом, в 5′-UTR-области генов с повышенной эффективностью трансляции после обработки растений elf18. Устранение репрессирующего эффекта R-мотива позволяет активировать гены иммунного ответа, поэтому R-мотив, по-видимому, важен для репрессии экспрессии генов в пути PTI у арабидопсиса (Xu et al., 2017). Эта работа подтвердила, что репрограммирование трансляции происходит на ранней стадии во время защитного ответа, скорее всего, до основных транскрипционных событий, и что в процессе активации PAMP-активируемого иммунитета транскрипционные и трансляционные изменения слабо коррелируют.

Регуляция трансляции в процессе иммунного ответа, активируемого эффектором (effector triggered immunity, ETI), пока еще плохо изучена. Поэтому, чтобы прояснить детали этого механизма и сравнить особенности трансляционной регуляции растений в процессе иммунитета, активируемого эффектором и паттернами, H. Yoo с коллегами (2019) выполнили полногеномное профилирование рибосом в ответ на бактериальный патоген *Pseudomonas syringae*

pv. *maculicola*, несущий эффекторный ген *AvrRpt2*, который кодирует белок, распознаваемый рецептором RPS2 у *A. thaliana*. Полученные данные демонстрируют, что в процессе паттерн-активируемого иммунного ответа регуляция трансляции четко скоординирована с ответом транскрипции. Отсутствие потенциальных мотивов консенсусной последовательности в нетранслируемых областях генов, меняющих свой экспрессию в ответ на стресс, предполагает, что скоординированное трансляционное изменение в процессе ETI происходит посредством модификаций трансляционного аппарата, а не через консенсусные последовательности в РНК. В исследовании также продемонстрировано, что регуляция трансляции в процессе ETI аналогична для различных иммунных рецепторов и что скоординированная регуляция генов, участвующих в нескольких метаболических путях, обеспечивает координацию метаболических изменений с иммунным ответом у растений (Yoo et al., 2019). Авторы показали, что, в распознавании *AvrRpt2* и *AvrRpm1* участвуют разные рецепторы хозяина (RPS2 и RPM1 соответственно), однако наборы дифференциально экспрессирующихся генов сигнальных путей, которые запускают эти рецепторы, существенно перекрываются. Так, общими для этих путей являются 50 % генов с повышенным уровнем транскрипции и 5 % – с пониженным. Что касается генов, у которых значимо меняется трансляционная активность, то их общая доля составляет выше 75 %.

Трансляционная регуляция экспрессии в процессе развития растений

Трансляционный контроль регуляции генов также важен для растений во время их развития для обеспечения синтеза онтогенетических и тканеспецифичных генных продуктов (Jiao, Meyerowitz, 2010; Mustroph, Bailey-Serres, 2010). Например, чтобы оценить трансляционную регуляцию экспрессии генов во время развития, опосредованного светом (фотоморфогенез) у *A. thaliana*, M.-J. Liu с коллегами (2013) использовали метод профилирования рибосом и картировали их положение в процессе трансляции на мРНК по всему геному в условиях света и темноты. Результаты показали, что трансляционная эффективность основных ORF была ниже в генах, содержащих транслируемые uORF, чем в генах без таковых. Авторы также сообщили, что у генов, являющихся мишенями миРНК, уровень трансляции существенно понижен в силу равномерного понижения покрытия рибосомами кодирующих последовательностей. Это исследование показало важную роль uORF и миРНК в регуляции трансляции в процессе фотоморфогенеза.

Основываясь на этом исследовании, Y. Kurihara с коллегами (2018) применили подход к рибосомному профилированию у *A. thaliana* и продемонстрировали, что после воздействия синего света растения используют альтернативные стартовые сайты транскрипции, чтобы обойти опосредованное uORF ингибирование экспрессии генов. Это позволяет поддерживать экспрессию генов, регулируемую светом, на высоком уровне.

Метод рибосомного профилирования был применен недавно для сравнения эффективности трансляции в связи с изменениями уровней транскрипции у сои на разных стадиях развития семян (Shamimuzzaman, Vodkin, 2018). Авторы выявили, что эффективность трансляции у многих генов в процессе развития семян изменяется. Повышенная эффективность трансляции на отдельных стадиях развития семян была обнаружена для генов, связанных с развитием семян (например, генов протеаз, пептидаз и 2S альбуминов), что представляет уникальные особенности изменения регуляции трансляции в процессе развития (Shamimuzzaman, Vodkin, 2018). Эти результаты подчеркивают важность изучения регуляции экспрессии генов на уровне трансляции, для того чтобы установить, как развитие органов растений контролируется на молекулярном уровне.

Обнаружение трансляционных событий из неаннотированных генов растений

Несмотря на значительные усилия по расшифровке и анализу геномов растений, их последовательности содержат множество ORF, которые остаются неаннотированными. Для таких ORF неизвестно, являются ли они псевдогенами, продуцируют ли некодирующие РНК (нкРНК) или функциональные белки. Оказалось, что технологии рибосомального профилирования можно применять и для улучшения аннотации генома. В экспериментах по профилированию рибосом у *A. thaliana* модификация буферов, используемых для выделения фрагментов мРНК, связанных с рибосомами, позволила определить новые sORF, которые ранее считались нкРНК. Интересно, что многие из этих sORF эволюционно консервативны и содержат неканонические кодоны инициации трансляции, такие как CUG или ACG (Hsu et al., 2016).

В исследовании (Wu et al., 2019) сборка транскриптомов на основе референсного генома и профилирование рибосом были использованы для улучшения аннотации генома томата (*Solanum lycopersicum*). Этот метод дал возможность идентифицировать сотни новых sORF из неаннотированных транскриптов, которые эволюционно консервативны. В данном исследовании также были сформированы транслированные последовательности для sORF, uORF из 5′-UTR-областей генов, кодирующих белки. Верификация результатов, полученных с помощью рибосомального профилирования методами протеомики, показала, что некоторые из этих sORF продуцируют стабильные пептиды. Анализ обогащения аннотации этих транскриптов терминами генной онтологии выявил, что их функция связана с фосфорилированием/дефосфорилированием белков, сигнальными путями. Все это демонстрирует, что регуляция трансляции выполняет высокоуровневую регуляторную функцию в процессе жизнедеятельности клетки. Интересно, что H.-Y.L. Wu с коллегами (2019) обнаружили, что у томата экспрессия генов также регулируется глобально как на транскрипционном, так и на трансляционном уровне за счет микроРНК. Таким образом, применение метода профилирования рибосом оказывается ценным для подтверждения предсказания новых открытых рамок считывания и сайтов инициации транскриптов в масштабе всего генома (Willems et al., 2017).

Указанные выше исследования установили неканонические трансляционные события и позволили идентифи-

цировать ранее не аннотированные малые белки, которые, вероятно, выполняют важные функции в растениях. Все это демонстрирует полезность экспериментов по рибосомному профилированию для более глубокой аннотации генов в геномах растений.

Проблемы и ограничения для применения методологии рибосомного профилирования у растений

Несмотря на множество преимуществ, технология Riboseq в настоящее время сталкивается с некоторыми серьезными техническими трудностями. Среди них – необходимость устранения потенциальных загрязнителей и надежное выравнивание коротких прочтений с эталонным геномом или транскриптомами. В этом разделе мы кратко обсудим эти проблемы и предложим возможные пути их решения.

Один из источников шума в результатах рибосомального профилирования – контаминация рибосомной РНК (рРНК) на стадии нуклеазной обработки. Она существенно затрудняет извлечение данных об информативных последовательностях, полученных в экспериментах по профилированию рибосом (Ingolia et al., 2009). Такая контаминация фрагментов мРНК фрагментами рРНК может быть устранена на основе специфической гибридизации или при помощи коммерческих наборов обеднения библиотек специфическими последовательностями РНК, а также и сочетанием обоих методов (McGlincy, Ingolia, 2017). С другой стороны, для предотвращения лигирования линкера с избытком 5.8S рРНК можно использовать маскирующие олигонуклеотиды (Faridani et al., 2016). Загрязнения рРНК можно устранить на этапе биоинформатической обработки, но этот подход требует получения большего покрытия при секвенировании.

Современные методы анализа транскриптомов, основанные на массовом секвенировании, обычно базируются на коротких прочтениях размером 100–150 п. н. для фрагментов с одиночным или парными концами (Wang et al., 2009), в то время как в более ранних подходах RNA-seq применяли прочтения длиной 25 и 32 п. н. (Marioni et al., 2008; Mortazavi et al., 2008). Так как секвенирование фрагментов, связанных с рибосомами, осуществляется короткими прочтениями, то это значительно увеличивает вероятность их выравнивания на множественные участки генома, что может приводить к возможным ошибкам в анализе результатов профилирования (Chhangawala et al., 2015). Кроме того, на уникальность выравнивания коротких прочтений более существенное влияние оказывают ошибки секвенирования.

Проблема множественного выравнивания прочтений в случае анализа транскриптомов растений усугубляется еще и тем, что большая часть существующих видов растений – полиплоиды (Wood et al., 2009). Например, у мягкой пшеницы (*Triticum aestivum*), аллогексаплоида, геном включает три субгенома, A, B и D, в которых всего ~100 000 генов. Около 85 % генома пшеницы – повторы, представленные многочисленными мобильными элементами (Ramírez-González et al., 2018). Во многих случаях последовательности генов гомеологов очень похожи, и чтобы получить выравнивание прочтений до совпадения, специфичного для гомеолога, требуется их большая длина (Pfeifer et al., 2014). Таким образом, изучение дифференциальной трансляции гомеологических генов затруднено, но может быть улучшено с помощью недавно разработанного подхода классификации субгеномов, при котором считывания сопоставляются с каждым субгеномом отдельно (Kuo et al., 2018). С другой стороны, проблема множественного картирования коротких прочтений может быть решена с помощью специальных инструментов, таких как алгоритм разрешения множественного картирования для фильтрации выравнивания (Kahles et al., 2016).

Эксперименты по профилированию рибосом генерируют огромное количество данных, что создает значительную проблему для их анализа (Calviello et al., 2016). В дополнение к существующим инструментам для обработки последовательностей и выравнивания были созданы пакеты последующего анализа для идентификации трехнуклеотидной периодичности, дифференциальной трансляции и занятости кодонов: riboSeqR (Chung et al., 2015), riboWaltz (Lauria et al., 2018). На сегодняшний день разработаны два веб-сервера, RiboGalaxy (Michel et al., 2016) и riboviz (Carja et al., 2017), объединяющие набор инструментов, необходимых для анализа данных рибосомального профилирования, начиная с контроля качества исходной последовательности до финальной визуализации данных (Wang et al., 2017).

Дальнейшее изучение методов анализа данных рибосомального профилирования и подбор параметров выравнивания и покрытия для улучшения точности определения дифференциальной трансляции в значительной степени помогут при анализе и интерпретации результатов этих экспериментов.

Заключение

Возможные перспективы методологии рибосомного профилирования у растений

Большая часть современной информации о профилировании рибосом получена из нерастительных видов. Следовательно, проведение подобных исследований на растениях в различных условиях необходимо для обнаружения генов, экспрессия которых регулируется только трансляционно и до сих пор не выявлялась при анализе транскриптомов. Хотя профилирование рибосом – относительно новый метод, улучшения в этой технологии уже появляются. Так, этот метод можно использовать для изучения специфической трансляции органелл, кодируемой ядерными генами (Jan et al., 2014). В настоящее время все большее внимание привлекают методы анализа транскриптомов единичных клеток (Saliba et al., 2014). Они широко применяются в исследованиях на животных и имеют большой потенциал для понимания функции генов для единичных клеток в растениях (Efroni, Birnbaum, 2016). Сочетание этой новой методики с профилированием рибосом и такими подходами, как аффинная очистка транслирующих рибосом (Heiman et al., 2014), может иметь огромное значение при изучении регуляции трансляции, специфичной для определенного типа клеток (Mironova, Xu, 2019).

Дальнейшее улучшение определенных технических этапов экспериментов по профилированию рибосом

могло бы способствовать более широкому внедрению этой техники. Например, в настоящее время подготовка библиотеки для глубокого секвенирования включает в себя несколько этапов лигирования и занимает много времени (обычно несколько дней). Кроме того, требование относительно больших количеств начальной РНК для подготовки библиотек может создавать проблемы.

Таким образом, новый подход, называемый профилированием рибосом без лигирования, разработанный на мышах, может быть исследован на растениях, поскольку этот метод не требует лигирования, а создание библиотеки производится в течение дня с использованием всего лишь 1 нг РНК (Hornstein et al., 2016). Профилирование рибосом перспективно для идентификации rQTL по всему геному (рибосомный QTL), подобно тому как проводится анализ eQTL (экспрессионные QTL).

Список литературы / References

Abeles F.B., Morgan P.W., Saltveit Jr. M.E. Ethylene in Plant Biology. San Diego, CA: Academic Press, 2012.

Alatorre-Cobos F., Cruz-Ramírez A., Hayden C.A., Pérez-Torres C.-A., Chauvin A.-L., Ibarra-Laclette E., Alva-Cortés E., Jorgensen R.A., Herrera-Estrella L. Translational regulation of *Arabidopsis XIPOTL1* is modulated by phosphocholine levels via the phylogenetically conserved upstream open reading frame. *J. Exp. Bot.* 2012;63(14): 5203-5221. DOI 10.1093/jxb/ers180.

Andreev D.E., O'Connor P.B., Loughran G., Dmitriev S.E., Baranov P.V., Shatsky I.N. Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res.* 2017; 45(2):513-526. DOI 10.1093/nar/gkw1190.

Archer S.K., Shirokikh N.E., Beilharz T.H., Preiss T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature.* 2016;535(7613):570-574. DOI 10.1038/nature18647.

Baek Y.S., Goodrich L.V., Brown P.J., James B.T., Moose S.P., Lambert K.N., Riechers D.E. Transcriptome profiling and genome-wide association studies reveal GSTs and other defense genes involved in multiple signaling pathways induced by herbicides Safener in grain sorghum. *Front. Plant Sci.* 2019;10:192. DOI 10.3389/fpls.2019.00192.

Brar G.A., Weissman J.S. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 2015; 16:651-664. DOI 10.1038/nrm4069.

Brar G.A., Yassour M., Friedman N., Regev A., Ingolia N.T., Weissman J.S. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science.* 2012;335(6068):552-557. DOI 10.1126/science.1215110.

Calviello L., Mukherjee N., Wyler E., Zauber H., Hirsekorn A., Selbach M., Landthaler M., Obermayer B., Ohler U. Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods.* 2016;13:165-170. DOI 10.1038/nmeth.3688.

Carja O., Xing T., Wallace E.W.J., Plotkin J.B., Shah P. Riboviz: analysis and visualization of ribosome profiling datasets. *BMC Bioinformatics.* 2017;18:461. DOI 10.1186/s12859-017-1873-8.

Chhangawala S., Rudy G., Mason C.E., Rosenfeld J.A. The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biol.* 2015;16:131. DOI 10.1186/s13059-015-0697-y.

Chung B.Y., Hardcastle T.J., Jones J.D., Irigoyen N., Firth A.E., Baulcombe D.C., Brierley I. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA.* 2015;21:1731-1745. DOI 10.1261/rna.052548.115.

Efroni I., Birnbaum K.D. The potential of single-cell profiling in plants. *Genome Biol.* 2016;17(1):65. DOI 10.1186/s13059-016-0931-2.

Faridani O.R., Abdullayev I., Hagemann-Jensen M., Schell J.P., Lanner F., Sandberg R. Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* 2016;34:1264-1266. DOI 10.1038/nbt.3701.

Gawroński P., Jensen P.E., Karpiński S., Leister D., Scharff L.B. Pausing of chloroplast ribosomes is induced by multiple features and is linked to the assembly of photosynthetic complexes. *Plant Physiol.* 2018;176:2557-2569. DOI 10.1104/pp.17.01564.

Hanfrey C., Franceschetti M., Mayer M.J., Illingworth C., Michael A.J. Abrogation of upstream open reading frame-mediated translational control of a plant S-adenosylmethionine decarboxylase results in polyamine disruption and growth perturbations. *J. Biol. Chem.* 2002; 277(46):44131-44139. DOI 10.1074/jbc.M206161200.

Heiman M., Kulicke R., Fenster R.J., Greengard P., Heintz N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* 2014;9(6):1282-1291. DOI 10.1038/nprot.2014.085.

Hornstein N., Torres D., Das Sharma S., Tang G., Canoll P., Sims P.A. Ligation-free ribosome profiling of cell type-specific translation in the brain. *Genome Biol.* 2016;17:149. DOI 10.1186/s13059-016-1005-1.

Hsu P.Y., Calviello L., Wu H.-Y.L., Li F.-W., Rothfels C.J., Ohler U., Benfey P.N. Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis. Proc. Natl. Acad. Sci. USA.* 2016;113(45):E7126-E7135. DOI 10.1073/pnas.1614788113.

Imai A., Hanzawa Y., Komura M., Yamamoto K.T., Komeda Y., Takahashi T. The dwarf phenotype of the *Arabidopsis acl5* mutant is suppressed by a mutation in an upstream ORF of a bHLH gene. *Development.* 2006;133:3575-3585. DOI 10.1242/dev.02535.

Ingolia N.T., Ghaemmaghami S., Newman J.R., Weissman J.S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science.* 2009;324(5924):218-223. DOI 10.1126/science.1168978.

Ingolia N.T., Lareau L.F., Weissman J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011;147:789-802. DOI 10.1016/j.cell.2011.10.002.

Jackson R., Standart N. The awesome power of ribosome profiling. *RNA.* 2015;21:652-654. DOI 10.1261/rna.049908.115.

Jan C.H., Williams C.C., Weissman J.S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science.* 2014;346(6210):1257521. DOI 10.1126/science.1257521.

Jiao Y., Meyerowitz E.M. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* 2010;6:419. DOI 10.1038/msb.2010.76.

Juntawong P., Girke T., Bazin J., Bailey-Serres J. Translational dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA.* 2014;111(1):E203-E212. DOI 10.1073/pnas.1317811111.

Kahles A., Behr J., Rätsch G. MMR: a tool for read multi-mapper resolution. *Bioinformatics.* 2016;32(5):770-772. DOI 10.1093/bioinformatics/btv624.

Kang W.-H., Sim Y.M., Koo N., Nam J.-Y., Lee J., Kim N., Jang H., Kim Y.-M., Yeom S.-I. Transcriptome profiling of abiotic responses to heat, cold, salt, and osmotic stress of *Capsicum annuum* L. *Sci. Data.* 2020;7:1-7. DOI 10.1038/s41597-020-0352-7.

Kazan K., Gardiner D.M. Transcriptomics of cereal – *Fusarium graminearum* interactions: what we have learned so far. *Mol. Plant Pathol.* 2018;19(3):764-778. DOI 10.1111/mpp.12561.

Kuo T.C.Y., Hatakeyama M., Tameshige T., Shimizu K.K., Sese J. Homeolog expression quantification methods for allopolyploids. *Brief. Bioinform.* 2018;21(2):395-407. DOI 10.1093/bib/bby121.

Kurihara Y., Makita Y., Kawashima M., Fujita T., Iwasaki S., Matsui M. Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis. Proc. Natl. Acad. Sci. USA.* 2018;115(30):7831-7836. DOI 10.1073/pnas.1804971115.

Lanver D., Müller A.N., Happel P., Schweizer G., Haas F.B., Franitza M., Pellegrin C., Reissmann S., Altmüller J., Rensing S.A. The biotrophic development of *Ustilago maydis* studied by RNA-

seq analysis. *Plant Cell.* 2018;30(2):300-323. DOI 10.1105/tpc.17.00764.

Lauria F., Tebaldi T., Bernabò P., Groen E.J.N., Gillingwater T.H., Viero G. riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput. Biol.* 2018;14. e1006169-e1006169.

Lei L., Shi J., Chen J., Zhang M., Sun S., Xie S., Li X., Zeng B., Peng L., Hauck A. Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J.* 2015;84(6):1206-1218. DOI 10.1111/tpj.13073.

Liu M.-J., Wu S.-H., Wu J.-F., Lin W.-D., Wu Y.-C., Tsai T.-Y., Tsai H.-L., Wu S.-H. Translational landscape of photomorphogenic *Arabidopsis*. *Plant Cell.* 2013;25(10):3699-3710. DOI 10.1105/tpc.113.114769.

Lukoszek R., Feist P., Ignatova Z. Insights into the adaptive response of *Arabidopsis thaliana* to prolonged thermal stress by ribosomal profiling and RNA-Seq. *BMC Plant Biol.* 2016;16:221. DOI 10.1186/s12870-016-0915-0.

Marioni J.C., Mason C.E., Mane S.M., Stephens M., Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 2008;18:1509-1517. DOI 10.1101/gr.079558.108.

McGlincy N.J., Ingolia N.T. Transcriptome-wide measurement of translation by ribosome profiling. *Methods.* 2017;126:112-129. DOI 10.1016/j.ymeth.2017.05.028.

Merchante C., Brumos J., Yun J., Hu Q., Spencer K.R., Enríquez P., Binder B.M., Heber S., Stepanova A.N., Alonso J.M. Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. *Cell.* 2015;163(3):684-697. DOI 10.1016/j.cell.2015.09.036.

Michel A.M., Mullan J.P.A., Velayudhan V., O'Connor P.B.F., Donohue C.A., Baranov P.V. RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol.* 2016;13(3):316-319. DOI 10.1080/15476286.2016.1141862.

Mironova V., Xu J. A single-cell view of tissue regeneration in plants. *Curr. Opin. Plant Biol.* 2019;52:149-154. DOI 10.1016/j.pbi.2019.09.003.

Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods.* 2008;5:621-628. DOI 10.1038/nmeth.1226.

Mustroph A., Bailey-Serres J. The *Arabidopsis* translatome cell-specific mRNA atlas: Mining suberin and cutin lipid monomer biosynthesis genes as an example for data application. *Plant Signal. Behav.* 2010;5(3):320-324. DOI 10.4161/psb.5.3.11187.

Pfeifer M., Kugler K.G., Sandve S.R., Zhan B., Rudi H., Hvidsten T.R., Mayer K.F.X., Olsen O.-A. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science.* 2014;345(6194):1250091. DOI 10.1126/science.1250091.

Ramírez-González R.H., Borrill P., Lang D., Harrington S.A., Brinton J., Venturini L., Davey M., Jacobs J., van Ex F., Pasha A., Khedikar Y., Robinson S.J., Cory A.T., Florio T., Concia L., Juery C., Schoonbeek H., Steuernagel B., Xiang D., Ridout C.J., Chalhoub B., Mayer K.F.X., Benhamed M., Latrasse D., Bendahmane A., International Wheat Genome Sequencing Consortium; Wulff B.B.H., Appels R., Tiwari V., Datla R., Choulet F., Pozniak C.J., Provart N.J., Sharpe A.G., Paux E., Spannagl M., Bräutigam A., Uauy C. The transcriptional landscape of polyploid wheat. *Science.* 2018;361(6403):eaar6089. DOI 10.1126/science.aar6089.

Rooijers K., Loayza-Puch F., Nijtmans L.G., Agami R. Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat. Commun.* 2013;4:2886. DOI 10.1038/ncomms3886.

Saliba A.E., Westermann A.J., Gorski S.A., Vogel J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* 2014;42(14):8845-8860. DOI 10.1093/nar/gku555.

Schenk P.M., Kazan K., Wilson I., Anderson J.P., Richmond T., Somerville S.C., Manners J.M. Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA.* 2000;97(21):11655-11660. DOI 10.1073/pnas.97.21.11655.

Shamimuzzaman M., Vodkin L. Ribosome profiling reveals changes in translational status of soybean transcripts during immature cotyledon development. *PLoS One.* 2018;13(3):e0194596. DOI 10.1371/journal.pone.0194596.

Stern-Ginossar N., Weisburd B., Michalski A., Le V.T.K., Hein M.Y., Huang S.-X., Ma M., Shen B., Qian S.-B., Hengel H. Decoding human cytomegalovirus. *Science.* 2012;338(6110):1088-1093. DOI 10.1126/science.1227919.

Wang H., Wang Y., Xie Z. Computational resources for ribosome profiling: from database to Web server and software. *Brief. Bioinform.* 2017;20(1):144-155. DOI 10.1093/bib/bbx093.

Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009;10:57-63. DOI 10.1038/nrg2484.

Willems P., Ndah E., Jonckheere V., Stael S., Sticker A., Martens L., van Breusegem F., Gevaert K., van Damme P. N-terminal proteomics assisted profiling of the unexplored translation initiation landscape in *Arabidopsis thaliana*. *Mol. Cell. Proteomics.* 2017;16(6):1064-1080. DOI 10.1074/mcp.M116.066662.

Wood T.E., Takebayashi N., Barker M.S., Mayrose I., Greenspoon P.B., Rieseberg L.H. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. USA.* 2009;106(33):13875-13879. DOI 10.1073/pnas.0811575106.

Wu H.-Y.L., Song G., Walley J.W., Hsu P.Y. The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiol.* 2019;181(1):367-380. DOI 10.1104/pp.19.00541.

Xu G., Greene G.H., Yoo H., Liu L., Marqués J., Motley J., Dong X. Global translational reprogramming is a fundamental layer of immune regulation in plants. *Nature.* 2017;545:487.

Yoo H., Greene G.H., Yuan M., Xu G., Burton D., Liu L., Marqués J., Dong X. Translational regulation of metabolic dynamics during effector-triggered immunity. *Mol. Plant.* 2019;13(1):88-98. DOI 10.1016/j.molp.2019.09.009.

Zoschke R., Watkins K.P., Barkan A. A rapid ribosome profiling method elucidates chloroplast ribosome behavior *in vivo*. *Plant Cell.* 2013;25(6):2265-2275. DOI 10.1105/tpc.113.111567.

Zumaquero A., Kanematsu S., Nakayashiki H., Matas A., Martínez-Ferri E., Barceló-Muñóz A., Pliego-Alfaro F., López-Herrera C., Cazorla F., Pliego C. Transcriptome analysis of the fungal pathogen *Rosellinia necatrix* during infection of a susceptible avocado rootstock identifies potential mechanisms of pathogenesis. *BMC Genomics.* 2019;20:1016. DOI 10.1186/s12864-019-6387-5.

ORCID ID

D.A. Afonnikov orcid.org/0000-0001-9738-1409
A.V. Kochetov orcid.org/0000-0003-3151-5181

Acknowledgements. The work was supported by the Russian Science Foundation, project 18-14-00293.

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

Received December 1, 2020. Revised December 8, 2020. Accepted December 8, 2020.

Вавиловский журнал генетики и селекции. 2021;25(3):260-268
DOI 10.18699/VJ21.029

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

Intraspecific diversity of durum wheat (*Triticum durum* Desf.): a unified classification

O.A. Lyapunova

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

Abstract. The Department of Wheat Genetic Resources of the All-Russian Research Institute of Plant Genetic Resources (VIR) had developed and published in 1979 a classification of the genus *Triticum* L., which is based on the genomic composition of species and the presence or absence of a number of main genes that govern the "classification" traits. The grounds have been laid by F. Körnicke and J. Percival, and supplemented by N.I. Vavilov and K.A. Flaksberger. The classification, which is most often referred to as the "Classification of *Triticum* by Dorofeev et al.", belongs to a number of the main modern classifications of the genus. This is the world's first standardized system that contains all known intraspecific (infraspecific) taxa of wild and cultivated wheat species. A detailed classification makes it possible to identify a wide variety of forms in the genus *Triticum* L. and its individual species, which is especially important for collections preserved in genetic seed banks. The use of the intraspecific classification of the genus *Triticum* L. greatly simplifies the identification of the VIR collection accessions introduced from various sources or checking accession identity after regeneration in the field. However, the direct use of such a voluminous classification meets several difficulties. Therefore, we propose a unified intraspecific classification of durum wheat, based on the description of only 16 main botanical varieties out of 131 described so far, which have complexes of morphological traits of the spike and kernel that occur most frequently in durum wheat collections. The remaining 115 botanical varieties, which have additional traits, get their name by the addition of the abbreviated Latin name of one or another additional trait to the main name. Having mastered this way of describing the morphological traits of accessions, any user can easily navigate oneself in the systematized intraspecific diversity of collections. The purpose of this work is to acquaint the reader with the intraspecific classification of durum wheat (*Triticum durum* Desf.) developed at VIR and to offer its simplified version, which is based on the identification of the main and additional morphological traits of the spike and kernel.

Key words: durum wheat (*Triticum durum*); intraspecific classification; complexes of morphological traits; inheritance of traits; botanical variety.

For citation: Lyapunova O.A. Intraspecific diversity of durum wheat (*Triticum durum* Desf.): a unified classification. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):260-268. DOI 10.18699/VJ21.029

Внутривидовое разнообразие твердой пшеницы (*Triticum durum* Desf.): унифицированная классификация

О.А. Ляпунова

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

Аннотация. В отделе генетических ресурсов пшеницы ВНИИ генетических ресурсов растений (ВИР) разработана и в 1979 г. опубликована система рода *Triticum* L., базирующаяся на учете геномного состава видов и наличии или отсутствии у них ряда главных генов, контролирующих «классификационные» признаки. Она основана на исследованиях F. Körnicke и J. Percival, дополненных Н.И. Вавиловым и К.А. Флаксбергером. Эту систему, известную как "Classification of *Triticum* by Dorofeev et al.", относят к ряду основных современных классификаций рода. Это первая в мире стандартизированная система, содержащая все известные внутривидовые (инфраспецифические) таксоны диких и культурных видов пшеницы. Она дает возможность идентификации большого разнообразия форм при работе с родом *Triticum* L. и отдельными его видами, что особенно важно для коллекций, сохраняемых в генетических банках семян. Применение внутривидовой классификации рода *Triticum* L. для идентификации образцов коллекции ВИР, интродуцированных из различных источников или поступивших после полевого размножения для по-

полнения репродукции образцов, значительно упрощает этот процесс. Однако прямое использование такой объемной классификации связано с рядом трудностей. Поэтому нами предлагается унифицированная внутривидовая классификация твердой пшеницы лишь 16 из 131 разновидности, описанной к настоящему времени, которые обладают наиболее часто встречающимися в коллекциях твердой пшеницы комплексами морфологических признаков колоса и зерновки. Остальные 115 разновидностей, имеющих дополнительные признаки, получают свое название путем добавления к основному названию сокращенного латинского названия того или иного дополнительного признака. Владение таким способом описания морфологических признаков образцов может помочь любому пользователю легко ориентироваться в систематизированном внутривидовом разнообразии коллекций. Цель данной статьи – познакомить читателя с внутривидовой классификацией твердой пшеницы (*Triticum durum* Desf.), разработанной в ВИР, и предложить ее упрощенный вариант, построенный на выделении главных и дополнительных морфологических признаков колоса и зерновки.
Ключевые слова: твердая пшеница (*Triticum durum*); внутривидовая классификация; комплексы морфологических признаков; наследование признаков; ботаническая разновидность.

Introduction

Durum wheat (*Triticum durum* Desf.) is characterized by a wide diversity of varieties and forms. Like any set, this diversity should be systematized to better understand the relationships between its constituent units. Classification (from the Latin word *classis* – category, class, and *facio* – do, make) is a method aimed at organizing a system of subordinate groups, in which these units are combined on the basis of similarity in certain essential properties (Subbotin, 2001). The product of the classification is a system. Plant systematics is a branch of botany that deals with the classification of plants. The term "systematic" (systematic botany) was introduced by the Swedish naturalist Carl von Linné in 1751 in his work "Philosophy of Botany" (Linnaeus, 1989). The term "taxonomy" was introduced by the Swiss botanist Augustin Pyrame de Candolle, the creator of the natural system of plants classification – the de Candolle system – and designated the theory of plant classification, according to the rules of which taxa are arranged in the system (de Candolle, 1813). In his treatise "On the Origin of Species…", the English naturalist Charles Robert Darwin considered the terms taxonomy and systematic as synonyms (Darwin, 1859). However, systematics studies not only the diversity of organisms, but also the causes and ways of its appearance, and includes taxonomy and nomenclature.

The history of the genus *Triticum* L. classification begins with C. Linnaeus (Linnaeus, 1737), who is considered by most triticologists as the author of the genus Wheat. Over 300 years of its existence, the Linnaeus classification has undergone numerous interpretations, which are associated with the inclusion or subsequent exclusion of certain cultivated and wild species from it.

The system of the genus *Triticum* L. developed at the Department of Wheat Genetic Resources of Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR) (Dorofeev et al., 1979), was built up on the research of such triticologists as F. Körnicke (1885) and J. Percival (1921), and further revised and supplemented by N.I. Vavilov (1935) and K.A. Flaksberger (Flaksberger, 1935; Flaksberger et al., 1939). The system is based on taking into account the genomic composition of species and the presence or absence of a number of major genes that govern systematically important traits.

In accordance with this system of the genus, durum wheat (*T. durum* Desf.) is treated as a separate species in the rank of the species, which was first described by the French botanist R.L. Desfontaines (1798). The species includes two subspecies: subsp. *durum* and subsp. *horanicum* Vav. The latter is a subspecies of the most dense-ear wheats, with a complex of specific morphological characters. Subsp. *durum* is a subspecies of durum wheat proper, within which six groups of botanical varieties (convarieties) are distinguished, namely convar. *durum*, *durocompactum* Flaksb., *aglossicon* Dorof. et A. Filat., *villosum* (Jakubz.) Dorof. et A. Filat., *falcatum* (Jakubz.) Dorof. et A. Filat., *caucasicum* (Dorof.) Dorof. In turn, convar. *durum* includes three subconvarieties: subconvar. *durum*, *muticum* (Orlov) Dorof. et A. Filat., and *duroramosum* Dorof. (Table 1). At the time of the creation of the classification by V.F. Dorofeev et al. (1979), the genus *T. durum* Desf. numbered 120 botanical varieties and 29 forms in 20 varieties. As a result of subsequent studies, 11 more botanical varieties and 12 forms were identified (Lyapunova, 2017, 2019).

The classification, which is most often referred to as the Classification of *Triticum* by Dorofeev et al., belongs to a number of the main modern classifications of the genus *Triticum* L. This was the first standardized classification that contained all known intraspecific taxa of wild and cultivated wheat species. A similar classification, a development of previous classifications based on the use of a comparative genetic approach, was proposed by N.P. Goncharov (Goncharov, 2002, 2005, 2009; Goncharov et al., 2007). In contrast to hexaploid wheats, the species classification of which can be constructed using only five main genes (Goncharov, 2011), in tetraploid species only Polish wheats and Ispahan emmer wheat can differ oligogenically (Watanabe et al., 1996; Watanabe, 1999). In all other species, only a part of taxonomically important traits has simple genetic control. This refers, e. g., to tetra-awnedness in the majority of *T. carthlicum* Nevski varieties (Haque et al., 2011)[1], purple grain of *T. aethiopicum* Jakubz. (Lachman et al., 2017), and ear branching in *T. turgidum* L. (Haque et al., 2012). At the same time, the botanical varieties identified by us have a simple control of characters. For example, ligulless-

[1] The gene has been recently introgressed into hexaploid wheat and mapped (Dobrovolskaya et al., 2020).

Table 1. Intraspecific differentiation of the species *Triticum durum* Desf.

Groups (convar.) and subgroups (subconvar.) of botanical varieties	No. of botanical varieties	Geographic distribution
Subsp. *durum*		
durum – group of botanical varieties of durum wheats proper, includes three subgroups:	65	Throughout the durum wheat area of distribution
durum – durum wheat proper	42	Throughout the durum wheat area of distribution
muticum – awnless durum wheat	17	Breeding organizations of Australia, Tunisia, Turkey, Kazakhstan, Azerbaijan, Russia
duroramosum – wheat durum with a branching spike	6	Foothill districts of Azerbaijan and Georgia
durocompactum – group of botanical varieties of durum wheats with a dense spike	21	Algeria, Tunisia, Morocco, Egypt, Syria, Jordan, Asia Minor, Azerbaijan
aglossicon – group of botanical varieties of liguleless durum wheats	10	Cyprus
villosum – the Palestinian group of botanical varieties of durum wheats with a rough spike and strong hairy of leaf	8	Coastal districts and foothills of Syria, Jordan, Lebanon, is very rare on Cyprus
falcatum – group of botanical varieties of falcatum durum wheats (grain has a sickle-shaped depression)	13	Greece, Sardinia, Sicily, Malta, Cyprus, Turkey, Iran, Afghanistan, China, Azerbaijan, Kirghizstan, Tadjikistan
caucasicum – the Caucasian groups of botanical varieties of durum wheats	6	Transcaucasia (foothill districts and lowlands – 100–600 m above sea level)
Subsp. *horanicum* Vav.		
–	8	Syria, Jordan, Asia Minor, is rare in Egypt and on islands of the Mediterranean Sea
Total botanical varieties	131	

ness (Barulina, 1937; Watanabe et al., 2004) or awnlessness (Goncharov et al., 2003).

Such a detailed classification makes it possible to identify a wide diversity when working with the genus *Triticum* L. as a whole and/or with its individual species, which is especially important for large-scale collections preserved in genetic seed banks.

The use of intraspecific classification of the genus *Triticum* L. greatly simplifies the identification of the VIR collection accessions introduced from various sources, or when checking accession identity after regeneration in the field. However, apart from the researchers at the Department of Wheat Genetic Resources of VIR, few people use this approach in their practical work, and there are several reasons for this. First, both the monograph itself (Dorofeev et al., 1979) and the accompanying "Identifier of Wheat" (Dorofeev et al., 1980) have not been reprinted for more than 40 years and became a bibliographic rarity, which makes it difficult for national breeders and other wheat researchers to use it (Chikida, 2020). After the collapse of the USSR, the genetic banks of the COMECON (Council for Mutual Economic Assistance – an economic organization from 1949 to 1991 under the leadership of the Soviet Union that comprised the countries of the Eastern Bloc along with a number of socialist states elsewhere in the world) countries stopped working according to a common pattern, although many of them continue to use the system developed by V.F. Dorofeev et al. (1979). Second, there is still no translation of these works into English, although there was an international project on the translation of this monograph (Knüpffer et al., 2003), which makes it impossible for the staff of foreign genetic seed banks to get acquainted with this classification. Third, only the long-term practice of identifying accessions by the name of a botanical variety makes it possible to carry out this laborious work promptly and without difficulty. For instance, durum wheat alone requires remembering names of 131 varieties and their meaning. One of the ways to reduce the number of hard-to-remember names may be unification as a standardization method aimed at reducing the number of objects by combining several characters. It assumes selection of the optimal number of objects, botanical varieties in our case, limited to a reasonable minimum and leads to a certain uniformity. This greatly simplifies the practical use of the classification.

The objective of this work is to acquaint the reader with the intraspecific classification of durum wheat (*Triticum durum* Desf.) developed at VIR, and to offer its simplified analog based on the identification and illustration of the main and additional morphological characters of the ear and kernel.

Materials and methods

Here, we propose a unified intraspecific classification of the durum wheat species, based on the description of only 16 main botanical varieties which have the most commonly occurring sets of morphological characters of the ear and kernel, and

Table 2. The main spike and kernel characters used to describe the intraspecific diversity of durum wheat

Character	Option	Abbreviated Latin designation
Glume color	White or stramineous (Fig. 1)	–
	Red (Fig. 2)	–
	White or red in combination with black (black-blue)* that shows up in the glume central part (Fig. 3)	*nigro-*
	White or red in combination with smoked-grayish (bluish-gray) that shows up in the glume central part (Fig. 4, 5)	*glauco-*
	White or red in combination with black along the edge of glume (form) (Fig. 6)	*triste-*
Glume pubescence	Glabrous glume (Fig. 7) Pubescent glume (Fig. 8)	–
Kernel color ** (Fig. 9)	White	–
	Red	–
	Purple	*violaceo-*
Awns presence	Normal awns (7 cm and longer) (Fig. 10)	–
	Awnless (awns are either absent or awn-like projections shorter than 2.0 cm are available) (Fig. 11)	*mutico-*
Awns color (see Fig. 10, 11)	Same color as of glumes	–
	Black	–

* The blue tint is due to the presence of a waxy coating on the spike glumes.

** Kernels with light yellow, yellow and amber-yellow color are attributed to the group of white-colored ones; while those with light brown, brown and amber-brown color are grouped as red-colored kernels (The International Comecon List of Descriptors…, 1984). Durum wheat kernels are mostly vitreous, therefore the color that is defined as white, is in fact amber-yellow.

Table 3. The most frequent complexes of spike and kernel characters in durum wheat and their Latin name

Name of the complex of characters	Kernel color	Glume color	Awn color	Name of the complex of characters	Kernel color	Glume color	Awn color
Glabrous glume				Pubescent glume			
leucurum (Alef.) Koern.	White	White	White	*valenciae* Koern.	White	White	White
leucomelan (Alef.) Koern.			Black	*melanopus* (Alef.) Koern.			Black
hordeiforme (Host.) Koern.		Red	Red	*italicum* (Alef.) Koern.		Red	Red
erythromelan Koern.			Black	*apulicum* Koern.			Black
affine Koern.	Red	White	White	*durum*	Red	White	White
reichenbachii Koern.			Black	*africanum* Koern.			Black
murciense Koern.		Red	Red	*aegyptiacum* Koern.		Red	Red
pseudoalexandrinum Flaksb.			Black	*niloticum* Koern.			Black

retain their author's name (Table 2). The remaining botanical varieties, which have additional characters, get their name by the addition of this or that additional character to the main abbreviated Latin name (Table 3). Such a way of describing and quickly memorizing intraspecific diversity was proposed for common wheat in (Zuev et al., 2019). This work has been successfully published twice and is in great demand both domestically and among employees of foreign genetic seed banks.

Results

Basic and additional morphological characters of durum wheat

The intraspecific description system is based on botanical varieties, the names of which are determined by a set of morphological characters of the ear and kernel. These sets were distinguished by a combination of such features as the presence or absence of glume pubescence, glume color

Fig. 1. White glume color.

Fig. 2. Red glume color.

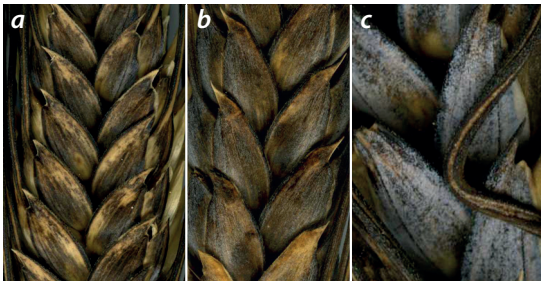


Fig. 3. White (*a*) or red (*b*) color of glume in combination with black color or blue-black color (*c*).

Fig. 4. White glume color in combination with smoked-grayish color (*glauco*-).

Fig. 5. Red glume color in combination with smoked-grayish color (*glauco*-).

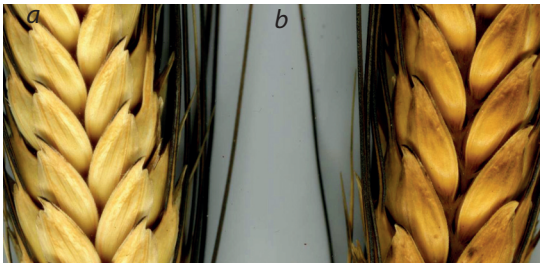


Fig. 6. White (*a*) or red (*b*) glume color in combination with black along edge (*triste*-).

Fig. 7. Glabrous glume.

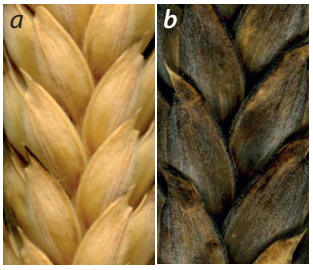


Fig. 8. Pubescent glume: *a* – white spike, *b* – black spike.

Fig. 9. Color of durum wheat kernels: white (*a*), red (*b*), purple (в).

Fig. 10. Awned durum wheat.

Fig. 11. Awnless durum wheat.

(white, red, smoky gray, or black on a white or red background), the presence or absence of awns on the lemma, their color (matching the color of the glume, or black), and the kernel shape and color (white, red or purple) (see Table 2).

A description of each botanical variety must include a set of main features: the presence/absence of glume pubescence, the color of the glume and kernel, the presence/absence of awns on the lemma, and the color of awns. The set of characters revealed by a specimen is designated by the corresponding Latin name given by the author (see Table 3).

To describe a specimen that possesses one of these sets of characters, but in combination with an additional character, like color of the glume, different length of awns, their color, etc., abbreviated Latin names of these characters are used (Table 4). In the case of durum wheat, these names are added to the name of the main set in the case of peduncle pubescence (*piloso*-) or awns smoothness (*levi*-), or when they determine the names of groups or subgroups of botanical varieties, i. e., dense-eared (-*compactus*); with the crescent-shaped kernel (*falcato*-); with the branching ear (*ramoso*-), non-ligulate (*quasi*-), with the densely pubescent leaf blade and sheath of the leaf, and with the hard glume (*villoso*-). Along with the characters of the ear and kernel, Table 4 contains that of the ligula absence, which is the only character of the leaf taken into account when describing botanical varieties.

Table 4. Additional spike and kernel characters used to describe the intraspecific diversity of durum wheat

Character	Option	Abbreviated Latin designation
Spike density	Spike lax or medium density (Fig. 12)	–
	Spike dense (d ≥ 40) (Fig. 13)	*-compactus*
Kernel shape	Round or elongated (see Fig. 9)	–
	Falcate (elongated with a transverse deepening the middle) (Fig. 14)	*falcato-*
Secondary rachis of the spike	Presence of the secondary rachis with spikelets, or of double or triple spikelets per node of the main rachis (axis) (branching spike) (Fig. 15)	*ramoso-*
Peduncle	Glabrous	–
	Pubescent (Fig. 16)	*piloso-*
Presence of ligule (Fig. 17)	Yes (ligule)	–
	Absent (no ligule)	*quasi-*
Roughness of the awns (Fig. 18)	Rough	–
	Weakly rough	*fere-*
	Smooth	*levi-*
Pubescence of the leaf blade and glume structure	Densely pubescent blade and vagina of the leaf, coarse glume	*villoso-*

Fig. 12. Lax spike of durum wheat.

Fig. 13. Dense spike of durum wheat.

Fig. 14. Falcate kernel shape.

Fig. 15. Branching spike of durum wheat.

Fig. 16. Pubescent peduncle.

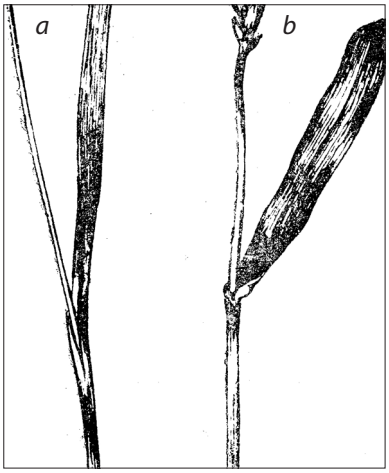


Fig. 17. Non-ligulate (*a*) and ligulatee (*b*) plant of durum wheat (from: Flaksberger, 1935).

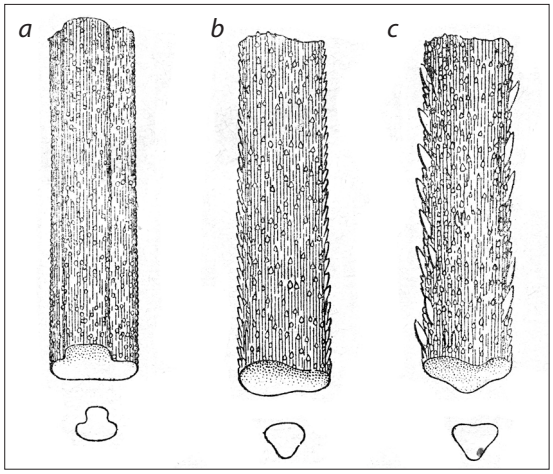


Fig. 18. Awn: smooth (*a*), weakly rough (*b*), strongly rough (*c*) (from: Flaksberger, 1935).

Table 5. Character complexes found in the awnless durum wheat accessions

Latin names of character complexes	Glume color	Color of awn-like projections	Botanical variety name	
			by K.A. Flaksberger (1935)	by V.F. Dorofeev et al. (1979)
Glabrous glume				
1. White spike, white kernel				
mutico-leucurum	White	White	var. *candicans* Meist.	*candicans* Meist.
mutico-leucomelan		Black	–	*muticoleucomelan* Lyapun.
mutico-nigro-leucomelan	Black on the white background	–	f. *quatuor-unum* Flaksb.	*muticalbiprovinciale* Flaksb.
2. Red spike, white kernel				
mutico-hordeiforme	Red	Red	var. *sub-australe* Perciv.	*subaustrale* Perciv.
mutico-hordeiforme-compactus		–	–	*yesilkoense* Gökg.
mutico-erythromelan		Black	–	*muticerythromelan* Lyapun.
mutico-nigro-erythromelan			var. *australe* Perciv.	*australe* Perciv.
3. White spike, red kernel				
mutico-affine	White	White	var. *schechurdini* Meist.	*schechurdinii* Meist.
mutico-nigro-reichenbachii		–	f. *quatuordecim-unum* Flaksb.	*muticalbobscurum* Flaksb.
4. Red spike, red kernel				
mutico-murciense	Red	Red	var. *Stebuti* Flaksb.	*stebutii* Meist.
mutico-nigro-alexandrinum	Black on the red background	Black	f. *quindecim-unum* Flaksb.	*muticobscurum* Dorof. et A. Filat.
Pubescent glume				
5. White spike, white kernel				
mutico-valenciae	White	–	f. *unum-quindecim* Flaksb.	*muticovalenciae* Dorof. et A. Filat.
mutico-melanopus		Black	–	*muticomelanopus* (A. Filat. et Schaid.) Lyapun.
mutico-nigro-melanopus	Black on the white background	–	f. *quatuor-quinque* Flaksb.	*muticoboeufii* Flaksb.
6. Red spike, white kernel				
mutico-italicum	Red	–	f. *duo-quinque* Flaksb.	*muticitalicum* Dorof. et A. Filat.
mutico-apulicum		Black	–	*muticapulicum* Lyapun.
mutico-nigro-apulicum	Black on the red background	–	f. *sex-quinque* Flaksb.	*muticocaerulescens* Flaksb.
7. White spike, red kernel				
mutico-nigro-africanum	Black on the white background	–	–	*muticonazilliense* Gökg.

Unified intraspecific classification of durum wheat (*Triticum durum* Desf.)

The proposed unified intraspecific classification is a simplified analog of the durum wheat key (Dorofeev et al., 1980). The whole diversity is arranged in the form of tables, where the names of varieties according to K.A. Flaksberger (1935) and V.F. Dorofeev et al. (1979) are given for comparison, which allows a user to establish a correspondence between the form being described and the botanical variety. The botanical varieties are presented in accordance with the main characters in the following sequence: the awned and awnless forms are presented in Table 5 and Supplementary[2].

In the first place, these tables present botanical varieties with non-pubescent glumes and different color combinations of the glume and kernel, and then those with the pubescent glumes in the same order.

[2] Supplementary material is available at: http://vavilov.elpub.ru/jour/manager/files/Suppl_Lyapunova_Engl.pdf

Glabrous glume
1. White spike, white kernel.
2. Red spike, white kernel.
3. White spike, red kernel.
4. Red spike, red kernel.
5. White spike, purple kernel.
6. Red spike, purple kernel.

Pubescent glume
7. White spike, white kernel.
8. Red spike, white kernel.
9. White spike, red kernel.
10. Red spike, red kernel.

All of the above main characters have simple genetic control (McIntosh et al., 2020).

Conclusion

Acquaintance with the durum wheat intraspecific classification, which was created at VIR and contained all the known intraspecific taxa of the time as well as the subsequently added ones, will make it possible to analyze all the intraspecific diversity of the main cultivated tetraploid species *Triticum durum* Desf. The proposed simplified analog version, based on the identification of the main and additional morphological characters of the ear and kernel, can help any user simplify the systematization of the intraspecific diversity contained in any collection and easily navigate it.

References

Barulina E.N. Comparative genetic study of *Triticum* species: I. Genetics of the *ligula* trait in different chromosomal wheat species: *T. vulgare* Vill., *T. compactum* Host, *T. durum* Desf. *Trudy po Prikladnoy Botanike, Genetike i Selektsii = Proceedings on Applied Botany, Genetics, and Breeding*. 1937;I(5):127-166. (in Russian)

Chikida N.N. Contribution of academician Vladimir Filimonovich Dorofeev to the development of agricultural and biological sciences. *Pisma v Vavilovskii Zhurnal Genetiki i Selektsii = Letters to Vavilov Journal of Genetics and Breeding*. 2020;6(1):18-36. DOI 10.18699/Letters2020-6-04. (in Russian)

Darwin Ch. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray, 1859.

De Candolle A.P. Théorie Elémentaire de Botanique. Paris, 1813.

Desfontaines R.L. Flora Atlantica: Sive Historia Plantarum Quae in Atlante, Agro Tunetano et Algeriensi Crescent. Parisiis: L.G. Desgranges, 1798; Vol. 1;114.

Dobrovolskaya O.B., Dresvyannikova A.E., Badaeva E.D., Popova K.I., Trávníčková M., Martinek P. The study of genetic factors that determine the awned glume trait in bread wheat. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(6):568-574. DOI 10.18699/VJ20.650. (in Russian)

Dorofeev V.F., Filatenko A.A., Migushova E.F. Identification Keys to Wheats. Leningrad: VIR, 1980. (in Russian)

Dorofeev V.F., Filatenko A.A., Migushova E.F., Udachin R.A., Jakubziner M.M. Cultivated Flora of the USSR. Wheat. Leningrad: Kolos Publ., 1979. (in Russian)

Flaksberger K.A. Cultural Flora of the USSR. Cereals. Vol. 1. Wheat. Moscow; Leningrad: Sel'khozgiz Publ., 1935;150-223. (in Russian)

Flaksberger K.A., Antropov V.I., Antropov V.F., Bakhteev F.Kh. Manual Book to True Cereals: Wheat, Rye, Barley, Oat. Moscow; Leningrad: Sel'khozgiz Publ., 1939;63-92. (in Russian)

Goncharov N.P. Comparative Genetics of Wheat and their Related Species. Novosibirsk: Siberian University Press, 2002. (in Russian)

Goncharov N.P. Comparative-genetic analysis – a base for wheat taxonomy revision. *Czech J. Genet. Plant Breed*. 2005;41;52-55.

Goncharov N.P. Manual Book of Common and Durum Wheat Varieties. Novosibirsk: Publishing House SB RAS, 2009. (in Russian)

Goncharov N.P. Genus *Triticum* L. taxonomy: the present and the future. *Plant Syst. Evol.* 2011;295:1-11. DOI 10.1007/s00606-011-0480-9.

Goncharov N.P., Kondratenko E.Ja., Bannikova S.V., Konovalov A.A., Golovnina K.A. Comparative genetic analysis of diploid naked wheat *Triticum sinskajae* and the progenitor *T. monococcum* accession. *Russ. J. Genet.* 2007;43(11):1248-1256. DOI 10.1134/S1022795407110075.

Goncharov N.P., Mitina R.L., Anfilova N.A. Inheritance of awnlessness in tetraploid wheat species. *Russ. J. Genet*. 2003;39:463-466.

Haque M.A., Martinek P., Kobayashi S., Kita I., Ohwaku K., Watanabe N., Kuboyama T. Microsatellite mapping of genes for semidwarfism and branched spike in *Triticum durum* Desf. var. *ramosoobscurum* Jakubz. "Vetvistokoloskaya". *Genet. Resour. Crop Evol.* 2012;59:831-837. DOI 10.1007/s10722-011-9722-5.

Haque M.A., Takayama A., Watanabe N., Kuboyama T. Cytological and genetic mapping of the gene for four-awned phenotype in *Triticum carthlicum* Nevski. *Genet. Resour. Crop Evol.* 2011;58(7):1087-1093. DOI 10.1007/s10722-010-9644-7.

Knüpffer H., Morrison L.A., Filatenko A.A., Hammer K., Morgounov A., Faberová I. English translation of the 1979 Russian taxonomic monograph of *Triticum* L. by Dorofeev et al.: project progress report. 2003. Available at: https://wheat.pw.usda.gov/ggpages/GrainTax/anuscript_Dorofeev_extended.html

Körnicke F. Der Weizen. In: Körnicke F., Werner H. (Eds.). Hundbuch des Getreidebaus. Bd. 1. Berlin: Verlag von Paul Parey, 1885; 22-114.

Lachman J., Martinek P., Kotikova Z., Šulc's M. Genetics and chemistry of pigments in wheat grain – A review. *J. Cereal Sci.* 2017;74:145-154. DOI 10.1016/j. jcs.2017.02.007.

Linnaeus C. Critica Botanica. Leiden. 1737. Available at: http://ru.knowledgr.com/09901965/CriticaBotanica

Linnaeus C. Philosophy of Botany. Moscow: Nauka Publ., 1989. (in Russian)

Lyapunova O.A. Intraspecific classification of durum wheat: New botanical varieties and forms. *Russ. J. Genet. Appl. Res.* 2017;7(7):757-762. DOI 10.1134/S2079059717070048.

Lyapunova O.A. Types of *Triticum durum* Desf. intraspecific taxa preserved in the VIR herbarium (WIR). *Vavilovia*. 2019;2(3):23-38. DOI 10.30901/2658-3860-2019-3-23-38. (in Russian)

McIntosh R.A., Yamazaki Y., Dubcovsky J., Rogers J., Morris C., Appels R., Xia X.C. Catalogue of gene symbols for wheat. Available at: http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp

Percival J. The Wheat Plants. London: Duckworth and Co., 1921.

Subbotin A.L. Classification: New Philosophical Encyclopedia. M.: Mysl' Publ., 2001. Available at: https://iphlib.ru/library/collection/newphilenc/document/HASH01b0ea7a433b0feb3ff681b0. (in Russian)

The International Comecon List of Descriptors for the Genus *Triticum* L. Leningrad: VIR, 1984. (in Russian)

Vavilov N.I. Botanical and geographical bases of breeding. In: Theoretical Foundations of Plant Breeding. Vol. 1. General Plant Genetics. Moscow: Sel'khozgiz Publ., 1935;17-75. (in Russian)

Watanabe N. Genetic control of the long glume phenotype in tetraploid wheat by homoeologous chromosomes. *Euphytica*. 1999; 106:39-43. DOI 10.1023/A:1003589117853.

Watanabe N., Nakayama A., Ban T. Cytological and microsatellite mapping of the genes determining liguleless phenotype in durum wheat. *Euphytica*. 2004;140:163-170. DOI 10.1007/s10681-004-2425-7.

Watanabe N., Yotani Y., Furuta Y. The inheritance and chromosomal location of a gene for long glume in durum wheat. *Euphytica*. 1996;91:235-239. DOI 10.1007/BF00021076.

Zuev E.V., Amri A., Brykova A.N., Pyukkenen V.P., Mitrofanova O.P. Atlas of Bread Wheat (*Triticum aestivum* L.) Genetic Diversity Based on Spike and Kernel Characters. 2nd ed. St. Petersburg: Kopi-R Publ., 2019. (in Russian)

ORCID ID

O.A. Lyapunova orcid.org/0000-0003-2164-4510

Acknowledgements. The work was performed within the framework of the State Task according to the theme plan of VIR, Project No. 0662-2019-0006 "Search for and viability maintenance, and disclosing the potential of hereditary variation in the global collection of cereal and groat crops at VIR for the development of an optimized genebank and its sustainable utilization in plant breeding and crop production".
The author thanks the reviewers for their contribution to the peer review of this work.

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

Received October 23, 2020. Revised December 10, 2020. Accepted December 10, 2020.

Вавиловский журнал генетики и селекции. 2021;25(3):269-275
DOI 10.18699/VJ21.030

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

Potato spindle tuber viroid

A.V. Kochetov[1, 2] ✉, A.Y. Pronozin[1], N.V. Shatskaya[1], D.A. Afonnikov[1, 2], O.S. Afanasenko[1, 3]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] All-Russian Institute of Plant Protection, Pushkin, St. Petersburg, Russia
✉ ak@bionet.nsc.ru

Abstract. Viroids belong to a very interesting class of molecules attracting researchers in phytopathology and molecular evolution. Here we review recent literature data concerning the genetics of *Potato spindle tuber viroid* (PSTVd) and the mechanisms related to its pathological effect on the host plants. PSTVd can be transmitted vertically through microspores and macrospores, but not with pollen from another infected plant. The 359 nucleotide-long genomic RNA of PSTVd is highly structured and its 3D-conformation is responsible for interaction with host cellular factors to mediate replication, transport between tissues during systemic infection and the severity of pathological symptoms. RNA replication is prone to errors and infected plants contain a population of mutated forms of the PSTVd genome. Interestingly, at 7 DAI, only 25 % of the newly synthesized RNAs were identical to the master copy, but this proportion increased to up to 70 % at 14 DAI and remained the same afterwards. PSTVd infection induces the immune response in host plants. There are PSTVd strains with a severe, a moderate or a mild pathological effect. Interestingly, viroid replication itself does not necessarily induce strong morphological or physiological symptoms. In the case of PSTVd, disease symptoms may occur due to RNA-interference, which decreases the expression levels of some important cellular regulatory factors, such as, for example, potato StTCP23 from the gibberellic acid pathway with a role in tuber morphogenesis or tomato FRIGIDA-like protein 3 with an early flowering phenotype. This association between the small segments of viroid genomic RNAs complementary to the untranslated regions of cellular mRNAs and disease symptoms provides a way for new resistant cultivars to be developed by genetic editing. To conclude, viroids provide a unique model to reveal the fundamental features of living systems, which appeared early in evolution and still remain undiscovered.
Key words: viroids; plants; pathogenesis.

For citation: Kochetov A.V., Pronozin A.Y., Shatskaya N.V., Afonnikov D.A., Afanasenko O.S. Potato spindle tuber viroid. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):269-275. DOI 10.18699/VJ21.030

Вироид веретеновидности клубней картофеля

А.В. Кочетов[1, 2] ✉, А.Ю. Пронозин[1], Н.В. Шацкая[1], Д.А. Афонников[1, 2], О.С. Афанасенко[1, 3]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Всероссийский научно-исследовательский институт защиты растений, Пушкин, Санкт-Петербург, Россия
✉ ak@bionet.nsc.ru

Аннотация. Вироиды, небольшие кольцевые молекулы РНК, которые вызывают патогенез у растений, остаются одним из самых необычных биологических объектов, привлекающих внимание не только фитопатологов, но и специалистов в области молекулярной эволюции. В статье приведен обзор последних литературных данных о генетике вироида веретеновидности клубней картофеля (ВВКК) и генетических механизмах формирования патологических состояний у растений-хозяев. ВВКК способен передаваться вертикально (через генеративные клетки зараженного растения), но, в отличие от некоторых других вироидов, не передается через пыльцу от зараженного растения. Большой интерес у исследователей вызывает структура геномной РНК вироида размером 359 нуклеотидов: хорошо известно, что особенности 3D конформации определяют основные параметры взаимодействия с клеточными факторами на стадии репликации, транспорта между различными тканями в процессе системной инфекции, а также степень выраженности симптомов заболевания. При репликации геномной РНК вироидов часто происходят ошибки, приводящие к появлению гетерогенной популяции молекул РНК в тканях зараженного растения. Примечательно, что через 7 дней после инокуляции только 25 % молекул геномной РНК ВВКК соответствовали исходной матрице, использованной для инокуляции, однако эта доля увеличилась до 70 % через 14 дней и далее оставалась на том же уровне. По-видимому, при сохранении у мутантных вариантов геномной РНК способности к репликации вироид обладает высоким потенциалом к отбору эффективных инфекционных

форм. ВВКК вызывает у пораженных растений развитие иммунного ответа, механизмы индукции которого недостаточно изучены. Известны сильно- и слабопатогенные штаммы ВВКК, вызывающие разные проявления болезни, фенотипические проявления от которых у пораженных растений в значительной мере различны. Сама по себе репликация вироида не обязательно приводит к выраженным фенотипическим проявлениям, в случае ВВКК они могут быть связаны с участками гомологии между геномной РНК и мРНК некоторых регуляторных генов, например транскрипционного фактора StTCP23 картофеля, участвующего в регуляторном контуре гиббереллиновой кислоты и в контроле морфогенеза клубня. Другой пример – индукция РНК-интерференции против мРНК гена FRIGIDA-like protein 3 у томата, что приводит к раннему цветению. В связи с этим обсуждаются потенциальные способы борьбы с вироидом, основанные на удалении из генома растений таких участков гомологии, расположенных в нетранслируемых областях мРНК и не выполняющих каких-либо функций. В целом вироиды представляют собой уникальную модель для исследования основ организации живых систем, многие из которых возникли на ранних этапах эволюции и остаются до сих пор не выявленными.
Ключевые слова: генетика вироида; патогенез растений.

Introduction

Viroids are highly structured circular single-stranded RNAs, which are able to replicate in infected organism and cause diseases from symptomless to lethal. Viroid genomes vary in size from 250 to 400 nucleotides: for example, it is 246 nt for Avocado sunblotch viroid (ASBVd) and Coconut cadang-cadang viroid, and 401 nt for Chrysanthemum chlorotic mottle viroid (CchMVd) (Srivastava, Prasad, 2020). Viroid genomes encode no proteins and the mechanisms of their replication and interactions with host cells are of great interest for specialists in phytopathology, molecular biology and molecular evolution.

Potato spindle tuber viroid (PSTVd) has been intensively investigated because of the adverse effects it has on potato yield. *Solanum tuberosum* is a vegetatively propagated crop, and thus it is especially vulnerable to viruses and viroids. PSTVd alone or in combination with viruses can decrease the yield of susceptible potato varieties from 40 to 70 % (Annenkov, 2000). The list of symptoms commonly includes leaf deformation, irregular tuber shape and development, slower sprouting, stunted phenotype, etc. Here we review recently published data on the PSTVd molecular genetics and the mechanisms mediating its specific pathological phenotypes.

PSTVd genome structure, RNA population, and quasispecies

PSTVd strains can cause different symptoms. For instance, inoculation of tomato with PSTVd-Dahlia strain results in mild symptoms, while PSTVd-Intermediate strain causes a severe phenotype. Their genomes differ in nine positions, six of which are located in structured RNA parts (the left terminal domain and the pathogenicity domain). It was reported that mutation at pos. 42 decreases severity and viroid synthesis and mutation at pos. 64 affects stunting. In general, enrichment of mutations in genomic RNAs revealed positions with importance to symptom severity and viroid replication intensity (Kitabayashi et al., 2020). The viroid genome encodes no proteins and its replication depends on the host cellular machinery. The 3D-structure of the viroid RNA genome mediates its interaction with cellular proteins, for example, loop 27 (pos. 177–182) similar to a structural element in the 3′-UTR of animal histone mRNAs was found to be important for PSTVd replication and transport through host tissues.

Viroids are considered appropriate models for studying regulatory and catalytic RNAs as well as RNA-mediated control of cellular processes. Interestingly, the PSTVd genome contains 17 G/U complementary interactions and some of them are conserved and functionally important for replication and systemic spreading throughout plant tissues (Wu J. et al., 2020). Artificial RNA constructs derived from viroid genomes can form circular molecules even in *Saccharomyces cerevisiae*, suggesting that their processing mechanisms are highly conservative (Friday et al., 2017).

Interestingly, viroid replication is error-prone, which results in a population of diverse genomic RNA molecules (“quasispecies”). The most frequent PSTVd genomic variants were analyzed at different time points after inoculation of tomato plants. It was found that at seven DAI only 25 % of the sequenced PSTVd genomes were identical to the master copy, but its frequency grew up to 70 % at 14 DAI and remained the same at 28 DAI (Adkar-Purushothama et al., 2020). It is likely that viroid replication produces a large variety of structural variants with different effects on host defense and the severity of disease symptoms. Viroids that replicated in plastids had higher mutation rates (1/800–1/1000 nucleotides)

than those that replicated in nuclei (e. g., the mutation frequency for PSTVd varies between 1/3800 and 1/7000) (López-Carrasco et al., 2017).

Mutations in the viroid RNA genome can cause different effects on its replication and life cycle. Three engineered PSTVd genomes with small deletions or insertions were characterized (Więsyk et al., 2017). In two cases, the viroid lost the replication ability, while, in one case, it was still replicating, but its genome stability was decreased. Further analysis of inoculated tomato plants revealed frequent cases of reversion to the master copy and a variety of new stable genomic variants. It is likely that viroid genomes can evolve rapidly (Więsyk et al., 2017).

Viroids may be transmitted both vertically (through micro- or macrospores of infected plants) and horizontally (with pollen of infected plants). PSTVd is transmitted to future generations through macrospores, while some other viroids (e. g., tomato planta macho viroid, TPMVd) spread with pollen (Matsushita et al., 2018). The terminal left (TL) and pathogenicity (P) domains of the TPMVd genome were found to be responsible for pollen-mediated transmission (Yanagisawa et al., 2019). Interestingly, pollen grains from infected Petunia plants could infect tomato, i. e. viroid transmission does not require fertilization (Yanagisawa, Matsushita, 2018).

Pathogenesis mechanisms, plant defense mechanisms, and RNA interference

Inoculation of potato plants with PSTVd resulted in the accumulation of jasmonic acid in leaves, castasterone in leaves and roots, indole-3-acetic acid in tubers and no increase in salicylic or abscisic acids. In addition, viroid infection induced accumulation of reactive oxygen species (ROS) and enhanced the activity of antioxidants (Milanović et al., 2019). A metabolomic analysis revealed considerable changes in the content of 79 substances associated with 23 metabolic chains (Bagherian et al., 2016).

The molecular mechanisms associated with the severity of symptoms caused by different viroid strains remain underinvestigated. It is considered that symptoms depend on the host genotype and plant growing conditions (temperature, humidity, etc.). A comparison of the transcriptomes of tomato leaves after inoculation with either severe or mild PSTVd strains revealed more than 3000 DEGs; however, most of them were specific for the severe strain. Symptom severity were likely correlated with the expression of the genes coding for the C2C2-GATA transcription factor and the growth regulatory factor (GRF) (Więsyk et al., 2020). Similar results were obtained for tomato roots inoculated with the mild and severe PSTVd strains: in addition to differences in expression between the genes controlling the induction of defense response, significant differences in expression were found between the genes for lignin biosynthesis and cell wall formation, and the genes of the auxin and cytokinin transduction pathways (Góra-Sochacka et al., 2019). Viroids produce neither proteins nor typical pathogen-associated molecular patterns (PAMP), and so the mechanisms of the induction of defense response remain unclear (Zheng et al., 2017; Nath et al., 2020).

The replication of viroids, their accumulation and traffic between plant tissues do not necessarily result in the development of disease symptoms. The severity of symptoms strongly depends on the viroid strain and host genotype. It is quite likely that some viroids infect plants without producing visible symptoms (it is possible that there are many undiscovered symptomless replicons persisting in the populations of host organisms). Interestingly, the disease symptoms may result from the RNA-interference induced by highly structured viroid genomic RNAs or replication intermediates. The viroid-derived siRNAs may target some host mRNAs and change the expression levels of the corresponding genes. If these genes participate in the control of plant development, physiological or biochemical processes, their suppression may result in the disease phenotype.

PSTVd infection stunts potato growth, results in aberrations in leave and tuber morphology and decreased yield. It was found that PSTVd induces siRNAs related to the *StTCP23* transcription factor (the teosinte branched1/Cycloidea/Proliferating cell factor). The *StTCP23* mRNA 3′-UTR contains a 21-nucleotide-long segment complementary to the VMR (virulence-modulating region) of PSTVd strain RG1. Experimental suppression of *StTCP23* with artificial microRNAs resulted in a potato phenotype similar to PSTVd disease symptoms. The functions of this gene are related to the gibberellic acid signal transduction pathway controlling plant growth and tuber development (Bao et al., 2019).

Flores et al. (2020) demonstrated that infected plants contain vd-sRNAs (viroid-derived small RNAs) able to interact with Argonaute proteins. PSTVd replication takes place in the nucleus and in this case vd-sRNAs appear at later stages of infection when systemic defense response is already a factor. Peach latent mosaic viroid (PLMVd) replicates in plastids and in this case vd-sRNAs appear at an early infection stage locally (Flores et al., 2020).

In another study (Adkar-Purushothama et al., 2018), one of the PSTVd-induced vd-sRNAs targeted the mRNA of tomato FRIGIDA-like protein 3. Tomato plants infected with severe strains of PSTVd are characterized by early flowering. Experimental suppression of FRIGIDA-like protein 3 results in a similar phenotype (Adkar-Purushothama, Perreault, 2018; Adkar-Purushothama et al., 2018).

RNA-interference is the mechanism controlling host defense against viruses and viroids. Experimental suppression of the Dicer 2 and Dicer 4 genes in tomato resulted in viroid accumulation and more severe symptoms. Interestingly, these proteins also take part in ROS generation and their absence interferes with the general defense response (Suzuki et al., 2019).

Eukaryotic mRNAs consist of the CDS and the 5′- and 3′-untranslated regions (UTRs). The 5′-UTR is responsible for translation initiation, while the 3′-UTR can influence mRNA cytoplasmic stability (Kochetov et al., 2002, 2004a, b; Kochetov, Sarai, 2004; Volkova, Kochetov, 2010; Ventoso et al., 2012). Self-complementary double-stranded RNAs (dsRNAs) are commonly used for engineered RNA-interference. Unlike CDS, most 3′-UTRs are not evolutionarily conserved and can be used as regions of choice for dsRNA-mediated selective gene suppression. Indeed, gene editing may be applied to remove the segment homologous to the viroid genome from the 3′-UTR regions of the host genes related to disease symptoms. Making viroid infection symptomless may strongly decrease yield loss and provides a new way for the molecular breeding of resistant cultivars (Kochetov et al., 2004b).

Bioinformatics methods for viroid detection and analysis

NGS techniques have provided a way for a systemic large-scale analysis of transcriptomes and revealed a large variety of new viruses and viroids. Computational methods for viroid detection are commonly similar to those developed for viruses (Burger, Maree, 2015; Pecman et al., 2017).

It should be noted that the population of virus or viroid genomic molecules in the tissues of the infected organism may be heterogeneous because of frequent replication errors. These genomic variants are considered quasispecies (Brass et al., 2017). Thus, computational identification of viroids is frequently based on a combination of mapping reads to the known viroid sequences from databases and *de novo* sequence assembly. Viral RNAs may consist of a small part of the cellular transcriptome, and for that reason high coverage sequencing data are needed for a reliable detection of viroids – or special methods for viral RNA enrichment should be applied (Roossinck, 2012). In other words, the search for viruses and viroids in metatranscriptomes is relatively expensive.

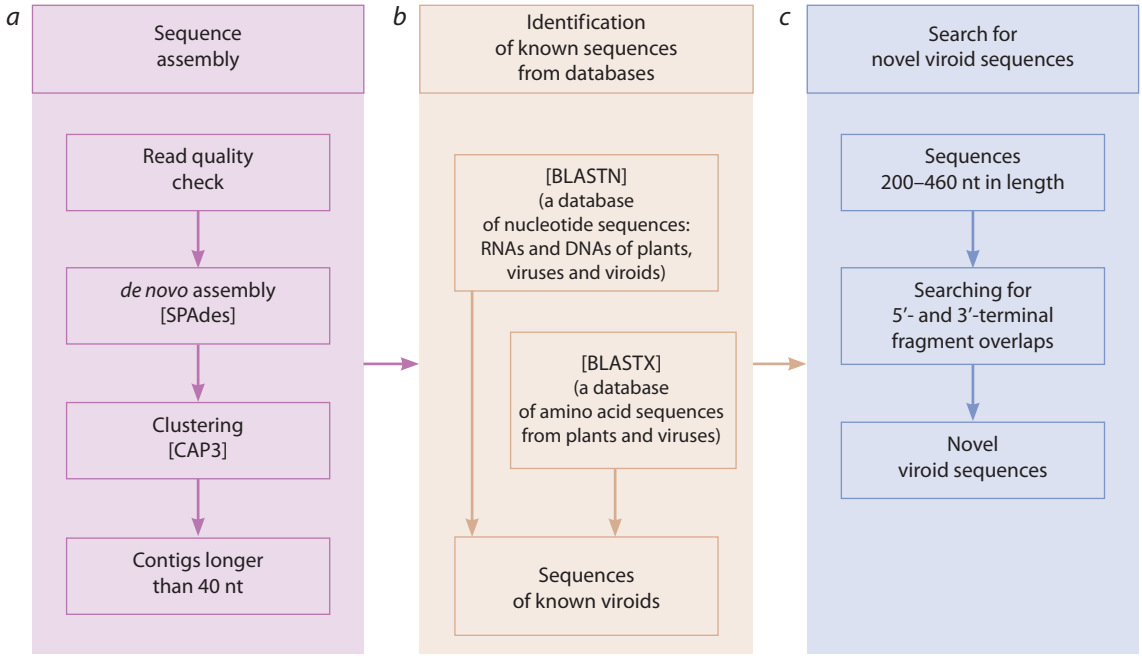
Computational analysis for viroid detection may include (Wu Q. et al., 2015):

1. Analysis of transcriptomes to reveal and describe viral/viroid RNAs.
2. Characterization of the population of viroid RNA molecules in tissues of an infected organism.
3. Characterization of viroid RNA genomes in a natural or a model population of host plants.

To increase sensitivity, it seems reasonable to deplete mRNA libraries of fractions of small or ribosomal RNAs. A comparative analysis of these approaches on nine transcriptomes of infected plants demonstrated that the depletion of small RNAs in the libraries increased the sensitivity of detection of viroids and viruses with single-stranded DNA genomes, while the depletion of ribosomal RNAs was efficient for viruses with RNA genomes (Pecman et al., 2017).

The Viral Surveillance and Diagnosis (VSD) pipeline was developed for viroid identification (Barrero et al., 2017). Short (21–24 nt) sequencing reads are taken as input. The pipeline consists of several modules (see Figure). The first module (*a*) assembles the genomes from short reads and performs adapter trimming, a quality check and removal of poorly assembled RNAs. SPAdes (Bankevich et al., 2012) and CAP3 (Huang, Madan, 1999), genome assembly tools, are used to select contigs longer than 40 nucleotides. The second module (*b*) contains programs for identification of transcripts corresponding to the host genome, viruses and viroids through a comparison with nucleotide and protein sequence databases with the aid of BLASTN and BLASTX (Altschul et al., 1997). The third module (*c*) provides tools for prediction of new viroids. For this purpose, RNA contigs between 200 and 460 nucleotides are checked for the presence of 5′- and 3′-end overlaps characteristic of circular molecules.

The method of viroid identification on the basis of comparison with reference sequences was developed by Brass et al. (2017). It consists of modules for executing a standard protocol: removal of adapters and poly(A)-tails with PrinSeq (Schmieder, Edwards, 2011) and TRIMMOMATIC (Bolger et al., 2014) followed by quality checking and filtering with SEGEMEHL, and alignment onto the reference database (Otto et al., 2014). This tool supports identification of viroid quasispecies in RNA libraries. It was used for analysis of tomato



The VSD pipeline for identification of viral and viroid sequences in plant transcriptomic data.

The pipeline consists of three modules: *a*, *de novo* sequence assembly; *b*, identification of the known sequences of plant viruses and viroids; *c*, the search for novel viroid sequences. Adapted from Barrero et al. (2017) Fig. 1.

cultivar ‘Heinz 1706’ transcriptomes bearing PSTVd strains QFA, C3 and AS1, ‘Rutgers’ transcriptomes with PSTVd strains M and I, and transcriptomes of four tomato cultivars infected with PSTVd strain RG. The results obtained were useful for evaluation of viroid evolutionary dynamics (Brass et al., 2017).

References

Adkar-Purushothama C.R., Bolduc F., Bru P., Perreault J.-P. Insights into potato spindle tuber viroid quasi-species from infection to disease. *Front. Microbiol*. 2020;11:1235. DOI 10.3389/fmicb.2020.01235.

Adkar-Purushothama C.R., Perreault J.-P. Alterations of the viroid regions that interact with the host defense genes attenuate viroid infection in host plant. *RNA Biol*. 2018;15(7):955-966. DOI 10.1080/15476286.2018.1462653.

Adkar-Purushothama C.R., Sano T., Perreault J.-P. Viroid-derived small RNA induces early flowering in tomato plants by RNA silencing. *Mol. Plant Pathol*. 2018;19(11):2446-2458. DOI 10.1111/mpp.12721.

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389-3402. DOI 10.1093/nar/25.17.3389.

Annenkov B.G. Problems and prospects of the influence of curative measures on the potato industry performance in the Russian Far East. In: Topical Issues of Crop Immunity, Sanitation, and Protection in the Russian Amur Region. Khabarovsk, 2000;26-42. (in Russian)

Bagherian S.A.A., Hamzehzarghani H., Izadpanah K., Djavaheri M. Effects of potato spindle tuber viroid infection on tomato metabolic profile. *J. Plant Physiol.* 2016;201:42-53. DOI 10.1016/j.jplph.2016.06.014. Epub 2016 Jun 30. PMID: 27393919.

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 2012;19(5):455-477. DOI 10.1089/cmb.2012.0021.

Bao S., Owens R.A., Sun Q., Song H., Liu Y., Eamens A.L., Feng H., Tian H., Wang M.-B., Zhang R. Silencing of transcription factor encoding gene *StTCP23* by small RNAs derived from the virulence modulating region of *potato spindle tuber viroid* is associated with symptom development in potato. *PLoS Pathog*. 2019;15(12):e1008110. DOI 10.1371/journal.ppat.1008110.

Barrero R.A., Napier K.R., Cunnington J., Liefting L., Keenan S., Frampton R.A., Szabo T., Bulman S., Hunter A., Ward L., Whattam M., Bellgard M.I. An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. *BMC Bioinform*. 2017;18(1):26. DOI 10.1186/s12859-016-1428-4.

Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. DOI 10.1093/bioinformatics/btu170.

Brass J.R., Owens R.A., Matoušek J., Steger G. Viroid quasi-species revealed by deep sequencing. *RNA Biol*. 2017;14(3):317-325. DOI 10.1080/15476286.2016.1272745.

Burger J.T., Maree H.J. Metagenomic next-generation sequencing of viruses infecting grapevines. *Methods Mol. Biol*. 2015; 1302:315-330. DOI 10.1007/978-1-4939-2620-6_23. PMID: 25981264.

Flores R., Navarro B., Delgado S., Serra P., Di Serio F. Viroid pathogenesis: a critical appraisal of the role of RNA silencing in triggering the initial molecular lesion. *FEMS Microbiol. Rev*. 2020;44(3):386-398. DOI 10.1093/femsre/fuaa011.

Friday D., Mukkara P., Owens R.A., Baumstark T., Bruist M.F. Processing of potato spindle tuber viroid RNAs in yeast, a nonconventional host. *J. Virol*. 2017;91(24):e01078-17. DOI 10.1128/JVI.01078-17.

Góra-Sochacka A., Więsyk A., Fogtman A., Lirski M., Zagórski-Ostoja W. Root transcriptomic analysis reveals global changes induced by systemic infection of *Solanum lycopersicum* with mild and severe variants of potato spindle tuber viroid. *Viruses*. 2019;11(11):992. DOI 10.3390/v1111 0992.

Huang X., Madan A. CAP3: a DNA sequence assembly program. *Genome Res.* 1999;9:868-877. DOI 10.1101/gr.9.9.868.

Kitabayashi S., Tsushima D., Adkar-Purushothama C.R., Sano T. Identification and molecular mechanisms of key nucleotides causing attenuation in pathogenicity of dahlia isolate of potato spindle tuber viroid. *Int. J. Mol. Sci*. 2020;21(19):7352. DOI 10.3390/ijms21197352.

Kochetov A.V., Kolchanov N.A., Sarai A. Interrelations between the efficiency of translation start sites and other sequence features of yeast mRNAs. *Mol. Genet. Genom*. 2004a; 270(5):442-447. DOI 10.1007/s00438-003-0941-0.

Kochetov A.V., Sarai A. Translational polymorphism as a potential source of plant proteins variety in *Arabidopsis thaliana*. *Bioinformatics*. 2004;20(4):445-447. DOI 10.1093/bioinformatics/btg443.

Kochetov A.V., Sirnik O.A., Rogozin I.B., Glazko G.V., Komarova M.L., Shumny V.K. Contextual features of higher plant mRNA 5′-untranslated regions. *Mol. Biol.* 2002;36(4):510-516. DOI 10.1023/A:1019852427093.

Kochetov A.V., Titov S.E., Kolodiazhnaia Ya.S., Komarova M.L., Koval V.S., Makarova N.N., Ily'inskyi Yu.Yu., Trifonova E.A., Shumny V.K. Tobacco transformants bearing antisense suppressor of proline dehydrogenase gene, are characterized by higher proline content and cytoplasm osmotic pressure. *Russ. J. Genet*. 2004b;40(2):216-218. DOI 10.1023/B:RUGE.0000016999.53466.e1.

López-Carrasco A., Ballesteros C., Sentandreu V., Delgado S., Gago-Zachert S., Flores R., Sanjuán R. Different rates of spontaneous mutation of chloroplastic and nuclear viroids as determined by high-fidelity ultra-deep sequencing. *PLoS Pathog*. 2017;13(9):e1006547. DOI 10.1371/journal.ppat. 1006547.

Matsushita Y., Yanagisawa H., Sano T. Vertical and horizontal transmission of Pospiviroids. *Viruses*. 2018;10(12):706. DOI 10.3390/v10120706.

Milanović J., Oklestkova J., Majdandžić A., Novák O., Mihaljević S. Organ-specific differences in endogenous phytohormone and antioxidative responses in potato upon PSTVd infection. *J. Plant Physiol*. 2019;232:107-114. DOI 10.1016/j.jplph.2018.10.027.

Nath V.S., Shrestha A., Awasthi P., Mishra A.K., Kocábek T., Matoušek J., Sečnik A., Jakše J., Radišek S., Hallan V. Mapping the gene expression spectrum of mediator subunits in response to viroid infection in plants. *Int. J. Mol. Sci*. 2020; 21(7):2498. DOI 10.3390/ijms21072498.

Otto C., Stadler P.F., Hoffmann S. Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics.* 2014; 30(13):1837-1843. DOI 10.1093/bio informatics/btu146.

Pecman A., Kutnjak D., Gutiérrez-Aguirre I., Adams I., Fox A., Boonham N., Ravnikar M. Next generation sequencing for detection and discovery of plant viruses and viroids: comparison of two approaches. *Front. Microbiol.* 2017;8:1998. DOI 10.3389/fmicb.2017.01998.

Roossinck M.J. Plant virus metagenomics: biodiversity and ecology. *Annu. Rev. Genet*. 2012;46:359-369. DOI 10.1146/ annurev-genet-110711-155600.

Schmieder R., Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863-864. DOI 10.1093/bioinformatics/btr026.

Srivastava S., Prasad V. Viroids: small entities with a mean punch. In: Awasthi L.P. (Ed.) Applied Plant Virology: Advances, Detection, and Antiviral Strategies. Academic Press, 2020;209-226. DOI 10.1016/B978-0-12-818654-1. 00016-5.

Suzuki T., Ikeda S., Kasai A., Taneda A., Fujibayashi M., Sugawara K., Okuta M., Maeda H., Sano T. RNAi-mediated down-regulation of Dicer-like 2 and 4 changes the response of 'moneymaker' tomato to potato spindle tuber viroid infection from tolerance to lethal systemic necrosis, accompanied by up-regulation of miR398, 398a-3p and production of excessive amount of reactive oxygen species. *Viruses.* 2019; 11(4):344. DOI 10.3390/v11040344.

Ventoso I., Kochetov A., Montaner D., Dopazo J., Santoyo J. Extensive translatome remodeling during ER stress response in mammalian cells. *PLoS One*. 2012;7(5):e35915.

Volkova O.A., Kochetov A.V. Interrelations between the nucleotide context of human start AUG codon, N-end amino acids of the encoded protein and initiation of translation. *J. Biomol. Struct. Dyn.* 2010;27(5):611-618. DOI 10.1080/ 07391102.2010.10508575.

Więsyk A., Candresse T., Zagórski-Ostoja W., Góra-Sochacka A. Viability and genetic stability of potato spindle tuber viroid mutants with indels in specific loops of the rod-like secondary structure. *Virus Res*. 2017;240:94-100. DOI 10.1016/j.virusres.2017.07.024.

Więsyk A., Lirski M., Fogtman A., Zagórski-Ostoja W., Góra-Sochacka A. Differences in gene expression profiles at the early stage of *Solanum lycopersicum* infection with mild and severe variants of potato spindle tuber viroid. *Virus Res*. 2020;286:198090. DOI 10.1016/j.virusres.2020.198090.

Wu J., Leontis N.B., Zirbel C.L., Bisaro D.M., Ding B. A three-dimensional RNA motif mediates directional trafficking of *Potato spindle tuber viroid* from epidermal to palisade me-

sophyll cells in *Nicotiana benthamiana*. *PLoS Pathog*. 2019; 15(10):e1008147. DOI 10.1371/journal.ppat.1008147.

Wu J., Zhou C., Li J., Li C., Tao X., Leontis N.B., Zirbel C.L., Bisaro D.M., Ding B. Functional analysis reveals G/U pairs critical for replication and trafficking of an infectious non-coding viroid RNA. *Nucleic Acids Res*. 2020;48(6):3134-3155. DOI 10.1093/nar/gkaa100.

Wu Q., Ding S.W., Zhang Y., Zhu S. Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. *Annu. Rev. Phytopathol*. 2015;53:425-444. DOI 10.1146/annurev-phyto-080614-120030.

Yanagisawa H., Matsushita Y. Differences in dynamics of horizontal transmission of *Tomato planta macho viroid* and *Potato spindle tuber viroid* after pollination with viroid-infected pollen. *Virology*. 2018;516:258-264. DOI 10.1016/j.virol.2018.01.023.

Yanagisawa H., Sano T., Hase S., Matsushita Y. Influence of the terminal left domain on horizontal and vertical transmissions of tomato planta macho viroid and potato spindle tuber viroid through pollen. *Virology*. 2019;526:22-31. DOI 10.1016/j.virol.2018.09.021.

Zheng Y., Wang Y., Ding B., Fei Z. Comprehensive transcriptome analyses reveal that potato spindle tuber viroid triggers genome-wide changes in alternative splicing, inducible trans-acting activity of phased secondary small interfering RNAs, and immune responses. *J. Virol*. 2017;91(11):e00247-17. DOI 10.1128/JVI.00247-17.

ORCID ID

A.V. Kochetov orcid.org/0000-0003-3151-5181
A.Y. Pronozin orcid.org/0000-0002-3011-6288
D.A. Afonnikov orcid.org/0000-0001-9738-1409
O.S. Afanasenko orcid.org/0000-0001-7368-0797

Acknowledgements. The work was supported by a grant from the Russian Science Foundation (Project 20-46-07001).

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

Received November 9, 2020. Revised January 25, 2021. Accepted January 28, 2021.

Вавиловский журнал генетики и селекции. 2021;25(3):276-283
DOI 10.18699/VJ21.031

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

Methodological approaches for producing doubled haploids in sugar beet and red beet (*Beta vulgaris* L.)

T.R. Grigolava ✉, A.V. Vishnyakova, A.A. Sinitsyna, A.V. Voronina, O.N. Zubko, O.V. Zudova, S.G. Monakhos

Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia
✉ grigolava1@gmail.com

Abstract. The *in vitro* production of doubled haploids is a biotechnological path of an accelerated development of parental lines in F1-hybrid breeding programs. Unlike the traditional inbreeding method requiring 5 to 6 generations to reach a sufficient homozygosity of lines, the number of generations to produce pure lines of beet by haploid technologies is reduced to 2. The production of doubled haploids by gynogenesis is the most common biotechnological approach in sugar and red beets. Protocols for the production of doubled haploids for *B. vulgaris* species are few and have been developed mainly for sugar beets. There are no protocols for the production of doubled haploids for red beet (*B. vulgaris* convar. *esculenta* Salisb.), and the protocols developed for sugar beet (*B. vulgaris* convar. *saccharifera* Alef.) are ineffective for red beet, even though these two crops belong to the same species. The greatest success has been achieved in the production of doubled haploids by gynogenesis through isolated ovule culture, especially in sugar beet. Studies on the production of doubled haploids by androgenesis were actively carried out in the 1970s and 1980s and did not lead to the production of regenerated plants. However, at present, there is renewed interest among researchers in this approach, and scientists in different countries are conducting studies of *Beta vulgaris* androgenesis through isolated microspore culture. This article provides an overview of studies devoted to the production of doubled haploids, addressing the main problems of doubled haploid technologies, and methods to increase the frequency of embryogenesis and doubled haploid plant formation in *B. vulgaris* crops.
Key words: *Beta vulgaris*; haploid technology; gynogenesis; microspore culture; embryogenesis; doubled haploids.

For citation: Grigolava T.R., Vishnyakova A.V., Sinitsyna A.A., Voronina A.V., Zubko O.N., Zudova O.V., Monakhos S.G. Methodological approaches for producing doubled haploids in sugar beet and red beet (*Beta vulgaris* L.). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021. DOI 10.18699/VJ21.031

Методические подходы создания удвоенных гаплоидов сахарной и столовой свеклы (*Beta vulgaris* L.)

Т.Р. Григолава ✉, А.В. Вишнякова, А.А. Синицына, А.В. Воронина, О.Н. Зубко, О.В. Зудова, С.Г. Монахос

Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, Россия
✉ grigolava1@gmail.com

Аннотация. Производство удвоенных гаплоидов *in vitro* – актуальный биотехнологический способ ускоренного создания родительских линий для селекции гибридов F1. В отличие от классического инбридинга время создания гомозиготных линий свеклы (*Beta vulgaris*) с помощью технологии удвоенных гаплоидов сокращается с пяти–шести до двух поколений. Гиногенез является наиболее распространенным биотехнологическим методом производства удвоенных гаплоидов сахарной и столовой свеклы. Протоколы производства удвоенных гаплоидов для видов *B. vulgaris* немногочисленны и разработаны в основном для сахарной свеклы (*B. vulgaris* convar. *saccharifera* Alef.). Наибольший успех достигнут в производстве удвоенных гаплоидов сахарной свеклы гиногенезом в культуре изолированных семязачатков. Для столовой свеклы (*B. vulgaris* convar. *esculenta* Salisb.) проведены единичные исследования с показанной низкой эффективностью производства гаплоидных растений андро- и гиногенезом. В итоге протоколы производства удвоенных гаплоидов столовой свеклы отсутствуют, а протоколы, разработанные для сахарной свеклы, неэффективны для столовой, несмотря на принадлежность к одному виду. Исследования производства удвоенных гаплоидов путем андрогенеза у представителей рода *Beta* активно проводились в 70–80-х гг. прошлого столетия и не закончились получением растений-регенерантов, однако в настоящее время среди ученых снова возник интерес к данному методу и в разных странах возобновлены работы по изучению андрогенеза у представителей рода *Beta*. Статья содержит обзор исследований, посвященных созданию удвоенных гаплоидов; обсуждение подходов решения основных проблем при получении удвоенных гаплоидов и методов, позволяющих повысить выход эмбриоидов и растений-регенерантов, а также удвоенных гаплоидов у растений вида *B. vulgaris*.
Ключевые слова: *Beta vulgaris*; гаплоидные технологии; гиногенез; культура микроспор; эмбриогенез; удвоенные гаплоиды.

Introduction

Subspecies of *B. vulgaris* is a valuable vegetable (red beet), fodder (feed beet) and technical (sugar beet) crops. Currently, the main trend in the breeding of these crops is development of F1 hybrids, based on the crossing of two homozygous parental lines. The traditional method to develop homozygous lines is inbreeding by self-pollination followed by selection of phenotypically uniform families over a minimum of 4–6 generations; this procedure takes 8–12 years for biennial crops (De La Fuente et al., 2013). The long-term development of parental lines by inbreeding is one of the limiting factors and a major problem in competitive F1 hybrids breeding. Doubled haploids (DH) technologies offer a time-saving approach to obtaining pure breeding lines by reducing the period to approximately 3–5 years (Zhuzhzhalova et al., 2020) and even less. A significant advantage of the doubled haploid production technology is the ability to achieve fully homozygous plant genotype in one generation. Another advantage of DH technology is the manifestation of recessive alleles in haploid plants masked in a heterozygous state in diploid plants, which facilitates the identification, assessment, and selection of plants with traits of agronomic importance (Doctrinal et al., 1989; Klimek-Chodacka, Baranski, 2013).

Doubled haploids of agricultural plants are produced *in vivo* by parthenogenesis or *in vitro* by isolated microspores culture, anthers, unfertilized ovules, etc. (Palmer, Keller, 2005). Among available DH technologies, the isolated unfertilized ovules culture (gynogenesis) is commonly used for doubled haploid production in *B. vulgaris*, particularly in sugar beet.

Sugar beet isolated unfertilized ovules culture is a simple but also a laborious technology. The frequency of embryos formation of the most responsive genotypes could reach 15 embryos per 100 cultivated ovules (Wremerth, Levall, 2003). Besides, the forming of mother plant clones from the somatic cells surrounding the embryo sac is not excluded. That makes it necessary to develop reliable methods for homo- and heterozygotes differentiation with the selection of the homozygotes among regenerated plants. Isolated microspores culture is a technology that avoids the culturing of somatic cells and somatic embryo formation.

Isolated microspores culture technology and isolated anthers culture of beets was considered ineffective for a long time, and researchers managed to obtain only callus or proembryo structures without further regeneration, or somatic clones (Banba, Tanabe, 1972; Goska, Rogozinska, 1981; Van Geyt et al., 1985; Herrmann, Lux, 1988a). In 2017 Polish researchers managed to obtain embryos of red beet in microspores and anthers culture, however, the unrooted rosettes obtained from them died (Gorecka et al., 2017).

The purpose of this review is to collect and summarize data by various DH production approaches in *B. vulgaris* cultures, as well as outline the main problems and the ways to solve them.

Historical overview of *Beta* DH technology development

The first haploids of sugar beet were discovered in 1945 by A. Levan (1945), later similar discoveries were reported by K. Zimmermann (1953), H.E. Fischer (1956), Th. Butterfass (1959), A. Kruse (1961) and B.L. Hammond (1966). Haploids were obtained from four types of material: 1) progeny of plants treated with polyploidizing agents; 2) offspring obtained from seeds of diploid or anisoploid varieties; 3) offspring obtained by vegetative propagation of diploid, cytoplasmically male sterile plants; 4) from anisoploid sugar beet plants.

Work on the experimental production of sugar beet haploids was started by N. Bosemark (1971): after pollinating sterile diploid beet plants with tetraploid plant pollen, he found about 0.2 % of haploids in the offspring. Attempts to produce haploids using distant hybridization were carried out in 1983 in Czechoslovakia: I. Seman crossed male sterile sugar beet plants with salad beets: the haploid yield was 0.013 % (Seman, 1983). A. Buchter-Larsen (1986) proposed an original method for the production of haploids – he combined pollination with irradiated pollen and the subsequent rescue of embryos, however, almost all plants turned out to be heterozygous for one or more genes. The obtained embryo rescue from pollination with pollen from other *Beta* species resulted in a relatively high yield of non-haploid plants (Buchter-Larsen, 1986). Since this method provided a low yield of haploids and was extremely laborious, the researchers abandoned attempts to produce doubled haploids *in vivo*.

Androgenesis is a simple and effective way of producing haploids *in vitro* in many crops. In the genus *Beta*, researchers have also attempted to produce haploids using isolated anther cultures. The first attempts to produce sugar beet haploids *in vitro* were undertaken by H. Banba and H. Tanabe (1972); when cultivating isolated anthers, they obtained one plant, the origin (somatic clone or haploid) of which is not indicated. M. Goska and J.H. Rogozinska (1981) continued their research on the cultivation of isolated sugar beet anthers and produced plants that were not androgenic.

D. Hosemans and D. Bossoutrot (1983) developed nutrient media for the cultivation of isolated ovules and carried out cytological studies of ovules developing on culture media. Their results showed that ovules containing a mature 7-nuclear embryo sac are most prone to *in vitro* development. The researchers obtained 17 haploid plants from 7237 isolated sugar beet ovules from male sterile donor plants.

Researchers have repeatedly attempted to obtain haploid beet plants by cultivating anthers or isolated microspores, however, they either failed to obtain full-fledged regenerant plants (Van Geyt et al., 1985; Gorecka et al., 2017; Gontarenko, Gerasimenko, 2018), or they were not haploids (Herrmann, Lux, 1988a). Therefore, gynogenesis

became the main haploids producing method in sugar and red beets.

The bulk of research on the study of gynogenesis was carried out on sugar beet. Researchers studied the influence on the frequency of embryogenesis and embryo germination/regeneration into a haploid plant of such factors as the type and concentration of growth regulators (D'Halluin, Keimer, 1986; Herrmann, Lux, 1988b; Ferrant, Bouharmont, 1994; Podvigina, 2003; Wremerth, Levall, 2003; Pazuki et al., 2018b), cold pretreatment of buds and ovules (Herrmann, Lux, 1988b; Svirshchevskaya, Dolezel, 2000; Podvigina, 2003; Pazuki et al., 2018b), shock stimulation by high temperatures of isolated ovules (Baranski, 1996; Wremerth, Levall, 2003), the time of ovule introduction into the culture *in vitro* (Lux et al., 1990; Baranski, 1996), the arrangement of buds on the inflorescence (D'Halluin, Keimer, 1986; Doctrinal et al., 1989; Podvigina, 2003), etc.

The gynogenesis of red beet is less studied. This topic was studied by R. Baranski (1996), who deliberated the effect of media growth regulators and the cultivation temperature of isolated ovules on the yield of embryos and callus, as well as the effect of the season of the year and mother plants growing condition (greenhouse/field) on the yield of regenerants. R. Baranski showed that the yield of regenerants is higher from the ovules taken from donor plants grown in the greenhouse compared to the field-grown donor plants. However, there was no significant difference in culturing efficiency of ovules obtained from donor plants grown in spring and summer seasons.

Factors affecting DH production efficiency by isolated ovules culture

The process of haploid production is determined at the genetic level, but it is implemented depending on physiological conditions and inducing factors, which directly affects the frequency of haploid regenerants (Baranski, 1996; Podvigina, 2003). Major factors are the genotype of donor plant, the developmental stage of the female gametophyte, the location of the bud on the inflorescence.

Exogenous factors are of high importance and they affect the regenerative ability of cultivated ovules. These factors include season and duration of growing (age) of donor plants, cold and X-ray treatment of flower buds, the percentage ratio of growth regulators in culture media, temperature and other cultivation conditions of isolated ovules (Van Geyt et al., 1987; Lux et al., 1990; Gurel et al., 2000).

Among the factors influencing the embryogenesis efficiency, genotype is considered the most significant. Researchers pointed out that the most responsive to the inoculation of isolated ovules into the culture media are the hybrids and inbreed lines, but the CMS lines and variety-population have the lowest regenerative ability (Gurel et al., 2000; Podvigina, 2003).

In the genus *Beta*, a study of the number of genes, which controls the ability of genotypes in terms of forming embryos *in vitro*, was not carried out, probably, due to the low responsiveness of genotypes and also the technical laboriousness which is conjugated with haploids production. Responsive genotypes can be identified only experimentally. The responsive marker of sugar beet genotype may be the presence of abnormal structures in the male gametophyte (the presence of abnormal pollen grains and microspores) caused by dysfunction of the spindle apparatus, irregular formation of a callose wall, and the absence of cytokinesis during meiosis (Podvigina, 2003).

The selected buds location on inflorescence and the embryo sacs' development stage are essentially important in sugar beet haploids induction. The highest regenerative activity is characterized by unfertilized ovules from buds 1 to 25 starting from bottom to top, the flower in the middle part of the inflorescence. In addition, the maximum yield of haploids is noted through the central shoot and first-order shoots compared to second-order branches (D'Halluin, Keimer, 1986; Doctrinal et al., 1989; Podvigina, 2003). The embryogenesis ability of isolated ovules is saved at all stages of the female gametophyte development; however, the 7th and 8th nuclear embryo sacs are the most responsive to embryogenesis and more easily pass from the gametophytic developmental pathway to the sporophyte one (Van Geyt et al., 1987; Podvigina, 2003). Markers of the 7th and 8th nuclear development stages of the ovule embryo sac for isolation and *in vitro* culture inoculation are the presence of mononuclear microspores and two-to three-core pollen in the anthers with ovules in one bud (Podvigina, 2003). Buds containing ovules at the appropriate stage of development can be found 1–5 days before flowering.

Donor plants growing

Donor plants preparation is one of the most fundamental stages in doubled haploids production technology. Donor plants must be vigorous and healthy enough to produce high quality explants. W.E. Wremerth and M.W. Levall (2003) recommend adding solutions of macro- and microfertilizers weekly in order to grow healthy. Most researchers recommend growing donor plants in greenhouses or climatic chambers to minimize the impact of unfavorable weather factors and pest damage (Lux et al., 1990; Baranski, 1996; Gurel, 2000; Wremerth, Levall, 2003). However, in O.A. Podvigina studies (2003), the ovules which were taken from plants in the field had the highest regenerative capacity.

It is recommended to grow donor plants in summer; since the ovules from such plants are more responsive to *in vitro* cultivation compared to the ovules from those grown in autumn-winter season (Lux et al., 1990; Baranski, 1996). O.A. Podvigina (2003) made an interesting observation in terms of the relationship between the regenerative ability of ovules and climatic conditions during introduction period into culture – with sharp fluctuations in day-night air temperature, the yield of haploid seedlings has been increased.

Embryogenesis induction

In *B. vulgaris*, heat treatment of buds and isolated ovules is used to stimulate embryogenesis: most often, the buds are pretreated at low temperatures 4–6 ° C for up to 5 days, followed by the cultivation of isolated ovules at 28–32 ° C (Lux et al., 1990; Baranski, 1996; Gurel et al., 2000; Podvigina, 2003; Wremerth, Levall, 2003). A. Pazuki et al. (2018b) showed that treatment of inflorescences for 7 days at 4 °C can stimulate embryogenesis in ovules; then, isolated ovules were cultivated at 27±2 °C in a climatic chamber with an 18-hour photoperiod. O.A. Podvigina (2003) directly cultivated isolated ovules of sugar beet at 4 °C for 5 days, which stimulated their development even on B5 hormone-free medium, however, the highest yield was observed on B5 medium supplemented with 2 mg/L gibberellin.

O.A. Podvigina (2003) used ovules pretreated with X-rays to stimulate embryogenesis. Studies showed that the yield of haploid regenerants depended on the X-ray dose, the maximum frequency was 5.3 % at a treatment dose of 3000 roentgens, an increase in the radiation dose to 5000 roentgens did not have a stimulating effect and led to the appearance of unwanted mutations.

Many authors recommended isolated ovules cultivation at high temperature before the embryoids appearance (Lux et al., 1990; Baranski, 1996; Wremerth, Levall, 2003). However, there are also studies on lower temperature cultivation of isolated ovules (Baranski, 1996; Podvigina, 2003). O.A. Podvigina (2003) cultivated isolated sugar beet ovules at a temperature of 21–26 °C and showed that the optimum temperature is 23–25 °C. W.E. Wremerth and M.W. Levall (2003) developed a protocol to produce doubled sugar beet haploids, and according to the authors, the optimal temperature for isolated ovules cultivation is 30±2 °C; the maximum frequency of embryo formation in the most responsive genotypes was 15 %. R. Baranski (1996) conducted research on the influence of the temperature of isolated ovules cultivation in beet on the yield of embryoids. It was found that the temperature of 25 °C was the least favorable for development, and regenerant yield from the ovules was 4 %; there were no significant differences in regenerant yield between temperatures of 27 and 32 °C and the yield was 12.7 and 11.3 %, respectively.

The isolated ovules are usually incubated in the dark until the appearance of embryoids/callus, after which they are placed in separate culture vessels and cultured in the light.

Culture medium

The cultivation conditions of isolated ovules affect both the number of regenerants (embryoids and callus) and their quality. The correct selection of culture medium is important for the production of haploid plants in the culture of isolated ovules.

Researchers use different nutrient media to cultivate isolated ovules, MS, N6, B5. The most commonly used solid media are MS and B5 with the addition of various growth regulators. Ovules were cultured on liquid media by H. Lux et al. (1990) and E.N. Vasilchenko et al. (2017). E.N. Vasilchenko et al., studying the effect of nutrition media concentration, showed that the proliferation of nuclei and cells of female gametophyte was activated on a liquid media, which had an effect on the initiation of neoplasms, and after transferring the resulting structures to solid media, the induction of haploid regenerants was observed.

Agar, agarose, and phytagel are usually used for culture media gelation in case of beet isolated ovules culture. Basically, for the culture of isolated ovules, media with the addition of agar or agarose are used (Baranski, 1996; Podvigina, 2003; Wremerth, Levall, 2003). W.E. Wremerth and M.W. Levall (2003) recommended using agarose as a gelling agent for embryo induction and doubled haploids production in sugar beet as well as agar in shoot and root induction media. In the studies (Gurel et al., 2000; Pazuki et al., 2017; Vasilchenko et al., 2017) was noted a positive effect of phytagel at a concentration of 2–3 g/L on embryogenesis and regeneration; its advantages are low consumption, low cost and impact similar to agar.

Growth regulators have the most significant influence on the development of explants (Seman, Farago, 1990; Gurel et al., 2000; Podvigina, 2003). During the culture, there are five pathways of development of unfertilized isolated ovules:

1. One embryoid is formed from the ovule (direct regeneration).
2. The ovule cells divide disorganized, resulting in the formation of callus tissue, from which secondary embryoids are formed.
3. Degeneration of the primary regenerant into a callus-like structure and further secondary regeneration through the formation of adventive shoots.
4. Formation of non-morphogenic callus.
5. Formation of amorphous structures, transformation of the primary regenerant into a callus-like formation without further regeneration (Seman, Farago, 1990; Podvigina, 2003).

O.A. Podvigina (2003) and E.N. Vasilchenko et al. (2017) indicated that adding gibberellin (2 mg/L) to the culture medium causes embryogenesis, adding of auxins (IBA) and cytokinins (6-BAP and kinetin) to gibberellin stimulates the growth of callus along with embryoids and morphogenesis through all possible pathways of development of isolated ovules.

W.E. Wremerth and M.W. Levall (2003) recommended using stepwise cultivation of isolated ovules on media with various combinations and concentrations of growth regulators. At the first stages of cultivation, media containing 2.4-D 0.5 mg/L and 6-BAP 0.3 mg/L are used in order to induce embryo or callusogenesis.

Sucrose is added to the medium as a source of carbohydrates. However, there is no consensus on the amount of

sucrose in the culture medium, sucrose concentration varies within 30–100 g/L, depending on the technology. S. Gurel and his colleagues recommended adding sucrose to the medium for sugar beet embryogenesis at a concentration of 100 g/L (Gurel et al., 2000; Pazuki et al., 2018a, b). R. Baranski (1996) used culture media with the addition of 60 g/L sucrose to induce gynogenesis in red beet. W.E. Wremerth and M.W. Levall (2003) added 80 g/L sucrose to the embryo-induction medium and 20 g/L to the shoot-induction medium to produce sugar beet doubled haploids. H. Lux et al. (1990) added sucrose 100 g/L to the embryo induction medium and 20 g/L to the regeneration medium. The yield of embryoids varies greatly (widely) in each study, which makes it difficult to choose the optimal concentration of carbohydrates in the medium. However, in all studies, a general tendency is observed to induce embryogenesis, so that media with an increased content of carbohydrates are used, and for regeneration with a reduced one.

S. Gurel et al. (2000) have demonstrated that the addition of 0.5 % activated charcoal to the culture medium significantly increases the yield of embryoids (on average for genotypes from 3.3 to 12.8 %), and E.N. Vasilchenko et al. (2017) have indicated that the addition of 3 g/L of activated charcoal negatively affects the development of sugar beet regenerant plants, which may be due to the adsorption of hormones from the media. At the same time, E.N. Vasilchenko recommends the use of activated charcoal at the rooting stage, which can crucially increase the yield of rooted plants due to the adsorption of phenolic compounds that inhibit root formation.

While studying the induced embryogenesis of red beet, R. Baranski (1996) found that the highest yield of regenerants was observed on N6 medium (by Chu) when using a combination of 0.5 mg/L IAA and 0.2 mg/L 6-BAP; the yield of regenerants amounted to a maximum of 8.3 %.

Plant regeneration

Plant regeneration from embryos and/or callus is one of the most significant stages in doubled haploids development. Researchers obtain regenerated plants either on the same culture medium on which ovules are cultivated (Baranski, 1996), or using media with different concentrations and growth regulators types (Podvigina, 2003; Wremerth, Levall, 2003; Vasilchenko et al., 2017).

O.A. Podvigina (2003) indicated that haploid plants at the first stages of development are characterized by poor development and viability, due to their haploid status. The death of plants at this stage can reach 45.5 %, depending on the genotype of the donor plant. To increase the yield of regenerant plants, it was proposed to introduce a stage of stabilization of haploids, including sequential plant-regenerants cultivation on media without the addition of growth regulators, gibberellin, 6-BAP, and IBA. Several passages with alternating cultivation on media with growth regulators and without growth regulators made it possible to reduce the excess hormones concentration in the haploids tissues and stimulate them for further regeneration. It was also noted during the study that the ability of regenerated haploids to adapt to changing environment depends on the genetic characteristics of donor plants.

W.E. Wremerth and M.W. Levall (2003) have developed a technology of stepwise cultivation of sugar beet regenerants obtained from isolated ovules to solve the problem of viability of developing regenerant plants. For the shoot regeneration from sugar beet embryoids, it is recommended to use MS culture media containing kinetin (0.2 mg/L) and IAA (0.1 mg/L), the pre-root MS medium contains kinetin at a concentration of 0.5 mg/L, IBA at a concentration of 0.55 mg/L. For rooting it is recommended to use 1/2 MS medium with a high concentration of IBA (5.5 mg/L). A. Pazuki et al. (2018a) studied the effect of adding an anti-stress agent, the amino acid proline, on shoot and root formation in regenerated sugar beet plants obtained from isolated ovules. The addition of 0.2 or 0.3 mM proline to media stimulated active shoot formation and faster rooting of plants in comparison with the complete absence of proline or its concentration in the medium of 0.1 and 0.4 mM.

Polyploidization

The cells of plants regenerating from ovules can be haploid, diploid, polyploid and occur in one regenerant in different proportions. The level of spontaneous diploidization in the studied accessions of sugar beet varies greatly: S. Gurel et al. (2000) recorded that only 5 % of plants underwent spontaneous diploidization; M. Goska (1997) obtained from 2 to 10 % of diploid sugar beet plants, and according to M. Tomaszewska-Sowa (2010), adding kinetin to the media gives diploid plants up to 93.8 %.

There are no reliable data on the factors that affect the degree of spontaneous diploidization of haploids; therefore, it is necessary to transfer the obtained haploid plants to the diploid level. Colchicine is usually used to double the chromosome number of haploid regenerant plants (Gurel et al., 2000; Podvigina, 2003). Haploid plants treatment is carried out by treating the meristematic tissues of a root crop or inflorescences with a solution of colchicine, by immersing the roots in a solution of colchicine or in a culture medium containing an antimitotic agent *in vitro* (Podvigina, 2003). Concentration and exposure may vary greatly: S. Gurel et al. (2000) recommended doubling the number of chromosomes by placing haploid plants on a medium supplemented with colchicine at a concentration of 5 g/L for 5 min. O.A. Podvigina (2003) pointed out 83.3 % the rate of diploidization when adding 0.05 % colchicine to the medium and exposure for 2 days.

Mixoploidy is a common phenomenon for sugar beet meristems (Kharechko-Savitskaya, 1940; Yudanova et al., 2004; Lukaszewska, Sliwinska, 2007); this complicates determining ploidy level of the developed plants.

The ploidy level of plants can be determined not only by direct counting of chromosomes number in the meristematic cells under microscope, but also by an indirect indicator – the number of chloroplasts in guard cells of the stomata (Yudanova et al., 2004). However, the chloroplasts number in stomatal guard cells may depend not only on the ploidy level, but also on the mode of reproduction (self-pollination or crossing). For example, in a diploid heterotic hybrid of sugar beet, the average number of chloroplasts can be 12–15 pcs. per cell, while in self-pollinated samples the average number of chloroplasts per cell is about 18, which makes this indirect method unreliable (Maletskiy et al., 2013).

The problem of determining the ploidy level of regenerated beet plants is solved by using flow cytometry (Gorecka et al., 2017; Vasilchenko et al., 2017) or absorption cytometry (Yudanova et al., 2004).

Some researchers (Podvigina, 2003; Tomaszewska-Sowa, 2010) considered it possible to distinguish haploid plants according to their phenotype: haploid plants are very different from diploid ones – they have small, numerous narrow leaves and a smaller habitus compared to diploids.

Microspores and anthers culture

Compared to gynogenesis, androgenesis is a less laborious method, because it does not require the isolation of small ovules. Isolated microspores culture allows to avoid formation of somatic clones as in the case of gynogenesis (from tissue surrounding embryo sac). In this connection, research in this area is promising, but in the genus *Beta*, such studies were rarely carried out and were not very successful.

Culture of isolated microspores and anthers culture are successfully used in many plant species to produce doubled haploids. However, haploid induction by isolated microspores and anthers in sugar beet and beetroot leads to the formation of proembryo structures, which sometimes form callus and/or roots. Early attempts to produce sugar beet haploids in anthers culture did not lead to obtaining androgenic plants (Banba, Tanabe, 1972; Goska, Rogozinska, 1981; Van Geyt et al., 1985; Herrmann, Lux, 1988a). One of the probable reasons for the failure of all researchers may be the presence of amyloplasts in pollen grains, which inhibits androgenesis due to the increased starch content in the plastids of mononuclear microspores (Sangwan, Sangwan-Norreel, 1987). The possibility of elimination starch grains in beetroot microspores was studied by K. Gorecka et al. (2017), for which two processing options were investigated. In the first case, donor plants were irrigated with a solution of gibberellin at a concentration of 50 mg/L, 250 ml per plant twice a week, which led to an increase in the yield of androgenic regenerants in only one of the accessions, while the other genotypes did not show any change in the yield of embryoids, and in one genotype there was no regeneration at all. In the second case, isolated anthers were kept in a solution of alpha-amylase at a concentration of 3 mg per 80 ml of water (from barley malt type VIII-A, Sigma-Aldrich) for 2 minutes, and then transferred to a medium for the induction of androgenesis, which led to the formation of 2 embryos in two genotypes. The researchers failed to obtain the plants, because regenerated rosettes turned black and died. M. Klimek-Chodacka and R. Baranski (2013) faced the same problem in some genotypes of beetroot in unfertilized ovules culture, which is connected with genotype-specificity.

K. Gorecka et al. (2017) obtained callus culturing isolated microspores and anthers in beetroot. The authors found that buds 1.3–1.5 mm long contain about 80 % of microspores of the mononuclear stage of development and about 15 % of the binuclear stage, which is the most optimal for the culture of isolated microspores and anthers in most crops. The authors indicated B5 with the addition of 100 g/L sucrose and 100 mg/L 2.4-D as the best culture medium for the cultivation of anthers and microspores. In the cytological examination of the obtained samples of callus and rosettes, the ploidy level was 4x, which indicates repeated endoreduplication in the callus tissue.

S.M. Gontarenko and G.M. Gerasimenko (2018) managed to obtain embryoids in the culture of isolated anthers in sugar beet; the embryoid yield was 0.15–0.92 %. They determined that the optimal stage of microspore development for the anther culture is the mononuclear stage. They showed that pretreatment of explants using low-temperature stress (4–8 °C) for 3–15 days is a factor initiating the transition of microspores from gametophytic to sporophytic development pathway, while pretreatment with high temperatures (30–32 °C) does not give positive effect. The best culture medium for the anther culture turned out to be half-concentration MS with the addition of a number of vitamins (B1 10 mg/L, B6 1 mg/L, PP 1 mg/L, C 1 mg/L) and amino acids (glutamic acid 250–500 mg/L, aspartic 30–50 mg/L, tyrosine 1–10 mg/L, arginine 2–10 mg/L, hydroxyproline 2–4 mg/L).

Conclusion

The development of homozygous lines using haploid technologies remains in demand by breeders around the world. The production of pure lines using doubled haploids has several advantages over conventional methods: homozygosity is achieved in one generation, eliminating the need for several generations of self-pollination. The time saved is substantial, particularly in biennial crops. The most developed technology for DH lines production in beetroot nowadays is the technology of isolated ovules, which has been developed mainly on sugar beets. This technology for developing pure lines is rather laborious in comparison with the technology of isolated microspores culture. However, the latter is practically not applicable due to the

lack of research and efficient protocol. In this regard, the study and development of the sugar and red beets doubled haploids technologies based on isolated microspores culture approach should be recommended.

References

Banba H., Tanabe H. A study of anther culture in sugar beet. *Bull. Sugar Beet Res.* 1972;14:9-16.

Baranski R. *In vitro* gynogenesis in red beet (*Beta vulgaris* L.): effects of ovule culture conditions. *Acta Soc. Bot. Pol. Tow. Bot.* 1996; 65(1-2):57-60. DOI 10.5586/asbp.1996.010.

Bosemark N. Haploids and homozygous diploids, triploids and tetraploids in sugar beet. *Hereditas.* 1971;69:193-204.

Buchter-Larsen A. *In situ* Induction af Haploide Embryoner pa Sukkerroer efterfulgt af *In vitro* Embryokultur. Thesis, Acad. Technical Sciences. Copenhagen, 1986.

Butterfass Th. Ploidie und Chloroplastenzahlen. *Ber. Dtsch. bot. ges*. 1959;72:440-451.

De La Fuente G.N., Frei U.K., Lübberstedt T. Accelerating plant breeding. *Trends Plant Sci.* 2013;18:667-672. DOI 10.1016/j. tplants.2013. 09.001.

Doctrinal M., Sangwan R.S., Sangwan-Norreel B.S. *In vitro* gynogenesis in *Beta vulgaris* L.: Effects of plant growth regulators, temperature, genotypes and season. *Plant Cell Tissue Organ Cult*. 1989; 17:1-12. DOI 10.1007/BF00042276.

D'Halluin K., Keimer B. Production of haploid sugar beets (*Beta vulgaris* L.) by ovule culture. In: Genetic Manipulation in Plant Breeding. Berlin, 1986. DOI 10.1515/9783110871944.

Ferrant V., Bouharmont J. Origin of gynogenetic embryos of *Beta vulgaris L. Sex. Plant Reprod.* 1994;7:12-16. DOI 10.1007/ BF0024 1883.

Fischer H.E. Untersuchungen an Zwillingen von *Beta vulgaris* L. *Der Züchter.* 1956;26:136-152. DOI 10.1007/BF00710616.

Gontarenko S.M., Gerasimenko G.M. Direct induced androgenesis in culture *in vitro* in sugar beet (*Beta vulgaris* L.). *Plant Varieties Studying and Protection.* 2018;14(4):375-381. DOI 10.21498/2518-1017.14.4.2018.151900. (in Ukrainian)

Gorecka K., Kryzanowska D., Kowalska U., Kiszczak W., Podwyszynska M. Development of embryoids by microspore and anther cultures of red beet (*Beta vulgaris* L. *subsp. vulgaris*). *J. Cent. Eur. Agric.* 2017;18(1):185-195. DOI 10.5513/ JCEA01/18.1.1877.

Goska M. Monografie i Rozprawy Naukowe. Haploidy i Podwojone Haploidy Buraka Cukrowego (*Beta vulgaris* L.) oraz Możliwości Ich Wykorzystania w Hodowli. Radzików: Instytut Hodowli i Aklimatyzacji Roślin, 1997.

Goska M., Rogozinska J.H. Recent results on obtaining beet haploids through *in vitro* culture of anthers. *Biuletyn Inst. Hodowli I Aklimatyzacji Roslin.* 1981;145:141-143.

Gurel S., Gurel E., Kaya Z. Doubled haploid plant production from unpollinated ovules of sugar beet (*Beta vulgaris* L.). *Plant Cell Rep.* 2000;19:1155-1159. DOI 10.1007/s002990000248.

Hammond B.L. Homozygous diploid sugar beets. *J. Am. Soc. Sugar Beet Technol.* 1966;14:75-78.

Herrmann L., Lux H. Antherenkultur bei Zuckerruben, *Beta vulgaris* L. var altissima. *Arch. Zuchtungsforsch. Berlin*. 1988a;18(6):375-383.

Herrmann L., Wetzel C., Lux H. Haploiden-technik bei der zuckerrube. *Potsdam Forsch. B.* 1988b;57:95-99.

Hosemans D., Bossoutrot D. Induction of haploid plant from *in vitro* culture of unpollinated beet ovules (*Beta vulgaris* L.). *Z. Pflanzenzucht.* 1983;91(1):74-77.

Kharechko-Savitskaya E.I. Cytology and embryology of sugar beet. Beet growing. Kiev, 1940. (in Russian)

Klimek-Chodacka M., Baranski R. Comparison of haploid and doubled haploid sugar beet clones in their ability to micropropagate and regenerate. *Electron. J. Biotechnol*. 2013;16(2):1-1. DOI 10.2225/vol16-issue2-fulltext-3.

Kruse A. Haploids in polyembryos of beet, *Beta vulgaris* L. *Roy. Vet. Agric. Coll.* Copenhagen: Yearb, 1961;87-98.

Levan A. A haploid sugar beet after colchicine treatment. *Hereditas.* 1945;31(3-4):399-410. DOI 10.1111/j.1601-5223.1945. tb02760.x.

Lukaszewska E., Sliwinska E. Most organs of sugar-beet (*Beta vulgaris* L.) plants at the vegetative and reproductive stages of development are polysomatic. *Sex. Plant Reprod.* 2007;20:99-107. DOI 10.1007/s00497-007-0047-7.

Lux H., Herrmann L., Wetzel C. Production on haploid sugar beet (*Beta vulgaris* L.) by culturing unpollinated ovules. *Plant Breed.* 1990; 104:177-183. DOI 10.1111/j.1439-0523.1990.tb00420.x.

Maletskii S.I., Yudanova S.S., Maletskaya E.I. Harmonic proportions of chloroplast number in stomata of guard cell populations in sugar beet (*Beta vulgaris* L.). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2013;17(1):72-80. (in Russian)

Palmer C.D., Keller W.A. Overview of haploidy. In: Haploids in Crop Improvement II. Biotechnology in Agriculture and Forestry. Springer, Berlin, Heidelberg, 2005;56:3-9. DOI 10.1007/3-540-26889-8_1.

Pazuki A., Aflaki F., Gürel S., Ergül A., Ekrem G. The effects of proline on *in vitro* proliferation and propagation of doubled haploid sugar beet (*Beta vulgaris*). *Turkish J. Bot.* 2018a;42(3):280-288. DOI 10.3906/bot-1709-14.

Pazuki A., Aflaki F., Gürel E., Ergül A., Gürel S. Gynogenesis Induction in Sugar Beet (*Beta vulgaris*) Improved by 6-Benzylaminopurine (BAP) and Synergized with Cold Pretreatment. *Sugar Tech.* 2018b;20:69-77. DOI 10.1007/s12355-017-0522-x.

Podvigina O.A. Theoretical substantiation and biotechnology methods in sugar beet breeding: Dr Sci. (Agricult.) Dissertation. Voronezh, 2003. (in Russian)

Sangwan R.S., Sangwan-Norreel B.S. Ultrastructural cytology of plastids in pollen grains of certain androgenic and nonandrogenic plants. *Protoplasma.* 1987;138:11-22. DOI 10.1007/ BF01281180.

Seman I. Possibilities of detection and induction of haploids in *Beta vulgaris* L. II *Biologia* (*Bratislava*). 1983;38:1113-1122.

Seman I., Farago J. *In vitro* cultivation of unfertilized ovules of sugar beet. Embryology and seed reproduction: XI Int. Symp. Leningrad, USSR, July 3–4, 1990. Leningrad, 1990;146.

Svirshchevskaya A., Dolezel J. Production and Performance of Gynogenetic Sugarbeet Lines. *J. Sugar Beet Res. USA.* 2000;4(37):117-133.

Śliwińska E., Łukaszewska E. Polysomaty in growing *in vitro* sugarbeet (*Beta vulgaris* L.) seedlings of different ploidy level. *Plant Sci.* 2005;168:1067-1074. DOI 10.1016/j.plantsci.2004.12.003.

Tomaszewska-Sowa M. Cytometric analyses of sugar beet (*Beta vulgaris* L.) plants regenerated from unfertilized ovules cultured *in vitro*. *Plant Breed.* 2010;2:231-235.

Van Geyt J., D'Halluin K., Jacobs M. Induction of nuclear and cell divisions in microspores of sugar beet (*Beta vulgaris* L.). *Z. Pflanzenzuecht.* 1985;95:325-335.

Van Geyt J., Speckmann G.J., D'Halluin K., Jacobs M. *In vitro* induction of haploid plants from unpollinated ovules and ovaries of the sugarbeet (*Beta vulgaris* L.). *Theor. Appl. Genet.* 1987;73:920-925. DOI 10.1007/BF00289399.

Vasilchenko E.N., Zhuzhzhalova T.P., Vashchenko T.G., Zemlyanukhina O.A., Karpechenko N.A., Podvigina O.A. Peculiarities of *in vitro* reproduction of sugar beet haploid regenerants. *Vestnik Voronezhskogo Gosudarstvennogo Agrarnogo Universiteta = Vestnik of Voronezh State Agrarian University.* 2017;3(54):57-66. DOI 10.17238/issn2071-2243.2017.3.57. (in Russian)

Wremerth W.E., Levall M.W. Doubled haploid production of sugar beet (*Beta vulgaris* L.): Published protocols for other crop plant species. Doubled Haploid Production in Crop Plants: A Manual. Kluwer Academic Publishers, 2003. DOI 10.1007/978-94-017-1293-4_38.

Yudanova S.S., Maletskaya E.I., Maletskii S.I. Epiplastome Variation of the Number of Chloroplasts in Stomata Guard Cells of Sugar Beet (*Beta vulgaris* L.) *Russ. J. Genet.* 2004;40(7):930-939. DOI 10.1023/B:RUGE.0000036525.14205.3f.

Zhuzhzhalova T.P., Kolesnikova E.O., Vasilchenko E.N., Cherkasova N.N. Biotechnological methods as a tool for efficient sugar beet breeding. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2020;24(1):40-47. DOI 10.18699/VJ20.593.

Zimmermann K. Vermendung haploider Pflanzen in der Zuchtung. *Ber. Deut. Bot. Ges.* 1953;66:28-30. DOI 10.1007/BF00710616.

ORCID ID

T.R. Grigolava orcid.org/0000-0002-1594-8430
A.V. Vishnyakova orcid.org/0000-0002-9160-1164
A.A. Sinitsyna orcid.org/0000-0002-9470-3676
A.V. Voronina orcid.org/0000-0003-0249-246X
O.N. Zubko orcid.org/0000-0001-9701-6647
O.V. Zudova orcid.org/0000-0001-5405-9460
S.G. Monakhos orcid.org/0000-0001-9404-8862

Acknowledgements. This work was supported by the Ministry of Science and Higher Education of the Russian Federation by the Agreement No. 075-15-2020-905, Nov 16, 2020, for the establishment and development of a World-class Scientific Center "Agrotechnologies for the Future."

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

Received October 5, 2020. Revised December 28, 2020. Accepted January 3, 2021.

Вавиловский журнал генетики и селекции. 2021;25(3):284-291
DOI 10.18699/VJ21.032

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

Problems of mini-pig breeding

K.S. Shatokhin

Novosibirsk State Agrarian University, Novosibirsk, Russia
true_genetic@mail.ru

Abstract. This article provides an overview of some problems of the breeding and reproduction of laboratory mini-pigs. The most obvious of these are the lack of centralized accounting of breeding groups, uniform selection standards for reproduction and evaluation of breeding animals, as well as minimizing the accumulation of fitness-reducing mutations and maintaining genetic diversity. According to the latest estimates, there are at least 30 breeding groups of mini-pigs systematically used as laboratory animals in the world. Among them, there are both breed formations represented by several colonies, and breeding groups consisting of a single herd. It was shown that the main selection strategy is selection for the live weight of adults of 50–80 kg and the adaptation of animals to a specific type of biomedical experiments. For its implementation in the breeding of foreign mini-pigs, selection by live weight is practiced at 140- and 154-day-old age. It was indicated that different herds of mini-pigs have their own breeding methods to counteract inbred depression and maintain genetic diversity. Examples are the maximization of coat color phenotypes, the cyclical system of matching parent pairs, and the structuring of herds into subpopulations. In addition, in the breeding of foreign mini-pigs, molecular genetic methods are used to monitor heterozygosity. Every effort is made to keep the number of inbred crosses in the breeding of laboratory mini-pigs to a minimum, which is not always possible due to their small number. It is estimated that to avoid close inbreeding, the number of breeding groups should be at least 28 individuals, including boars of at least 4 genealogical lines and at least 4 families of sows. The accumulation of genetic cargo in herds of mini-pigs takes place, but the harmful effect is rather the result of erroneous decisions of breeders. Despite the fact that when breeding a number of mini-pigs, the goal was to complete the herds with exclusively white animals, in most breeding groups there is a polymorphism in the phenotype of the coat color.
Key words: laboratory mini-pigs; inbreeding; genetic diversity; recessive mutations; selection; lines; families; agriculture.

For citation: Shatokhin K.S. Problems of mini-pig breeding. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):284-291. DOI 10.18699/VJ21.032

Проблемы селекции лабораторных мини-свиней

К.С. Шатохин

Новосибирский государственный аграрный университет, Новосибирск, Россия
true_genetic@mail.ru

Аннотация. В статье представлен обзор проблем разведения и селекции лабораторных мини-свиней. Наиболее очевидные из них – отсутствие централизованного учета селекционных групп, единых стандартов отбора для воспроизводства и оценки племенных животных, а также минимизация накопления снижающих приспособленность мутаций и поддержание генетического разнообразия. По последним данным, в мире насчитывают не менее 30 селекционных групп мини-свиней, систематически используемых в качестве лабораторных животных. Среди них существуют как породные образования, представленные несколькими колониями, так и селекционные группы, состоящие из одного стада. Показано, что основная стратегия отбора включает селекцию на живую массу взрослых особей 50–80 кг и приспособленность животных к конкретному типу биомедицинских экспериментов. Для ее реализации в разведении зарубежных мини-свиней практикуют отбор по живой массе в 140- и 154-дневном возрасте. Указано, что в стадах мини-свиней представлены разные селекционные методы противодействия инбредной депрессии и поддержания генетического разнообразия. Примерами служат максимизация фенотипов масти, цикличная система подбора родительских пар и структурирование стад на субпопуляции. Кроме того, в разведении зарубежных мини-свиней для мониторинга гетерозиготности используют молекулярно-генетические методы. Количество инбредных скрещиваний в разведении лабораторных мини-свиней стараются минимизировать, что не всегда возможно из-за их малочисленности. Подсчитано, что во избежание тесного инбридинга численность селекционной группы должна быть не менее 28 особей, включающих хряков как минимум четырех генеалогических линий и свиноматок из не менее четырех семейств. Накопление генетического груза в стадах мини-свиней возможно, но вредоносный эффект является скорее следствием ошибочных решений селекционеров. Несмотря на то что при выведении ряда мини-свиней стояла цель укомплектовать стада исключительно белыми животными, в большинстве селекционных групп наблюдается полиморфизм по фенотипу масти.
Ключевые слова: лабораторные мини-свиньи; инбридинг; генетическое разнообразие; рецессивные мутации; отбор; линии; семейства; сельское хозяйство.

Background

Despite the practicality of laboratory use in comparison with primates and several morphophysiological advantages over other laboratory animals (Tikhonov, 2010; Shatokhin et al., 2019), mini-pigs are still not the most popular biological model, second not only to rodents but also to dogs, cats and monkeys (Heining, Ruysschaert, 2016). However, according to various estimates, there are from 21 to 45 breeding groups of mini-pigs globally (Smith, Swindle, 2006; Köhn, 2011), of which two are bred in Russia (Stankova et al., 2017; Shatokhin et al., 2019). Although, despite the importance of understanding the breeding of any animal species, regardless of their use, the problems of breeding laboratory mini-pigs are shown in a fairly small number of scientific papers. The apparently insufficient attention to the breeding and selection of mini-pigs resulted in some problems and the lack of a unified concept for their solution. The main ones are:

1) the lack of centralized accounting of the number of laboratory mini-pigs and the registration system of specialized herds as breeding achievements;
2) the lack of generally accepted standards for the selection of animals for reproduction. This also includes the lack of regulatory documents for the evaluation of breeding animals;
3) maximizing herds' genetic diversity under conditions of gene pool depletion vectors (gene drift, bottleneck effect), optimization of monitoring and selection management methods;
4) minimizing the accumulation of fitness-reducing mutations;
5) the creation of herds of laboratory mini-pigs, staffed exclusively from animals of white coat color;

The purpose of this paper is to describe the listed problems and suggest some ways to solve them.

The global genetic fund of laboratory mini-pigs

To date, it is difficult to estimate the number of the world's population of laboratory mini-pigs and the exact number of their breeds, herds, and breeding groups. The main difficulty lies in the absence of a single body for recording laboratory mini-pigs as objects of breeding. For example, according to Russian legislation, the registration of laboratory mini-pigs is difficult due to their formal non-compliance with the criteria for evaluating breeds and breed groups of pigs as breeding achievements, particularly according to the uniformity of the breeding stock (Method of testing for distinctness..., 2007). No special standards have been developed for them. Registration is possible on the website of the American Mini-pig Association (https://americanminipigassociation.com). However, out of 14 registered breeds, only four breeding groups were reliably used as laboratory animals.

The only available accounting tool is scientific publications, but the number of breeding groups of laboratory mini-pigs varies from 21 to 45 (Smith, Swindle, 2006; Tikhonov, 2010). One of the reasons for the discrepancy in the calculation results is the presence of more than one name for the same breed formation. Our own count of laboratory mini-pigs indicated 31 breeding groups in the world (Table 1). Both breed formations are represented by several colonies (Hormel, Hanford, Göttingen, NIH, Yucatan) and breeding groups consisting of a single herd (NIBS; mini-pigs of the Institute of Cytology and Genetics SB RAS, ICG SB RAS; Svetlogorsk). Representatives of the species *Sus scrofa* L. were taken into account with a live weight of no more than 150 kg and an indication of systematic use as a model object over the past 10 years.

Selection principles of breeding animals

In the breeding of laboratory mini-pigs, there are two main selection vectors: for small size and low live weight and suitability for laboratory use. However, in the breeding of mini-pigs, there are no uniform specially developed standards for evaluating animals by live weight at an early age, exterior, coat color and a set of characteristics necessary for use in the most common types of biomedical experiments (Helke et al., 2016). Simultaneously, almost every herd has a systematic approach to breeding with its own specific methods (Itoh et al., 2016; Nikitin et al., 2018). Animals are often evaluated at an early age, for example, 140–154 days (Miniature Swine Book of Normals, 2019; Simon, 2019). Some private farms practice selection of the smallest animals from each nest[1], which in the defunct selection group Minisibs had such consequences as lowering the safety of piglets, sexual activity of boars and destroying the complex of maternal qualities of sows (Nikitin et al., 2014).

The only general principle is selecting the most robust, healthy and proportionally developed animals with a live weight of adults from 50 to 80 kg (Nunoya et al., 2007; Tikhonov2010; Miniature Swine Book of Normals, 2019). Vietnamese mini-pigs' exterior traits such as a weak back or early obesity are not welcomed by Russian, European and American specialists. Russian mini-pigs and several foreign breeding groups meet the accepted standards, but there are deviations, both in larger and smaller directions (Table 2). Recently, the breeding of herds of tiny pigs weighing 30–50 kg, for example, German mini-pigs Aachen, American Panepinto and Korean Micro-Pig®, is gaining popularity (see Table 2).

The preservation of genetic diversity

The problem of preserving genetic diversity in populations is one of the most discussed issues in animal genetics (Peripolli et al., 2017; Mable, 2019) and, for several reasons, is particularly relevant for laboratory mini-pigs. The first reason is the low population of herds. The risk of depletion of the gene pool due to stochastic processes is significantly higher than in large structured subpopulations communities (Mariani et al., 2020). The second reason is the existence of several breeding groups of laboratory mini-pigs in the singular, which deprives them of such a powerful resource for controlling heterozygosity as the periodic exchange of the gene pool between different herds (Mariani et al., 2020). The third reason is that creating new herds of laboratory mini-pigs from a small number of progenitors (see Table 1) creates a risk of depleting the gene

[1] Erasmus D. Pigs as pets: Breeding teacup pigs. Farmer's Weekly. 2013. https://www.farmersweekly.co.za/animals/pigs-as-pets-breeding-teacup-pigs/

Table 1. List of the breeding groups of laboratory mini-pigs

No.	Name	Origin	Time and place of origin	Literature source
1	Aachen	Mini-Lewe × Vietnamese potbelly pig × Schwäbisch Hällisch Landpig × Hormel	Rheinisch Westfälische Technische Hochschule (RWTH Aachen University), Germany	Pawlowsky et al., 2017
2	Bama	Native small breed	China	Zhang et al., 2016
3	Banna	Native small breed	China	Xin et al., 2013
4	Br1	Hormel	1964, Universidade de Sao Paolo, Brasil	Scheffer et al., 2013
5	Clawn Miniature Swine	Göttingen × Ohminy × Large white × Landrace	1978, CLAWN Institute, Kagoshima University, Japan	Köhn, 2011
6	Diannan	Native small breed	Yunnan Agricultural University, China	Cheng et al., 2016
7	Fuji Micra Inc.	Other mini-pigs (not specified)	2009, Miyahara, Fujinomiya, Shizuoka, Japan	Maeda et al., 2016
8	Göttingen	Vietnamese potbelly pigs (gray) × Hormel × Vietnamese native spotted × Landrace	1960–1964, Göttingen University, Germany	Simon, 2019
9	Guizhou	Native small breed	Laboratory Animal Center of Chongqing Medical University, Chongqing, China	Xia et al., 2014
10	Hanford	Palose × Pitman-Moore	1958, Hanford laboratory, Washington, United States	Köhn, 2011
11	Hormel (Sinclair, Minnesota)	Piney Wood × Ras-n-Lansa × Catalina × Guam	1949, Hormel Institute, Minnesota University, United States	Köhn, 2011; Miniature Swine Book of Normals, 2019
12	KCG	Kogata Chinese × Clawn × Göttingen	1991, National Livestock Breeding Center, Ibaraki Station, Independent Administration Institution of Japan, Japan	Kobayashi et al., 2012
13	Lanyu	Native small breed	Taitung Animal Propagation Station, Livestock Research Institute, Taiwan, China	Chu, 2010; Chien et al., 2017
14	Lee-Sung	Lanyu × Landrace	1975, Department of Animal Science and Technology, National Taiwan University, Taiwan	Ju et al., 2019
15	MeLiM	Hormel × Landrace, Large White × Cornwall × Vietnamese pigs × Göttingen	1967–2000, Institute of Animal Physiology and Genetics of the Academy of Sciences of the Czech Republic, Libechov, Czech Republic	Horak et al., 2019
16	Mexican hairless mini	Feral hog from Mexico	–	Kobayashi et al., 2012
17	Micro-Pig®	Native small breed × Yucatan × Vietnamese potbellied pig × Pygmy pig × Meishan	Medi Kinetics Co., Ltd., Pyeongtaek, Republic of Korea	Jo et al., 2017
18	Micro-Yucatan	Yucatan	1982, Charles River Laboratories, United States	Köhn, 2011
19	Mini-Lewe (Berlin mini-pigs)	Vietnamese Pot Belly Pigs × Saddle Back Pigs × Landrace	1966, Außenstelle Lehnitz der Humboldt Universität, Germany	Schachler et al., 2020
20	Mini-Pig®	Native small breed	Cronex Co., Ltd., Hwaseong, Republic of Korea	Jo et al., 2017
21	Mini-pigs of ICG SB RAS	Large White × Svetlogorsk × Landrace × Vietnamese native breed	1990–1992, Institute of Cytology and Genetics SB RAS (ICG SB RAS), Novosibirsk region, Russia	Nikitin et al., 2014
22	Munich miniature (Troll)	Hanford × Columbian Miniature Swine	1993, Munich, Germany	Köhn, 2011; Bourneuf, 2017
23	NIBS	Pitman-Moore × Taiwanese small-ear pigs × Göttingen	1993, Nippon Institute for Biological Science, Tokyo, Japan	Yoshimatsu et al., 2016

Table 1 (end)

No.	Name	Origin	Time and place of origin	Literature source
24	NIH	Feral hog from Indiana state × Hanford	1972, National Institute of Health, Bethesda, Maryland, United States	Sachs et al., 1976; Nicholls et al., 2012
25	Ossabaw	Feral hog from Ossabaw island	2001–2002, Indiana University, United States	McKenney-Drake et al., 2016
26	Panepinto	Yucatan × Vietnamese	1990, Colorado State University, United States	Köhn, 2011
27	Pitman-Moore	Feral hog from Florida state	1969, Pitman-Moore Pharmaceutical Company, Indianapolis, United States	Val-Laillet et al., 2013
28	Svetlogorsk	Minisib × Göttingen	1974, Scientific Center of Biomedical Technologies, Moscow region, Russia	Stankova et al., 2017
29	Westran	Feral hog from Kangaroo island	1976, Commonwealth Scientific and Research Organization, Australia	Köhn, 2011
30	Wuzhishan	Native small breed	1990, Wuzhishan pig breeding farm of Academy of Agricultural Sciences, Hainan province, China	Song et al., 2014
31	Yucatan	Feral hog from Mexico	1960, Colorado StateUniversity, United States	Miniature Swine Book of Normals, 2019

Table 2. Live weight of adult laboratory mini-pigs from different breeding groups

Breeding group	Live weight, kg	Literature source
Aachen	45–50	Pawlowsky et al., 2017
Bama	~50	Zhang et al., 2016
Br1	30–70	Mariano, 2003
Clawn	~40	Köhn, 2011
Göttingen	25–50	Simon, 2019
Hanford	80–95	Köhn, 2011
Hormel	55–70	Miniature Swine Book of Normals, 2019
Micro-Pig®	30–35	Jo et al., 2017
Micro-Yucatan	55–70	Köhn, 2011
Mini-Lewe	45–60	Schachler et al., 2020
Mini-Pig®	57–64	Jo et al., 2017
Mini-pigs of ICG SB RAS	60–70	Shatokhin et al., 2019
Munich miniature (Troll)	60–100	Köhn, 2011; Bourneuf, 2017
Ossabaw	72–116	McKenney-Drake et al., 2016
Panepinto	25–30	Köhn, 2011
Pitman-Moore	40–69	Tikhonov, 2010
Svetlogorsk	35–50	Stankova et al., 2017
Westran	80–93	Köhn, 2011
Wuzhishan	35–40	Song et al., 2014
Yucatan	70–80	Köhn, 2011

pool due to the bottleneck effect (Ji et al., 2011). Interestingly, according to various estimates, the genetic diversity of laboratory pigs can be both greater and lower compared with similar parameters of pigs of factory breeds and wild boar (Nikitin et al., 2010; Heckel et al., 2015).

Several publications mentioned the existence of natural "contr inbred" mechanisms in natural populations (Charlesworth, Willis, 2009; Cheptou, Donohue, 2011; Mable, 2019), which is indirectly confirmed by the existence of the short populations of feral pigs with no signs of inbreeding depression on small islands throughout the centuries (Köhn, 2011; McKenney-Drake et al., 2016). In the conditions of farms for breeding of laboratory mini-pigs, the formation of the composition of the reproductive group and the choice of parent pairs during the breeding campaign is carried out by the breeder. Therefore, the question about the full functioning of such mechanisms arises. Thus, there is a need to analyze the methods available to humans to control herds' heterozygosity of laboratory mini-pigs. The first method is monitoring genetic diversity using molecular genetic methods, which is used to select some of the mini-pigs (Chang et al., 2009). A limiting factor in further implementing this method is the lack of data on its economic feasibility in routine use.

The second way to control heterozygosity is to use breeding techniques and methods, for example, to minimize inbred crosses (Simianer, Köhn, 2010). In the breeding of mini-pigs of the ICG SB RAS, the conservation of the maximum possible number of color phenotypes and inbreeding mainly on the progenitors is used to preserve genetic diversity (Nikitin et al., 2018). Given that the mammalian suit is controlled by 120 to 350 genes (Cieslak et al., 2011; Chandramohan et al., 2013), the number of possible genotypes can be in the thousands. Another breeding method for maximizing genetic diversity is dividing an array of animals into subpopulations with a limited gene flow between them (Mariani et al.,

Table 3. Conditional scheme of the selection of boars and sows during one cycle

Familia	Line				Minimum number of sows in each cycle
	B_1	B_2	B_3	B_4	
S_1	I	II	III	IV	5
S_2	II	I	IV	III	5
S_3	III	IV	I	II	5
S_4	IV	III	II	I	5
Minimum number of hogs in every cycle	2	2	2	2	–

Note. The cells at the intersection of lines (columns) and families (rows) indicate the generations of descendants.

2020). However, due to the low number of rock formations, partial genealogical separation of lines, with rare exceptions (Stankova et al., 2017), is practically impossible to implement. Instead, a cyclical selection system is practiced (Chu, 2010; Schachler et al., 2020) based on periodically repeated crosses of lines and families (Table 3). According to the calculations, to avoid close inbreeding, the minimum number of the reproductive group should be at least 28 individuals, of which boars should be represented by at least four lines and sows – by four families. Each line should include at least one main and one checked boar, and the family should consist of at least 5 main and checked sows.

Accumulation of genetic cargo

In the 1970s, it was reported that in populations of less than 2,000 individuals, the probability of accumulation of fitness-reducing mutations is quite high (Nei, Roychoudhury, 1973). Even earlier, it was established that recessive semi-lethal mutations could persist in a population for up to 99 generations even with targeted culling of homozygotes (Dubinin, Glembotsky, 1967), which is generally not refuted by later mathematical modelling (Johnsson et al., 2019). It is considered that the elimination of harmful recessive mutations is a difficult task for the breeder, even if he uses modern genotyping methods (Derks et al., 2017). Given that the reproductive number of individual herds of laboratory mini-pigs does not exceed 30–40 individuals reducing sustainability, semi-lethal and lethal recessive mutations pose a danger in breeding these animals. At the same time, in the entire history of breeding laboratory mini-pigs, only in the extinct breeding group Minisibs a decrease in the viability of young animals and the reproductive qualities of adults was described, the alleged cause of which was the accumulation of recessive mutations due to unilateral selection (Nikitin et al., 2014). Thus, laboratory mini-pigs' breeding system should include measures to purify the herd from harmful mutations, leading to strict selection in the reproductive group (Nikitin et al., 2018, 2020). Another method of cleaning herds from unwanted mutations is to assess the progeny in the inbred cross. This method was proposed for various farm animals' species in the 1950s and 1970s (Robertson, Rendel, 1950; Serebrovsky, 1970). However, despite its simplicity, the method has a serious drawback – it is the duration of the assessment and, accordingly, the cost of feeding and maintaining the tested boar and its descendants.

However, there are cases where breeders have benefited from the emergence of viability-reducing mutations in the herd in the form of creating model objects to optimize specific medical methods or treat strictly defined pathologies. Examples are the creation of mini-pigs by MeLiM and NIH (Sachs et al., 1976; Horak et al., 2019). Thus, it can be argued that the very fact of the occurrence of mutations that reduce viability, of course, is a danger. But much more important is breeders' ability to prioritize the selection of animals for reproduction and to carry out measures to clear the herds of genetic cargo; and if necessary, to consolidate the carriers of mutations in the form of a new selection group that is of value as a model object.

The problem of white coat color in the breeding of laboratory mini-pigs

It is known that when creating the first breeding groups of laboratory mini-pigs, the task was to create white-colored animals (Pond, Houpt, 1978), which were planned to be used as a biological model for studying the effects of radioactive radiation on the skin. However, despite the "influx of blood" of factory breeds of white color, attempts to consolidate it in herds of laboratory mini-pigs, as a rule, did not succeed. The exceptions are the Mini-Lewe pigs (Schachler et al., 2020) and the Bintang line (Lanyu 400) in the Lanyu mini-pig breeding group (Chu, 2010), but most herds have polymorphism by suit type (Mariano, 2003; Tikhonov, 2010; https://americanminipigassociation.com). Thus, the question arises about the factors that prevent the breeding of herds fully equipped with white individuals. It can be assumed that this is due to the dominant control of the most common type of white coat color (Pielberg et al., 2002), which is why there is a regular cleavage of pigmented piglets. Another explanation is that white piglets are born smaller and, therefore, less viable than colored animals (Nikitin et al., 2019). Despite this, the white coat color was successfully consolidated in a factory breeds series (Porter et al., 2016). It should be noted that the factory breeds of white-colored pigs were obtained by the method of more than 70 years of selection of white individuals in each generation with a preference for those animals in whose offspring there was no splitting according to the color phenotype (Porter et al., 2016). And this, in turn, is comparable to the duration of the oldest breeding groups of laboratory mini-pigs (Tikhonov, 2010). Thus, it can be assumed that the breeders of most breeding groups of mini-pigs simply did not have enough time to consolidate the white suit.

Molecular genetic typing of white animals would significantly speed up the process of fixing the white suit. It is known that the dominant white color of pigs is controlled by allele *I* of the *KIT* gene (Pielberg et al., 2002; Wu et al., 2019). Thus, the first step to create a breeding group complete with all-white animals should be to cross white sows with white

boars. All-white offspring from such crosses will need to be genotyped according to the *KIT* gene with the setting of homozygotes (*I/I*) for rearing. The method of determining the *KIT* gene's alleles using real-time PCR is described in detail in the literature (Pielberg et al., 2002).

Another way is to consolidate the recessive white suit's phenotype, as demonstrated by the Lanyu 400 line (Chi, 2010) and the Chinese Rongchang breed (Lai et al., 2007). However, a rather serious restriction on using this method may be the low frequency of cleavage of recessive white color individuals, which in the herd of mini-pigs of the Institute of Cytology and Genetics SB RAS, according to zootechnical accounting, is about 1 %.

Conclusion

Over the past 10 years, facts have been discovered confirming the existence of 31 breeding groups of mini-pigs. Despite the lack of uniform selection standards in breeding laboratory mini-pigs, they adhere to such general criteria as a live weight of 50–80 kg, normal viability, and the strength of the animals' constitution and exterior. Maintaining genetic diversity in herds of laboratory mini-pigs is possible both with the use of molecular genetic monitoring and purely selective methods. The minimization of the negative effect of genetic cargo accumulation in the herds of mini-pigs should be implemented mainly through a strict selection for fitness in the reproductive group. If necessary, due to the need for a specific type of biomedical experiments, it is possible to fix external and physiological characteristics in the herd, controlled by recessive mutations that reduce viability. Consolidation of white individuals is possible, which is proved by the examples of the Bintang line and the Mini-Lewe breeding group.

References

Bourneuf E. The MeLiM Minipig: An Original Spontaneous Model to Explore Cutaneous Melanoma Genetic Basis. *Front. Genet.* 2017; 8:146. DOI 10.3389/fgene.2017.00146.

Chandramohan B., Renieri C., La Manna V., La Terza A. The alpaca agouti gene: genomic locus, transcripts and causative mutations of eumelanic and pheomelanic coat color. *Gene*. 2013;521(2):303-310. DOI 10.1016/j.gene.2013.03.060.

Chang W.H., Chu H.P., Jiang Y.N., Li S.H., Wang Y., Chen C.H., Chen K.J., Lin C.Y., Ju Y.T. Genetic variation and phylogenetics of Lanyu and exotic pig breeds in Taiwan analyzed by nineteen microsatellite markers. *J. Anim. Sci*. 2009;87(1):1-8. DOI 10.2527/jas.2007-0562.

Charlesworth D., Willis J.H. The genetics of inbreeding depression. *Nat. Rev. Genet*. 2009;10:783-796. DOI 10.1038/nrg2664.

Cheng W., Zhao H., Yu H., Xin J., Wang J., Zeng L., Yuan Z., Qing Y., Li H., Jia B., Yang C., Shen Y., Zhao L., Pan W., Zhao H.Y., Wang W., Wei H.J. Efficient generation of GGTA1-null Diannan miniature pigs using TALENs combined with somatic cell nuclear transfer. *Reprod. Biol. Endocrinol*. 2016;14(1):77. DOI 10.1186/s12958-016-0212-7.

Cheptou P.O., Donohue K. Environment-dependent inbreeding depression: its ecological and evolutionary significance. *New Phytol*. 2011;189(2):395-407. DOI 10.1111/j.1469-8137.2010.03541.x.

Chien S., Hsue S., Lin C., Kuo T., Wang D.J., Yang j.C., Lee S. Influence of Thread Design on Dental Implant Osseointegration Assayed Using the Lan-Yu Mini-Pig Model. *J. Med. Biol. Eng*. 2017; 37:627-638. DOI 10.1007/s40846-017-0240-6.

Chu H.-P. Selection and Utilization of Minipigs for Biomedical Research. Taitung Animal Propagation Station, Livestock Research Institute, 2010;50. Available at: https://www.angrin.tlri.gov.tw/meeting/2010training/12.Mr.H.P.Chu.pdf.

Cieslak M., Reissmann M., Hofreiter M., Ludwig A. Colours of domestication. *Biol. Rev. Camb. Philos. Soc*. 2011;86(4):885-899. DOI 10.1111/j.1469-185X.2011.00177.x.

Derks M.F.L., Megens H.J., Bosse M., Lopes M.S., Harlizius B., Groenen M.A.M. A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics*. 2017; 18(1):858. DOI 10.1186/s12864-017-4278-1.

Dubinin N.P., Glembotskiy Ya.L. Population Genetics and Selection. Moscow: Nauka Publ., 1967. (in Russian)

Heckel T., Schmucki R., Berrera M., Ringshandl S., Badi L., Steiner G., Ravon M., Küng E., Kuhn B., Nicole A Kratochwil N.A, Schmitt G., Kiialainen A., Nowaczyk C., Daff H., Khan A.P., Lekolool I., Pelle R., Okoth E., Bishop R., Daubenberger C., Ebeling M., Certa U. Functional analysis and transcriptional output of the Göttingenminipig genome. *BMC Genomics*. 2015;16:932. DOI 10.1186/s12864-015-2119-7.

Heining P., Ruysschaert T. The use of minipig in drug discovery and development: pros and cons of minipig selection and strategies to use as a preferred nonrodent species. *Toxicol. Pathol.* 2016;44(3): 467-473. DOI 10.1177/0192623315610823.

Helke K.L., Nelson K.N., Sargeant A.M., Jacob B., McKeag S., Haruna J., Vemireddi V., Greeley M., Brocksmith D., Navratil N., Stricker-Krongrad A., Hollinger C. Background Pathological Changes in Minipigs: A Comparison of the Incidence and Nature among Different Breeds and Populations of Minipigs. *Toxicol. Pathol*. 2016;44(3):325-337. DOI 10.1177/0192623315611762.

Horak V., Palanova A., Cizkova J., Miltrova V., Vodicka P., Kupcova Skalnikova H. Melanoma-Bearing LibechovMinipig (MeLiM): The Unique Swine Model of Hereditary Metastatic Melanoma. *Genes* (*Basel*). 2019;10(11):915. DOI 10.3390/genes10110915.

Itoh T., Kabawe M., Nagase T., Endo K., Miyoshi M., Miyahara M. Body surface area measurement in laboratory miniature pigs using a computer tomography scanner. *J. Toxicol. Sci.* 2016;41(5):637-644. DOI 10.2131/jts.41.637.

Ji Y.Q., Wu D.D., Wu G.S., Wang G.D., Zhang Y.P. Multi-locus analysis reveals a different pattern of genetic diversity for mitochondrial and nuclear DNA between wild and domestic pigs in East Asia. *PLoS One*. 2011;6(10):e26416. DOI 10.1371/journal.pone.0026416.

Jo H., Lee S.C., Kim D.C., Kim B.G. Estimation of Growth Model Parameters for Mini Pig Breeds. *The Faseb Journal.* 2017;31(S1): ib299. DOI 10.1096/fasebj.31.1_supplement.lb299.

Johnsson M., Gaynor R.C., Jenko J., Gorjanc G., de Koning D.J., Hickey J.M. Removal of alleles by genome editing (RAGE) against deleterious load. *Genet. Sel. Evol.* 2019;51(1):14. DOI 10.1186/s12711-019-0456-8.

Ju Y.T., Pan Y.T., Tu C.F., Hsiao J., Lin Y.H., Yu P.J., Yu P.H., Chi C.H., Liu I.L. Growth and Behavior of Congenitally Anophthalmic Lee-Sung Pigs. *Comp. Med.* 2019;69(3):212-220. DOI 10.30802/AALAS-CM-18-000095.

Kobayashi E., Hishikawa S., Teratani T., Lefor A.T. The pig as a model for translational research: overview of porcine animal models at Jichi Medical University. *Transplant. Res.* 2012;1:8. DOI 10.1186/2047-1440-1-8.

Köhn F. History and Development of Miniature, Micro- and Minipigs. The minipig in biomedical research. P.A. Mcanulty, A.D. Dayan,

N.C. Ganderup, K.L. Hastings (Eds.). Boca Raton, FL: CRC Press, 2011;3-16. DOI 10.1201/b11356-3.

Lai F., Ren J., Ai H., Ding N., Ma J., Zeng D., Chen C., Guo Y., Huang L. Chinese white Rongchang pig does not have the dominant white allele of *KIT* but has the dominant black allele of *MC1R*. *J. Hered.* 2007;98(1):84-87. DOI 10.1093/jhered/esl053.

Mable B.K. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation Genetics*. 2019;20:89-100. DOI 10.1007/s10592-018-1129-9.

Maeda M., Takashina S., Takasu M., Mori T., Goto N., Matsubara T., Almunia J., Imaeda N., Ando A., Kitagawa H. Magnetic Resonance Imaging of Ovarian Activity in Microminipigs Showing Normal Estrous Cycles. *In vivo*. 2016;30:35-40. Available at: http://iv.iiarjournals.org/content/30/1/35.full.

Mariani E., Summer A., Ablondi M., Sabbioni A. Genetic Variability and Management in Nero di Parma Swine Breed to Preserve Local Diversity. *Animals* (*Basel*). 2020;10(3):538. DOI 10.3390/ani10030538.

Mariano M. Minisuíno (minipig) na pesquis abiomédica experimental. O Minipig br1. *Acta Cir. Bras*. 2003;18(5):387-391. DOI 10.1590/S0102-86502003000500003. (in Portuguese)

McKenney-Drake M.L., Rodenbeck S.D., Owen M.K., Schultz K.A., Alloosh M., Tune J.D., Sturek M. Repeat cross-sectional data on the progression of the metabolic syndrome in Ossabaw miniature swine. *Data Brief*. 2016;7:1393-1395. DOI 10.1016/j.dib.2016.04.023.

Miniature Swine Book of Normals. Sinclair Bio Resources, 2019. Available at: https://sinclairresearch.com/bio-resources/.

Nei M., Roychoudhury A.K. Probability of Fixation of Nonfunctional Genes at Duplicate Loci. *Amer. Nat*. 1973;107(955):362-372. DOI 10.1086/282840.

Nicholls S.M., Mitchard L.K., Laycock G.M., Harley R., Murrell J.C., Dick A.D., Bailey M. A Model of Corneal Graft Rejection in Semi-Inbred NIH Miniature Swine: Significant T-Cell Infiltration of Clinically Accepted Allografts. Invest. Ophthalmol. *Vis. Sci*. 2012;53(6):3183-3192. DOI 10.1167/iovs.11-9106.

Nikitin S.V., Knyazev S.P., Shatokhin K.S. Miniature pigs of ICG as a model object for morphogenetic research. *Russ. J. Genet.: Appl. Res*. 2014;4(6):511-522. DOI 10.1134/S207905971406015X.

Nikitin S.V., Knyazev S.P., Shatokhin K.S., Zaporozhets V.I., Bashur D.S., Ermolaev V.I. Prenatal growth of live weight and coat color of the minipigs of ICG SB RAS. *Sel'skokhozyaystvennye Tekhnologii = Agricultural Technologies*. 2019;1(3):21-34. DOI 10.35599/agritech/01.03.04. (in Russian)

Nikitin S.V., Knyazev S.P., Shatokhin K.S., Zaporozhets V.I., Ermolaev V.I. Breeding and selection of mini-pigs in the ICG SB RAS. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2018;22(8):922-930. DOI 10.18699/VJ18.434. (in Russian)

Nikitin S.V., Shatokhin K.S., Zaporozhets V.I., Kochnev N.N., Bashur D.S., Khodakova A.V., Ermolaev V.I. 'Genetic load' and changes in the chronology of early mortality in mini-pigs of ICG SB RAS. *Agronomy Research*. 2020;18(3):2156-2165. DOI 10.15159/AR.20. 171.

Nikitin S.V., Yudin N.S., Knyazev S.P., Aitnazarov R.B., Bekenev V.A., Deeva V.S., Goncharenko G.M., Kobzev V.F., Savina M.A., Ermolaev V.I. Differentiation of wild boar and domestic pig population based on the frequency of chromosomes carring endogenous retroviruses. *Natural Science*. 2010;2(6):527-534. DOI 10.4236/ns.2010. 26066.

Nunoya T., Shibua K., Saitoh T., Yazawa H., Nakamura K., Baba Y., Hirai T. Use of miniature pig for biomedical research, with reference to toxicologic studies. *J. Toxicol. Pathol.* 2007;20:125-132. DOI 10.1293/TOX.20.125.

Pawlowsky K., Ernst L., Steitz J., Stopinski T., Kögel B., Henger A., Kluge R., Tolba R. The Aachen Minipig: Phenotype, Genotype, Hematological and Biochemical Characterization, and Comparison to the GöttingenMinipig. *Eur. Surg. Res*. 2017;58(5-6):193-203. DOI 10.1159/000471483.

Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R., Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.* 2017;48(3):255-271. DOI 10.1111/age.12526.

Pielberg G., Olsson C., Syvanen A.C., Andersson L. Unexpectedly high allelic diversity at the *KIT* locus causing dominant white color in the domestic pig. *Genetics*. 2002;160:305-311.

Pond W.G., Houpt K.A. The Biology of the Pig. Ithaca, New York: Cornell University Press, 1978.

Porter V., Alderson L., Hall S.J.G., Sponenberg D.P. Mason's world encyclopedia of livestock breeds and breeding. Boston, MA: CABI, 2016.

Robertson A., Rendel J.M. The use of progeny testing with artifcial insemination in dairy cattle. *J. Genet.* 1950;50:1-31.

Protocol of testing for distinctness, uniformity, and stability. Pigs (*Sus scrofa* L.). All-Russia Research Institute of Animal Breeding, 2007. (in Russian)

Sachs D.H., Leight G., Cone J., Schwarz S., Stuart L., Rosenberg S. Transplantation in miniature swine. I. Fixation of the major histocompatibility complex. *Transplantation*. 1976;22(6):559-567. DOI 10.1097/00007890-197612000-00004.

Schachler K., Minx J-O., Sürie C., Distl O., Metzger J. Genetic characterisation of Mini-LEWE as resource population for experimental research. *Berl. Munch. Tierarztl. Wochenschrt*. 2020;133. DOI 10.2376/1439-0299-2020-15.

Scheffer J.P., Oliveira A.L.A., Markoski M., Antunes F., Atallah F.A., Monteiro G.A.S., Carvalho C.B., Kalil R.A.K. Indução da miocardiopatia isquêmica em modelo experimental mini porco. *Rev. Bras. Med. Vet*. 2013;35(1):45-48. (in Portuguese)

Serebrovskiy A.S. Genetic analisys. Moscow: Nauka Publ., 1970. (in Russian)

Shatokhin K.S., Nikitin S.V., Knyazev S.P., Goncharenko G.M., Ermolaev V.I., ZaporozhetsV.I. Livestock, physiology, and genetics of the mini-pigs of the Institute of Cytology and Genetics. Novosibirsk: Institute of Cytology and Genetics, 2019. (in Russian)

Simianer H., Köhn F. Genetic management of the Göttingen Minipig. *J. Pharmacol. Toxicol. Methods*. 2010;62:221-226. DOI 10.1016/j.vascn.2010.05.004.

Simon C. Fachinformation aus dem Ausschuss für Ernährung der Versuchstiere. Fütterungskonzepte und -methoden in der Versuchstierhaltung und im Tierversuch – MINIPIG. Basel Reinhart Kluge, Potsdam-Rehbrücke. 2019;11. (in Germany). Available at: https://docplayer.org/170135602-Fachinformation-aus-dem-ausschuss-fuer-ernaehrung-der-versuchstiere.html.

Smith A.C., Swindle M.M. Preparation of swine for the laboratory. *ILAR J.* 2006;47(4):358-63. DOI 10.1093/ilar.47.4.358.

Song T., Wu T., Wei F., Li A., Wang F., Xie Y., Liu D., Fan Z., Wang X., Cheng S., Zhang C., He J., Wang S. Construction of a cDNA library for miniature pig mandibular deciduous molars. *BMC Dev. Biol*. 2014;14:16. DOI 10.1186/1471-213X-14-16.

Stankova N.V., Savina M.A., Kapanadze G.D. The formation of new lines of Svetlogorsk minipigs. *Biomeditsina = Biomedicine*. 2017;3:95-101. (in Russian)

Tikhonov V.N. Laboratory Minipigs. Genetics and Biomedical Use. Novosibirsk: Publishing House SB RAS, 2010. (in Russian)

Val-Laillet D., Céline Tallet C., CaroleGuérin C., Meunier-Salaün M.-C. Behavioural reactivity, social and cognitive abilities of Vietnamese and Pitman–Moore weaned piglets. *Appl. Anim. Behav. Sci*. 2013;148:108-119. DOI 10.1016/j.applanim.2013.06.003.

Wu Z., Deng Z., Huang M., Hou Y., Zhang H., Chen H., Ren J. Whole-Genome Resequencing Identifies *KIT* New Alleles That Affect Coat Color Phenotypes in Pigs. *Front. Genet*. 2019;10:218. DOI 10.3389/fgene.2019.00218.

Xia J., Liu Z., Xin,L. Wang Z., Qian L., Wu S., Y. Shulin, Li K. Molecular characterization of swine leukocyte antigen class I genes by sequence-based and PCR-SSP method in Guizhouminipigs. *Mol. Biol. Rep*. 2014;41:7775-7782. DOI 10.1007/s11033-014-3670-z.

Xin J., Yang H., Fan N., Zhao B., Ouyang Z., Liu Z., Zhao Y., Li X., Song J., Yang Y., Zou Q., Yan Q., Zeng Y., Lai L. Highly efficient generation of GGTA1 biallelic knockout inbred mini-pigs with TALENs. *PLoS One*. 2013;8(12):e84250. DOI 10.1371/journal.pone.0084250.

Yoshimatsu H., Konno Y., Ishii K., Satsukawa M., Yamashita S. Usefulness of minipigs for predicting human pharmacokinetics: Prediction of distribution volume and plasma clearance. *Drug. Metab. Pharmacokinet*. 2016;31(1):73-81. DOI 10.1016/j.dmpk.2015.11.001.

Zhang S., Guo K., Kong D., Liu Y., Shang H., Wei H. Semen quality and fertility of Bama miniature pig (*Sus scrofa domestica*). *Indian J. Anim. Res*. 2016;(50):434-437. DOI 10.18805/ijar.7494.

ORCID ID
K.S. Shatokhin orcid.org/0000-0002-0885-2772

Acknowledgements. The study was carried out with the financial support of the Novosibirsk State Agrarian University.
Conflict of interest. The author declares no conflict of interest.
Received October 6, 2020. Revised February 2, 2021. Accepted February 10, 2021.

Вавиловский журнал генетики и селекции. 2021;25(3):292-300
DOI 10.18699/VJ21.033

Недостаток белка GAGA у мутантов *Trl* вызывает массовую клеточную гибель в спермато- и оогенезе дрозофилы

Н.В. Дорогова[1] ✉, А.Е. Зубкова[1, 2], Е.В. Федорова[1], Е.У. Болоболова[1], Е.М. Баричева[1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ dorogova@bionet.nsc.ru

Аннотация. Белок дрозофилы GAGA (GAF) является фактором эпигенетической регуляции транскрипции большой группы генов с широким разнообразием клеточных функций. GAF кодируется геном *Trithorax-like* (*Trl*), который экспрессируется в различных органах и тканях на всех стадиях онтогенеза дрозофилы. Мутации этого гена вызывают множественные нарушения развития. В предыдущих работах мы показали, что этот белок необходим для развития половой системы как самцов, так и самок дрозофилы. Снижение экспрессии гена *Trl* приводило ко множественным нарушениям спермато- и оогенеза. Одно из значительных нарушений было связано с массовой деградацией и потерей клеток зародышевого пути, что позволило предположить, что этот белок вовлечен в регуляцию клеточной гибели. В представленной работе мы провели более детальное цитологическое исследование, чтобы определить, какой тип гибели клеток зародышевого пути характерен для *Trl*-мутантов, и происходят ли нарушения или изменения этого процесса по сравнению с нормой. Полученные результаты показали, что недостаток белка GAF вызывает массовую гибель клеток зародышевого пути как у самок, так и самцов дрозофилы, но проявляется эта гибель в зависимости от пола по-разному. У самок, мутантных по гену *Trl*, фенотипически этот процесс не отличается от нормы и в гибнущих яйцевых камерах выявлены признаки апоптоза и аутофагии клеток зародышевого пути. У самцов, мутантных по гену *Trl*, в отличие от самок, не обнаружены признаки апоптоза. У самцов мутации *Trl* индуцируют массовую гибель клеток через аутофагию, что не характерно для сперматогенеза дрозофилы и не описано ранее ни в норме, ни у мутаций по другим генам. Таким образом, недостаток GAF у мутантов *Trl* приводит к усилению апоптотической и аутофагической гибели клеток зародышевого пути. Эктопическая клеточная гибель и атрофия зародышевой линии, вероятно, связаны с нарушением экспрессии генов-мишеней GAGA-фактора, среди которых есть гены, регулирующие как апоптоз, так и аутофагию.
Ключевые слова: дрозофила; GAGA-фактор; клетки зародышевого пути; апоптоз; аутофагия; сперматогенез; оогенез.

Для цитирования: Дорогова Н.В., Зубкова А.Е., Федорова Е.В., Болоболова Е.У., Баричева Е.М. Недостаток белка GAGA у мутантов *Trl* вызывает массовую клеточную гибель в спермато- и оогенезе дрозофилы. *Вавиловский журнал генетики и селекции*. 2021;25(3):292-300. DOI 10.18699/VJ21.033

Lack of GAGA protein in *Trl* mutants causes massive cell death in *Drosophila* spermatogenesis and oogenesis

N.V. Dorogova[1] ✉, A.E. Zubkova[1, 2], E.V. Fedorova[1], E.U. Bolobolova[1], E.M. Baricheva[1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
✉ dorogova@bionet.nsc.ru

Abstract. *Drosophila* protein GAGA (GAF) is a factor of epigenetic transcription regulation of a large group of genes with a wide variety of cellular functions. GAF is encoded by the *Trithorax-like* (*Trl*) gene, which is important for the formation of various organs and tissues at all stages of ontogenesis. In our previous works, we showed that this protein is necessary for the development of the reproductive system, both in males and females of *Drosophila*. Decreased expression of the *Trl* gene led to multiple disorders of spermatogenesis and oogenesis. One of the significant disorders was associated with massive degradation and loss of cells in the germline. In this work, we carried out a more detailed cytological study to determine what type of germ cell death is characteristic of *Trl* mutants, and whether there are disturbances or changes in this process compared to the norm. The results obtained showed that the lack of GAF protein causes massive germ cell death in both females and males of *Drosophila*, but this death manifests itself in different ways, depending on the sex. In *Trl* females, this process does not differ

phenotypically from the norm. In the dying egg chambers, signs of apoptosis and autophagy were revealed, as well as morphological features that are characteristic of the wild type. In males, *Trl* mutations induce mass germ cell death through autophagy, which is not typical of *Drosophila* spermatogenesis, and has not been previously described, neither in the norm nor in other genes' mutations. Thus, GAF lack in *Trl* mutants leads to increased germ cell death through apoptosis and autophagy. Ectopic cell death and germ line atrophy are probably associated with impaired expression of the GAGA factor target genes, among which there are genes that regulate both apoptosis and autophagy.

Key words: *Drosophila*; GAGA factor; germ cells; apoptosis; autophagy; spermatogenesis; oogenesis.

For citation: Dorogova N.V., Zubkova A.E., Fedorova E.V., Bolobolova E.U., Baricheva E.M. Lack of GAGA protein in *Trl* mutants causes massive cell death in *Drosophila* spermatogenesis and oogenesis. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):292-300. DOI 10.18699/VJ21.033 (in Russian)

Введение

Белок дрозофилы, GAGA-фактор (GAGA-factor, GAF), кодируется геном *Trithorax-like* (*Trl*), который экспрессируется в различных органах и тканях на всех стадиях онтогенеза дрозофилы (Soeller et al., 1993; Baricheva et al., 1997; Karagodin et al., 2013). GAF имеет важную биологическую функцию, связанную как с позитивной, так и негативной регуляцией экспрессии большой группы генов, контролирующих основные этапы развития дрозофилы (Granok et al., 1995; van Steensel et al., 2001, 2003). GAF является эволюционно-консервативным белком, имеющим гомологию с белками многих эукариот, в том числе высших позвоночных (Matharu et al., 2010; Berger, Dubreucq, 2012). Гомологи GAF, как и сам белок, могут связываться с GA-последовательностями в регуляторных районах эволюционно-консервативных генов (Matharu et al., 2010). Результаты полногеномного анализа (профилирование хроматина с использованием Dam) показали, что мишенями GAF у дрозофилы могут быть около 250 генов, участвующих как минимум в 28 сигнальных путях (van Steensel et al., 2003). Анализ профилей связывания GAF, полученных в рамках проекта modENCODE (http://www.modencode.org/) (Roy et al., 2010), позволяет предполагать, что GAF может участвовать в регуляции более 3700 генов дрозофилы (неопубликованные данные).

Анализ *Trl*-мутантов продемонстрировал, что GAF необходим для эмбриогенеза, развития глаза и крыла дрозофилы (Bhat et al., 1996; Dos-Santos et al., 2008; Omelina et al., 2011; Bayarmagnai et al., 2012).

В предыдущих работах мы показали, что GAF принимает участие в развитии половой системы как самцов, так и самок дрозофилы, поскольку снижение экспрессии гена у мутантов *Trl* приводит к множественным нарушениям спермато- и оогенеза (Ogienko et al., 2006, 2008; Dorogova et al., 2014; Fedorova et al., 2019). Одно из значительных нарушений как оогенеза, так и сперматогенеза связано с массовой гибелью и потерей клеток зародышевого пути (КЗП) (Dorogova et al., 2014; Fedorova et al., 2019). Это позволило прийти к заключению, что GAF может регулировать активность генов, отвечающих за клеточную гибель. Однако для дальнейшего исследования роли GAF и поиска его генов-мишеней в клеточной гибели необходимо получить более полную информацию о проявлениях этого процесса у *Trl*-мутантов и соотнести данные с определенным типом регулируемой клеточной смерти согласно существующей в настоящее время классификации. Чтобы решить эту задачу, мы провели детальное цитологическое исследование характера гибели КЗП в яичниках и семенниках у *Trl*-мутантов дрозофилы.

В норме яичник дрозофилы состоит из 15–20 овариол, которые представляют собой цепочку прогрессивно развивающихся яйцевых камер (ЯК), состоящих из 16 КЗП, которые окружены монослоем соматических фолликулярных клеток. Одна из 16 клеток цисты становится яйцеклеткой, остальные – питающими клетками. По размеру и морфологии ЯК оогенез можно условно разделить на 14 стадий (King, 1957; Spradling, 1993; Ogienko et al., 2007). В норме онтогенетически запрограммированная смерть КЗП происходит на трех специфических стадиях: во вновь сформированных цистах (второй район гермария), во время среднего (стадии 7–9) и позднего (стадии 12, 13) оогенеза. При достаточном питании мух смерть клеток в гермарии и на стадиях 7–9 (называемых контрольными точками гибели клеток в оогенезе, англ. checkpoints) происходит в ответ на аномалии развития и резко увеличивается под влиянием различных стрессов (McCall, 2004; Jenkins et al., 2013*)*. Гибель питающих клеток в позднем оогенезе происходит как часть нормального развития каждого яйца (Jenkins et al., 2013; Peterson et al., 2015; Bolobolova et al., 2020).

Для оогенеза дрозофилы характерны два основных типа клеточной гибели: апоптоз и аутофагия (McCall, 2004; Barth et al., 2011; Jenkins et al., 2013; Bolobolova et al., 2020). Апоптоз является универсальным консервативным механизмом, при котором запускается программа саморазрушения клеток с участием протеолитических ферментов, относящихся к семейству каспаз (Kumar, 2007). В гибнущих ЯК проявляются характерные признаки апоптоза: конденсация хроматина, разрывы ДНК, фрагментация ядерного материала и цитоплазмы (Kihlmark et al., 2001; Greenwood, Gautier, 2005; Sarkissian et al., 2014). У дрозофилы инициация апоптоза связана с активностью эффекторной каспазы Dcp-1, и окраска антителами к этому белку является основным маркером каспаза-зависимого апоптоза (Sarkissian et al., 2014). Клеточная гибель через аутофагию сопровождается избыточным образованием аутофаго- и лизосом, что приводит к перевариванию всех клеточных органелл и закислению цитоплазмы. Вследствие этой особенности для выявления аутофагии используют лизотрекер (LysoTracker), ацидофильный краситель, который маркирует лизо- и аутофагосомы (DeVorkin, Gorski, 2014).

В семенниках КЗП расположены вдоль органа в соответствии со стадиями сперматогенеза. На апикальном конце – стволовые клетки, которые делятся с образованием гониальных клеток. Гониальные клетки вступают в сперматогенез, который включает митотическое и мейотическое деления, в результате чего образуются 64 синцитиальные сперматиды, которые затем дифференцируются в сперматозоиды (Fuller, 1993; Fabian, Brill, 2012). Клеточная гибель в сперматогенезе дрозофилы происходит крайне редко, и примеры ее исследования – единичны. Показано, что в процессе сперматогенеза дефектные сперматоциты элиминируются с помощью лизосомного механизма (без формирования аутофагосом) (Yacobi-Sharon et al., 2013), а также программируемого некроза, опосредованного белком p53 (Napoletano et al., 2017).

В данной работе мы показали, что снижение экспрессии GAF в яичниках у самок *Trl*-мутантов приводит к усилению гибели КЗП и деградации ЯК на 7–9-й стадиях оогенеза. При этом процесс клеточной гибели протекает так же, как в норме, и имеет признаки как аутофагии, так и апоптоза. В семенниках недостаток GAF индуцирует массовую гибель через аутофагию, что не характерно для сперматогенеза и не описано ранее ни в норме, ни у мутаций по другим генам.

Материал и методы

В экспериментах использовали следующие мутации *D. melanogaster*: мутация Trl^{R85} – нуль-аллель гена, любезно предоставлена Ф. Каршем (Женевский университет, Швейцария) (Farkas et al., 1994); гипоморфные мутации Trl^{362} и $Trl^{(ex)15}$, нарушающие 5′-область гена, получены в ИЦиГ СО РАН (Ogienko et al., 2007; Dorogova et al., 2014); *Oregon R* – дикий тип, из фонда лаборатории ИЦиГ СО РАН, использована в качестве контроля. Все скрещивания проводили на стандартной среде при температуре 25 °C.

Выделение, фиксацию и окраску гонад для электронной и флуоресцентной микроскопии производили согласно описанной ранее методике (Dorogova et al., 2014). В работе использованы первичные антитела: rabbit anti-Vasa (разведение 1:300; SC30210, Santa Cruz Biotechnology), rabbit anti-Dcp-1 (разведение 1:100; Asp216, Cell Signaling Technology). Вторичные антитела – anti-rabbit, конъюгированные с AlexaFluor-488 (1:500; A-11001, Thermo Fisher Scientific) и AlexaFluor-568 (1:500; A-11369, Thermo Fisher Scientific). Анализ с помощью LysoTracker выполняли, как описано в предыдущей работе (Dorogova et al., 2014) (LysoTracker Red DND-99, Thermo Fisher Scientific). После окраски яичники и семенники помещали в реагент против выцветания ProLong Gold с DAPI (Thermo Fisher Scientific). Изображения получены с помощью микроскопа AxioImager Z1 с приставкой ApoTome (Zeiss), программным обеспечением AxioCam MR и AxioVision (Zeiss).

Результаты

В предыдущих работах мы представили данные, позволяющие сделать вывод, что снижение экспрессии гена *Trl* у мутантов приводит к значительному усилению клеточной гибели в спермато- и среднем оогенезе дрозофилы (Dorogova et al., 2014; Fedorova et al., 2019). У самцов и самок дрозофилы, несущих мутантные аллели Trl^{362} и $Trl^{(ex)15}$ в сочетании с нуль-аллелем Trl^{R85} (Trl^{R85}/Trl^{362} и $Trl^{R85}/Trl^{(ex)15}$), наблюдалась массовая гибель КЗП (Dorogova et al., 2014; Fedorova et al., 2019). В этой работе, проведя более детальное цитологическое исследование данных аллельных комбинаций, мы определили, какой тип гибели КЗП характерен для *Trl*-мутантов и как нарушается или меняется этот процесс по сравнению с нормой.

Мутации *Trl* приводят к усилению клеточной гибели в оогенезе, но не влияют на цитологические проявления этого процесса

Для выявления аутофагии мы использовали лизотрекер (LysoTracker), ацидофильный краситель, который маркирует лизо- и аутофагосомы. После окрашивания сигнал лизотрекера был обнаружен во всех деградирующих ЯК, но его появление совпадало с началом ядерной конденсации как в яичниках мух дикого типа, так и у *Trl*-мутантов (рис. 1, *а, б*). Таким образом, появление аутофаго- и лизосом сопутствовало уже начавшемуся апоптозу, но не предшествовало ему. Это означает, что клеточная гибель в среднем оогенезе у мутантов *Trl* не связана с процессами, которые приводят к избыточной аутофагии, несовместимой с жизнеспособностью клеток.

Для дополнительной верификации апоптоза и выявления активности каспаз мы использовали антитела к эффекторной каспазе Dcp-1. Иммунофлуоресцентное окрашивание антителами к Dcp-1 показало, что этот белок выявляется в гибнущих яйцевых камерах *Trl*-мутантов и его паттерн не отличается от такового у самок дикого типа (см. рис. 1, *в*, *г*).

Поскольку при апоптозе разрушается и фрагментируется ядерная оболочка, с помощью антител к белку ядерной ламины дрозофилы, Lamin Dm0, мы проверили целостность ядерной оболочки в клетках гибнущих ЯК. Ядерная оболочка четко выявлялась антителами к белку ламины во всех клетках ЯК, не подверженных деградации (см. рис. 1, *д*). Однако с появлением признаков конденсации хроматина ядерная оболочка не визуализировалась в КЗП как у мутантов, так и в контроле. Эта структура оставалась видимой только в фолликулярных клетках, которые деградируют позже, после того как осуществят фагоцитоз погибших питающих клеток (см. рис. 1, *е–л*).

Мы также провели анализ деградирующих ЯК методами электронной микроскопии, который выявил специфические изменения в ядре и цитоплазме, характерные для нормального проявления клеточной гибели в среднем оогенезе, описанного ранее в других работах (Giorgi, Deri, 1976) (данные не представлены). То есть у мутантов этот процесс морфологически не отличался от контроля (*Oregon R*).

Таким образом, эффект мутации *Trl* связан с усилением клеточной гибели в среднем оогенезе, что соответствует ранее полученным результатам (Dorogova et al., 2014; Fedorova et al., 2019). Этот процесс становится более массовым у мутантов Trl^{R85}/Trl^{362} и $Trl^{R85}/Trl^{(ex)15}$, но основные критерии и морфологические характеристики соответствуют норме.

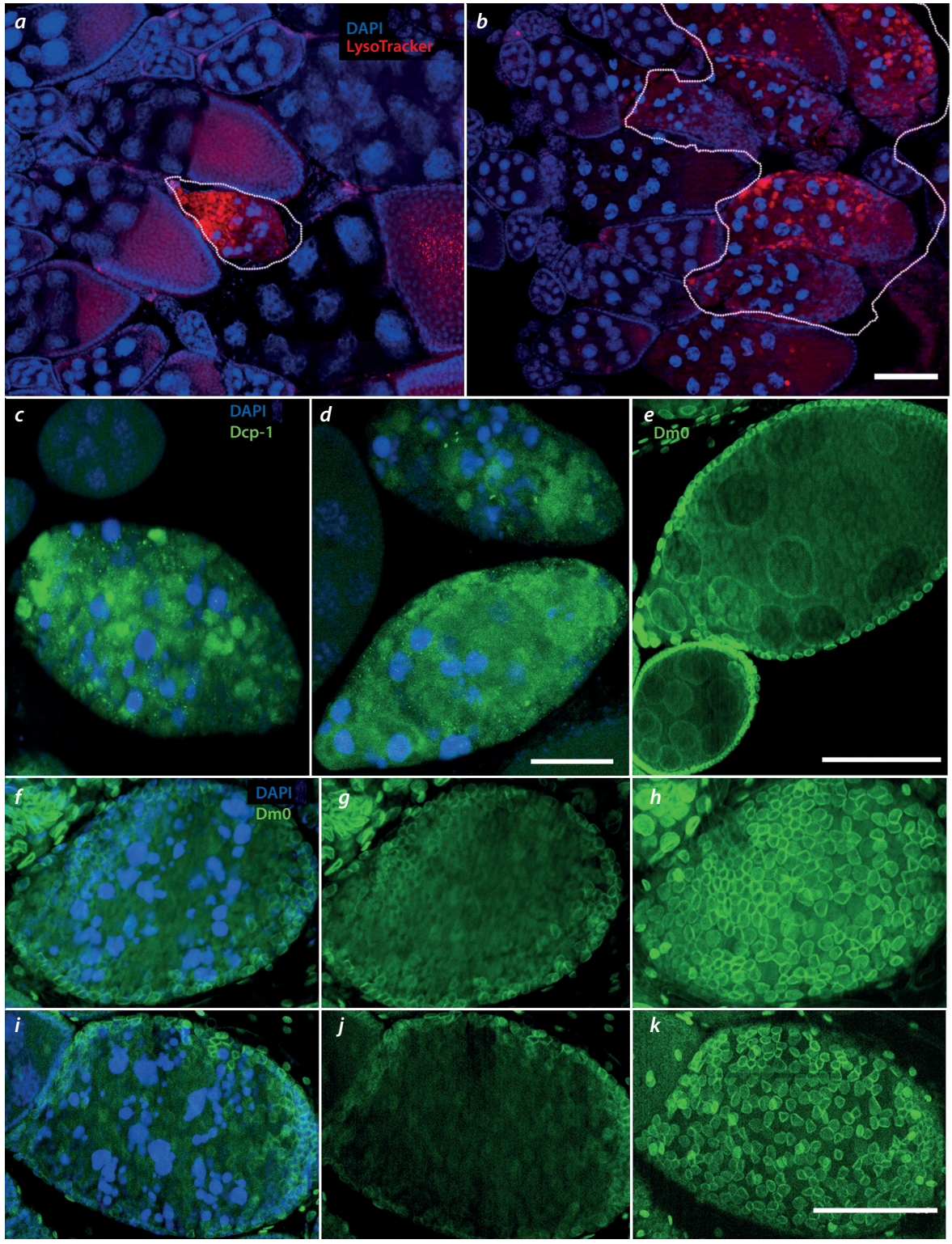


Fig. 1. Features of cell death manifestation in the oogenesis of females mutant for *Trl* (*Trl*^{R85}/*Trl*^{362}).

a, b – Identification of autophagosomes and lysosome in dying egg chambers using LysoTracker red. *a* – In *Oregon R* females ovaries, only few egg chambers undergo cell death (outlined), the rest develop normally and enter vitellogenesis. *b* – In *Trl* females, most egg chambers die at stages 7–9 of oogenesis (outlined), but the LysoTracker staining pattern in mutants does not differ from that in *Oregon R*. *c*, *d* – Staining with antibodies to the Dcp-1 effector caspase. Caspase Dcp-1 is detected in the same way in dying egg chambers of (*c*) *Oregon R* and (*d*) *Trl*. *e–k* – Staining with antibodies to the Lamin Dm0 protein, allowing the assessment of the nuclear envelope integrity during cell death. *e* – The nuclear envelope is clearly visible in all cells of the egg chambers that do not undergo degradation. *f–h* – In the degrading egg chambers of *Oregon R* females, the nurse cell nuclear membrane is destroyed (*g*), but it is preserved and visualized in follicular cells (*h*). *i–k* – In *Trl* females ovaries, the pattern of staining with antibodies to caspase Dcp-1 is the same as in *Oregon R* females. LysoTracker is red; caspase Dcp-1, green; and Lamin Dm0, green. Scale bars: *a*, *b* – 15 μm; *c–e* – 30 μm; *f–k* – 40 μm.

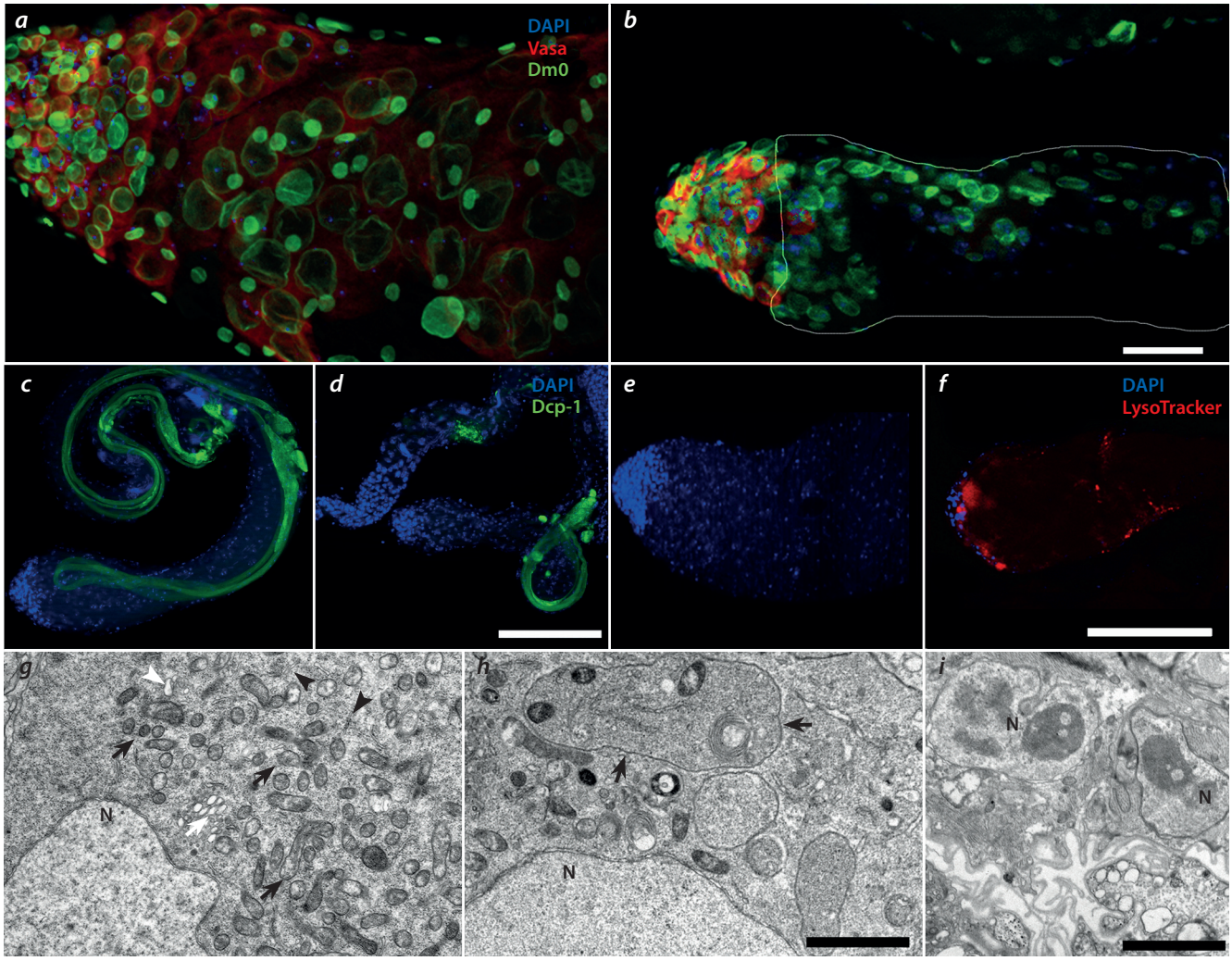


Fig. 2. Features of cell death in the spermatogenesis of *Trl* mutant males (в Trl^{R85}/Trl^{362}).

a, *b* – Immunostaining with antibodies to the Vasa and Lamin Dm0 proteins of the testicle regions where germ tract cells are located at the early stages of spermatogenesis. *a* – In wild-type males, the entire basal region of the testis is filled with spermatocytes at early stages of spermatogenesis, which are stained with antibodies to the Vasa protein. *b* – In mutants, germ cells are detected only in the apical region, and all of them are at the very beginning of spermatogenesis. The nuclear membrane is visualized with antibodies to Lamin Dm0 even in spermatocytes that are no longer stained with antibodies to Vasa. *c*, *d* – The staining pattern with antibodies to the Dcp-1 protein corresponds to the late stages of spermatogenesis, both in (*c*) *Oregon R* males and (*d*) mutants. In mutants, Dcp-1 is not detected in testis regions where mass-scale cell death occurs (*d*). *e*, *f* – Detection of lysosome and autophagosome activity with LysoTracker. The LysoTracker signal is much more intense in *Trl* (*f*) than in the wild type (*e*). *g*–*i* – Ultrastructure of spermatocytes at the interphase (before meiosis) in the wild type (*g*) and in the mutant (*h*, *i*). *g* – In normal spermatogenesis, numerous mitochondria (black arrow), membranes of the endoplasmic reticulum (black arrowhead), multivesicular bodies (white arrow), and occasional lysosomes (white arrowhead) are observed in the cells at this stage. *h* – Mutant cytoplasm is filled with autophagosomes (arrow). *i* – Spermatocytes at the lysis stage retain the internal structure of the nucleus and the nuclear membrane (arrow). Vasa and LysoTracker are red; caspase Dcp-1 and Lamin Dm0, green; N – nucleus. Scale bars: *a*, *b* – 20 μm; *c*, *d* – 30 μm; *e*, *f* – 20 μm; *g*, *h* – 1 μm; *i* – 3 μm.

Мутации *Trl* индуцируют массовую аутофагию в сперматогенезе

Используя такой же методологический подход, мы проанализировали, как происходит гибель КЗП у мутантов *Trl* в сперматогенезе. Предварительное иммуноокрашивание антителами к белку Vasa, специфичному для КЗП, показало, что у мутантов выявляются преимущественно ранние стадии сперматогенеза. Большинство цист на более поздних стадиях сперматогенеза элиминируются в процессе гибели (рис. 2, *а, б*). Однако особенностью гибнущих КЗП в сперматогенезе было сохранение целостности ядерной оболочки. Ядерная оболочка визуализируется антителами к Lamin Dm0 даже в сперматоцитах, которые уже не окрашиваются антителами к Vasa (см. рис. 2, *а, б*).

У мутантных самцов, в отличие от самок, в генеративной ткани не обнаружены признаки апоптоза. Окрашивание антителами к белку Dcp-1 показало активность эффекторной каспазы только на самых поздних стадиях сперматогенеза, во время которых в процессе, называемом индивидуализацией, образуются зрелые сперматозоиды. Такой же паттерн окрашивания наблюдался и в контроле (см. рис. 2, *в, г*). В районах семенника, где у мутантных самцов происходит гибель КЗП, каспаза Dcp-1 не выявлялась. Окрашивание DAPI также подтверждает, что хро-

матин в этих клетках не подвергается конденсации и фрагментации, характерных для апоптоза.

Окраска лизотрекером показала, что ацидофильные компартменты, соответствующие аутофаго- и лизосомам, в изобилии присутствуют во внутренней области семенников. Эти структуры располагаются преимущественно в зоне семенника, где находятся КЗП на ранних стадиях развития – сперматоциты первого порядка (см. рис. 2, *д*, *е*).

С помощью электронно-микроскопического анализа в цитоплазме гибнущих сперматоцитов выявлялись множественные аутофаго- и лизосомы, при этом ядерная оболочка не разрушалась и морфология ядер не отличалась от таковой в негибнущих КЗП (см. рис. 2, *ж–и*).

В результате сравнения проявлений клеточной гибели в сперматогенезе у самцов Trl^{R85}/Trl^{362} и $Trl^{R85}/Trl^{(ex)15}$ не обнаружено фенотипических отличий. Обе мутации вызывают массовую аутофагию и последующий лизис КЗП. Однако процесс гибели этих клеток не сопровождается апоптозом. В норме, у линии *Oregon R*, не выявлены ни признаки апоптоза, ни аутофагии.

Обсуждение

Полученные результаты показали, что недостаток белка GAF вызывает массовую гибель КЗП как у самок, так и самцов дрозофилы, но проявляется эта гибель в зависимости от пола по-разному.

В оогенезе у самок, мутантных по гену *Trl*, большинство ЯК имеют низкую жизнеспособность и не могут пройти через контрольную точку среднего оогенеза (mid oogenesis check point). Известно, что эта стадиоспецифичная контрольная точка активируется в ответ на неблагоприятные стимулы, физиологические нарушения или патологии развития. Яйцевые камеры, которые не пропускаются на следующий этап оогенеза (вителлогенез), подвергаются генетически регулируемой клеточной гибели (Pritchett et al., 2009; Jenkins et al., 2013; Peterson et al., 2015). Отличительной чертой мутантного фенотипа был только высокий показатель гибнущих ЯК, при этом деградируют они так же, как в диком типе. В гибнущих ЯК обнаружены признаки апоптоза и аутофагии, а также морфологические изменения, характерные для нормы.

Согласно данным литературы, подобный фенотип может возникать в ответ на недостаток питательных веществ или снижение активности компонентов инсулин/TOR-сигнального пути (Drummond-Barbosa, Spradling, 2001; Barth et al., 2011; Pritchett, McCall, 2012). Инсулин/TOR-сигнальный путь – консервативный механизм, ответственный за рост клеток и тканей. Он действует как сенсор доступности питательных веществ, способствуя метаболизму, росту и пролиферации клеток. В оогенезе *Drosophila* этот механизм является критически важным для развития КЗП и созревания ооцитов. При его нарушениях ЯК не могут вступать в энергоемкий вителлогенез и деградируют (LaFever et al., 2010; Laws, Drummond-Barbosa, 2017; Jeong et al., 2019). Массовая деградация ЯК в среднем оогенезе также наблюдалась при подавлении экспрессии генов, кодирующих белки, принадлежащие семейству ингибиторов апоптоза – Bruce и Diap, которые негативно регулируют активность каспаз. В таком случае происходило усиление клеточной гибели на фоне нарушения ее регуляции, при этом морфологические критерии этого процесса не менялись и не отличались от нормы (Rodriguez et al., 2002; Xu et al., 2005; Hou et al., 2008). У *Trl*-мутантов отмечен похожий фенотип в оогенезе, что позволяет предположить нарушение регуляторного механизма клеточной гибели.

В сперматогенезе у *Trl*-мутантов массовая гибель КЗП реализуется через механизм избыточной аутофагии. При нормальном развитии зародышевой линии дефектные сперматоциты периодически элиминируются накануне мейоза с помощью лизосом и катаболических ферментов без участия аутофагосом (Yacobi-Sharon et al., 2013). Этот механизм является одним из вариантов генетически регулируемой клеточной гибели и включен в последний каталог Комитета по номенклатуре клеточной смерти (Nomenclature Committee on Cell Death) (Galluzzi et al., 2018). У *Trl*-мутантов лизосомы также принимают участие в деградации сперматоцитов, однако появляются на заключительном этапе гибели, после того как цитоплазма клеток заполнится аутофагосомами.

Важно отметить, что основная функция аутофагии не убивать, а защищать клетки. С ее помощью происходит удаление из клеток поврежденных и состарившихся органелл, цитоплазматических фрагментов, неправильных или нефункциональных белков (Denton et al., 2013; Fitzwalter, Thorburn, 2015; Swart et al., 2016). Базальный (репаративный) уровень аутофагии необходим для поддержания нормальных физиологических условий функционирования клеток (Glick et al., 2010). При определенных условиях, связанных со спецификой развития или стрессовыми воздействиями, аутофагия становится массовой и вместо цитопротекторной функции индуцирует клеточную гибель (Fitzwalter, Thorburn, 2015; Swart et al., 2016). Поэтому аутофагия включена в каталог клеточной смерти как одна из основных форм (Galluzzi et al., 2012, 2018).

У дрозофилы клеточная гибель через аутофагию обнаружена при деградации слюнных желез и средней кишки на стадии метаморфоза «личинка–куколка». Атрофия этих органов является онтогенетически программируемой и регулируется одним и тем же стероидным гормоном – экдизоном (Berry, Baehrecke, 2007; Denton et al., 2012). Однако избыточная аутофагия в сперматогенезе у *Trl*-мутантов, вероятно, не связана с экдизоном, поскольку в сперматоцитах не были выявлены активные рецепторы к этому гормону (Schwedes et al., 2011). Аутофагии и лизису подвергались сперматоциты на стадии роста перед мейотическим делением. Для этой стадии характерен высокий уровень транскрипции генов и синтетической активности. В норме объем сперматоцитов возрастает в 25 раз, что требует значительного потребления энергии и ресурсов (Fuller, 1993). Можно предположить, что мутации *Trl* негативно влияют на клеточный метаболизм, не позволяя достичь необходимого уровня синтеза макромолекул и роста клеток. В результате активируется сигнальный путь, регулируемый TOR-киназой, который индуцирует клеточную гибель через аутофагию. Однако также можно предположить, что недостаток белка GAF приводит к нарушению экспрессии генов, кодирующих компоненты TOR-зависимого сигнального пути или/и

факторы, регулирующие аутофагию, что также может вызвать эктопическую гибель.

Известно, что GAGA-фактор участвует в регуляции транскрипции большой группы генов с различными клеточными функциями (van Steensel et al., 2003; Omelina et al., 2011). В базе данных Flybase аннотированы гены, участвующие в широком спектре процессов, связанных с клеточной гибелью (Gene Ontology terms: apoptotic process, autophagic cell death, salivary gland cell autophagic death). Всего в этой группе представлено около 400 генов, приблизительно 180 из них, согласно профилям связывания GAF из проекта modeNCODE, содержат в промоторных районах сайты связывания GAF и являются, таким образом, потенциальными мишенями этого белка. Наибольший интерес представляют присутствующие в этом списке консервативные гены аутофагии – *Atg2*, *Atg4*, *Atg5*, *Atg7*, *Atg8*, *Atg9*, *Atg16*, *Atg17*, *Atg18*. Большинство этих генов кодируют белки, непосредственно участвующие в формировании и созревании аутофагосом, а *Atg17* – белок, контролирующий инициацию аутофагии (Noda, Inagaki, 2015). Также потенциальным геном-мишенью GAF является *Tor*, кодирующий киназу, вовлеченную в регуляцию аутофагии. Инактивация TOR-киназы в ответ на недостаток питательных веществ или ростовых факторов стимулирует аутофагию (Levine, Klionsky, 2004; Das et al., 2012). Если предположить, что GAF регулирует *Tor* и группу генов *Atg* в сперматогенезе, то его недостаток может вызвать нарушение экспрессии этих генов и привести к массовой и неуправляемой аутофагии. При этом по отношению к *Atg*-генам GAF должен выполнять функцию негативной регуляции, а по отношению к *Tor* – позитивной.

В оогенезе у мутантов *Trl* массовая клеточная гибель происходит не только за счет аутофагии, но и апоптоза, поэтому очевидно, что в этом случае дополнительно может нарушаться активность генов, регулирующих апоптоз. Одним из вероятных кандидатов является ген *diap* (death-associated inhibitor of apoptosis), который, помимо того что является потенциальной мишенью GAF, в мутантной форме индуцирует такой же фенотип в оогенезе, как и *Trl* (Rodriguez et al., 2002; Xu et al., 2005).

Заключение

Фенотип массовой гибели КЗП у *Trl*-мутантов имеет определенную специфику проявления в спермато- и оогенезе. В яичниках аутофагия сопутствует апоптозу, что соответствует каноническому сценарию клеточной гибели в среднем оогенезе дрозофилы. В семенниках наблюдается не характерная для этого типа ткани гибель через аутофагию, которая становится экспансивной и является основной причиной атрофии зародышевой линии. Массовая потеря КЗП связана с недостатком GAGA-фактора, что, вероятно, приводит к нарушению экспрессии генов-мишеней этого белка, отвечающих за клеточную гибель. К его потенциальным мишеням относятся как гены аутофагии, так и апоптоза, и можно предположить, что обе эти группы зависят от активности GAF в оогенезе. Тогда как в сперматогенезе этот белок взаимодействует только с генами аутофагии. Чтобы установить, какие гены под контролем GAF вовлечены в разные механизмы клеточной гибели, необходимо привлечь транскриптомные технологии и проанализировать изменения профилей экспрессии генов в спермато- и оогенезе на фоне недостатка GAF у *Trl*-мутантов.

Список литературы / References

Baricheva E.M., Katokhin A.V., Perelygina L.M. Expression of Drosophila melanogaster gene encoding transcription factor GAGA is tissue-specific and temperature-dependent. *FEBS Lett.* 1997;414(2): 285-288. DOI 10.1016/s0014-5793(97)01010-7.

Barth J.M., Szabad J., Hafen E., Köhler K. Autophagy in *Drosophila* ovaries is induced by starvation and is required for oogenesis. *Cell Death Differ*. 2011;18(6):915-924. DOI 10.1038/cdd.2010.157.

Bayarmagnai B., Nicolay B.N., Islam A.B., Lopez-Bigas N., Frolov M.V. *Drosophila* GAGA factor is required for full activation of the dE2f1-Yki/Sd transcriptional program. *Cell Cycle*. 2012;11(22): 4191-4202. DOI 10.4161/cc.22486.

Berger N., Dubreucq B. Evolution goes GAGA: GAGA binding proteins across kingdoms. *Biochim. Biophys. Acta.* 2012;1819(8):863-868. DOI 10.1016/j.bbagrm.2012.02.022.

Berry D.L, Baehrecke E.H. Growth arrest and autophagy are required for salivary gland cell degradation in *Drosophila*. *Cell*. 2007;131: 1137-1148. DOI 10.1016/j.cell.2007.10.048.

Bhat K.M., Farkas G., Karch F., Gyurkovics H., Gausz J., Schedl P. The GAGA factor is required in the early *Drosophila* embryo not only for transcriptional regulation but also for nuclear division. *Development*. 1996;122;(4):1113-1124.

Bolobolova E.U., Dorogova N.V., Fedorova S.A. Major scenarios of genetically regulated cell death during oogenesis in *Drosophila melanogaster*. *Russ. J. Genet.* 2020;56:655-665. DOI 10.1134/S1022795420060034.

Das G., Shravage B.V., Baehrecke E.H. Regulation and function of autophagy during cell survival and cell death. *Cold Spring Harb. Perspect. Biol*. 2012;4(6):a008813. DOI 10.1101/cshperspect.a008813.

Denton D., Aung-Htut M.T., Lorensuhewa N., Nicolson S., Zhu W., Mills K., Cakouros D., Bergmann A., Kumar S. UTX coordinates steroid hormone-mediated autophagy and cell death. *Nat. Commun.* 2013;4:2916. DOI 10.1038/ncomms3916.

Denton D., Chang T.-K., Nicolson S., Shravage B., Simin R., Baehrecke E.H., Kumar S. Relationship between growth arrest and autophagy in midgut programmed cell death in *Drosophila*. *Cell Death Differ*. 2012;19:1299-1307. DOI 10.1038/cdd.2012.43.

DeVorkin L., Gorski S.M. LysoTracker staining to aid in monitoring autophagy in Drosophila. *Cold Spring Harb. Protoc.* 2014;9:951-958. DOI 10.1101/pdb.prot080325.

Dorogova N.V., Fedorova E.V., Bolobolova E.U., Ogienko A.A., Baricheva E.M. GAGA protein is essential for male germ cell development in *Drosophila*. *Genesis.* 2014;52(8):738-751. DOI 10.1002/dvg.22789.

Dorogova N.V., Khrushcheva A.S., Fedorova E.V., Ogienko A.A., Baricheva E.M. Role of GAGA factor in *Drosophila* primordial germ cell migration and gonad development. *Ontogenez.* 2016;47(1):40-48.

Dos-Santos N., Rubin T., Chalvet F., Gandille P., Cremazy F., Leroy J., Boissonneau E., Theodore L. *Drosophila* retinal pigment cell death is regulated in a position-dependent manner by a cell memory gene. *Int. J. Dev. Biol.* 2008;52(1):21-31. DOI 10.1387/ijdb.072406nd.

Drummond-Barbosa D., Spradling A.C. Stem cells and their progeny respond to nutritional changes during *Drosophila* oogenesis. *Dev. Biol*. 2001;231:265-278. DOI 10.1006/dbio.2000.0135.

Fabian L., Brill J.A. Drosophila spermiogenesis: Big things come from little packages. *Spermatogenesis.* 2012;2(3):197-212. DOI 10.4161/spmg.21798.

Farkas G., Gausz J., Galloni M., Reuter G., Gyurkovics H., Karch F. The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature*. 1994;371:806-808. DOI 10.1038/371806a0.

Fedorova E.V., Dorogova N.V., Bolobolova E.U., Fedorova S.A., Karagodin D.A., Ogienko A.A., Khruscheva A.S., Baricheva E.M. GAGA protein is required for multiple aspects of *Drosophila* oogenesis and female fertility. *Genesis*. 2019;57(2):e23269. DOI 10.1002/dvg.23269.

Fitzwalter B., Thorburn A. Recent insights into cell death and autophagy. *FEBS J.* 2015;282(22):4279-4288. DOI 10.1111/febs.13515.

Fuller M.T. Spermatogenesis In: The development of *Drosophila melanogaster*. *NY Cold Spring Harbor. Press,* 1993;71-147.

Galluzzi L., Vitale I., Aaronson S.A., Abrams J.M., Adam D., Agostinis P., Alnemri E.S., Altucci L., Amelio I., Andrews D.W., …, Zakeri Z., Zhivotovsky B., Zitvogel L., Melino G., Kroemer G. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ.* 2018; 25(3):486-541. DOI 10.1038/s41418-017-0012-4.

Galluzzi L., Vitale I., Abrams J.M., Alnemri E.S., Baehrecke E.H., Blagosklonny M.V., Dawson T.M., Dawson V.L., El-Deiry W.S., Fulda S., Gottlieb E., Green D.R., Hengartner M.O., Kepp O., Knight R.A., Kumar S., Lipton S.A., Lu X., Madeo F., Malorni W., Mehlen P., Nuñez G., Peter M.E., Piacentini M., Rubinsztein D.C., Shi Y., Simon H.-U., Vandenabeele P., White E., Yuan J., Zhivotovsky B., Melino G., Kroemer G. Molecular definitions of cell death subroutines: recommendations of the Nomenclature Committee on Cell Death 2012. *Cell Death Differ.* 2012;19:107-120. DOI 10.1038/cdd.2011.96.

Giorgi F., Deri P. Cell death in ovarian chambers of Drosophila melanogaster. *J. Embryol. Exp. Morphol.* 1976;35:521-533.

Glick D., Barth S., Macleod K.F. Autophagy: cellular and molecular mechanisms. *J. Pathol.* 2010;221(1):3-12. DOI 10.1002/path.2697.

Granok H., Leibovitch B.A., Shaffer C.D., Elgin S.C.R. Gaga over GAGA factor. *Curr. Biol.* 1995;5:238-241. DOI 10.1016/s0960-9822(95)00048-0.

Greenwood J., Gautier J. From oogenesis through gastrulation: developmental regulation of apoptosis. *Semin. Cell Dev. Biol*. 2005;16:215-224. DOI 10.1016/j.semcdb.2004.12.002.

Hay B.A., Guo M. Caspase-dependent cell death in *Drosophila*. *Annu. Rev. Cell Dev. Biol*. 2006;22:623-650. DOI 10.1146/annurev.cellbio.21.012804.093845.

Hou Y.C., Chittaranjan S., Barbosa S.G., McCall K., Gorski S.M. Effector caspase Dcp-1 and IAP protein Bruce regulate starvation-induced autophagy during *Drosophila melanogaster* oogenesis. *J. Cell Biol.* 2008;182:1127-1139. DOI 10.1083/jcb.200712091.

Jenkins V.K., Timmons A.K., McCall K. Diversity of cell death pathways: insight from the fly ovary. *Trends Cell Biol*. 2013;23(11):567-574. DOI 10.1016/j.tcb.2013.07.005.

Jeong E.B., Jeong S.S., Cho E., Kim E.Y. Makorin 1 is required for *Drosophila* oogenesis by regulating insulin/Tor signaling. *PLoS One.* 2019;14(4):e0215688. DOI 10.1371/journal.pone.0215688.

Karagodin D.A., Omelina E.S., Fedorova E.V., Baricheva E.M. Identification of functionally significant elements in the second intron of the *Drosophila melanogaster Trithorax-like* gene. *Gene.* 2013;520: 178-184. DOI 10.1016/j.gene.2013.02.012.

Kihlmark M., Imreh G., Hallberg E. Sequential degradation of proteins from the nuclear envelope during apoptosis. *J. Cell Sci.* 2001; 114(Pt 20):3643-3653.

King R.C. Ovarian development in *Drosophila melanogaster*. New York: Academic Press Inc., 1970.

Kumar S. Caspase function in programmed cell death. *Cell Death Differ.* 2007;14(1):32-43. DOI 10.1038/sj.cdd.4402060.

LaFever L., Feoktistov A., Hsu H.J., Drummond-Barbosa D. Specific roles of Target of rapamycin in the control of stem cells and their progeny in the *Drosophila* ovary. *Development*. 2010;137:2117-2126. DOI 10.1242/dev.050351.

Laws K.M., Drummond-Barbosa D. Control of germline stem cell lineages by diet and physiology. *Results Probl. Cell Differ*. 2017;59: 67-99. DOI 10.1007/978-3-319-44820-6_3.

Levine B., Klionsky D.J. Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev. Cell*. 2004; 6:463-477. DOI 10.1016/s1534-5807(04)00099-1.

Matharu N.K., Hussain T., Sankaranarayanan R., Mishra R.K. Vertebrate homologue of Drosophila GAGA factor. *J. Mol. Biol.* 2010; 400:434-447. DOI 10.1016/j.jmb.2010.05.010.

McCall K. Eggs over easy: cell death in the *Drosophila* ovary. *Dev. Biol.* 2004;274(1):3-14. DOI 10.1016/j.ydbio.2004.07.017.

Napoletano F., Gibert B., Yacobi-Sharon K., Vincent S., Favrot C., Mehlen P., Girard V., Teil M., Chatelain G., Walter L., Arama E., Mollereau B. p53-dependent programmed necrosis controls germ cell homeostasis during spermatogenesis. *PLoS Genet.* 2017;13(9): e1007024. DOI 10.1371/journal.pgen.1007024.

Noda N.N., Inagaki F. Mechanisms of autophagy *Annu. Rev. Biophys.* 2015;44:101-122. DOI 10.1146/annurev-biophys-060414-034248.

Ogienko A.A., Fedorova S.A., Baricheva E.M. Basic aspects of ovarian development in *Drosophila melanogaster*. *Russ. J. Genet.* 2007; 43(10):1120-1134.

Ogienko A.A., Karagodin D.A., Fedorova S.A., Fedorova E.V., Lashina V.V., Baricheva E.M. Effect of hypomorphic mutation in *Trithorax-like* gene on *Drosophila melanogaster* oogenesis. *Ontogenez.* 2006;37(3):211-220.

Ogienko A.A., Karagodin D.A., Pavlova N.V., Fedorova S.A., Voloshina M.A., Baricheva E.M. Molecular and genetic description of a new hypomorphic mutation of *Trithorax-like* gene and analysis of its effect on *Drosophila melanogaster* oogenesis. *Ontogenez*. 2008; 39(2):134-142.

Omelina E.S., Baricheva E.M., Oshchepkov D.Y., Merkulova T.I. Analysis and recognition of the GAGA transcription factor binding sites in *Drosophila* genes. *Comput. Biol. Chem*. 2011;35(6):363-370. DOI 10.1016/j.compbiolchem. 2011.10.008.

Omichinski J.G., Pedone P.V., Felsenfeld G., Gronenborn A.M., Clore G.M. The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat. Struct. Biol.* 1997;4: 122-132.

Peterson J.S., Timmons A.K., Mondragon A.A., McCall K. The end of the beginning: cell death in the germline. *Curr. Top. Dev. Biol.* 2015; 114:93-119. DOI 10.1016/bs.ctdb.2015.07.025.

Pritchett T.L., McCall K. Role of the insulin/Tor signaling network in starvation-induced programmed cell death in *Drosophila* oogenesis. *Cell Death Differ.* 2012;19:1069-1079. DOI 10.1038/cdd.2011.200.

Pritchett T.L., Tanner E.A., McCall K. Cracking open cell death in the *Drosophila* ovary. *Apoptosis.* 2009;14:969-979. DOI 10.1007/s10495-009-0369-z.

Rodriguez A., Chen P., Oliver H., Abrams J.M. Unrestrained caspase-dependent cell death caused by loss of *Diap1* function requires the *Drosophila* Apaf-1 homolog, *Dark*. *EMBO J*. 2002;21:2189-2197. DOI 10.1093/emboj/21.9.2189.

Roy S., Ernst J., Kharchenko P.V., Kheradpour P., Negre N., Eaton M.L., Landolin J.M., Bristow C.A., Ma L.J., Lin M.F., Kellis M. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*. 2010;330(6012):1787-1797. DOI 10.1126/science.1198374.

Sarkissian T., Timmons A., Arya R., Abdelwahid E., White K. Detecting apoptosis in *Drosophila* tissues and cells. *Methods.* 2014;68: 89-96. DOI 10.1016/j.ymeth.2014.02.033.

Schwedes C., Tulsiani S., Carney G.E. Ecdysone receptor expression and activity in adult *Drosophila* melanogaster. *J. Insect Physiol*. 2011;57(7):899-907. DOI 10.1016/j.jinsphys.2011.03.027.

Soeller W.C., Oh C.E., Kornberg T.B. Isolation of cDNAs encoding the Drosophila GAGA transcription factor. *Mol. Cell. Biol.* 1993;13: 7961-7970. DOI 10.1128/mcb.13.12.7961.

Spradling A.C. Developmental genetics of oogenesis. The development of *Drosophila melanogaster*. NY Cold Spring Harb. Lab. Press, 1993;1-70.

Swart Ch., Du Toit A., Loos B. Autophagy and the invisible line between life and death. *Eur. J. Cell Biol*. 2016;95(12):598-610. DOI 10.1016/j.ejcb.2016.10.005.

van Steensel B., Delrow J., Bussemaker H.J. Genomewide analysis of *Drosophila* GAGA factor target genes reveals context-dependent DNA binding. *Proc. Natl. Acad. Sci. USA.* 2003;100(5):2580-2580. DOI 10.1073/pnas.0438000100.

van Steensel B., Delrow J., Henikoff S. Supplementary information: Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.* 2001;27(3):304-308. DOI 10.1038/85871.

Wilkins R.C., Lis J.T. Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. *Nucleic Acids Res.* 1997;25(20):3963-3968. DOI 10.1093/nar/25.20.3963.

Xu D., Li Y., Arcaro M., Lackey M., Bergmann A. The CARD-carrying caspase Dronc is essential for most, but not all, developmental cell death in Drosophila. *Development.* 2005;132:2125-2134. DOI 10.1242/dev.01790.

Yacobi-Sharon K., Namdar Y., Arama E. Alternative germ cell death pathway in Drosophila involves HtrA2/Omi, lysosomes, and a caspase-9 counterpart. *Dev Cell*. 2013;25:29-42. DOI 10.1016/j.devcel.2013.02.002.

ORCID ID

N.V. Dorogova orcid.org/0000-0003-2119-541X

Acknowledgements. This study was supported by the Russian Foundation for Basic Research, project 20-04-00496a, and State Budgeted Project 0259-2021-0011. The authors are grateful to Dr. D.A. Karagodin for valuable discussion and help in the result interpretation.

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

Received November 9, 2020. Revised December 24, 2020. Accepted December 25, 2020.

Вавиловский журнал генетики и селекции. 2021;25(3):301-309
DOI 10.18699/VJ21.51-o

Polymorphism of genes associated with infectious lung diseases in Northern Asian populations and in patients with community-acquired pneumonia

S.V. Mikhailova[1] ✉, L.V. Shcherbakova[2], N.I. Logvinenko[3], I.I. Logvinenko[2, 3], M.I. Voevoda[1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Novosibirsk State Medical University, Novosibirsk, Russia
✉ mikhail@bionet.nsc.ru

Abstract: The innate immune system is the first to respond to invading pathogens. It is responsible for invader recognition, immune-cell recruitment, adaptive-immunity activation, and regulation of inflammation intensity. Previously, two single-nucleotide polymorphisms of innate-immunity genes – rs5743708 (Arg753Gln) of the *TLR2* gene and rs8177374 (Ser180Leu) of the *TIRAP* gene – have been shown to be associated with both pneumonia and tuberculosis in humans, but the data are contradictory among different ethnic groups. It has also been reported that rs10902158 at the *PKP3-SIGGIR-TMEM16J* genetic locus belongs to a haplotype race-specifically associated with tuberculosis. Meanwhile, a gradient of its frequency is observed in Asia. The aim of this work was to assess the effect of selection for the genotypes of the above-mentioned SNPs on the gene pools of populations living in harsh climatic conditions that contribute to the development of infectious lung diseases. We estimated the prevalence of these variants in white and Asian (Chukchis and Yakuts) population samples from Northern Asia and among patients with community-acquired pneumonia (CAP). Carriage of the rs5743708 A allele was found to predispose to severe CAP (odds ratio 2.77, $p = 0.021$), whereas the GG/CT genotype of rs5743708/rs8177374 proved to be protective against it (odds ratio 0.478, $p = 0.022$) in white patients. No association of rs10902158 with CAP (total or severe) was found among whites. Stratification of CAP by causative pathogen may help eliminate the current discrepancies between different studies. No significant difference in rs5743708 or rs8177374 was found between adolescent and long-lived white samples. Carriage of the alleles studied is probably not associated with predisposition to longevity among whites in Siberia. Both white and Asian populations studied were different from Western European and East Asian populations in the variants' prevalence. The frequency of the rs8177374 T (Ser180Leu) variant was significantly higher in the Chukchi sample ($p = 0$, $\chi^2 = 63.22$) relative to the East Asian populations. This result may confirm the hypothesis about the selection of this allele in the course of human migration into areas with unfavorable climatic conditions.
Key words: community-acquired pneumonia; pulmonary tuberculosis; genetic predisposition; genetic polymorphism; *TLR2*; *TIRAP*; *PKP3-SIGGIR-TMEM16J*; long-lived people.

For citation: Mikhailova S.V., Shcherbakova L.V., Logvinenko N.I., Logvinenko I.I., Voevoda M.I. Polymorphism of genes associated with infectious lung diseases in Northern Asian populations and in patients with community-acquired pneumonia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):301-309. DOI 10.18699/VJ21.51-o

Полиморфизм генов, ассоциированных с инфекционными заболеваниями легких, в популяциях Северной Азии и среди пациентов с внебольничными пневмониями

С.В. Михайлова[1] ✉, Л.В. Щербакова[2], Н.И. Логвиненко[3], И.И. Логвиненко[2, 3], М.И. Воевода[1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Научно-исследовательский институт терапии и профилактической медицины – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Новосибирский государственный медицинский университет Министерства здравоохранения Российской Федерации, Новосибирск, Россия
✉ mikhail@bionet.nsc.ru

Аннотация: Врожденный иммунитет первым отвечает на инфекцию. Он участвует в распознавании патогена, привлечении к месту заражения иммунных клеток, а также активирует адаптивный иммунитет и регулирует интенсивность воспалительного ответа. Для двух однонуклеотидных полиморфизмов (ОНП) генов врожденного иммунитета – rs5743708 (Arg753Gln) гена *TLR2* и rs8177374 (Ser180Leu) гена *TIRAP* – была показана ассоциация одновременно с пневмонией и туберкулезом, однако полученные данные различаются для разных этнических групп. Для rs10902158, расположенного в генетическом локусе *PKP3-SIGGIR-TMEM16J*, ранее было показано, что он входит в гаплотип, расоспецифически ассоциированный с туберкулезом. При этом на тер-

ритории Азии наблюдается градиент его частоты. Целью нашей работы была оценка влияния отбора по генотипам названных ОНП на генофонды популяций, живущих в климатических условиях, неблагоприятных по инфекционным заболеваниям легких. Мы оценили распространение этих ОНП в выборках европеоидного и монголоидного (чукчи и якуты) населения Северной Азии и среди пациентов с внебольничной пневмонией. Носительство аллеля A rs5743708 гена *TLR2* предрасполагало к развитию тяжелой внебольничной пневмонии (OR = 2.77, $p = 0.021$), а генотип GG/CT по rs5743708/rs8177374 оказался протективным против нее (OR = 0.478, $p = 0.022$) у европеоидов. Ассоциации rs10902158 с внебольничной пневмонией (как в целом, так и ее тяжелой формы) в европеоидной выборке обнаружено не было. Дифференцировка внебольничных пневмоний по их этиологии, возможно, позволит устранить наблюдаемые противоречия в данных разных исследователей об ассоциации исследованных ОНП с заболеванием. Не выявлено достоверных различий по частоте rs5743708 гена *TLR2* и rs8177374 гена *TIRAP* между европеоидными выборками подростков и долгожителей. Вероятно, эти ОНП не влияют на предрасположенность к долгожительству в европеоидных популяциях, проживающих на Севере Евразии. Исследованные нами европеоидные и монголоидные популяционные выборки отличались по частотам вышеперечисленных вариантов от западноевропейских и восточноазиатских популяций. Частота варианта rs8177374 T (Ser180Leu) гена *TIRAP* в выборке чукчей была достоверно выше ($p = 0$, $\chi^2 = 63.22$), чем в популяциях Восточной Азии, что может служить подтверждением нашего предположения об отборе этого варианта в ходе миграции человека в районы с неблагоприятными климатическими условиями.
Ключевые слова: внебольничная пневмония; туберкулез легких; генетическая предрасположенность; генетический полиморфизм; ген *TLR2*; ген *TIRAP*; генетический район *PKP3-SIGGIR-TMEM16J*; долгожители.

Introduction

Innate immunity constitutes the first barrier against microorganisms and viruses by destroying infected cells and activating adaptive immunity. Nonetheless, an excessive nonspecific immune reaction (inflammation) may be life threatening because it can completely disrupt the functioning of vital organs. Community-acquired pneumonia (CAP) and pulmonary tuberculosis (PTB) are infectious diseases characterized by high mortality, and according to WHO, are ranked consistently among the top 10 leading causes of death in the world (https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death).

Pneumonia is an inflammatory lower-respiratory-tract disease caused by viruses, bacteria, fungi, and parasites. In addition, it may be due to noninfectious processes or have a combined cause. For a long time, *Streptococcus pneumoniae* infection has been considered the main cause of CAP; however, it was shown recently that CAP develops mainly as a result of viral infections (influenza A and B viruses, parainfluenza viruses, adenovirus, respiratory syncytial virus, or coronaviruses) (Choi et al., 2012; Hong et al., 2014; Self et al., 2017). *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Staphylococcus aureus*, *Mycoplasma pneumoniae*, *Chlamydophila pneumoniae*, *Legionella pneumophila* and other microbes may be causative agents of bacterial pneumonia (Choi et al., 2012; Hong et al., 2014; Self et al., 2017). After invasion of the airway epithelium by pathogens, these cells start to produce reactive oxygen species, cytokines, and other mediators to recruit immune cells. Being most abundant in lungs, alveolar macrophages ingest bacteria and apoptotic cells and can present antigens on MHC II to other immune cells. Proinflammatory M1 macrophages produce cytokines TNFα, IL-6, IL-1β, IL-12, and IL-23 to enhance inflammation for elimination of the invaders. Anti-inflammatory M2 macrophages produce cytokines IL-4, IL-13, and TGF-β to induce completion of the inflammatory reaction and remodeling of damaged tissue (Moldoveanu et al., 2009; Arango Duque, Descoteaux, 2014; Kumar, 2019).

Depending on the set of present chemokines and cytokines, different cells responsible for humoral and cellular immunity are attracted to the site of infection (Kumar, 2019). Severe pneumonias are more likely to develop in coinfections; it has been demonstrated that a viral infection (in particular influenza) facilitates the development of pneumococcal infection by damaging the epithelium and reducing the amount of a surfactant (McCullers, 2014; Aguilera, Lenz, 2020). Human respiratory syncytial virus, metapneumovirus, adenovirus, and influenza viruses A and B prefer a cold season (Price et al., 2019). The seasonal increase in the incidence of pneumococcal pneumonia coincides with seasonal outbreaks of influenza; *S. pneumoniae*, *H. influenzae*, and *S. aureus* infections have been reported to be associated with significant influenza pandemics (McCullers, 2014; Bystritskaya, Bilichenko, 2017; Morris et al., 2017).

PTB is a pulmonary infectious disease caused mainly by *Mycobacterium tuberculosis* (*Mtb*). According to the WHO, approximately one-quarter of the world's population is estimated to be infected by *Mtb*, and 5–15 % of these people will fall ill with active tuberculosis. In Russia, most of these patients (95 %) have PTB (https://minzdrav.gov.ru/ministry/61/22/stranitsa-979/statisticheskie-i-informatsionnye-materialy/statisticheskiy-sbornik-2018-god). The pathogenesis of pulmonary tuberculosis is based on *Mtb* survival after phagocytosis by alveolar macrophages. These bacteria can modulate a host immune response to protect the infected cells, change their metabolism, induce IL-10, suppress IL-12 and TNFα synthesis, and to inhibit MHC II expression and antigen presentation. *Mtb* makes macrophages unresponsive to interferon (IFN) γ and inhibits autophagy. It allows the mycobacteria to establish a persistent or latent infection in macrophages. Mycobacteria are believed to use the general mechanism of negative feedback regulation that restricts excessive inflammation (Harding, Boom, 2010; Richardson et al., 2015; Gopalakrishnan, Salgame, 2016). With the loss of immunity-driven control over mycobacterial reproduction, foamy macrophages accumulate in granulomas, and lung tissue necrosis begins (Liu C.H. et al., 2017). Vitamin D deficiency is known to negatively affect the effectiveness of the immune response in tuberculosis (Wilkinson et al., 2000; Aibana et al., 2019).

These data suggest that in Northern Asia, a region with low temperature and reduced insolation during most of the year,

signs of purifying selection for genes associated with lung infections may be noticeable. For many genes of innate immunity, an association with viral, bacterial, and autoimmune diseases has been proven. Despite differences in the pathogenesis between CAP and PTB, it has been shown that minor alleles of rs5743708 (the *TLR2* gene) and of rs8177374 (the *TIRAP* gene) can have a pathogenic and protective effect, respectively, in both of these lung diseases in humans. Nevertheless, data obtained by different research groups are contradictory.

Toll-like receptors (TLRs) play a pivotal role in host defense. Being membrane-anchored (TLRs 1, 2, 4–6, and 10) or endosomal (TLRs 3 and 7–9) in human immune cells (e. g., macrophages, monocytes, dendritic cells, and some leukocytes), they are involved in the recognition of structurally conserved surface molecules of microorganisms and viruses as well as viral nucleic acids (Barbalat et al., 2009; Kawai, Akira, 2010; Kumar, 2019). TLR2 participates in the recognition of a large number of diverse lipoproteins and peptidoglycans of gram-positive and gram-negative bacteria, fungi, and virus-infected cells. After ligand binding to the receptor, TIR (Toll-interleukin 1 receptor) domains of TLR2 and TLR1 or TLR2 and TLR6 dimerize via the formation of an extensive hydrogen-bonding network and hydrophobic interactions (Jin et al., 2007; Takeda, Akira, 2015). Homodimerization of the cytoplasmic domains of TLR2 does not induce TNFα production *in vitro* in murine macrophages, and the formation of the TLR2–TLR2 dimer is not detectable even in the presence of an agonist (Ozinsky et al., 2000; Shukla et al., 2018). Therefore, the existence of TLR2–TLR2 homodimers *in vivo* is being questioned. After ligand binding, reorientation of the TIR domains and triggering of a cascade of intracellular reactions lead to the activation of proinflammatory NF-κB and MAPK pathways, synthesis and a release of proinflammatory cytokines (IL-1β, IL-12, TNF-α, and IL-6) and various chemokines into extracellular space, and the development of an inflammatory response at the pathogen entry site (Liu C.H. et al., 2017; Tapader et al., 2018). In inflammatory monocytes, TLR2 induces type I IFN production in response to a viral ligand (Barbalat et al., 2009). It is reported that prolonged stimulation of TLR2 (more than 24 h) causes PI3K/Akt pathway activation in alveolar macrophages. It limits the production of NF-κB, TNF-α, and IL-12 and activates the synthesis of anti-inflammatory IL-10. This mechanism is assumed to prevent excessive inflammation (Richardson et al., 2015; Liu Y. et al., 2016).

The *TLR2* gene is located in 4q31.3, has five exons, and expresses few splicing isoforms, but all of the coding sequences are contained within exon 3. The protein consists of 784 amino acid residues (aa) and includes extracellular leucine-rich repeat domains, which are primarily responsible for ligand recognition (aa 54–524), followed by the leucine-rich repeat C-terminal domain (aa 525–579) and intracellular TIR domain (aa 639–782), which mediates downstream signaling (https://www.uniprot.org/uniprot/O60603). It is expressed constitutively on macrophages and dendritic cells and can be induced in epithelial cells or B-cells. Its overexpression in patients with pneumococcal disease had been documented (Siebert et al., 2018).

Single-nucleotide polymorphism (SNP) rs5743708 (of the *TLR2* gene) causing the Arg753Gln substitution is located in the TIR domain of the protein. This SNP is associated simultaneously with resistance to Lyme disease (Schröder et al., 2005) and with predisposition to tuberculosis (Guo, Xia, 2015; Patarčić et al., 2015), whereas the association with predisposition to PTB is race-specific (Caws et al., 2008; Guo, Xia, 2015; Hu et al., 2019). It is believed that TLR2 signaling may be nonessential to control acute tuberculosis but important during chronic tuberculosis (Gopalakrishnan, Salgame, 2016). A meta-analysis has shown the *TLR2* rs5743708 minor allele to be associated with CAP, Legionnaires' disease, and pneumococcal disease; however, the data obtained in different studies are contradictory (Moens et al., 2007; Patarčić et al., 2015; Smelaya et al., 2016).

The *TIRAP* (TIR domain-containing adaptor protein) gene also known as *Mal* (MyD88 adapter-like) encodes one of the five adapter proteins that are involved in signal transduction from activated TLRs to protein kinases at the plasma membrane (Bonham et al., 2014). It is located in 11q24.2, consists of six exons, and encodes a protein of 221 aa. The TIRAP protein includes an N-terminal PEST domain (aa 15–35) responsible for binding to special sites in the plasma membrane, followed by an AB-loop mediating MyD88 and TLR4 binding. A binding site for TRAF6 (TNF receptor-associated factor 6) is located within the region aa 188–193 (Bernard, O'Neill, 2013). TIRAP is expressed in many cell types (Narayanan, Park, 2015), and its isoforms resulting from alternative splicing have unknown functions. There are different opinions about whether TIRAP forms a complex with the TIR domain of TLR6 for signal transmission; however, it has been proven that TIRAP mediates TLR2 and TLR4 signaling by facilitating the recruitment of the MyD88 adaptor protein to the TLRs (Nagpal et al., 2009; Bernard, O'Neill, 2013).

Activation of NF-κB, MAPK1, MAPK3, and JNK results in cytokine secretion and an inflammatory response. SNP rs8177374 (the *TIRAP* gene) is located in exon 5 and represents the Ser180Leu substitution in the encoded protein. It is located close to the TLR-binding site of TIRAP. In carriers of this substitution, modulation of TLR1, TLR2, TLR4, and TLR6 but not TLR9 signaling has been shown (Khor et al., 2007; Ferwerda et al., 2009; Siebert et al., 2018). Ser180Leu in a heterozygous state has a protective effect against PTB and invasive pneumococcal disease in white and African samples and against malaria in African and Asian samples (Khor et al., 2007; Panda et al., 2016). Carriage of heterozygous Ser180Leu protects children from pneumococcal lower-respiratory-tract infections, whereas carriers of the homozygous 180Leu polymorphism alone or in combination with some TLR1 and TLR6 polymorphisms may be susceptible to recurrent pneumococcal infections (Siebert et al., 2018). Simultaneous carriage of the TIRAP 180Leu variant and some SNPs in the *TLR4* gene as well as 180Leu homozygosity increases susceptibility to severe hospital-acquired infections (Kumpf et al., 2010).

The opposite results have been obtained as well. The rs8177374 T allele (180Leu) increases the risk of PTB in a sample of Iranian population (Naderi et al., 2014). A meta-analysis of nine published case-control studies did not reveal a significant association of 180L with tuberculosis risk (Miao et al., 2011). There are controversial opinions about the mechanism behind the observed protective effect of Ser180Leu heterozygosity. They are based on differences in observed

effects of the SNP at the level of proinflammatory cytokines. Depending on the model used, some research groups showed an increased level (Ferwerda et al., 2009; Panda et al., 2016) and others a decreased level (Khor et al., 2007; Kumpf et al., 2010; Siebert et al., 2018) of cytokines after their induction in 180 Leu/Leu carriers. Accordingly, homozygosity of the minor variant of rs8177374 is thought to cause either an excessive inflammatory reaction or the absence of an adequate immune response. It is supposed that selection pressure on the *TIRAP* gene provides a balance between protection against excessive inflammation and effective defense during infectious diseases (Khor et al., 2007; Ferwerda et al., 2009).

Besides the polymorphisms in genes *TLR2* and *TIRAP*, in this paper, we focused on the *PKP3-SIGGIR-TMEM16J* gene region. An association of its haplotypes with different types of tuberculosis has been shown among children in Vietnam and South Africa (Horne et al., 2012; Gupta et al., 2016). It is believed that the impact of the haplotypes on immunity is determined by SIGIRR (single immunoglobulin interleukin 1 receptor; synonym: IL-1R8), which is a negative regulator of TLR signaling (Molgora et al., 2016). Carriage of rs10902158 GG and rs7111432 AA in introns of *PKP3* and *TMEM16J*, respectively, acts additively with a vitamin D deficiency and "pathogenic" genotypes of rs5743708 (*TLR2*) and rs8177374 (*TIRAP*) on tuberculosis predisposition (Horne et al., 2012; Gupta et al., 2016). rs10902158 located in intron 2 of the *PKP3* gene has been analyzed. Encoded desmosomal plaque protein plakophilin 3 is involved in intracellular adhesion (Gurjar et al., 2018). Of note, rs10902158 has a frequency gradient in Asia; according to the Genome Aggregation Database (GnomAD) (https://gnomad.broadinstitute.org/), it is absent in South Asia and is found with a frequency of ~50 % in Southeast Asia. Nonetheless, functional significance of genetic variants in noncoding parts of the *PKP3-SIGGIR-TMEM16J* gene region, including rs10902158, is not clear.

In this work, we analyzed the frequencies of rs5743708, rs8177374, and rs10902158 (for which conflicting data on the association with respiratory infections have been reported previously) in white and Asian samples from Northern Asia and among CAP patients. According to the statistics of the Ministry of Health of the Russian Federation, Novosibirsk Oblast and Yakutia are characterized by an increased incidence of PTB, whereas Chukotka Autonomous Okrug is the leader in both pneumonia and PTB morbidity in Russia (https://minzdrav.gov.ru/ministry/61/22/stranitsa-979/statisticheskie-i-informatsionnye-materialy/statisticheskiy-sbornik-2017-god; https://minzdrav.gov.ru/ministry/61/22/stranitsa-979/statisticheskie-i-informatsionnye-materialy/statisticheskiy-sbornik-2018-god). According to the WHO, the highest death rate from pneumonia is observed before the age of 5 and after 75–80 years (https://www.who.int/medicines/areas/priority_medicines/Ch6_22Pneumo.pdf). Therefore, we assumed that long-lived people of the Siberian Federal District may differ from adolescents in the frequency of rs5743708 and rs8177374, and we assessed the prevalence of the pathogenic variants in the sample of long-lived people.

Materials and methods

The study protocol was approved by the local Ethics Committee of the Institute of Internal and Preventive Medicine (branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia; approval No. 22.06.2008). Written informed consent to be examined and to participate in the study was obtained from each patient. For individuals younger than 18 years, the informed consent was signed by a parent or legal guardian.

The white sample consisted of 451 adolescents (95 % Russians, 197 males, 253 females, aged 14–17) from Oktiabr'skii district of Novosibirsk (55°01′ N 82°55′ E) and 289 Russian settlers (120 males, 169 females, aged 45–64, mean age 52.5) in towns Tommot (58°58′00″ N 126°16′0″ E), Neryungri (56°39′30″N 124°43′30″ E), Ust-Nera (64°34′05″ N 143°14′10″ E), and Yakutsk (62°01′38″ N 129°43′55″ E), who lived in Yakutia for more than 16 years or were born there. The adolescent sample was described previously (Zavyalova et al., 2011). Asian samples (220 individuals) consisted of 130 Chukchis (66 males, 64 females, aged 18–73, mean age 40) from the Kanchalan village (65°10′41″ N 176°44′52″ E) of Chukotka Autonomous Okrug and 132 Yakuts (53 males, 79 females, aged 44–64, mean age 49) from the Kylayy village (63°13′34″ N 132°08′06″ E) and towns Tommot, Neryungri, and Ust-Nera of Yakutia. The sample of long-lived people was collected in cities Novosibirsk, Tomsk, and Tumen and consisted of 188 individuals (180 females, 8 males) aged 90–105, mean age 92.

Ethnicity of individuals was identified using questionnaires and additional cross-examination with elucidation of the nationality of ancestors (at least in three generations). Persons of mixed origin were excluded from the analysis. The CAP patient sample (406 whites) was collected in offices of pulmonary hospitals of Novosibirsk and Yakutsk in 2003–2005 before the COVID-19 outbreak. The sample consists of 120 patients with severe CAP (aged 18–80, mean age 52) and 286 patients with nonsevere (mild to moderate severity) CAP (aged 16–92, mean age 39). The diagnosis of pneumonia was made on the basis of radiologically confirmed "fresh" lung tissue infiltration and clinical data (fever, cough, sputum production, chest pain, and shortness of breath) in the absence of an obvious diagnostic alternative. CAP of various etiologies was regarded as severe if the CURB65 rating scale index was 4–5 points (Lim et al., 2003).

Genomic DNA was extracted from peripheral blood leucocytes by the standard phenol–chloroform method (Sambrook, Russell, 2006). Genotyping was performed using polymerase chain reaction (PCR) with an analysis of restriction fragment length polymorphism by electrophoresis in a 5 % polyacrylamide gel after visualization with an ethidium bromide solution.

The rs5743708 SNP (*TLR2*) was identified by amplification of DNA with primers 5′-GCCATTCTCATTCTTCTGG*AGC-3′ and 5′-GGGAACCTAGGACTTTATCGCA-3′ (* denotes a nucleotide changed for restriction site creation). The 168-bp PCR product was digested with the Pst I restriction endonuclease (SibEnzyme, Novosibirsk) for 2 h at 37 °C. The rs5743708 A (753Q) allele was revealed by the presence of fragments of 20, 45, and 103 bp, whereas the rs5743708 G allele by fragments of 20 and 148 bp.

The detection of rs8177374 (*TIRAP*) was performed by amplification of genomic DNA with primers 5′-GGCTGCACCATCCCCCA*GC-3′ and 5′-CCGTTCCCCTTCTCCCT

CCTGTAG-3′ (* denotes a nucleotide changed for restriction site creation). The 162-bp PCR product was digested with the AccB7 I restriction endonuclease (SibEnzyme, Novosibirsk) for 2 h at 37 °C. The rs8177374 T (180L) allele was identified by the presence of fragments of 21 and 141 bp, whereas in case of rs8177374 C, the PCR product was not cut.

Primers 5′-TGGCAAGGATTGGAGAACTC*C*TGTC-3′ and 5′-CAGGGCCAGTGCCTCCCC-3′ (* denotes nucleotides changed for restriction site creation) were used for the amplification of the *PKP3* intron 2 sequence containing rs10902158. The resulting 192-bp amplicon was digested with the BstEN I restriction endonuclease (Sibenzyme, Novosibirsk) for 2 h at 65 °C. In the presence of the rs10902158 A allele, the PCR product was not cut, whereas in the case of the rs10902158 G allele, fragments of 24 and 168 bp were observed.

Statistical analysis was performed in the SPSS 16.0 software.

Results

Genotype distributions were consistent with the Hardy–Weinberg equilibrium among all the population samples (data not shown). Minor allele frequencies for rs5743708, rs8177374, and rs10902158 are represented in Table 1.

In the sample of Novosibirsk adolescents, allele frequencies of rs5743708, rs8177374, and rs10902158 differed from those of the non-Finnish European sample in GnomAD ($\chi^2 = 6.621$, $p = 0.013$; $\chi^2 = 19.541$, $p = 0$; $\chi^2 = 54.554$, $p = 0$, respectively).

For the frequency of the rs5743708 A (Arg753Gln) allele, there was a tendency for a decrease in Russian settlers in Yakutia and among long-lived people compared with the Novosibirsk sample.

The rs8177374 T (Ser180Leu) frequency did not differ among the studied white samples.

In the sample of Russian settlers of Yakutia, males and females differed in the frequency of rs10902158 ($p = 0.037$, $\chi^2 = 4.86$). Moreover, the frequency of this SNP among males was closer to that observed in non-Finnish Europeans according to GnomAD data, and the frequency among females was closer to that observed in the Novosibirsk sample. Perhaps there were sex differences during recent migration to Yakutia from different regions of Russia. Genetic analysis of a larger sample and estimation of this SNP's frequency in western regions of Russia are required for explaining the observed differences.

The two analyzed Asian samples differed from each other and from GnomAD East Asian cohorts. Among Yakuts, the rs5743708 A (Arg753Gln) variant, which is very rare in other Asian populations, was found at a frequency of 0.015 ± 0.007 (mean ± SD). Chukchis differed significantly from GnomAD East Asians in rs10902158 allele frequency ($p = 0$, $\chi^2 = 63.22$) (see Table 1).

Frequencies of polymorphisms rs5743708, rs8177374, and rs10902158 were not different among adolescents and total white CAP patient samples. By contrast, after the sample was divided into patients with severe and nonsevere CAP, differences were found for rs5743708 ($p = 0.021$, $\chi^2 = 6.24$). Next, genotype frequencies were estimated for rs5743708, rs8177374, and rs10902158 in the examined samples (except for long-lived people regarding rs10902158). For the latter SNP (in the *PKP3* gene), no difference in frequency was detectable within any group (data not shown). The observed distribution of rs5743708 and rs8177374 genotypes among the studied samples is presented in Table 2.

Table 1. Minor allele frequencies for rs5743708, rs8177374, and rs10902158 in the studied samples and in the Genome Aggregation Database (GnomAD)

Group, sample size	rs5743708 A (*TLR2*)		rs8177374 T (*TIRAP*)		rs10902158 A (*PKP3*)	
	Allele number	Allele frequency	Allele number	Allele frequency	Allele number	Allele frequency
Asians						
Chukchis, 130	0	0	17	0.065	116	0.446
Yakuts, 132	4	0.015	6	0.023	100	0.379
East Asia (GnomAD)		0.0002		0.012		0.562
Whites						
Adolescents (Novosibirsk), 451	39	0.043	93	0.103	198	0.220
Long-lived people, 188	11	0.029	29	0.077	–	–
Russians (Yakutia) total sample, 289	13	0.022	50	0.087	119	0.206
Russians (Yakutia), females, 169	7	0.020	29	0.086	80	0.237
Russians (Yakutia), males, 120	6	0.025	21	0.088	39	0.163
Total CAP, 406	25	0.031	80	0.100	200	0.246
Severe CAP, 120	13	0.054	17	0.071	62	0.258
Nonsevere CAP, 286	12	0.021	63	0.109	138	0.241
Non-Finnish European (GnomAD)		0.029		0.157		0.153

Note: CAP, community-acquired pneumonia.

Table 2. Combined genotype frequencies of rs5743708 and rs8177374 in the studied populations and patient groups

Samples/genotypes	Sample size	GG/CC		GG/CT		GG/TT		AG/CC		AG/CT		AA/CC		AA/CT	
		n	*q*	*n*	*q*	*n*	*q*	*n*	*q*	*n*	*q*	*n*	*q*	*n*	*q*
Adolescents, Novosibirsk	451	335	0.742	75	0.166	4	0.009	27	0.060	8	0.018	1	0.002	1	0.002
Russians, Yakutia	289	230	0.796	44	0.152	2	0.007	10	0.035	3	0.010	0	0	0	0
Long-lived people	188	150	0.798	25	0.133	2	0.011	11	0.058	0	0	0	0	0	0
Total CAP	406	308	0.759	71	0.175	2	0.005	22	0.054	3	0.007	0	0	0	0
Severe CAP	120	93	0.775	13	0.108	1	0.008	11	0.092	2	0.017	0	0	0	0
Nonsevere CAP	286	215	0.752	58	0.203	1	0.003	11	0.038	1	0.003	0	0	0	0
Chukchis	130	113	0.869	17	0.131	0	0	0	0	0	0	0	0	0	0
Yakuts	132	122	0.924	6	0.045	0	0	4	0.030	0	0	0	0	0	0

Note: *n*, number of genotype carriers; *q*, genotype frequency.

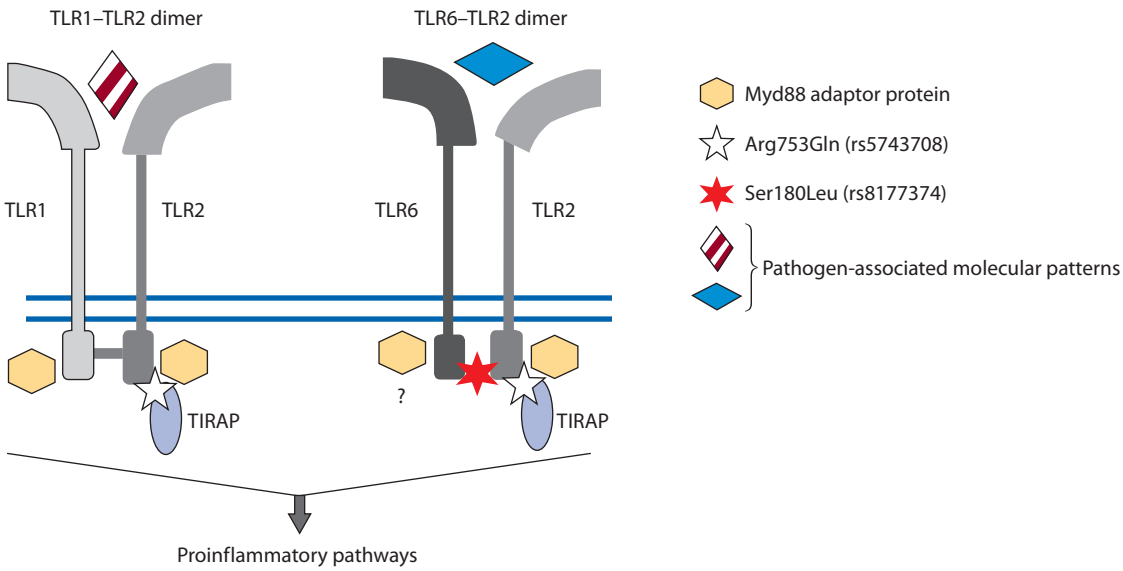
The frequencies of genotypes of rs5743708 and rs8177374 did not differ among the following samples: adolescents of Novosibirsk, long-lived people of Siberia, all patients with CAP, and all Russians in Yakutia. Possibly, the carriage of the studied alleles is not associated with predisposition to longevity in Siberia and does not significantly affect the probability of resettling of migrants in more unfavorable climatic conditions at present. This notion is consistent with WHO findings that in Eastern Europe, in contrast to Western and Central Europe, the pneumonia mortality rate does not increase significantly after age 80 (https://www.who.int/medicines/areas/priority_medicines/Ch6_22Pneumo.pdf). As for pneumonia, carriage of the rs5743708 A allele predisposed to severe CAP (AG/CC+AG/CT vs all: odds ratio 2.77, 95 % confidence interval 1.227–6.272, $p = 0.021$). The heterozygous genotype of rs8177374 in combination with the GG genotype of rs5743708 had a protective effect against severe CAP (GG/CT vs all: odds ratio 0.478, 95 % confidence interval 0.251–0.909, $p = 0.022$).

Discussion

It was shown here that carriage of none of the three studied SNPs, rs5743708, rs8177374, and rs10902158, is associated with the predisposition to CAP in total. By contrast, we found that the Arg753Gln variant of *TLR2* predisposes to severe CAP, and the heterozygous Ser180Leu variant of *TIRAP* in combination with the 753 Arg/Arg variant of *TLR2* has a protective effect against it in the white population. These data partially explain the contradictions in the data from different researchers. Most likely, the contribution of the alleles of genes *TLR2* and *TIRAP* to CAP predisposition is determined by pneumonia etiology. A substantial proportion of severe pneumonia cases are known to be caused by combined viral and bacterial infections (McCullers, 2014; Morris et al., 2017; Aguilera, Lenz, 2020). TLR2 is responsible mainly for the recognition of bacteria-associated molecular patterns; Ser180Leu of the *TIRAP* gene modulates signal transduction only from TLR2 and TLR4 recognizing molecular patterns of bacteria as well (Nagpal et al., 2009). Most likely, combined and bacterial but not viral pneumonias are associated with *TLR2* and *TIRAP* gene variants.

The studied Asian ethno-geographical groups showed an increased frequency of the protective rs8177374 T (Ser180Leu) variant of *TIRAP* relative to neighboring East Asian populations. In the Chukchi sample, the difference was significant ($p = 0$, $\chi^2 = 63.22$). It may be a consequence of the natural selection that has promoted protection from excessive inflammation during pulmonary diseases. The hypothesis about the selection of the Ser180Leu variant along with the out-of-Africa migration to a harsh environment has been advanced earlier (Khor et al., 2007; Ferwerda et al., 2009). An increased frequency of the Arg753Gln variant of *TLR2* and a decreased frequency of Ser180Leu of *TIRAP* as compared to non-Finnish Europeans may indicate higher genetic predisposition of the Siberian white population to PTB and severe CAP. Nevertheless, there are a lot of genes associated with CAP and PTB independently. Apparently, during the settlement of peoples in Northern Eurasia, the formation of gene pools had been determined by the selection that facilitated adaptation to specific infections (Lime disease among them), parasites, and the climate. It would be interesting to determine why two mutations changing the same TLR2 signaling have opposite effects on the predisposition to severe CAP. One possible explanation is the difference in the structure and functions of heterodimers TLR2–TLR1 and TLR2–TLR6 (see the Figure).

Existing data on the roles of TLRs 1, 2, and 6 in the activation of proinflammatory and anti-inflammatory signaling are conflicting. Overexpression of TLR2 carrying the Arg753Gln variant has been demonstrated to cause a significantly stronger impairment of cytokine induction by TLR2/TLR1 ligands as compared with TLR2/TLR6 ligands in the HEK293 cell line (Schröder et al., 2005). In later papers, it has been shown that the Arg-to-Gln substitution at position 753 of TLR2 changes the size, charge, and hydrophobic properties of this site and reduces the ability of TLR2 to form a heterodimer with TLR6 (Basith et al., 2011; Xiong et al., 2012). This SNP significantly alters agonist-inducible association of TLR2 with adaptor proteins TIRAP and MyD88 and impairs NF-κB signaling and IL-8 mRNA expression in the HEK293 cell line (Xiong et al., 2012). Genes *TLR1* and *TLR2* have different expression activators (Lancioni et al., 2011). In the TLR2–TLR1 dimer, TLR1 and TLR2 are responsible for NF-κB and MAPK pathways



Locations of the Arg753Gln substitution in TLR2 and Ser180Leu in TIRAP on the protein complexes.

and for PI3K pathway activation, respectively; therefore, upregulation of proinflammatory cytokines is TLR1-dependent, whereas upregulation of type I IFN is TLR2-dependent (Raieli et al., 2019). TLR2–TLR6 binding disruption possibly causes an increase in the number of TLR2–TLR1 dimers in which TLR1 drives the activation of the NF-κB inflammatory signaling cascade. On the contrary, the Ser180Leu variant of TIRAP weakens proinflammatory signal transmission. Besides, TIRAP acts as an adaptor protein for TLR4 homodimers. This receptor is primarily responsible for the recognition of lipopolysaccharides of gram-negative bacteria and fungi (Takeda, Akira, 2015). It is believed that TLR4 takes part not only in MyD88-dependent proinflammatory signaling but also in MyD88-independent anti-inflammatory signaling. Weakening of TLR4 signaling through TIRAP probably enhances the signaling through adapter proteins TRIF and TRAM, causing the secretion of anti-inflammatory cytokines (Li et al., 2013).

It remains unclear why the SNPs in *TLR2* and *TIRAP* have similar effects on the predisposition to or protection against acute (CAP) and chronic (PTB) lung infections. Perhaps this phenomenon is due to an impact on inflammation in both diseases. The severity of CAP is determined by life-threatening acute inflammation; in PTB, the development of chronic inflammation masks the infection from the host immune system (Liu C.H. et al., 2017).

Conclusions

In Northern Asian populations, the observed difference in rs8177374 frequency may reflect consequences of natural selection during the settlement of peoples on territories with unfavorable climatic conditions. As for pneumonia, carriage of the rs5743708 A allele (the *TLR2* gene) predisposes to severe CAP; the heterozygous genotype of rs8177374 (the *TIRAP* gene) in combination with the GG genotype of rs5743708 (the *TLR2* gene) has a protective effect against it. Stratification of CAP by causative pathogen may help to eliminate the current discrepancies among research groups. Regional differences in a set of pathogens along with the genetic characteristics of ethno-geographical groups can determine the associations of various genetic variants of innate immunity with the prevalence and severity of pneumonia.

References

Aguilera E.R., Lenz L.L. Inflammation as a modulator of host susceptibility to pulmonary influenza, pneumococcal, and co-infections. *Front. Immunol.* 2020;11:105. DOI 10.3389/fimmu.2020.00105.

Aibana O., Huang C.C., Aboud S., Arnedo-Pena A., Becerra M.C., Bellido-Blasco J.B., Bhosale R., Calderon R., Chiang S., Contreras C., Davaasambuu G., Fawzi W.W., Franke M.F., Galea J.T., Garcia-Ferrer D., Gil-Fortuño M., Gomila-Sard B., Gupta A., Gupte N., Hussain R., Iborra-Millet J., Iqbal N.T., Juan-Cerdán J.V., Kinikar A., Lecca L., Mave V., Meseguer-Ferrer N., Montepiedra G., Mugusi F.M., Owolabi O.A., Parsonnet J., Roach-Poblete F., Romeu-García M.A., Spector S.A., Sudfeld C.R., Tenforde M.W., Togun T.O., Yataco R., Zhang Z., Murray M.B. Vitamin D status and risk of incident tuberculosis disease: a nested case-control study, systematic review, and individual-participant data meta-analysis. *PLoS Med.* 2019;16(9):e1002907. DOI 10.1371/journal.pmed.1002907.

Arango Duque G., Descoteaux A. Macrophage cytokines: involvement in immunity and infectious diseases. *Front. Immunol.* 2014;5:491. DOI 10.3389/fimmu.2014.00491.

Barbalat R., Lau L., Locksley R.M., Barton G.M. Toll-like receptor 2 on inflammatory monocytes induces type I interferon in response to viral but not bacterial ligands. *Nat. Immunol.* 2009;10(11):1200-1207. DOI 10.1038/ni.1792.

Basith S., Manavalan B., Govindaraj R.G., Choi S. *In silico* approach to inhibition of signaling pathways of Toll-like receptors 2 and 4 by ST2L. *PLoS One.* 2011;6(8):e23989. DOI 10.1371/journal.pone.0023989.

Bernard N.J., O'Neill L.A. Mal, more than a bridge to MyD88. *IUBMB Life.* 2013;65(9):777-786. DOI 10.1002/iub.1201.

Bonham K.S., Orzalli M.H., Hayashi K., Wolf A.I., Glanemann C., Weninger W., Iwasaki A., Knipe D.M., Kagan J.C. A promiscuous lipid-binding protein diversifies the subcellular sites of toll-like receptor signal transduction. *Cell.* 2014;156(4):705-716. DOI 10.1016/j.cell.2014.01.019.

Bystritskaya E.V., Bilichenko T.N. An analysis of pneumonia morbidity in adults and children at Russian Federation, 2010–2014. *Pulmonologiya = Russian Pulmonology.* 2017;27(2):173-178. DOI 10.18093/0869-0189-2017-27-2-173-178. (in Russian)

Caws M., Thwaites G., Dunstan S., Hawn T.R., Lan N.T., Thuong N.T., Stepniewska K., Huyen M.N., Bang N.D., Loc T.H., Gagneux S., van Soolingen D., Kremer K., van der Sande M., Small P., Anh P.T., Chinh N.T., Quy H.T., Duyen N.T., Tho D.Q., Hieu N.T., Torok E., Hien T.T., Dung N.H., Nhu N.T., Duy P.M., van Vinh Chau N., Farrar J. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis. PLoS Pathog.* 2008;4(3):e1000034. DOI 10.1371/journal.ppat.1000034.

Choi S.H., Hong S.B., Ko G.B., Lee Y., Park H.J., Park S.Y., Moon S.M., Cho O.H., Park K.H., Chong Y.P., Kim S.H., Huh J.W., Sung H., Do K.H., Lee S.O., Kim M.N., Jeong J.Y., Lim C.M., Kim Y.S., Woo J.H., Koh Y. Viral infection in patients with severe pneumonia requiring intensive care unit admission. *Am. J. Respir. Crit. Care Med.* 2012;186(4):325-332. DOI 10.1164/rccm.201112-2240OC.

Ferwerda B., Alonso S., Banahan K., McCall M.B.B., Giamarellos-Bourboulis E.J., Ramakers B.P., Mouktaroudi M., Fain P.R., Izagirre N., Syafruddin D., Cristea T., Mockenhaupt F.P., Troye-Blomberg M., Kumpf O., Maiga B., Dolo A., Doumbo O., Sundaresan S., Bedu-Addo G., van Crevel R., Hamann L., Oh D.-Y., Schumann R.R., Joosten L.A.B., de la Rúa C., Sauerwein R., Drenth J.P.H., Kullberg B.-J., van der Ven A.J.A.M., Hill A.V., Pickkers P., van der Meer J.W.M., O'Neill L.A.J., Netea M.G. Functional and genetic evidence that the Mal/TIRAP allele variant 180L has been selected by providing protection against septic shock. *Proc. Natl. Acad. Sci. USA.* 2009;106(25):10272-10277. DOI 10.1073/pnas.0811273106.

Gopalakrishnan A., Salgame P. Toll-like receptor 2 in host defense against *Mycobacterium tuberculosis*: to be or not to be – that is the question. *Curr. Opin. Immunol.* 2016;42:76-82. DOI 10.1016/j.coi.2016.06.003.

Guo X.G., Xia Y. The rs5743708 gene polymorphism in the TLR2 gene contributes to the risk of tuberculosis disease. *Int. J. Clin. Exp. Pathol.* 2015;8(9):11921-11928.

Gupta A., Montepiedra G., Gupte A., Zeldow B., Jubulis J., Detrick B., Violari A., Madhi S., Bobat R., Cotton M., Mitchell C., Spector S., IMPAACT NWCS113 and P1041 Study Team. Low vitamin-D levels combined with *PKP3-SIGIRR-TMEM16J* host variants is associated with tuberculosis and death in HIV-infected and -exposed infants. *PLoS One.* 2016;11(2):e0148649. DOI 10.1371/journal.pone.0148649.

Gurjar M., Raychaudhuri K., Mahadik S., Reddy D., Atak A., Shetty T., Rao K., Karkhanis M.S., Gosavi P., Sehgal L., Gupta S., Dalal S.N. Plakophilin3 increases desmosome assembly, size and stability by increasing expression of desmocollin2. *Biochem. Biophys. Res. Commun.* 2018;495(1):768-774. DOI 10.1016/j.bbrc.2017.11.085.

Harding C.V., Boom W.H. Regulation of antigen presentation by *Mycobacterium tuberculosis*: a role for Toll-like receptors. *Nat. Rev. Microbiol*. 2010;8(4):296-307. DOI 10.1038/nrmicro2321.

Hong H.L., Hong S.B., Ko G.B., Huh J.W., Sung H., Do K.H., Kim S.H., Lee S.O., Kim M.N., Jeong J.Y., Lim C.M., Kim Y.S., Woo J.H., Koh Y., Choi S.H. Viral infection is not uncommon in adult patients with severe hospital-acquired pneumonia. *PLoS One.* 2014;9(4):e95865. DOI 10.1371/journal.pone.0095865.

Horne D.J., Randhawa A.K., Chau T.T., Bang N.D., Yen N.T., Farrar J.J., Dunstan S.J., Hawn T.R. Common polymorphisms in the *PKP3-SIGIRR-TMEM16J* gene region are associated with susceptibility to tuberculosis. *J. Infect Dis*. 2012;205(4):586-594. DOI 10.1093/infdis/jir785.

Hu L., Tao H., Tao X., Tang X., Xu C. TLR2 Arg753Gln gene polymorphism associated with tuberculosis susceptibility: an updated meta-analysis. *Biomed. Res. Int.* 2019;2628101. DOI 10.1155/2019/2628101.

Jin M.S., Kim S.E., Heo J.Y., Lee M.E., Kim H.M., Paik S.G., Lee H., Lee J.O. Crystal structure of the TLR1–TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell.* 2007;130(6):1071-1082. DOI 10.1016/j.cell.2007.09.008.

Kawai T., Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat. Immunol.* 2010;11(5):373-384. DOI 10.1038/ni.1863.

Khor C.C., Chapman S.J., Vannberg F.O., Dunne A., Murphy C., Ling E.Y., Frodsham A.J., Walley A.J., Kyrieleis O., Khan A., Aucan C., Segal S., Moore C.E., Knox K., Campbell S.J., Lienhardt C., Scott A., Aaby P., Sow O.Y., Grignani R.T., Sillah J., Sirugo G., Peshu N., Williams T.N., Maitland K., Davies R.J., Kwiatkowski D.P., Day N.P., Yala D., Crook D.W., Marsh K., Berkley J.A., O'Neill L.A., Hill A.V. A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat. Genet.* 2007;39(4):523-528. DOI 10.1038/ng1976.

Kumar V. Inflammation research sails through the sea of immunology to reach immunometabolism. *Int. Immunopharmacol.* 2019;73:128-145. DOI 10.1016/j.intimp.2019.05.002.

Kumpf O., Giamarellos-Bourboulis E.J., Koch A., Hamann L., Mouktaroudi M., Oh D.Y., Latz E., Lorenz E., Schwartz D.A., Ferwerda B., Routsi C., Skalioti C., Kullberg B.J., van der Meer J.W., Schlag P.M., Netea M.G., Zacharowski K., Schumann R.R. Influence of genetic variations in *TLR4* and *TIRAP/Mal* on the course of sepsis and pneumonia and cytokine release: an observational study in three cohorts. *Crit. Care.* 2010;14(3):R103. DOI 10.1186/cc9047.

Lancioni C.L., Li Q., Thomas J.J., Ding X., Thiel B., Drage M.G., Pecora N.D., Ziady A.G., Shank S., Harding C.V., Boom W.H., Rojas R.E. *Mycobacterium tuberculosis* lipoproteins directly regulate human memory $CD4^+$ T cell activation via Toll-like receptors 1 and 2. *Infect. Immun.* 2011;79(2):663-673. DOI 10.1128/IAI.00806-10.

Li J., Lee D.S., Madrenas J. Evolving bacterial envelopes and plasticity of TLR2-dependent responses: basic research and translational opportunities. *Front. Immunol*. 2013;4:347. DOI 10.3389/fimmu.2013.00347.

Lim W.S., van der Eerden M.M., Laing R., Boersma W.G., Karalus N., Town G.I., Lewis S.A., Macfarlane J.T. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax.* 2003;58:377-382. DOI 10.1136/thorax.58.5.377.

Liu C.H., Liu H., Ge B. Innate immunity in tuberculosis: host defense vs pathogen evasion. *Cell Mol. Immunol.* 2017;14(12):963-975. DOI 10.1038/cmi.2017.88.

Liu Y., Li J.Y., Chen S.T., Huang H.R., Cai H. The rLrp of *Mycobacterium tuberculosis* inhibits proinflammatory cytokine production and downregulates APC function in mouse macrophages via a TLR2-mediated PI3K/Akt pathway activation-dependent mechanism. *Cell. Mol. Immunol.* 2016;13(6):729-746. DOI 10.1038/cmi.2015.58.

McCullers J.A. The co-pathogenesis of influenza viruses with bacteria in the lung. *Nat. Rev. Microbiol.* 2014;12(4):252-262. DOI 10.1038/nrmicro3231.

Miao R., Li J., Sun Z., Xu F., Shen H. Meta-analysis on the association of TIRAP S180L variant and tuberculosis susceptibility. *Tuberculosis* (*Edinb*.). 2011;91(3):268-272. DOI 10.1016/j.tube.2011.01.006.

Moens L., Verhaegen J., Pierik M., Vermeire S., De Boeck K., Peetermans W.E., Bossuyt X. Toll-like receptor 2 and Toll-like receptor 4 polymorphisms in invasive pneumococcal disease. *Microbes Infect.* 2007;9(1):15-20. DOI 10.1016/j.micinf.2006.10.002.

Moldoveanu B., Otmishi P., Jani P., Walker J., Sarmiento X., Guardiola J., Saad M., Yu J. Inflammatory mechanisms in the lung. *J. Inflamm. Res.* 2009;2:1-11.

Molgora M., Barajon I., Mantovani A., Garlanda C. Regulatory role of IL-1R8 in immunity and disease. *Front. Immunol.* 2016;7:149. DOI 10.3389/fimmu.2016.00149.

Morris D.E., Cleary D.W., Clarke S.C. Secondary bacterial infections associated with influenza pandemics. *Front. Microbiol.* 2017;8:1041. DOI 10.3389/fmicb.2017.01041.

Naderi M., Hashemi M., Pourmontaseri Z., Eskandari-Nasab E., Bahari G., Taheri M. TIRAP rs8177374 gene polymorphism increased the risk of pulmonary tuberculosis in Zahedan, southeast Iran. *Asian Pac. J. Trop. Med.* 2014;7(6):451-455. DOI 10.1016/S1995-7645(14)60073-0.

Nagpal K., Plantinga T.S., Wong J., Monks B.G., Gay N.J., Netea M.G., Fitzgerald K.A., Golenbock D.T. A TIR domain variant of MyD88

adapter-like (Mal)/TIRAP results in loss of MyD88 binding and reduced TLR2/TLR4 signaling. *J. Biol. Chem.* 2009;284(38):25742-25748. DOI 10.1074/jbc.M109.014886.

Narayanan K.B., Park H.H. Toll/interleukin-1 receptor (TIR) domain-mediated cellular signaling pathways. *Apoptosis.* 2015;20(2):196-209. DOI 10.1007/s10495-014-1073-1.

Ozinsky A., Underhill D.M., Fontenot J.D., Hajjar A.M., Smith K.D., Wilson C.B., Schroeder L., Aderem A. The repertoire forpattern recognition of pathogens by the innate immune system is defined by cooperation between Toll-like receptors. *Proc. Natl. Acad. Sci. USA.* 2000;97:13766-13771. DOI 10.1073/pnas.250476497.

Panda A.K., Das B.K., Panda A., Tripathy R., Pattnaik S.S., Mahto H., Pied S., Pathak S., Sharma S., Ravindran B. Heterozygous mutants of TIRAP (S180L) polymorphism protect adult patients with *Plasmodium falciparum* infection against severe disease and mortality. *Infect. Genet. Evol.* 2016;43:146-150. DOI 10.1016/j.meegid.2016.04.035.

Pataržić I., Gelemanović A., Kirin M., Kolčić I., Theodoratou E., Baillie K.J., de Jong M.D., Rudan I., Campbell H., Polašek O. The role of host genetic factors in respiratory tract infectious diseases: systematic review, meta-analyses and field synopsis. *Sci. Rep.* 2015;5:16119. DOI 10.1038/srep16119.

Price R.H.M., Graham C., Ramalingam S. Association between viral seasonality and meteorological factors. *Sci. Rep.* 2019;9(1):929. DOI 10.1038/s41598-018-37481-y.

Raieli S., Trichot C., Korniotis S., Pattarini L., Soumelis V. TLR1/2 orchestrate human plasmacytoid predendritic cell response to gram + bacteria. *PLoS Biol.* 2019;17(4):e3000209. DOI 10.1371/journal.pbio.3000209.

Richardson E.T., Shukla S., Sweet D.R., Wearsch P.A., Tsichlis P.N., Boom W.H., Harding C.V. Toll-like receptor 2-dependent extracellular signal-regulated kinase signaling in *Mycobacterium tuberculosis*-infected macrophages drives anti-inflammatory responses and inhibits Th1 polarization of responding T cells. *Infect. Immun.* 2015;83(6):2242-2254. DOI 10.1128/IAI.00135-15.

Sambrook J., Russell D.W. Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harbor Protoc.* 2006;2006(1):4455. DOI 10.1101/pdb.prot4455.

Schröder N.W.J., Diterich I., Zinke A., Eckert J., Draing C., von Baehr V., Hassler D., Priem S., Hahn K., Michelsen K.S., Hartung T., Burmester G.R., Göbel U.B., Hermann C., Schumann R.R. Heterozygous Arg753Gln polymorphism of human TLR-2 impairs immune activation by *Borrelia burgdorferi* and protects from late stage Lyme disease. *J. Immunol.* 2005;175(4):2534-2540. DOI 10.4049/jimmunol.175.4.2534.

Self W.H., Balk R.A., Grijalva C.G., Williams D.J., Zhu Y., Anderson E.J., Waterer G.W., Courtney D.M., Bramley A.M., Trabue C., Fakhran S., Blaschke A.J., Jain S., Edwards K.M., Wunderink R.G. Procalcitonin as a marker of etiology in adults hospitalized with community-acquired pneumonia. *Clin. Infect. Dis.* 2017;65(2):183-190. DOI 10.1093/cid/cix317.

Shukla S., Richardson E.T., Drage M.G., Boom W.H., Harding C.V. *Mycobacterium tuberculosis* lipoprotein and lipoglycan binding to Toll-like receptor 2 correlates with agonist activity and functional outcomes. *Infect. Immun.* 2018;86(10). pii: e00450-18. DOI 10.1128/IAI.00450-18.

Siebert J.N., Hamann L., Verolet C.M., Gameiro C., Grillet S., Siegrist C.A., Posfay-Barbe K.M. Toll-interleukin 1 receptor domain-containing adaptor protein 180L single-nucleotide polymorphism is associated with susceptibility to recurrent pneumococcal lower respiratory tract infections in children. *Front. Immunol.* 2018;9:1780. DOI 10.3389/fimmu.2018.01780.

Smelaya T.V., Belopolskaya O.B., Smirnova S.V., Kuzovlev A.N., Moroz V.V., Golubev A.M., Pabalan N.A., Salnikova L.E. Genetic dissection of host immune response in pneumonia development and progression. *Sci. Rep.* 2016;6:35021. DOI 10.1038/srep35021.

Takeda K., Akira S. Toll-like receptors. *Curr. Protoc. Immunol.* 2015;109:14.12.1-14.12.10. DOI 10.1002/0471142735.im1412s109.

Tapader R., Bose D., Dutta P., Das S., Pal A. SslE (YghJ), a cell-associated and secreted lipoprotein of neonatal septicemic *Escherichia coli*, induces Toll-like receptor 2-dependent macrophage activation and proinflammation through NF-kB and MAP kinase signaling. *Infect. Immun.* 2018;86(9):e00399-18. DOI 10.1128/IAI.00399-18.

Wilkinson R.J., Llewelyn M., Toossi Z., Patel P., Pasvol G., Lalvani A., Wright D., Latif M., Davidson R.N. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. *Lancet.* 2000;355(9204):618-621. DOI 10.1016/S0140-6736(99)02301-6.

Xiong Y., Song C., Snyder G.A., Sundberg E.J., Medvedev A.E. R753Q polymorphism inhibits Toll-like receptor (TLR) 2 tyrosine phosphorylation, dimerization with TLR6, and recruitment of myeloid differentiation primary response protein 88. *J. Biol. Chem.* 2012;287(45):38327-38337. DOI 10.1074/jbc.M112.375493.

Zavyalova L.G., Denisova D.V., Simonova G.I., Orlov P.S., Voevoda M.I. Association of polymorphisms of genes *FTO* and *TCF7L2* with cadiometabolic parameters of the adolescents in Siberia. *Bulleten SO RAMN = Bulleten SB RAMS.* 2011;31(5):5-13. (in Russian).

ORCID ID

S.V. Mikhailova orcid.org/0000-0002-0897-5473
L.V. Shcherbakova orcid.org/0000-0001-9270-9188
I.I. Logvinenko orcid.org/0000-0003-1348-0253
M.I. Voevoda orcid.org/0000-0001-9425-413X

Acknowledgements. This study was supported by Russian State Budjet program No. 0259-2021-0014. The authors are grateful to E.S. Merekina, L.A. Ponikarovskaya, and N.E. Egorova for the help with the collection of the clinical material.
The English language was corrected and certified by shevchuk-editing.com.

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

Received June 5, 2020. Revised August 12, 2020. Accepted September 28, 2020. Published online March 29, 2021.

Вавиловский журнал генетики и селекции. 2021;25(3):310-317
DOI 10.18699/VJ21.034

The expression of apoptosis-regulating proteins Bcl-2 and Bad in liver cells of C57Bl/6 mice under light-induced functional pinealectomy and after correction with melatonin

S.V. Michurina[1], I.Yu. Ishchenko[1] ✉, S.A. Arkhipov[1], A.Yu. Letyagin[1, 2], M.A. Korolev[1], E.L. Zavjalov[2]

[1] Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ irenisch@mail.ru

Abstract. The presence of humans and animals under long-term continuous lighting leads to a suppression of melatonin synthesis, that is, to light-induced functional pinealectomy (LIFP), and the development of desynchronosis. To create LIFP, C57Bl/6 mice were kept under 24-hour lighting (24hL) for 14 days. The animals in the control group were kept under standard lighting conditions. In the next series of experiments, mice with LIFP received daily intragastrically either melatonin (1 mg/kg body weight in 200 µl of distilled water) or 200 µl of water as a placebo. The comparison group consisted of intact animals that received placebo under standard lighting conditions. Immunohistochemical analysis (using an indirect avidin-biotin peroxidase method) revealed the expression of the antiapoptotic protein Bcl-2 and the proapoptotic protein Bad in sinusoid liver cells (a heterogeneous population consisting of the endotheliocytes, Kupffer cells, Ito cells, and Pit cells) and in individual hepatocytes. The Bad expression area in the liver of LIFP mice increased 4 times against a background of the unchanged Bcl-2 expression area. Changes in the brightness (a parameter inversely proportional to the marker concentration) of Bad and Bcl-2 areas did not reach significance. Our results indicate a weakening of the antiapoptotic protection of liver cells of LIFP animals, which creates conditions for activation of the "mitochondrial branch" of apoptosis. Melatonin treatment of LIFP mice resulted in a 3.3-fold increase in Bcl-2 expression area and a 2.7 % decrease in Bcl-2 region brightness compared with the experimental untreated group. Bad protein parameters were unreliable. Thus, melatonin treatment of animals cancels the effect of LIFP, restoring the Bcl-2 expression area and increasing this protein concentration, which indicates an increase in antiapoptotic protection and creates conditions for blocking the development of the "mitochondrial branch" of apoptosis in liver cells.
Key words: melatonin; 24-hour lighting; light-induced functional pinealectomy; liver; Bad; Bcl-2.

For citation: Michurina S.V., Ishchenko I.Yu., Arkhipov S.A., Letyagin A.Yu., Korolev M.A., Zavjalov E.L. The expression of apoptosis-regulating proteins Bcl-2 and Bad in liver cells of C57Bl/6 mice under light-induced functional pinealectomy and after correction with melatonin. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):310-317. DOI 10.18699/VJ21.034

Экспрессия белков-регуляторов апоптоза Bcl-2 и Bad в клетках печени мышей C57Bl/6 в условиях светоиндуцированной функциональной эпифизэктомии и после коррекции мелатонином

С.В. Мичурина[1], И.Ю. Ищенко[1] ✉, С.А. Архипов[1], А.Ю. Летягин[1, 2], М.А. Королев[1], Е.Л. Завьялов[2]

[1] Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ irenisch@mail.ru

Аннотация. Пребывание человека и животных в условиях длительного непрерывного освещения приводит к подавлению синтеза мелатонина, т. е. к светоиндуцированной функциональной эпифизэктомии (СФЭ), и развитию десинхроноза. Для создания СФЭ мыши линии C57Bl/6 содержались в условиях круглосуточного освещения в течение 14 суток. Животные контрольной группы находились при стандартном режиме освещения. В следующей серии экспериментов мыши с СФЭ получали ежедневно внутрижелудочно либо мелатонин (1 мг/кг массы тела в 200 мкл воды), либо в качестве плацебо 200 мкл дистиллированной воды. Группой сравнения служили интактные животные, получавшие плацебо при стандартном режиме освещения. В результате иммуногистохимического анализа (непрямым авидин-биотиновым пероксидазным методом) в си-

нусоидных клетках печени (гетерогенная популяция, состоящая из эндотелиоцитов, клеток Купфера, клеток Ито и pit-клеток) и в отдельных гепатоцитах была выявлена экспрессия антиапоптотического белка Bcl-2 и проапоптотического протеина Bad. В клетках печени мышей с моделью СФЭ обнаружено четырехкратное увеличение площади экспрессии Bad на фоне неизменившейся площади экспрессии Bcl-2. Достоверных изменений яркости (параметр, обратно пропорциональный концентрации маркера) участков, окрашенных на Bcl-2 и Bad, отмечено не было. Полученные данные свидетельствуют об ослаблении антиапоптотической защиты клеток печени животных, содержавшихся при круглосуточном освещении, что создает условия для активации «митохондриальной ветви» апоптоза, наиболее выраженной в синусоидных клетках печени. Введение мелатонина мышам с моделью СФЭ привело к возрастанию в 3.3 раза площади экспрессии Bcl-2 и снижению на 2.7 % яркости (т. е. увеличению концентрации) участков, окрашенных на Bcl-2, по сравнению с опытной группой без лечения. Для Bad изменения исследуемых параметров имели характер тенденций. Таким образом, интрагастральное введение мелатонина животным аннулирует эффект светоиндуцированной функциональной пинеалэктомии, восстанавливая площадь экспрессии и увеличивая концентрацию антиапоптотического белка Bcl-2 в клетках печени, что свидетельствует об усилении антиапоптотической защиты клеток органа и создает условия для блокирования развития «митохондриальной ветви» апоптоза.
Ключевые слова: мелатонин; круглосуточное освещение; светоиндуцированная функциональная эпифизэктомия; печень; Bad; Bcl-2.

Introduction

At present, human activities are often associated with a change in the natural rhythm of life and with working in artificial lighting conditions, which leads to an increase in the light day period. Lighting at night is considered by scientists as "light pollution"; it is attributed to non-chemical endocrine disruptors affecting both human and animal health, including violations of circadian regulation of melatonin (MT) synthesis, metabolism and other hormone-controlled systems, and the cancer risk (Michurina et al., 2005; Borodin et al., 2012; Russart, Nelson, 2018). By now, scientists have concluded that melatonin is not "a sleep hormone, but a dark hormone" (Reiter et al., 2013; Arendt, 2019). It's known that light suppresses melatonin production, and darkness weakens this suppression, stimulating the synthesis and release of this hormone into the bloodstream. Of particular importance is the fact that MT suppression in nocturnal rodents is initiated by light. A light pulse lasting only 15 min is sufficient to induce locomotor suppression that endures for more than an hour, and a 1-min light pulse also suppresses MT synthesis for about the same amount of time (Morin, 2013). As a result of long-term stay of humans and animals under 24-hour lighting (24hL) conditions, a decrease/cessation of hormone production leads to the development of light-induced functional pinealectomy (LIFP) (Delibas et al., 2002) and desynchronosis (Reiter et al., 2017; Arendt, 2019). Under these conditions, a significant load falls on the homeostatic systems providing the body resistance (lymphatic, immune and endocrine systems), which are in an integral relationship with the liver, which is the main organ of homeostasis. The study of the structural and functional features of liver cells showed that exactly the cooperative interactions of highly specialized parenchymal liver cells (hepatocytes) and sinusoidal cells (a heterogeneous population of cells consisting of endotheliocytes, Kupffer cells, Ito cells and Pit cells), and their work in a strictly defined rhythm, help the organ to perform numerous functions.

Apoptosis is a fundamental biological mechanism, which causes a clean, non-inflammatory form of cell death and helps the body get rid of unnecessary and defective cells. The ratio of antiapoptotic (Bcl-2, Bcl-XL) and proapoptotic proteins (Bad, Bax, etc.) is considered to be a "molecular switch", which determines whether tissue growth or atrophy will occur (Willis et al., 2003; Polčic, Mentel, 2020). The features of Bcl-2 family protein expression in liver cells under light-induced functional pinealectomy remain largely unexplored.

Based on the above, the aim of the study was to evaluate the expression of antiapoptotic Bcl-2 protein and proapoptotic Bad protein in the liver cells of C57Bl/6 mice under light-induced functional pinealectomy and after melatonin treatment.

Materials and methods

The experiments were carried out in the SPF Vivarium of the Institute of Cytology and Genetics, SB RAS (RFMEFI61914 X0005 and RFMEFI62114X0010). C57Bl/6 mice (male, aged 10–12 weeks) were kept in controlled barrier rooms with free access to water and food (Ssniff, Germany).

Two series of experiments were carried out. In the first event, mice were kept under 24-hour lighting (24hL) for 14 days (light/dark photoperiod 24:0 h) to create light-induced functional pinealectomy (the "24hL" group, $n = 6$). The comparison group consisted of intact animals (the "Control" group, $n = 5$) kept under standard lighting conditions (14:10 h). At the same time a smooth increase in illumination to daytime values within 1 hour (dawn) and a smooth decrease in illumination values until complete shutdown within 1 hour (sunset) were assigned to the light phase of the day. In the second series of experiments mice were kept under 24hL for 14 days and received daily intragastrically either melatonin at a dose of 1 mg/kg of body weight in 200 µl of distilled water (the "24hL+MT" group, $n = 5$) or 200 µl of water (the "24hL+Placebo" group, $n = 6$). The comparison group consisted of animals (the "Placebo" group, $n = 6$) kept under standard lighting conditions (14:10 h) and received daily intragastrically 200 µl of distilled water.

Animals were removed from the experiment by the craniocervical dislocation method and liver samples were taken for light-optical and immunohistochemical studies. All experiments were performed in accordance with humanity principles and were carried out in compliance with "Rules for working with experimental animals" (The Annex to the Order of the Ministry of Health of the USSR No. 755 of 12.08.1977) and Council Directive 86/609/EEC. Experiments were approved by the local ethical committee (The Protocol No. 128 of 15.03.2017).

Liver samples were fixed in 10 % buffered formalin (BioVitrum, Russia) for 48 hours, dehydrated in a series of alcohols of increasing concentrations and embedded in Histomix (BioVitrum, Russia). Tissue sections with a thickness of 3 μm were prepared on a microtome HM 340E (Thermo Fisher Scientific, USA). Immunohistochemical study of the expression of the antiapoptotic Bcl-2 protein and the proapoptotic Bad protein was performed on liver paraffin sections by means of indirect avidin-biotin peroxidase method (ABC-method) using the Vectastain Universal ABC-Peroxidase Kit (Vector Laboratories, Catalog Number PK-7200). At the last stage, immunohistochemical staining was carried out in a chromogenic substrate containing diaminobenzidine (the solution is prepared *ex tempore* from the components of the set "ImmPACT DAB"; Vector Laboratories, Catalog Number SK-4105).

For quantification of Bcl-2 and Bad expression in the mouse liver, a computer morphometric analysis of digital photographs obtained using a LEICA DM 2500 microscope with a LEICA DFC425C video camera (Germany, Switzerland) at ×400 magnification was performed. The relative area and the brightness of intermediate zones of the hepatic lobules staining for Bcl-2 and Bad were determined in digital images using the program ImageJ. The significance of differences between the compared values was determined using the nonparametric Mann–Whitney test. Differences of compared values were considered statistically significant at $p < 0.05$.

Results

The expression of Bcl-2 and Bad proteins in liver cells of mice under light-induced functional pinealectomy

A study of Bcl-2 family protein expression in the liver of mice kept under 24-hour lighting (light/dark photoperiod 24:0 h) revealed the pronounced immunohistochemical staining of the proapoptotic Bad protein in sinusoidal cells of blood sinusoid capillaries (Fig. 1). The Bad-positive signal was detected in the endothelium of interlobular veins and in the ductal epithelium of triad bile ducts, and it was also sometimes found in single hepatocytes. At the same time, weak immunohistochemical staining of the antiapoptotic Bcl-2 protein was revealed in sinusoidal liver cells and in single hepatocytes of "24hL" mice liver (see Fig. 1). Staining of Bcl-2 wasn't determined in the ductal epithelium of triad bile ducts.

Morphometric analysis of liver preparations of the "24hL" animals confirmed the results of the light-optical study. An increase in the Bad expression area was found to be 4.1 times greater than in animals under natural light conditions (Fig. 2, *a*). At the same time, the brightness (a parameter inverse to the concentration) of the areas stained of that protein did not change significantly (see Fig. 2, *b*). Changes in the relative area and the brightness of zones stained for the antiapoptotic Bcl-2 protein were in the nature of a trend and reflected a slight decrease in the expression area and

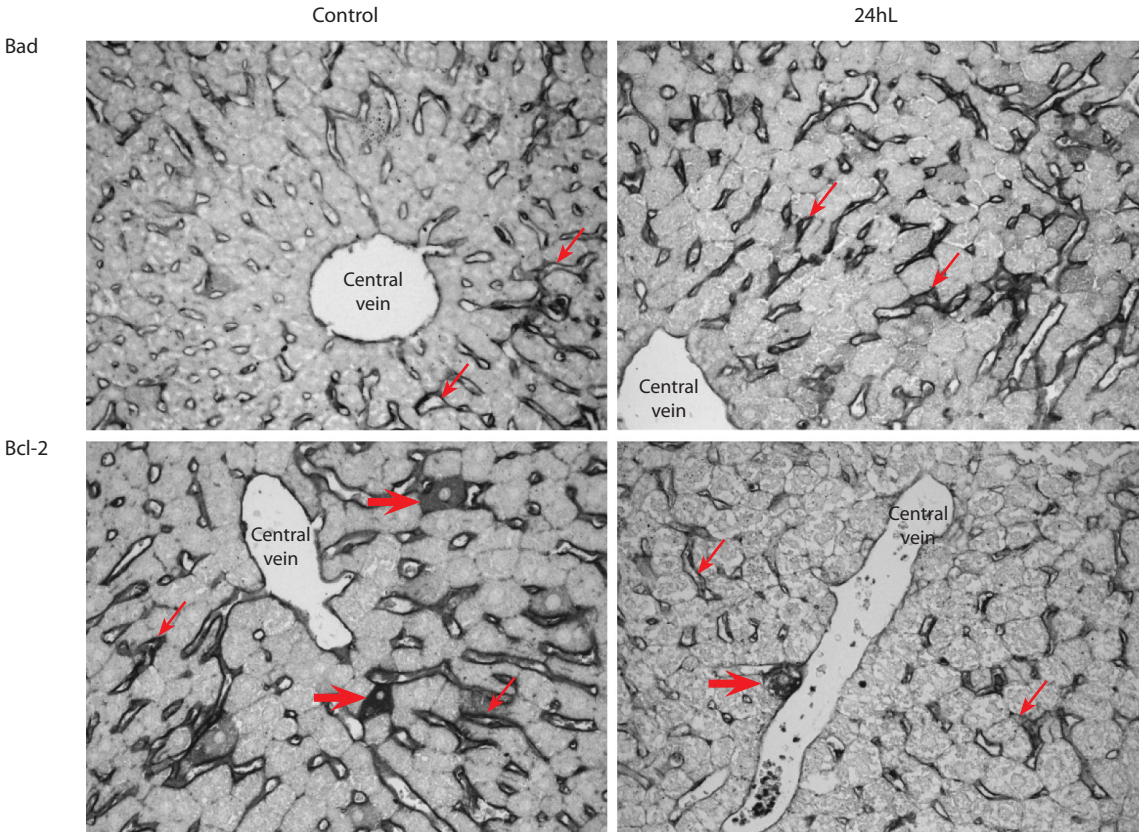


Fig. 1. The expression of the proapoptotic Bad protein and the antiapoptotic Bcl-2 protein in mouse liver cells in a light-induced functional pinealectomy model (the "24hL").

Immunohistochemical staining by the indirect ABC method. There is a pronounced Bad coloration and a less pronounced Bcl-2 coloration in sinusoidal cells (thin arrows) of blood sinusoidal capillaries in 24hL mouse liver. Thick arrows indicate separately found stained hepatocytes. Magnification ×400.

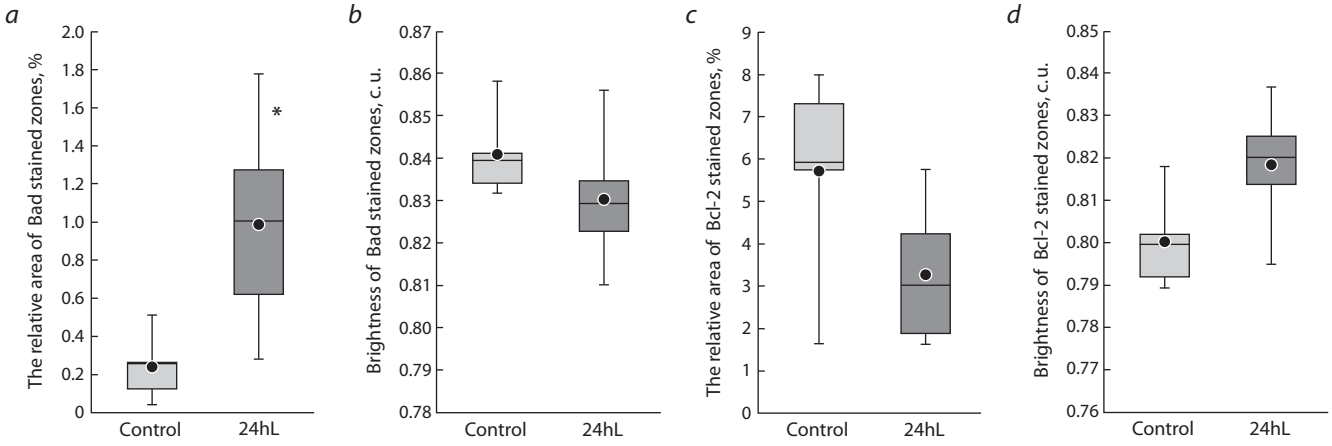


Fig. 2. Relative areas of Bad (*a*) and Bcl-2 (*c*) protein expressions and the brightnesses of zones stained with these proteins (*b* – Bad, *d* – Bcl-2) in the liver of "Control" and "24hL" mice.

Notations on the box diagrams: lines – median, boxes – 25–75 %, ● arithmetic mean, * differences are statistically significant between the "Control" and "24hL" groups; the Mann–Whitney U-test ($p < 0.05$).

concentration of this protein (see Fig. 2, *c*, *d*) in the liver of mice kept under 24-hour lighting.

Thus, it can be concluded that the antiapoptotic protection was weakened and the conditions for apoptosis mitochondrial pathway activation in liver cells of animals with light-induced functional pinealectomy were created.

Melatonin effect on the expression of Bad and Bcl-2 proteins in mouse liver cells under light-induced functional pinealectomy

MT treatment of the 24hL mice led to the pronounced Bcl-2 protein expression in a heterogeneous population of sinusoidal cells in intra-lobular blood liver capillaries and in single hepatocytes compared to the group without hormone treatment (the "24hL+Placebo" group) (Fig. 3). The immunohistochemical reaction to the Bad protein revealed in all three groups ("Placebo", "24hL+Placebo", "24hL+MT") the staining of sinusoidal capillary lining in intermediate zones and portal tracts, portal vein endothelium and bile duct epithelium in portal tracts (Fig. 4). Bad-staining was more significant in the "24hL+Placebo" group compared to "Placebo". Bad expression after MT administration wasn't as pronounced as Bcl-2 expression (see Fig. 3) in the same animals.

Morphometric analysis found a 3.3-fold increase in Bcl-2 expression area in 24hL-animals treated with MT compared with the group without treatment "24hL +Placebo" (Fig. 5, *a*). At the same time, the studied parameter reached the initial level of the "Placebo" group. The use of MT also led to a significant decrease in brightness (see Fig. 5, *b*) of stained areas compared with the comparison groups (by 2.7 % – compared with the "24hL+Placebo", by 2.1 % – compared with the "Placebo"), which reflects an increase in the Bcl-2 concentration in the "24hL+MT" animals. MT intragastric administration contributed to a tendency for an increase in the Bad relative area and a tendency for a decrease in the stained zone brightness compared to animals without hormone treatment. As a result, the use of MT led to a significant increase in the area and concentration of the studied protein compared to the "Placebo" group (see Fig. 5, *c*, *d*).

Thus, MT administration to mice under two-week 24-hour lighting led to a significant increase in the expression area and concentration of the Bcl-2 protein in liver cells against the background of unchanged expression area and concentration of the Bad protein compared to the "24hL+Placebo" group. The obtained results indicate that intragastric administration of MT physiological doses to C57Bl/6 mice cancels the effect of light-induced functional pinealectomy, restoring the expression area of the antiapoptotic Bcl-2 protein and increasing its concentration in liver cells, which indicates increased antiapoptotic protection of organ cells and creates conditions for blocking the apoptosis "mitochondrial branch" development.

Discussion

Violation of melatonin production is a starting point, leading at the initial stages to the appearance of desynchronosis followed by the development of organic pathology. Our previous studies showed that 24-hour lighting for two weeks has a modulating effect on all elements of the lymphatic region of the liver. There is a migration of lymphocytes, macrophages into the expanded interstitial non-vascular pathways and lymphatic vessels, and a formation of lymphoid nodules, which are considered temporary accumulations of lymphoid tissue that form in response to injury. The unbalancing of the roots of the lymphatic system leads to the disconnection of contacts between the endothelial cells of the liver sinusoids, as well as to a violation of contacts between the parenchymal cells of the organ. The overflow of Disse spaces with fragments of necrotically altered cells, collagen fibers, lymphoid cells, erythrocytes contributes to the lymph stagnation, and as a result leads to the development of tissue hypoxia, which is an inducer of cell death. This adversely affects the structure and functions of mitochondria, the protein-synthesizing apparatus of cells, causes stress in the endoplasmic reticulum (Ishchenko, Michurina, 2014; Michurina et al., 2018). Under these conditions, a significant burden falls on the intracellular detoxification systems, in particular on the cytochrome P450 system (Woolbright, Jaeschke, 2015). The enzymes of this family can produce reactive oxygen species (ROS), leading to

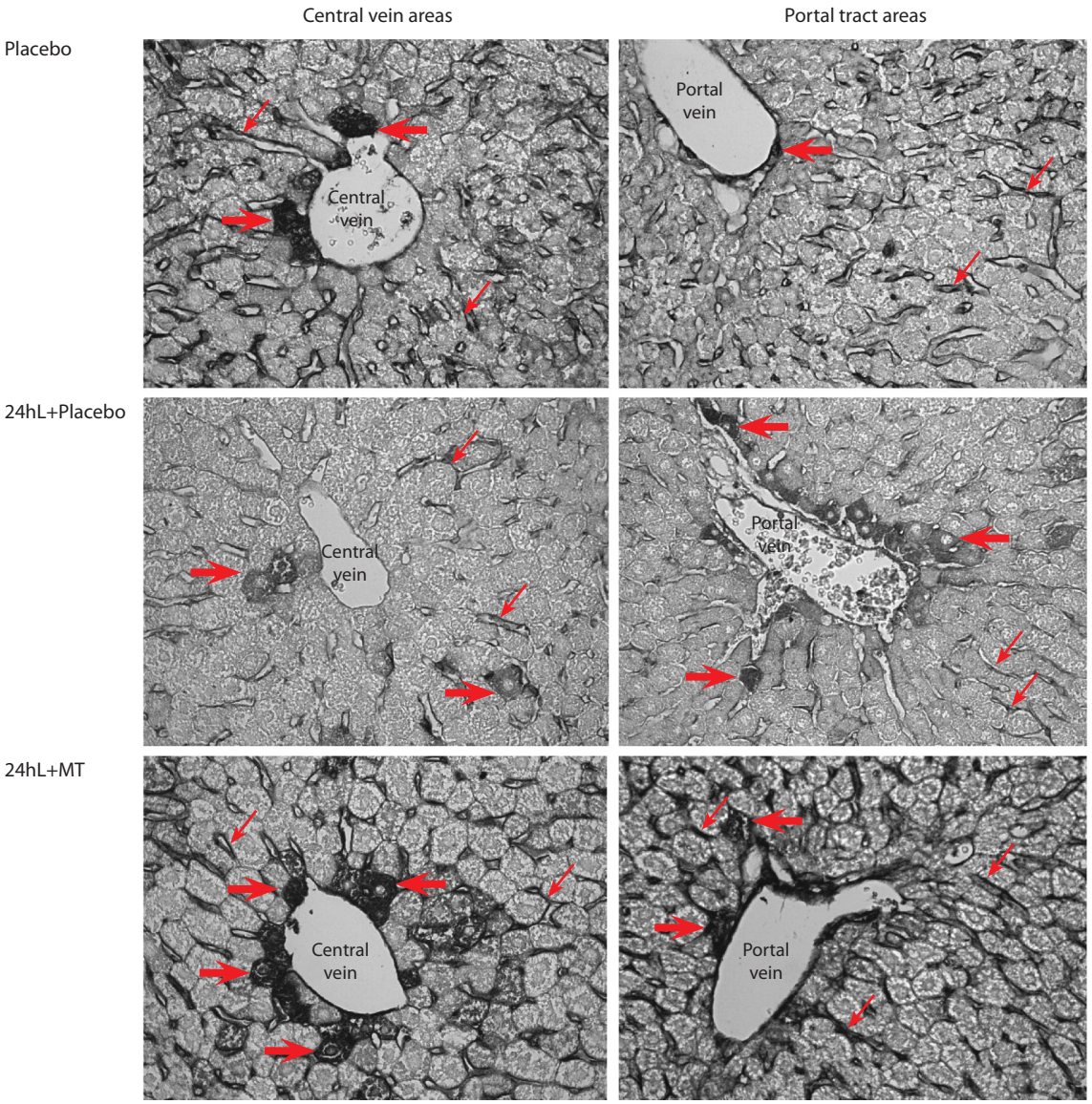


Fig. 3. The MT treatment influence on the Bcl-2 expression in liver cells of mice with the LIFP model.

Immunohistochemical staining by the indirect ABC method. There is weak Bcl-2-signal in the sinusoidal cells of the liver blood capillaries in "24hL+Placebo" mice compared with "Placebo" animals. MT treatment leads to a pronounced staining of the lining of the blood sinusoidal capillaries, endothelial cells of the portal tract veins, and individual hepatocytes. Thin arrows – the sinusoidal cells, thick arrows – single stained hepatocytes. Magnification ×400.

the activation of apoptosis. Excessive and uncontrolled ROS production in mitochondria leads to damage to mitochondrial membranes, proteins, and mitochondrial DNA (mtDNA) and triggers the mitochondrial apoptosis pathway (Li et al., 2020).

In our study, the greatest changes were found in sinusoidal cells of hepatic lobule blood capillaries. This is consistent with the data of Motoyama S. et al. (2000, 2003), who showed the predominant apoptosis development in liver sinusoidal endothelial cells compared to hepatocytes in male Sprague-Dawley rats with a hypoxia model. Currently, it has been proven that these cells, dynamically regulating the expression of angiopoietin-2, govern their own regeneration, and not only control the proliferation of hepatocytes, but also support the restoration of connective tissue, regulate the maturation and resting state of blood vessels (Hu et al., 2014). Since apoptosis is triggered by the inactivation of Bcl-2 when binding to the Bad protein, the fourfold increase revealed by us in the expression area of the proapoptotic protein Bad against the background of the unchanged expression area of the antiapoptotic protein Bcl-2 in mice with LIFP model indicates a decrease in antiapoptotic protection and the apoptosis development along the mitochondrial pathway in liver cells.

It's found that when melatonin synthesis is disrupted by night lighting, there is a decrease in the activity of its MT1 and MT2 membrane receptors, through which the hormone has its effect on cells (Gupta, Haldar, 2014; Jockers et al., 2016). Due to the non-receptor mechanism using the oligopeptide transporter-1/2 (PEPT-1/2) and organic anion transporter-3 (OAT-3) (Huo et al., 2017) MT penetrates cells and binds free oxygen radicals, protecting macromolecules (proteins, fats, nuclear and mitochondrial DNA) from oxidative damage in all subcellular structures. Currently, numerous data indicate

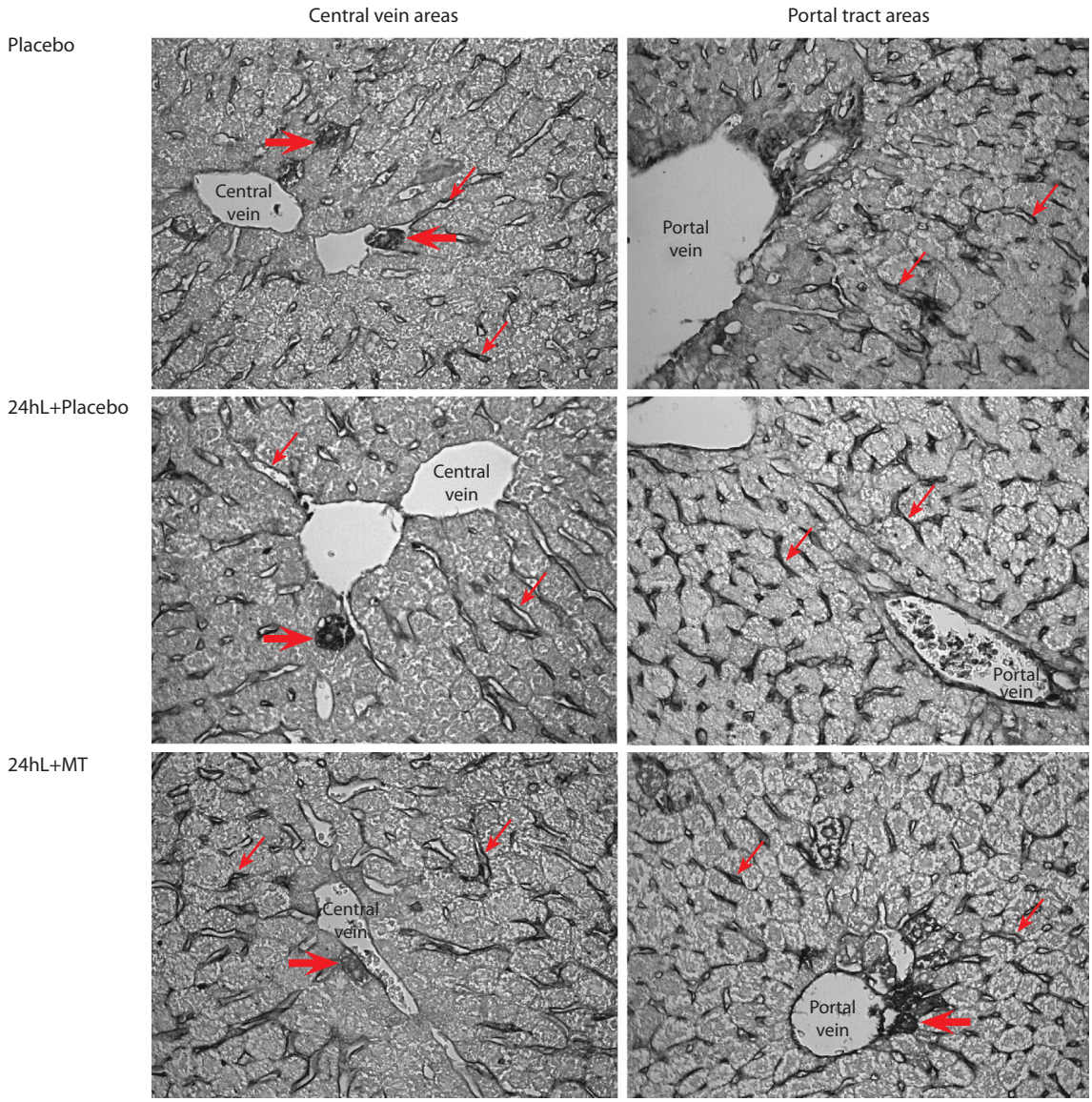


Fig. 4. The expression of the proapoptotic Bad protein in liver cells of mice with the LIFP model – in the sinusoidal cells of the liver blood capillaries, the vein endothelium and the ductal epithelium of the bile ducts of the liver portal tracts.

Immunohistochemical staining by the indirect ABC method. Bad-staining was more significant in the "24hL+Placebo" group compared to "Placebo". The Bad expression after MT administration wasn't as pronounced as Bcl-2 expression (see Fig. 3) in the same animals. Thin arrows – the sinusoidal cells, thick arrows – single stained hepatocytes. Magnification ×400.

that mitochondria is the main target of MT action: enzymes N-acetyltransferase and hydroxyindole-O-methyltransferase are present in mitochondria and these important subcellular organelles are the place of synthesis of melatonin itself (Hardeland, 2017; Reiter et al., 2018).

There are numerous ways in which MT destroys ROS: starting an antioxidant cascade with the formation of melatonin metabolites detoxifying free radicals; chelating metal ions involved in the Haber–Weiss and Fenton reactions to prevent the formation of a destructive •OH; stimulating antioxidant and inhibition of pro-oxidant enzymes; increasing the efficiency of electron transfer between mitochondrial respiratory complexes and reducing electron leakage and free radical formation. Studies have shown that MT reduces the rate of apoptosis, prevents the opening of mitochondrial pores and the release of cytochrome *c*, and preserves mitochondrial functions. In addition, mitochondrial biogenesis and dynamics are also regulated by MT (Hardeland, 2017; Reiter et al., 2018; Jou et al., 2019). The effectiveness of MT as a means of protection against oxidative stress and structural changes in the liver and pancreas tissue was revealed in rats with surgical pinealectomy (Sahna et al., 2004; Col et al., 2010). There is strong evidence that MT has the ability to prevent oxidative damage to liver cell mitochondria in rats with diabetes and obesity (Agil et al., 2015). The question of the effect of this unique hormone on apoptosis is extremely interesting. MT treatment of rats kept under 24-hour lighting during two weeks leads to an increase in the antiapoptotic Bcl-2 protein in the liver (Borodin et al., 2012).

Our use of the melatonin-containing complex in the treatment of animals with a model of obesity and type 2 diabetes mellitus showed its pronounced hepatotropic, lymphotropic

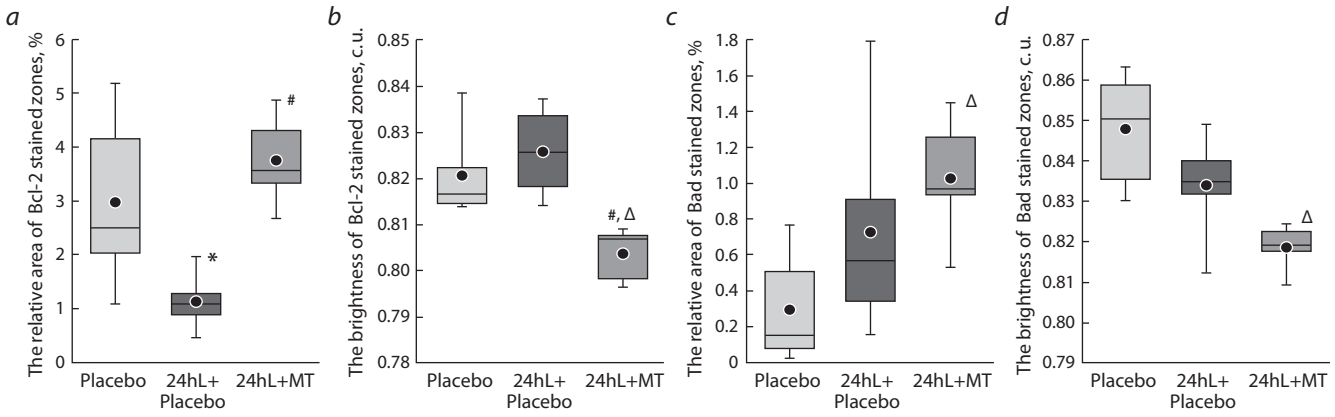


Fig. 5. Relative areas of Bcl-2 (*a*) and Bad (*c*) protein expressions and the brightnesses of zones stained for these proteins (*b* – Bcl-2, *d* – Bad) in the liver of the "Placebo", "24hL+Placebo" and "24hL+MT" mice.

Notations on the box diagrams: lines – median, boxes – 25–75 %, ● arithmetic mean, * differences are statistically significant between the "24hL+Placebo" and "Placebo" groups, # differences are statistically significant between the "24hL+MT" and "24hL+Placebo" groups, Δ the differences are statistically significant between the "24hL+MT" and "Placebo" groups; the Mann–Whitney U-test ($p < 0.05$).

action and cytoprotective effect, which consists in stimulating the expression of the antiapoptotic Bcl-2 protein in liver cells against the background of a decrease in the proapoptotic Bad protein activity (Michurina et al., 2017, 2020). In the present study the revealed predominance of the antiapoptotic Bcl-2 protein over the proapoptotic Bad protein, induced by the use of MT, indicates an increase in the antiapoptotic protection of liver cells, which blocks the development of the apoptosis "mitochondrial branch". This is facilitated by the previously established ability of MT to increase the expression of the lymphatic vascular endothelial LYVE-1 marker in the liver sinusoid endothelial cells of *db*/*db* mice, which creates conditions for improving lymph drainage and prevents the development of tissue hypoxia and apoptosis of organ cells (Michurina et al., 2016). The protective properties of MT, largely based on its antioxidant, antiapoptotic, and immunomodulatory activity, place this hormone among the most effective lympho- and angioprotectors (Jing et al., 2017; Chen et al., 2020), which is especially important in the prevention and treatment of new coronavirus infection (Darenskaya et al., 2020; El-Missiry et al., 2020). Thus, melatonin cytoprotective effect revealed by us in the liver cells of C57Bl/6 mice in the model of light-induced functional pinealectomy may be a consequence of reduced damage to mitochondria and other intracellular structures.

Conclusion

Thus our results indicate a weakening of the antiapoptotic protection of liver cells of LIFP animals that creates conditions for activation of the "mitochondrial branch" of apoptosis. Melatonin treatment of animals cancels the effect of LIFP, restoring the Bcl-2 expression area and increasing this protein concentration, which indicates an increase in antiapoptotic protection and creates conditions for blocking the development of the "mitochondrial branch" of apoptosis in liver cells.

References

Agil A., El-Hammadi M., Jiménez-Aranda A., Tassi M., Abdo W., Fernández-Vázquez G., Reiter R.J. Melatonin reduces hepatic mitochondrial dysfunction in diabetic obese rats. *J. Pineal. Res*. 2015; 59(1):70-79. DOI 10.1111/jpi.12241.

Arendt J. Melatonin: countering chaotic time cues. *Front. Endocrinol.* (*Lausanne*). 2019;10:391. DOI 10.3389/fendo.2019.00391.

Borodin Yu.I., Trufakin V.A., Michurina S.V., Shurlygina A.V. Structural and Temporal Organization of the Liver, Lymphatic, Immune, and Endocrine Systems in Violation of the Light Regime and Melatonin Treatement. Novosibirsk: Manuscript Publ., 2012. (in Russian)

Chen W.R., Yang J.Q., Liu F., Shen X.Q., Zhou Y.J. Melatonin attenuates vascular calcification by activating autophagy via an AMPK/mTOR/ULK1 signaling pathway. *Exp. Cell. Res.* 2020;389(1):111883. DOI 10.1016/j.yexcr.2020.111883.

Col C., Dinler K., Hasdemir O., Buyukasik O., Bugdayci G. Oxidative stress and lipid peroxidation products: effect of pinealectomy or exogenous melatonin injections on biomarkers of tissue damage during acute pancreatitis. *Hepatobiliary Pancreat. Dis. Int*. 2010; 9(1):78-82. PMID: 20133234.

Darenskaya M.A., Kolesnikova L.I., Kolesnikov S.I. COVID-19: oxidative stress and the relevance of antioxidant therapy. *Vestnik Rossijskoj Akademii Meditsynskikh Nauk = Annals of the Russian Academy of Medical Sciences*. 2020;75(4):318-325. DOI 10.15690/vramn1360. (in Russian)

Delibas N., Tuzmen N., Yonden Z., Altuntas I. Effect of functional pinealectomy on hippocampal lipid peroxidation, antioxidant enzymes and N-methyl-D-aspartate receptor subunits 2A and 2B in young and old rats. *Neuro Endocrinol. Lett*. 2002;23(4):345-350. PMID: 12195239.

El-Missiry M.A., El-Missiry Z.M.A., Othman A.I. Melatonin is a potential adjuvant to improve clinical outcomes in individuals with obesity and diabetes with coexistence of Covid-19. *Eur. J. Pharmacol*. 2020;882:173329. DOI 10.1016/j.ejphar.2020.173329.

Gupta S., Haldar C. Nycthemeral variation in melatonin receptor expression in the lymphoid organs of a tropical seasonal breeder *Funambulus pennanti*. *J. Comp. Physiol. A*. 2014;200(12):1045-1055. DOI 10.1007/s00359-014-0959-2.

Hardeland R. Melatonin and the electron transport chain. *Cell. Mol. Life Sci.* 2017;74(21):3883-3896. DOI 10.1007/s00018-017-2615-9.

Hu J., Srivastava K., Wieland M., Runge A., Mogler C., Besemfelder E., Terhardt D., Vogel M.J., Cao L., Korn C., Bartels S., Thomas M., Augustin H.G. Endothelial cell-derived angiopoietin-2 controls liver regeneration as a spatiotemporal rheostat. *Science*. 2014; 343(6169):416-419. DOI 10.1126/science.1244880.

Huo X., Wang C., Yu Z., Peng Y., Wang S., Feng S., Zhang S., Tian X., Sun C., Liu K., Deng S., Ma X. Human transporters, PEPT1/2, facilitate melatonin transportation into mitochondria of cancer cells: an

implication of the therapeutic potential. *J. Pineal Res.* 2017;62(4): e12390. DOI 10.1111/jpi.12390.

Ishchenko I.Y., Michurina S.V. Regional lymph nodes in the liver of rats in functional pinealectomy. *Bull. Exp. Biol. Med*. 2014;157(5): 671-676. DOI 10.1007/s10517-014-2636-4.

Jing Y., Bai F., Chen H., Dong H. Melatonin prevents blood vessel loss and neurological impairment induced by spinal cord injury in rats. *J. Spinal. Cord. Med*. 2017;40(2):222-229. DOI 10.1080/10790268.2016.1227912.

Jockers R., Delagrange P., Dubocovich M.L., Markus R.P., Renault N., Tosini G., Cecon E., Zlotos D.P. Update on melatonin receptors: IUPHAR Review 20. *Br. J. Pharmacol.* 2016;173(18):2702-2725. DOI 10.1111/bph.13536.

Jou M.J., Peng T.I., Reiter R.J. Protective stabilization of mitochondrial permeability transition and mitochondrial oxidation during mitochondrial Ca^{2+} stress by melatonin's cascade metabolites C3-OHM and AFMK in RBA1 astrocytes. *J. Pineal Res.* 2019;66(1):e12538. DOI 10.1111/jpi.12538.

Li R., Toan S., Zhou H. Role of mitochondrial quality control in the pathogenesis of nonalcoholic fatty liver disease. *Aging* (*Albany NY*). 2020;12(7):6467-6485. DOI 10.18632/aging.102972.

Michurina S.V., Ishchenko I.Yu., Arkhipov S.A., Cherepanova M.A., Vasendin D.V., Zavjalov E.L. Apoptosis in the liver of male *db/db* mice during the development of obesity and type 2 diabetes. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2020;24(4):435-440. DOI 10.18699/VJ20.43-o.

Michurina S.V., Ischenko I.Yu., Arkhipov S.A., Klimontov V.V., Cherepanova M.A., Korolev M.A., Rachkovskaya L.N., Zav'yalov E.L., Konenkov V.I. Melatonin–aluminum oxide–polymethylsiloxane complex on apoptosis of liver cells in a model of obesity and type 2 diabetes mellitus. *Bull. Exp. Biol. Med*. 2017;164(2):165-169. DOI 10.1007/s10517-017-3949-x.

Michurina S.V., Ishchenko I.Yu., Arkhipov S.A., Klimontov V.V., Rachkovskaya L.N., Konenkov V.I., Zavyalov E.L. Effects of melatonin, aluminum oxide, and polymethylsiloxane complex on the expression of LYVE-1 in the liver of mice with obesity and type 2 diabetes mellitus. *Bull. Exp. Biol. Med*. 2016;162(2):269-272. DOI 10.1007/s10517-016-3592-y.

Michurina S.V., Shurlygina A.V., Belkin A.D., Vakulin G.M., Verbitskaia L.V., Trufakin V.A. Changes in liver and in some organs of immune system of animals exposed to twenty-four-hour illumination. *Morfologiia*. 2005;128(4):65-68. PMID: 16400925. (in Russian)

Michurina S.V., Vasendin D.V., Ishchenko I.Yu. Physiological and biological effects of melatonin: some results and prospects for the study. *Rossiyskiy Fiziologicheskiy Zhurnal im. I.M. Sechenova = I.M. Sechenov Physiological Journal*. 2018;104(3):257-271. (in Russian)

Morin L.P. Nocturnal light and nocturnal rodents: similar regulation of disparate functions? *J. Biol. Rhythms*. 2013;28(2):95-106. DOI 10.1177/0748730413481921.

Motoyama S., Saito S., Alojado M.E., Itoh H., Kitamura M., Suzuki H., Saito R., Momiyama H., Nakae H., Ogawa J., Inaba H. Hydrogen peroxide induces midzonal heat shock protein 72 and apoptosis in sinusoidal endothelial cells of hypoxic rat liver. *Crit. Care Med.* 2000; 28(5):1509-1514. DOI 10.1097/00003246-200005000-00042.

Motoyama S., Saito S., Saito R., Minamiya Y., Nakamura M., Okuyama M., Imano H., Ogawa J. Hydrogen peroxide-dependent declines in Bcl-2 induces apoptosis in hypoxic liver. *J. Surg. Res*. 2003; 110(1):211-216. DOI 10.1016/s0022-4804(03)00006-4.

Polčic P., Mentel M. Reconstituting the mammalian apoptotic switch in yeast. *Genes* (*Basel*). 2020;11(2):145. DOI 10.3390/genes110 20145.

Reiter R.J., Rosales-Corral S.A., Tan D.X., Alatorre-Jimenez M., Lopez C. Circadian dysregulation and melatonin rhythm suppression in the context of aging. In: Jazwinski S., Belancio V., Hill S. (Eds). Circadian Rhythms and Their Impact on Aging. (Ser. Healthy Ageing and Longevity. Vol. 7). Springer, Cham, 2017;1-25. DOI 10.1007/978-3-319-64543-8_1.

Reiter R.J., Tan D.X., Rosales-Corral S., Galano A., Jou M.J., Acuna-Castroviejo D. Melatonin mitigates mitochondrial meltdown: interactions with SIRT3. *Int. J. Mol. Sci.* 2018;19(8):2439. DOI 10.3390/ijms19082439.

Reiter R.J., Tan D.X., Rosales-Corral S., Manchester L.C. The universal nature, unequal distribution and antioxidant functions of melatonin and its derivatives. *Mini-Rev. Med. Chem.* 2013;13(3):373-384. DOI 10.2174/1389557511313030006.

Russart K.L.G., Nelson R.J. Light at night as an environmental endocrine disruptor. *Physiol. Behav*. 2018;190:82-89. DOI 10.1016/j.physbeh.2017.08.029.

Sahna E., Parlakpinar H., Vardi N., Ciğremis Y., Acet A. Efficacy of melatonin as protectant against oxidative stress and structural changes in liver tissue in pinealectomized rats. *Acta Histochem*. 2004;106(5):331-336. DOI 10.1016/j.acthis.2004.07.006.

Willis S., Day C.L., Hinds M.G., Huang D.C. The Bcl-2-regulated apoptotic pathway. *J. Cell. Sci*. 2003;116(Pt.20):4053-4056. DOI 10.1242/jcs.00754.

Woolbright B.L., Jaeschke H. Xenobiotic and endobiotic mediated interactions between the cytochrome P450 system and the inflammatory response in the liver. *Adv. Pharmacol.* 2015;74:131-161. DOI 10.1016/bs.apha.2015.04.001.

ORCID ID

I.Yu. Ishchenko orcid.org/0000-0001-6281-0402
S.A. Arkhipov orcid.org/0000-0002-1390-4426
A.Yu. Letyagin orcid.org/0000-0002-9293-4083
E.L. Zavjalov orcid.org/0000-0002-9412-3874

Acknowledgements. The work was carried out within the framework of the budget project No. 0324-2019-0046. The study was performed using the equipment of the Center for genetic resources of laboratory animals of the Institute of Cytology and Genetics SB RAS, supported by the Ministry of Education and Science of Russia (Unique identifier of the project RFMEFI62119X0023).

Conflict of interest. The authors have no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Received December 17, 2020. Revised March 29, 2021. Accepted March 30, 2021.

Вавиловский журнал генетики и селекции. 2021;25(3):318-330
DOI 10.18699/VJ21.035

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

The finding and researching algorithm for potentially oscillating enzymatic systems

T.N. Lakhova[1] ✉, F.V. Kazantsev[1], S.A. Lashin[1, 3], Yu.G. Matushkin[2, 3]

[1] Kurchatov Genomics Center of ICG SB RAS, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
✉ tlakhova@bionet.nsc.ru

Abstract. Many processes in living organisms are subject to periodic oscillations at different hierarchical levels of their organization: from molecular-genetic to population and ecological. Oscillatory processes are responsible for cell cycles in both prokaryotes and eukaryotes, for circadian rhythms, for synchronous coupling of respiration with cardiac contractions, etc. Fluctuations in the numbers of organisms in natural populations can be caused by the populations' own properties, their age structure, and ecological relationships with other species. Along with experimental approaches, mathematical and computer modeling is widely used to study oscillating biological systems. This paper presents classical mathematical models that describe oscillatory behavior in biological systems. Methods for the search for oscillatory molecular-genetic systems are presented by the example of their special case – oscillatory enzymatic systems. Factors influencing the cyclic dynamics in living systems, typical not only of the molecular-genetic level, but of higher levels of organization as well, are considered. Application of different ways to describe gene networks for modeling oscillatory molecular-genetic systems is considered, where the most important factor for the emergence of cyclic behavior is the presence of feedback. Techniques for finding potentially oscillatory enzymatic systems are presented. Using the method described in the article, we present and analyze, in a step-by-step manner, first the structural models (graphs) of gene networks and then the reconstruction of the mathematical models and computational experiments with them. Structural models are ideally suited for the tasks of an automatic search for potential oscillating contours (linked subgraphs), whose structure can correspond to the mathematical model of the molecular-genetic system that demonstrates oscillatory behavior in dynamics. At the same time, it is the numerical study of mathematical models for the selected contours that makes it possible to confirm the presence of stable limit cycles in them. As an example of application of the technology, a network of 300 metabolic reactions of the bacterium *Escherichia coli* was analyzed using mathematical and computer modeling tools. In particular, oscillatory behavior was shown for a loop whose reactions are part of the tryptophan biosynthesis pathway.
Key words: oscillations; feedback; cyclic processes; modelling of biological systems.

For citation: Lakhova T.N., Kazantsev F.V., Lashin S.A., Matushkin Yu.G. The finding and researching algorithm for potentially oscillating enzymatic systems. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):318-330. DOI 10.18699/VJ21.035

Технологии поиска и исследования потенциально осциллирующих ферментативных систем

Т.Н. Лахова[1] ✉, Ф.В. Казанцев[1], С.А. Лашин[1, 3], Ю.Г. Матушкин[2, 3]

[1] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ tlakhova@bionet.nsc.ru

Аннотация. Многие процессы в живых организмах подвержены периодическим колебаниям на различных иерархических уровнях их организации: от молекулярного-генетического до популяционного и экологического. Осциллирующие процессы отвечают за клеточные циклы как у прокариот, так и у эукариот, за циркадные ритмы, синхронную связь дыхания с сердечными сокращениями и др. Колебания численностей организмов в природных популяциях могут быть обусловлены собственными свойствами популяций, их возрастной структурой, а также экологическими взаимоотношениями с другими видами. Наряду с экспериментальными подходами, для исследования осциллирующих биологических систем широко применяется

математическое и компьютерное моделирование. В данной статье представлены классические математические модели, которые описывают осциллирующее поведение в биологических системах. Приведены методы поиска осциллирующих молекулярно-генетических систем на примере их частного случая – осциллирующих ферментативных систем. Рассмотрены факторы, влияющие на циклическую динамику в живых системах, характерные не только для молекулярно-генетического уровня, но и для более высоких уровней организации. Обсуждается применение различных способов описания генных сетей для моделирования осциллирующих молекулярно-генетических систем, где важнейшим фактором возникновения циклического поведения является наличие обратных связей. Представлены технологии поиска потенциально осциллирующих ферментативных систем. С помощью метода, описанного в статье, проводится поэтапный процесс построения и анализа сначала структурных моделей (графов) генных сетей, а затем реконструкции математических моделей и вычислительных экспериментов с ними. Структурные модели идеально подходят для задач автоматического поиска потенциальных осциллирующих контуров (связных подграфов), структура которых может соответствовать математической модели молекулярно-генетической системы, демонстрирующей осциллирующее поведение в динамике. При этом именно численное исследование математических моделей для отобранных контуров позволяет подтвердить наличие в них устойчивых предельных циклов. В качестве примера применения технологии проанализирована сеть из 300 метаболических реакций бактерии *Escherichia coli* с использованием инструментов математического и компьютерного моделирования. В частности, показано осциллирующее поведение для контура, реакции которого входят в путь биосинтеза триптофана.

Ключевые слова: осцилляции; обратная связь; циклические процессы; моделирование биологических систем.

Introduction

Many processes in living organisms are subject to periodic oscillations at different hierarchical levels of their organization: from the molecular-genetic to the population and ecological levels. For example, at the molecular-genetic level, there are oscillations in the concentrations of p53, a protein involved in apoptosis or cell cycle delay in DNA damage, and its inhibitor Mdm2 (Prives, 1998; Lahav et al., 2004). There are also fluctuations in concentrations of hormones in the cell, such as melatonin (Boccalandro et al., 2011), prolactin, total cholesterol (Garde et al., 2000), etc.; concentrations of low molecular weight compounds, such as intracellular and intercellular calcium ion concentrations, can also oscillate (Pasti et al., 1997; Allen et al., 2000).

One well-known example of organism-wide periodic processes is circadian rhythms, for the work on which the 2017 Nobel Prize in Physiology or Medicine was awarded (Young et al., 1984; Siwicki et al., 1988; Hardin et al., 1990; Price et al., 1998). Jeffrey C. Hall, Michael Rosbash, and Michael W. Young discovered the *period* gene in *Drosophila melanogaster*, which is regulated through feedback by the PER protein underlying circadian rhythm.

In the article (Podkolodnyy et al., 2017), the authors considered genes located in liver and kidney cells that are overexpressed with a certain periodicity during the 24-hour cycle. In a subsequent paper, the authors provided an overview of various mathematical models used to model the autonomous circadian clock in mammalian cells (Podkolodnaya et al., 2017).

At the cellular level, cyclic processes can include cell cycles in both prokaryotes and eukaryotes (Cooper, 1991). Such important cyclic processes as heartbeat (Ashkenazy et al., 2001), respiration, as well as synchronous relationship between respiration and heartbeat (Yasuma, Hayano, 2004), photosynthesis (Holtum, Winter, 2003) and other similar processes occur at the level of an individual organ or functional systems of an organism.

Population waves (Chetverikov, 2009) are a classic example of cyclic processes at the population level of organization. Fluctuations in the number of organisms in natural populations can be caused both by external environmental factors and by the population's own properties, its age structure, and ecological relationships with other species. A natural factor such as seasonal periodicity plays an important role in the cyclic processes of the population level, influencing the migration of birds, falling into anabiosis, the appearance and fall of leaves, etc.

Thus, in the article (Erdakov, Moroldoev, 2017), the authors considered the cyclicity in the population dynamics of the red vole, which varies depending on the geographical habitat and external conditions in the area. And in the paper (Pertsev, Loginov, 2011), using a stochastic model, the authors considered how the population size changes when harmful food resources are consumed. The investigation of population dynamics, often cyclical, is one of the most studied processes, both by empirical methods and with the help of mathematical methods, including modeling (Volterra, 1928; Bazykin, 2003; Riznichenko, 2017).

Finally, biogeochemical cycles, i. e., processes of dynamic exchange of chemicals between organisms from prokaryotes to higher animals and plants and elements of the biosphere (soil, water and air) can be classified as cyclic processes at the ecological level (Van Cappellen, 2003; Zavarzin, 2003, 2011; Struyf et al., 2009).

Cyclical processes in biology are investigated using experimental and theoretical methods. Mathematical modeling is one of the main methods for their investigation, particularly in finding areas of stationary, oscillatory, and possibly chaotic behavior (Romanovsky et al., 1975; Schnol, 1996; Becks, Arndt, 2013).

The first works devoted to oscillatory biochemical processes belong to Alfred Lotka (Lotka, 1910). Lotka described the dynamics of biochemical processes using systems of nonlinear ordinary differential equations. Around the same time, Vito Volterra, independently of Lotka, developed the same models, but in application to population-ecological problems. These models were later called the "Lotka–Volterra models". Further study of oscillatory chemical processes led to the discovery of Belousov–Zhabotinsky type systems, in which oscillations occur not only in time but also in space, and, therefore, can be described not only by ordinary differential equations, but also by partial differential equations (Zhabotinsky, 1974; Field, Burger, 1988; Mushtakova, 1997; Shnol, 2009).

This article presents a review of classical mathematical models that describe oscillatory behavior in biological systems and gives illustrations of methods for finding such systems using enzymatic oscillatory systems as an example. The role of gene networks in modeling oscillatory molecular-genetic systems is discussed. The factors influencing the presence or absence of oscillatory behavior in various molecular-genetic systems are given.

Classical models and methods for modeling oscillatory processes

Among the first mathematical approaches describing oscillatory processes are models that have already become classical in the field of mathematical biology (Riznichenko, 2002). In one of the papers devoted to the theory of periodic reactions, Lotka studied a chemical reaction of the form:

$$A \to X \to Y \to B, \tag{1}$$

where $X \to Y$ is an autocatalytic process. Based on the law of mass action, Lotka described this reaction by the following differential equations (Lotka, 1910):

$$\begin{aligned} \frac{dx}{dt} &= k_0 - k_1 xy, \\ \frac{dy}{dt} &= k_1 xy - k_2 y, \end{aligned} \tag{2}$$

where k_0, k_1, k_2 are the constant parameters and x, y are the concentrations of chemicals.

The following model, described by Lotka (Lotka, 1920) and then independently formulated by Volterra (Volterra, 1928), expresses two autocatalytic reactions (i. e. $A \to X$ and $X \to Y$). The Lotka–Volterra model has the following form:

$$\begin{aligned} \frac{dx}{dt} &= ax - bxy, \\ \frac{dy}{dt} &= cxy - dy, \end{aligned} \tag{3}$$

where a, b, c, d are the rates of transformation of some substances into others, x, y are the concentrations of chemicals. This model is also known as the "predator–prey system", which is used in population dynamics to explain periodic fluctuations in the abundance of individuals in populations.

In the same period a paper with the van der Pol and van der Mark oscillator model was published (van der Pol, van der Mark, 1928). They modeled the heart as three connected relaxation systems: sinus, atrium and ventricle. As such a system, the authors chose a system consisting of a neon lamp, a condenser, a resistance, and a battery, which is capable of producing relaxation oscillations. However, this system simulates only some modes of heart operation due to the complexity of the object under study. The model is described by an equation of the form:

$$\frac{d^2v}{dt^2} - \alpha(1 - v^2)\frac{dv}{dt} + \omega^2 v = 0, \tag{4}$$

where α is a positive value, which is an oscillator parameter (responsible for non-linearity and damping of oscillations), ω – oscillation frequency, v – the value corresponding to the heart rhythm signal.

This model is noteworthy because it has found an application not only in biology problems, but also in physics and other sciences. For example, the review (Kuznetsov et al., 2014) presented a number of problems in which this oscillator was applied; in particular, the authors gave details on modeling human body processes, such as colonic myoelectric activity and processes of excitation and inhibition of neurons. In the paper (Rompala et al., 2007), the authors considered three van der Pol oscillators to study the in-phase mode, which corresponds to the synchronized periodic behavior of circadian rhythms. Moreover, two of them correspond to the eye models, and the third oscillator is a model of the brain (mainly represented by the pineal gland), through which the interconnection of the first two is performed. They considered the periodic change of melatonin concentration under the influence of circadian rhythms as a possible scheme of connection between the eyes and the pineal gland.

In 1965, an article by Brian Goodwin (Goodwin, 1965) was published, which raised the question of the oscillatory motion role in the organization of cellular processes over time. For the mathematical study of oscillatory behavior in model systems involving enzyme regulation processes, he introduced certain concepts of thermodynamic nature. In the article, the author cited a model of the process of genetic control of enzyme synthesis:

$$\begin{aligned} \frac{dX_i}{dt} &= \frac{a_i}{A_i + k_i Y_i} - b_i, \\ \frac{dY_i}{dt} &= a_i X_i - b_i, \end{aligned} \tag{5}$$

where X_i is an mRNA concentration of the ith species, Y_i is a protein (repressor) concentration of the ith species, k_i – parameter, which describes the interactions between the DNA and the repressor.

Another classic example is the Higgins model (Higgins, 1964) of oscillatory reactions in the glycolysis system, the scheme of which is shown below:

$$\begin{aligned} &\text{GLU} \to \text{F6P}, \\ &\text{F6P} + E_1^* \to E_1^* \cdot \text{F6P}, \\ &E_1^* \cdot \text{F6P} \to E_1^* + \text{FDP}, \\ &\text{FDP} + E_1^+ \leftrightarrow E_1^*, \\ &\text{FDP} + E_2 \to E_2 \cdot \text{FDP}, \\ &E_2 \cdot \text{FDP} \to E_2 + \text{GAP}. \end{aligned} \tag{6}$$

Here GLU, F6P, FDP, GAP are designations of biochemical substances that enter into reactions, E_1^* – the active form of the enzyme (phosphofructokinase), E_1^+ – the inactive form of the enzyme, E_2 – the enzyme that is a combination of aldolase and triose phosphate isomerase.

Higgins considered general pathway types of enzymatic reactions in glycolysis in which the chemical mechanism exhibits oscillatory behavior. Therefore, in his work, he takes into account the following conditions: (1) one of the chemicals must activate its own production (assuming the concentration of the second substance is constant); (2) the second substance must tend to inactivate its own net production; (3) there must be a cross-coupling of the interaction of substances. If an increase in the first substance activates the production of the second substance, then an increase in the second substance inhibits the production of the first, and vice versa.

Sel'kov in his classic article (Sel'kov, 1968), in accordance with the mass action law, gave a mathematical model of the glycolytic system based on the phosphofructokinase (PFK) transformations:

$$\begin{aligned}
\frac{ds_1}{dt} &= v_1 - k_{+1}s_1x_1 + k_{-1}x_2,\\
\frac{ds_2}{dt} &= k_{+2}x_2 - k_{+3}s_2^{\gamma}e + k_{-3}x_1 - k_2s_2,\\
\frac{dx_1}{dt} &= -k_{+1}s_1x_1 + (k_{-1} + k_{+2})x_2 + k_{+3}s_2^{\gamma}e - k_{-3}x_1, \quad (7)\\
\frac{dx_2}{dt} &= k_{+1}s_1x_1 - (k_{-1} + k_{+2})x_2,\\
\frac{de}{dt} &= -k_{+3}s_2^{\gamma}e - k_{-3}x_1,
\end{aligned}$$

where s_1 – the substrate (ATP), v_1 – the inflow rate of the substrate from some source, s_2 – the product (ADP), $v_2 = k_2s_2$ – the outflow rate of the product from the system, e – free enzyme (phosphofructokinase), which is inactive on its own, but becomes active when combined with product molecules as a complex – ES_2^{γ}, x_1 – the molecule of the complex (ES_2^{γ}), x_2 – the molecule of enzyme-substrate complex ($S_1ES_2^{\gamma}$), s_2^{γ} – product molecules that enter into a complex with the free enzyme, $\gamma > 1$ – the parameter responsible for the number of the product molecules, k_{+1}, k_{+2}, k_{+3} – rates of direct reactions, k_{-1}, k_{-3} – rates of reverse reactions, t – time.

Goldbeter and Lefever (Goldbeter, Lefever, 1972) presented a model of the glycolytic system, which is a generalization of the models presented by Higgins (Higgins, 1964, 1967) and Sel'kov (Sel'kov, 1968). The model is based on the mechanism of positive feedback, namely, the activation of the product by the enzyme PFK.

In the article (Boiteux et al., 1975), the authors not only analyzed the allosteric model of the oscillatory reaction of phosphofructokinase, but also made experimental verification of theoretical predictions. The data obtained for the model agreed well with the experimental data.

In 2000, a model of a yeast population consisting of a small ensemble of individual cells was presented to describe the phenomenon of synchronization of glycolytic oscillations. In this case, the communication between the cells was performed through the exchange of acetaldehyde (Bier et al., 2000). Glycolytic oscillations were also studied using stochastic methods and chaos theory in (Bashkirtseva, Ryashko, 2017); Selkov's minimal model was taken as the basis, and in the article (Ryashko, 2018) a two-dimensional Higgins model was used.

In biochemistry, the processes of changing the concentration of ions in cells, which can increase or decrease the activity of enzymes, participate in the metabolism of carbohydrates, lipids and proteins, as well as play an important role in signal transduction through signaling pathways and are responsible for cell excitability, are actively studied. One of such processes is periodic changes in calcium ion concentrations. A number of mathematical models have been developed to study these periodic processes. A model describing calcium ion concentration fluctuations was first proposed in (Dupont, Goldbeter, 1989):

$$\begin{aligned}
\frac{dZ}{dt} &= v_0 + v_1\beta - v_2 + v_3 - kZ,\\
\frac{dY}{dt} &= v_2 - v_3,
\end{aligned} \quad (8)$$

where Z – the cytosolic calcium concentration, Y – the calcium concentration in IP_3 (inositol-1,4,5-triphosphate) endoplasmic reticulum, v_i ($i = 0, \ldots, 3$) – reaction rates.

They analyzed the conditions for the emergence of stable fluctuations based on the mechanism of calcium-induced calcium release (CICR). In a number of studies (Goldbeter et al., 1990; Dupont et al., 1991; Dupont, Goldbeter, 1993), the authors continued their researches of calcium concentration fluctuations based on the same minimal model.

At the same time, papers (Meyer, Stryer, 1988; Meyer, 1991) in which the authors investigated fluctuations in calcium concentrations by considering the mechanism of inositol cross-coupling (ICC) IP_3 with extracellular, cytosolic, and endoplasmic Ca^{2+} have been coming out. Lavrentovich and Hemkin (Lavrentovich, Hemkin, 2008) proposed a model for spontaneous Ca^{2+} oscillations in astrocytes that takes into account the mechanisms presented above as well as IP_3 production in a receptor-independent manner.

After Goldbeter and Dupont had published their results, the authors of the article (Kraus et al., 1996) tested the hypothesis that in unexcited cells the amplitudes of oscillatory processes can be cell type-specific and vary with Ca^{2+} diffusion. They performed their study using stochastic computer modeling on a two-dimensional Ca^{2+} oscillation model.

Analysis of oscillatory processes in living systems shows that the most important factor in the emergence of cyclic behavior is a feedback in the system (Kolchanov et al., 2000). A distinction is made between positive and negative feedbacks, which was once discussed by Goodwin, Walter, Cardon, Iberall and other researchers (Goodwin, 1965; Walter, 1969, 1970; Cardon, Iberall, 1970). Both types of these feedbacks can influence the emergence of cyclic dynamics in the system, as has been shown in works (Likhoshvai et al., 2001; Goldbeter, 2002; Tyson et al., 2003).

At the molecular level, the principle of feedback regulates a huge number of enzymatic reactions simultaneously going on in a living cell, the rate of which can be affected by such compounds as inhibitors, activators, cofactors, allosteric ef-

Table 1. Brief characteristics of a number of classical models with oscillatory behavior

Model name	Modeled biological process	Model class	Type and number of feedbacks
Lotka model (Lotka, 1910)	Biochemical reaction	Nonlinear inhomogeneous SODE with constant coefficients	Positive (1) and Negative (2)
Lotka–Volterra model ("predator–prey") (Lotka, 1920; Volterra, 1928)	Biochemical reaction; population dynamics	Nonlinear homogeneous SODE with constant coefficients	Positive (2) and Negative (2)
van der Pol oscillator (van der Pol, van der Mark, 1928)	Heart function, colonic myoelectric activity, excitation and inhibition of neurons, etc.	Nonlinear homogeneous second-order ODE – the Lienar equation, which can be reduced to a first-order ODE	Positive (1) and Negative (2)
Goodwin oscillator (Goodwin, 1965)	Genetic control of enzyme synthesis	Nonlinear inhomogeneous SODE with constant coefficients	Positive (1) and Negative (1)
Sel'kov model (Sel'kov, 1968)	Enzyme reaction	Linear inhomogeneous SODE with constant coefficients	Negative (substrate inhibition), Positive (product activation). Negative (9), Positive (7)
Dupont–Goldbeter model (Dupont, Goldbeter, 1989)	Calcium concentration oscillation	Linear inhomogeneous SODE with constant coefficients	Positive (4), and Negative (3)

Note. ODE – ordinary differential equation, SODE – system of ODEs.

fectors, etc. As early as 1913 an article (Michaelis, Menten, 1913) by biochemists Michaelis and Menten was published, in which the scientists derived the equation for the dependence of the reaction rates catalyzed by the enzyme on the concentration of the substrate. Later, the researchers have showed that, using computational methods, optimizing the parameters of the equation by approximating the model data to the experimental data corresponds to the results, which were obtained manually by Michaelis and Menten for their constant.

Not long ago, a review was conducted of how methods for quantitative analysis of enzyme kinetics have emerged, changed, and been modified over a century (Johnson, 2013). In the same year an article (Goldbeter, 2013) examined the influence of Michaelis–Menten kinetics on oscillatory behavior in enzymatic systems, namely, in glycolysis from phosphofructokinase activity and in the cell cycle from cyclin-dependent kinases.

Novák and Tyson reviewed examples of oscillatory processes and formulated the necessary conditions for oscillations in the system: negative feedback, time delay, sufficient 'nonlinearity' of the reaction kinetics and proper balancing of the timescales of opposing chemical reactions (Novák, Tyson, 2008).

In a recent review (Tyson et al., 2019), the authors compiled various approaches to modeling the dynamics of the behavior of biochemical regulatory networks that have been developed over the past 50 years. Models such as Boolean (logical) ones, models consisting of piecewise-linear or fully nonlinear ordinary differential equations, and stochastic models (including hybrid deterministic/stochastic approaches) are considered. The authors focused on two approaches: modeling genetic control systems as networks of Boolean switches and metabolic and signaling networks using systems of nonlinear ordinary differential equations. They considered only spatially homogeneous systems. The authors showed the advantages and disadvantages of each method depending on the type and amount of available experimental information.

The models, which we reviewed in this section, are summarized in Table 1.

Application of gene networks in the modeling of oscillatory systems

Modeling of metabolism is often associated with modeling of genetic regulation (Smolen et al., 2000; Hecker et al., 2009). The concept of gene networks plays an integrative role in this case (Kolchanov et al., 2013; Ocone et al., 2013).

The main task of the theory of gene networks is to identify causal relationships between the structural and functional organization of gene networks and their dynamic properties (Chen et al., 2010; Kolchanov et al., 2013). The structural and functional organization of gene networks is understood as a set of molecular-genetic and biochemical processes, while the dynamic properties are understood as the kinetics of changes in the concentrations of end products over time.

Computer analysis and modeling of small gene networks, especially hypothetical gene networks, provides very valuable information for understanding the fundamental features of the dynamics of regulatory gene networks. Likhoshvai and his colleagues developed a theory linking the structural and functional organization of hypothetical gene networks with their dynamics (Likhoshvai et al., 2001, 2003, 2004; Fadeev, Likhoshvai, 2003; Demidenko et al., 2004). Namely, the concept of a hypothetical gene network was defined; rules for formalizing the description and assembling mathematical models from them are given. The (n, k)-criterion for predicting some properties of the models by the structure of the network

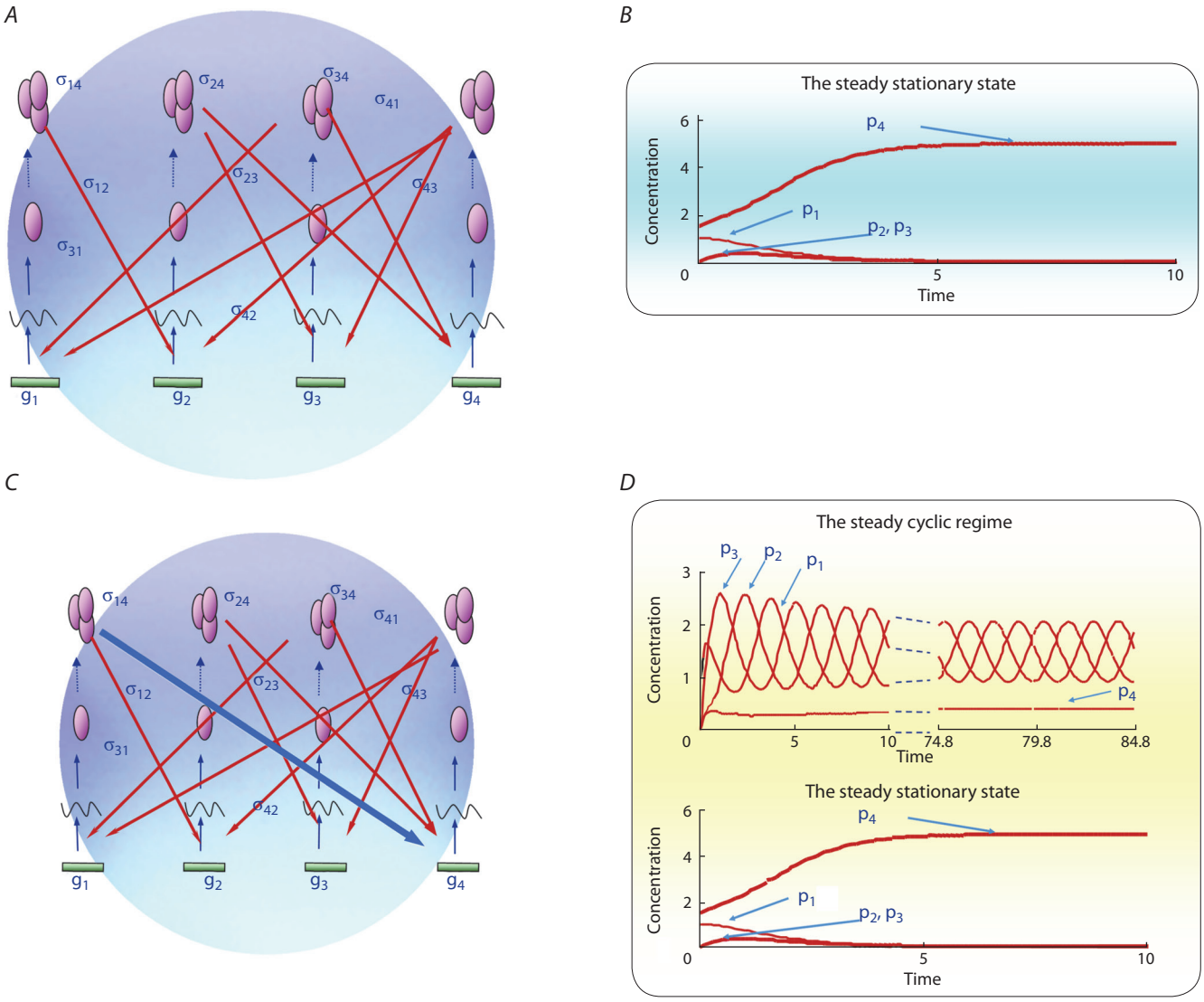


Fig. 1. Relationship between the structural model (graph) of the hypothetical gene network and its dynamics: *A*, structure of the hypothetical gene network of 4 genes and 8 negative feedbacks; *B*, dynamics of the hypothetical gene network *A*; *C*, structure of the modified gene network *A* – to which an additional negative regulatory link was added – inhibition of gene g_4 expression by the product of gene g_1 (marked by a blue arrow); *D*, dynamics of the modified gene network *B*.

Here, green rectangles (g_i) are genes, broken line is RNA corresponding to a certain gene, pink ellipse is polypeptide chain of protein, several pink ellipses are a complex of proteins performing gene regulation (regulation is shown by a red arrow). Modified according to (Kolchanov et al., 2008).

graph is formulated; 4 classes of the hypothetical gene network are introduced according to the types of regulatory links in the network; and analytical and numerical studies of the models for each class of the hypothetical gene network are given.

In particular, it was first theoretically and then numerically demonstrated how the appearance of a new regulatory link leads to a qualitative change in the dynamics of the gene network (Fig. 1). Thus, the addition of another regulatory link in the gene network cardinally changes the possible modes of functioning of this network – if only one stationary state was possible in the initial network, then after adding another regulatory link, there are already two possible states – stationary (as in the previous case) and cyclic mode.

The connection between the structures of gene networks and the presence of dynamic cycles in them has been studied for many years. In particular, the connection between network structure and cyclic dynamics has been theoretically shown (Likhoshvai et al., 2003; Demidenko et al., 2004; Novák, Tyson, 2008). Elowitz and Leibler designed and studied a genetic network of a repressilator, in which the network under study is locked into a cycle of interactions based on the principle of negative feedbacks. The authors experimentally showed that this type of network has an oscillatory mode of behavior (Elowitz, Leibler, 2000).

In the Sobolev Institute of Mathematics SB RAS is studied the qualitative theory of dynamical systems describing various gene networks that are regulated by feedbacks. Golubyatnikov and his colleagues have studied in their works (Gaidov, Golubyatnikov, 2007; Golubyatnikov et al., 2010; Akinshin, Golubyatnikov, 2012; Golubyatnikov, Kazantsev, 2016; Golu-

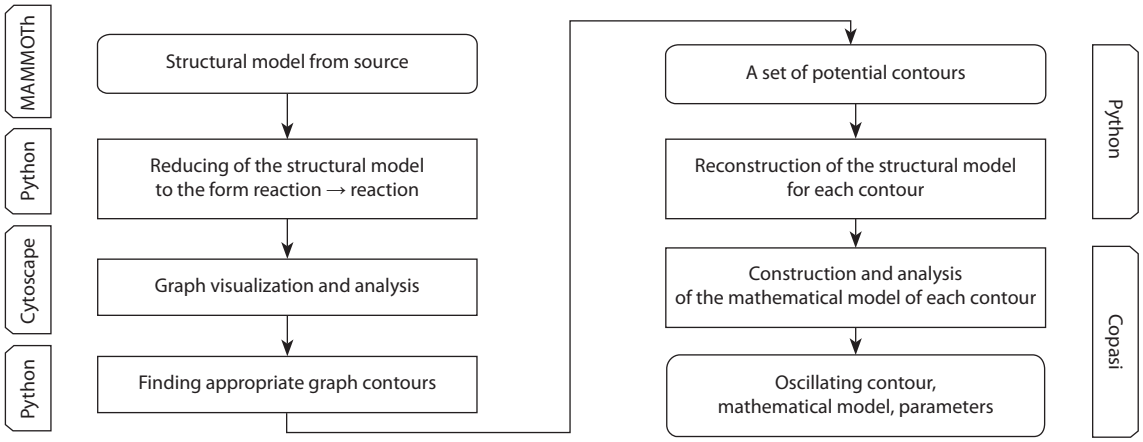


Fig. 2. Scheme of the algorithm for searching oscillatory enzymatic systems.

byatnikov, Kirillova, 2018) the existence and uniqueness of periodic solutions, existence of closed trajectories, cycle stability, etc. in such systems. The interest in the analysis of the behavior of such trajectories is to correspond them to the modes of functioning of gene networks. An article (Likhoshvai et al., 2020) showed that oscillatory trajectories are present in models of the simplest circular gene networks and they are stable.

The method for finding oscillating molecular-genetic systems

In this paragraph, we describe the algorithm for searching the oscillatory molecular-genetic systems (the algorithm scheme is shown in Fig. 2). It uses information resources both developed by the authors and widely known in systems biology. In particular, the MAMMOTh database is a source of structural and mathematical models of *Escherichia coli* metabolic reactions (Kazantsev et al., 2018). Cytoscape (cytoscape.org) is a tool for working with structural models and Copasi (Hoops et al., 2006) is a tool for reconstructing and investigating mathematical models. Python (python.org) is both a data processing tool and a link between the steps.

The input of the algorithm takes a structural model – a gene network graph with typing of model elements and their relations. There are two types of nodes in the graph: biological substances (molecules and their groups) and processes (or reactions). The edges specify the following relations between the nodes: substance is a substrate in a reaction, substance is a product of a reaction, and substance is a regulator of a reaction. This information can be obtained directly from models in SBML (Hucka et al., 2003), SBGN (Le Novère et al., 2009), from other tools for building structural models, or from Python scripts. To date, any database that has information on metabolic pathways and molecular-genetic systems models can be used as a data source. The best-known databases are KEGG (Kanehisa, Goto, 2000), GeneNet (Ananko, 2002), MetaCyc (Caspi et al., 2016), EcoCyc (Keseler et al., 2017), BioModels (Le Novere et al., 2006; Malik-Sheriff et al., 2019), etc.

In this article, we considered a special case of molecular-genetic systems – the oscillatory enzymatic systems. Analysis of the literature (Likhoshvai et al., 2001; Novák, Tyson, 2008; Tyson, Novák, 2010; Wong, Huck, 2017) allows us to identify the following key characteristics of potentially oscillating contours: (1) the closure of the contour (oriented path from node A to it, through N nodes, where $N > 3$); (2) the orientation of the contour in one direction, with the last node having an edge of regulatory inhibitory influence on the first node in the contour (as in the contour in Fig. 4, *a*, for example).

A graph of 300 subsystems (Fig. 3) representing models of *E. coli* metabolic reactions taken from the MAMMOTh database was taken as initial data.

The construction of a mathematical model of a potentially oscillating contour can be performed both in general-purpose engineering simulation environments (Matlab, Mathematica or Scilab) or in specialized environments designed for the simulation of molecular-genetic systems (Copasi, CellDesigner (Funahashi et al., 2003), VCELL (Schaff et al., 1997; Cowan et al., 2012), etc.). The advantage of the latter is the ready library of tools for reconstruction, computational experiments and model analysis.

Six potentially oscillating contours were found in the analyzed graph, and during the numerical analysis of the reconstructed mathematical model oscillatory behavior was shown for only one of them (Fig. 4). The mathematical model of the contour was constructed based on the reactions related to the metabolic pathway of tryptophan biosynthesis:

$$\begin{aligned} &\mathrm{CHOR} + \mathrm{GLN} \rightarrow \mathrm{PYR} + \mathrm{GLU} + \mathrm{AN};\ \mathrm{AnthS},\ \mathrm{Trp},\ \mathrm{Mg},\\ &\mathrm{AN} + \mathrm{PRPP} \rightarrow \mathrm{NPRAN} + \mathrm{PPI};\ \mathrm{AnthSII},\\ &\mathrm{NPRAN} \rightarrow \mathrm{CPAD5P};\ \text{Phosphoribosyl anthranilate isomerase},\\ &\mathrm{CPAD5P} \rightarrow \mathrm{IGP} + \mathrm{CO_2};\ \text{Indoleglycerol phosphate synthase},\\ &\mathrm{IGP} + \mathrm{SER} \rightarrow \mathrm{T3P1} + \mathrm{TRP};\ \mathrm{TryptS}. \end{aligned} \quad (9)$$

Here CHOR, GLN, PYR, GLU, AN, PRPP, NPRAN, PPI, CPAD5P, IGP, SER, T3P1, TRP – before the semicolon are the designations of the biochemical substances involved in the reaction, and after that are regulators of reactions. Full names of substances are given in Table 2.

The model was built in Copasi and consists of 5 differential equations.

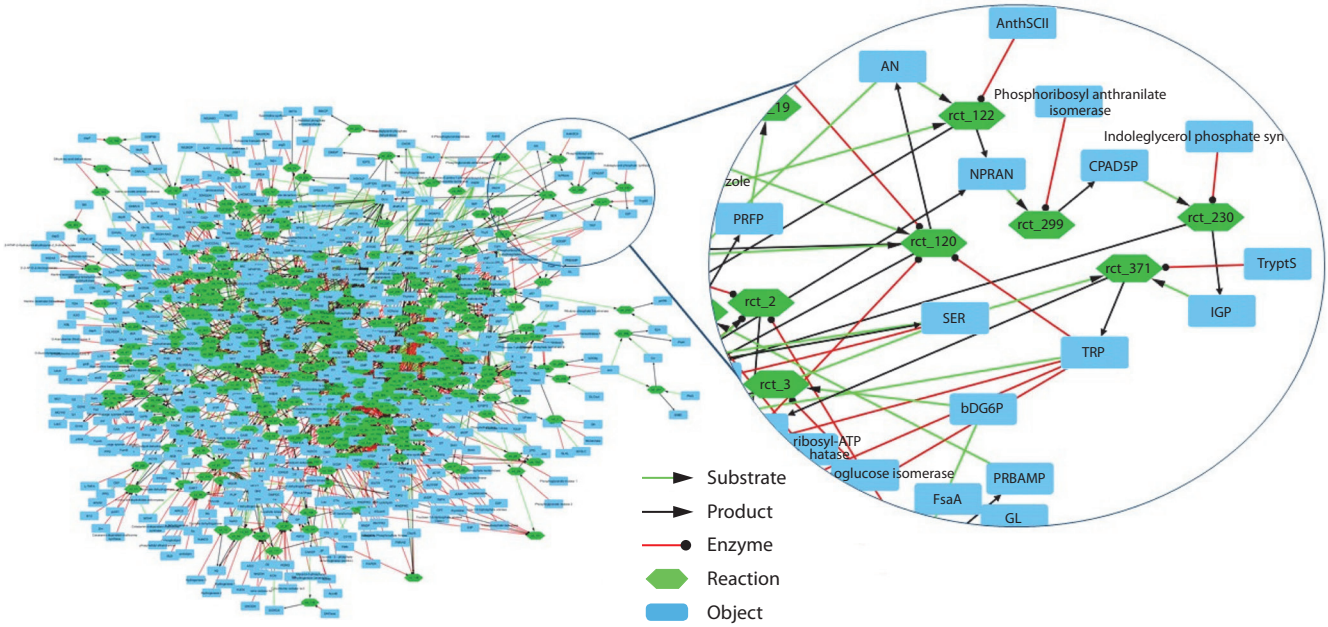


Fig. 3. Structural model (graph *G*) constructed from 300 subsystems of *E. coli* metabolic pathways taken from the MAMMOTh database (Kazantsev et al., 2018).

Here and in the Fig. 4 the following notation is used: Blue squares represent substances involved in metabolic reactions. Green hexagons indicate reactions, with arrows in/out specifying the relations of the interacting substances: green arrows specify the reaction substrates; black arrows specify the reaction products; red arrows specify the regulatory effects of the substances on the reactions.

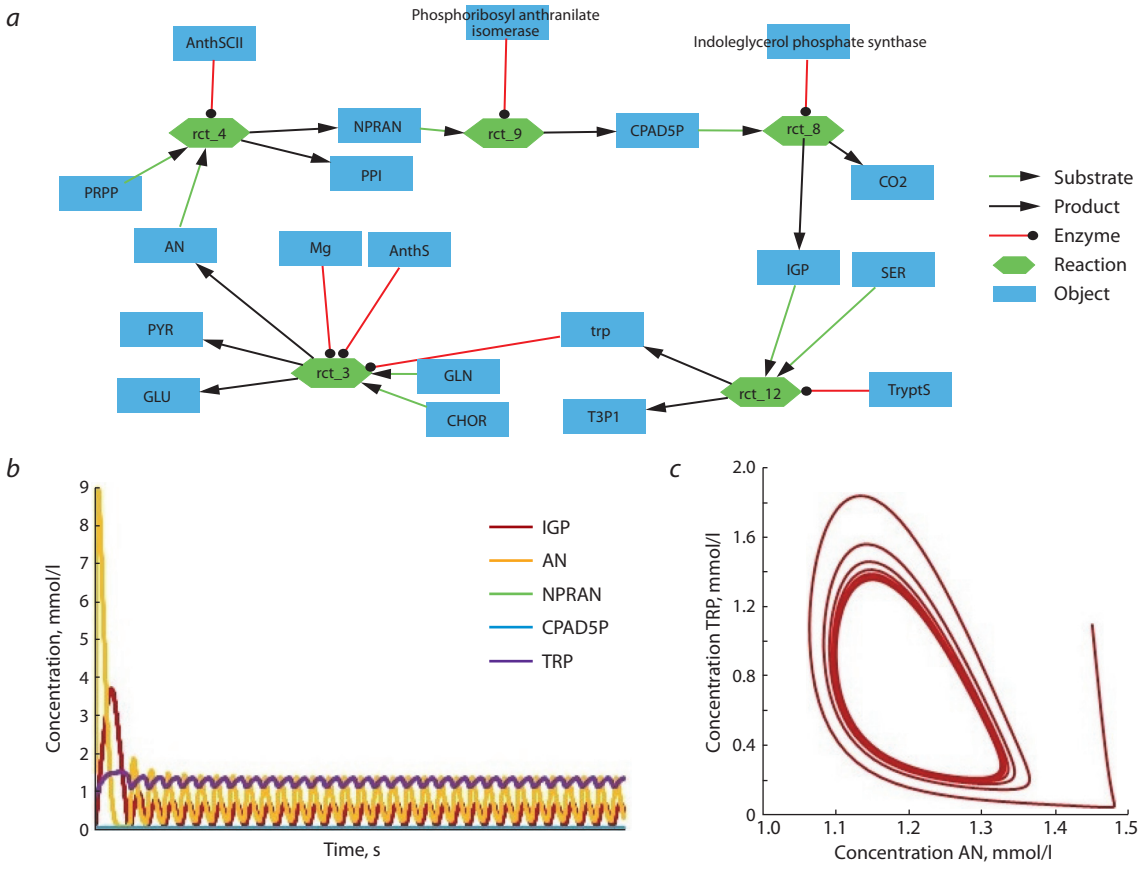


Fig. 4. Potentially oscillating contour and its numeric analysis.

a, studied contour that is a part of the metabolic pathway of tryptophan biosynthesis; *b*, the plot with the results of the simulation, the dependence of the concentration of the specified substances on time; *c*, phase trajectory plot based on simulation results, where the abscissa and ordinate axes are the concentrations of anthranilate (AN) and L-tryptophan (TRP), respectively.

$$\frac{d([\text{IGP}])}{dt} = +\left(\frac{3.1\cdot[\text{"Indoleglycerol phosphate synthase"}]\cdot\frac{[\text{CPAD5P}]}{0.0012}}{1+\frac{[\text{CPAD5P}]}{0.0012}+\frac{[\text{IGP}]}{0.12}}\right)$$
$$-\left(\frac{1.4\cdot[\text{TryptS}]\cdot\frac{[\text{IGP}]}{0.05}\cdot\frac{[\text{SER}]}{0.4}}{\left(1+\frac{[\text{IGP}]}{0.05}+\frac{[\text{T3P1}]}{5}\right)\cdot\left(1+\frac{[\text{SER}]}{0.4}+\frac{[\text{TRP}]}{40}\right)}\right)-(\text{kD\_IGP}\cdot[\text{IGP}]),$$

$$\frac{d([\text{TRP}])}{dt} = -(\text{kD\_TRP}\cdot[\text{TRP}])+\left(\frac{1.4\cdot[\text{TryptS}]\cdot\frac{[\text{IGP}]}{0.05}\cdot\frac{[\text{SER}]}{0.4}}{\left(1+\frac{[\text{IGP}]}{0.05}+\frac{[\text{T3P1}]}{5}\right)\cdot\left(1+\frac{[\text{SER}]}{0.4}+\frac{[\text{TRP}]}{40}\right)}\right),$$

$$\frac{d([\text{AN}])}{dt} = +\left(\frac{260\cdot[\text{AnthS}]\cdot\frac{[\text{CHOR}]}{1.5}\cdot\frac{[\text{GLN}]}{0.2}}{\left(1+\frac{[\text{CHOR}]}{1.5}+\frac{[\text{PYR}]}{150}+\frac{[\text{TRP}]}{0.6}\right)\cdot\left(1+\frac{[\text{GLN}]}{0.2}+\frac{[\text{GLU}]}{20}\right)}\cdot\frac{1}{1+\left(\frac{[\text{TRP}]}{\text{TRP\_denominator}}\right)^{\text{TRP\_power}}}\cdot\frac{\frac{[\text{Mg}]}{1}}{1+\frac{[\text{Mg}]}{1}}\right)$$
$$-\left(\frac{\text{AN\_PRPP\_kf}\cdot[\text{AnthSCII}]\cdot\frac{[\text{AN}]}{1.1}\cdot\frac{[\text{PRPP}]}{2.9}}{\left(1+\frac{[\text{AN}]}{1.1}\right)\cdot\left(1+\frac{[\text{PRPP}]}{2.9}\right)}\right)-(\text{kD\_AN}\cdot[\text{AN}]),$$

$$\frac{d([\text{NPRAN}\cdot])}{dt} = +\left(\frac{\text{AN\_PRPP\_kf}\cdot[\text{AnthSCII}]\cdot\frac{[\text{AN}]}{1.1}\cdot\frac{[\text{PRPP}]}{2.9}}{\left(1+\frac{[\text{AN}]}{1.1}\right)\cdot\left(1+\frac{[\text{PRPP}]}{2.9}\right)}\right)$$
$$-\left(\frac{\text{AN\_PRPP\_kf}\cdot[\text{"Phosphoribosyl anthranilate isomerase"}]\cdot\frac{[\text{NPRAN}]}{0.007}}{1+\frac{[\text{NPRAN}]}{0.007}+\frac{[\text{CPAD5P}]}{0.7}}\right)-(\text{kD\_NPRAN}\cdot[\text{NPRAN}]),$$

$$\frac{d([\text{CPAD5P}])}{dt} = +\left(\frac{\text{AN\_PRPP\_kf}\cdot[\text{"Phosphoribosyl anthranilate isomerase"}]\cdot\frac{[\text{NPRAN}]}{0.007}}{1+\frac{[\text{NPRAN}]}{0.007}+\frac{[\text{CPAD5P}]}{0.7}}\right)$$
$$-\left(\frac{3.1\cdot[\text{"Indoleglycerol phosphate synthase"}]\cdot\frac{[\text{CPAD5P}]}{0.0012}}{1+\frac{[\text{CPAD5P}]}{0.0012}+\frac{[\text{IGP}]}{0.12}}\right)-(\text{kD\_CPAD5P}\cdot[\text{CPAD5P}]),$$

where kD_"substance name" are degradation constants of corresponding substances, parameters TRP_power and TRP_denominator varied in the process of searching for oscillatory behavior of the system. The given numerical parameters were taken from the MAMMOTh database.

The mathematical model of only one of the six contours found exhibits with oscillatory behavior. As we considered, a network consisting of only 300 enzymatic reactions, which had mathematical models adapted to the experimental data, may explain such a small number of contours. In turn, there are currently not many such mathematical models for describing the enzymatic reactions of biological systems. Thousands of existing models presented in databases are often automatically generated, as in the Path2Models project for the biomodels.net database, for example. Experimentally measured kinetic parameters of biochemical reactions are becoming increasingly scarce. Using graphs with higher dimensionality (full-genome models) to study oscillatory behavior will increase the number of variants to be tested, but this will require additional consideration of the regulatory component of genetic synthesis. All of these things present additional challenges in the study of this problem.

Conclusion

The article gives an overview of a number of biological processes of oscillatory nature, as well as mathematical models

Table 2. List of full names of biochemical substances used in the model

Abbreviation	Full name
CHOR	Chorismate
GLN	L-glutamine
PYR	Pyruvate
GLU	L-glutamate
AN	Anthranilate
AnthS	Anthranilate synthase
TRP	L-tryptophan
PRPP	5-Phospho-α-D-ribose 1-diphosphate
NPRAN	N-(5-phosphoribosyl)-anthranilate
PPI	Diphosphate
AnthSCII	Anthranilate synthase component II
CPAD5P	1-(o-carboxyphenylamino)-1′-deoxyribulose-5′-phosphate
PRAI	Phosphoribosyl anthranilate isomerase
IGP	Indole-3-glycerol-phosphate
SER	L-serine
T3P1	D-glyceraldehyde 3-phosphate
TryptS	Tryptophan synthase

of these processes. It is noted that the most important factor for the emergence of cyclic behavior is feedbacks in the system. Based on the analysis of these factors, an algorithm for finding cyclic modes of functioning of molecular-biological systems is given.

References

Akinshin A.A., Golubyatnikov V.P. Cycles in symmetric dynamical systems. *Vestnik NGU = Herald of the Novosibirsk State University.* 2012;12(2):3-12. (in Russian)

Allen G.J., Chu S.P., Schumacher K., Shimazaki C.T., Vafeados D., Kemper A., Hawke S.D., Tallman G., Tsien R.Y., Harper J.F., Chory J., Schroeder J.I. Alteration of stimulus-specific guard cell calcium oscillations and stomatal closing in *Arabidopsis det3* mutant. *Science*. 2000;289:2338-2342. DOI 10.1126/science.289.5488.2338.

Ananko E.A. GeneNet: a database on structure and functional organisation of gene networks. *Nucleic Acids Res*. 2002;30:398-401. DOI 10.1093/nar/30.1.398.

Ashkenazy Y., Ivanov P.C., Havlin S., Peng C.K., Goldberger A.L., Stanley H.E. Magnitude and sign correlations in heartbeat fluctuations. *Phys. Rev. Lett*. 2001;86:1900-1903. DOI 10.1103/PhysRevLett.86.1900.

Bashkirtseva I., Ryashko L. Stochastic sensitivity and variability of glycolytic oscillations in the randomly forced Sel'kov model. *Eur. Phys. J. B*. 2017;90:17. DOI 10.1140/epjb/e2016-70674-4.

Bazykin A.D. Nonlinear Dynamics of Interacting Populations. Moscow–Izhevsk, 2003. (in Russian)

Becks L., Arndt H. Different types of synchrony in chaotic and cyclic communities. *Nat. Commun.* 2013;4:1359. DOI 10.1038/ncomms2355.

Bier M., Bakker B.M., Westerhoff H.V. How yeast cells synchronize their glycolytic oscillations: a perturbation analytic treatment. *Biophys. J.* 2000;78:1087-1093. DOI 10.1016/S0006-3495(00)76667-7.

Boccalandro H.E., González C.V., Wunderlin D.A., Silva M.F. Melatonin levels, determined by LC-ESI-MS/MS, fluctuate during the day/night cycle in *Vitis vinifera* cv Malbec: evidence of its antioxidant role in fruits. *J. Pineal Res*. 2011;51:226-232. DOI 10.1111/j.1600-079X.2011.00884.x.

Boiteux A., Goldbeter A., Hess B. Control of oscillating glycolysis of yeast by stochastic, periodic, and steady source of substrate: a model and experimental study. *Proc. Natl. Acad. Sci. USA.* 1975;72:3829-3833. DOI 10.1073/pnas.72.10.3829.

Cardon S.Z., Iberall A.S. Oscillations in biological systems. *Biosystems*. 1970;3:237-249. DOI 10.1016/0303-2647(70)90004-3.

Caspi R., Billington R., Ferrer L., Foerster H., Fulcher C.A., Keseler I.M., Kothari A., Krummenacker M., Latendresse M., Mueller L.A., Ong Q., Paley S., Subhraveti P., Weaver D.S., Karp P.D. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. 2016;44:D471-D480. DOI 10.1093/nar/gkv1164.

Chen W.W., Niepel M., Sorger P.K. Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev*. 2010;24:1861-1875. DOI 10.1101/gad.1945410.

Chetverikov S.S. The waves of life. *Russkii Ornitologicheskii Zhurnal = The Russian Journal of Ornithology.* 2009;18:1822-1829. (in Russian)

Cooper S. Bacterial Growth and Division. Acad. Press, Inc., 1991.

Cowan A.E., Moraru I.I., Schaff J.C., Slepchenko B.M., Loew L.M. Spatial modeling of cell signaling networks. In: Asthagiri A.R., Arkin A.P. (Eds.). Methods in Cell Biology. Acad. Press, 2012;195-221. DOI 10.1016/B978-0-12-388403-9.00008-4.

Demidenko G.V., Kolchanov N.A., Likhoshvai V.A., Matushkin Yu.G., Fadeev S.I. Mathematical modeling of gene networks regulation contours. *Zhurnal Vychislitel'noy Matematiki i Matematicheskoy Fiziki = Computational Mathematics and Mathematical Physics.* 2004;44:2276-2295. (in Russian)

Dupont G., Berridge M.J., Goldbeter A. Signal-induced Ca^{2+} oscillations: properties of a model based on Ca^{2+}-induced Ca^{2+} release. *Cell Calcium*. 1991;12:73-85. DOI 10.1016/0143-4160(91)90010-C.

Dupont G., Goldbeter A. Theoretical insights into the origin of signal-induced calcium oscillations. In: Cell to Cell Signalling. Elsevier, 1989;461-474. DOI 10.1016/B978-0-12-287960-9.50040-X

Dupont G., Goldbeter A. One-pool model for Ca^{2+} oscillations involving Ca^{2+} and inositol 1,4,5-trisphosphate as co-agonists for Ca^{2+} release. *Cell Calcium*. 1993;14:311-322. DOI 10.1016/0143-4160(93)90052-8.

Elowitz M.B., Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000;403:335-338. DOI 10.1038/35002125.

Erdakov L.N., Moroldoev I.V. Variability of long-term cyclicity in the population dynamics of the northern red-backed vole (Myodes Rutilus (Pallas, 1779)). *Principles Ecol*. 2017;25:26-36. DOI 10.15393/j1.art.2017.7342.

Fadeev S.I., Likhoshvai V.A. On hypothetical gene networks. *Sibirskii Zhurnal Industrial'noi Matematiki = Journal of Applied and Industrial Mathematics.* 2003;6(3):134-153. (in Russian)

Field R., Burger M. (Eds.) Oscillations and Traveling Waves in Chemical Systems. Chichester: John Wiley Publ., 1985. (Russ. ed.: Fild R., Burger M. Kolebaniya i begushchie volny v khimicheskikh sistemakh. Moscow: Mir Publ., 1988. (in Russian))

Funahashi A., Morohashi M., Kitano H., Tanimura N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*. 2003;1:159-162. DOI 10.1016/S1478-5382(03)02370-9.

Gaidov Yu.A., Golubyatnikov V.P. On some nonlinear dynamical systems modelling asymmetric gene networks. *Vestnik NGU = Herald of the Novosibirsk State University*. 2007;7(2):19-27. (in Russian)

Garde A.H., Hansen A.M., Skovgaard L.T., Christensen J.M. Seasonal and biological variation of blood concentrations of total cholesterol, dehydroepiandrosterone sulfate, hemoglobin A_{1c}, IgA, prolactin, and free testosterone in healthy women. *Clin. Chem*. 2000;46(4):551-559. DOI 10.1093/clinchem/46.4.551.

Goldbeter A. Computational approaches to cellular rhythms. *Nature*. 2002;420:238-245. DOI 10.1038/nature01259.

Goldbeter A. Oscillatory enzyme reactions and Michaelis–Menten kinetics. *FEBS Lett*. 2013;587:2778-2784. DOI 10.1016/j.febslet.2013.07.031.

Goldbeter A., Dupont G., Berridge M.J. Minimal model for signal-induced Ca^{2+} oscillations and for their frequency encoding through protein phosphorylation. *Proc. Natl. Acad. Sci. USA*. 1990;87:1461-1465. DOI 10.1073/pnas.87.4.1461.

Goldbeter A., Lefever R. Dissipative structures for an allosteric model. *Biophys. J*. 1972;12:1302-1315. DOI 10.1016/S0006-3495(72)86164-2.

Golubyatnikov V.P., Golubyatnikov I.V., Likhoshvai V.A. On the existence and stability of cycles in five-dimensional models of gene networks. *Numer. Anal. Appl*. 2010;3:329-335. DOI 10.1134/S199542391004004X.

Golubyatnikov V.P., Kazantsev M.V. On a piecewise linear dynamical system that models a gene network with variable feedback. *Sibirskiy Zhurnal Chistoy i Prikladnoy Matematiki = Siberian Journal of Pure and Applied Mathematics.* 2016;16(4):28-37. DOI 10.17377/PAM.2016.16.404. (in Russian)

Golubyatnikov V.P., Kirillova N.E. On cycles in models of functioning of circular gene networks. *Sibirskiy Zhurnal Chistoy i Prikladnoy Matematiki = Siberian Journal of Pure and Applied Mathematics.* 2018;18(1):54-63. DOI 10.17377/PAM.2018.18.5. (in Russian)

Goodwin B.C. Oscillatory behavior in enzymatic control processes. *Adv. Enzyme Regul.* 1965;3:425-437. DOI 10.1016/0065-2571(65)90067-1.

Hardin P.E., Hall J.C., Rosbash M. Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature*. 1990;343:536-540. DOI 10.1038/343536a0.

Hecker M., Lambeck S., Toepfer S., van Someren E., Guthke R. Gene regulatory network inference: data integration in dynamic models – a review. *Biosystems*. 2009;96:86-103. DOI 10.1016/j.biosystems.2008.12.004.

Higgins J. A chemical mechanism for oscillation of glycolytic intermediates in yeast cells. *Proc. Natl. Acad. Sci. USA*. 1964;51:989-994. DOI 10.1073/pnas.51.6.989.

Higgins J. The theory of oscillating reactions – kinetics symposium. *Ind. Eng. Chem*. 1967;59:18-62. DOI 10.1021/ie50689a006.

Holtum J.A.M., Winter K. Photosynthetic CO_2 uptake in seedlings of two tropical tree species exposed to oscillating elevated concentrations of CO_2. *Planta*. 2003;218:152-158. DOI 10.1007/s00425-003-1089-1.

Hoops S., Sahle S., Gauges R., Lee C., Pahle J., Simus N., Singhal M., Xu L., Mendes P., Kummer U. COPASI – a COmplex PAthway SImulator. *Bioinformatics*. 2006;22:3067-3074. DOI 10.1093/bioinformatics/btl485.

Hucka M., Finney A., Sauro H.M., Bolouri H., Doyle J.C., Kitano H., and the rest of the SBML Forum: Arkin A.P., Bornstein B.J., Bray D., Cornish-Bowden A., Cuellar A.A., Dronov S., Gilles E.D., Ginkel M., Gor V., Goryanin I.I., Hedley W.J., Hodgman T.C., Hofmeyr J.-H., Hunter P.J., Juty N.S., Kasberger J.L., Kremling A., Kummer U., Le Novere N., Loew L.M., Lucio D., Mendes P., Minch E., Mjolsness E.D., Nakayama Y., Nelson M.R., Nielsen P.F., Sakurada T., Schaff J.C., Shapiro B.E., Shimizu T.S., Spence H.D., Stelling J., Takahashi K., Tomita M., Wagner J., Wang J. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 2003;19:524-531. DOI 10.1093/bioinformatics/btg015.

Johnson K.A. A century of enzyme kinetic analysis, 1913 to 2013. *FEBS Lett*. 2013;587:2753-2766. DOI 10.1016/j.febslet.2013.07.012.

Kanehisa M., Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28:27-30. DOI 10.1093/nar/28.1.27.

Kazantsev F.V., Akberdin I.R., Lashin S.A., Ree N.A., Timonov V.S., Ratushnyi A.V., Khlebodarova T.M., Likhoshvai V.A. MAMMOTh: a new database for curated mathematical models of biomolecular systems. *J. Bioinf. Comput. Biol*. 2018;16:1740010. DOI 10.1142/S0219720017400108.

Keseler I.M., Mackie A., Santos-Zavaleta A., Billington R., Bonavides-Martínez C., Caspi R., Fulcher C., Gama-Castro S., Kothari A., Krummenacker M., Latendresse M., Muñiz-Rascado L., Ong Q., Paley S., Peralta-Gil M., Subhraveti P., Velázquez-Ramírez D.A., Weaver D., Collado-Vides J., Paulsen I., Karp P.D. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 2017;45:D543-D550. DOI 10.1093/nar/gkw1003.

Kolchanov N.A., Goncharov S.S., Likhoshvai V.A., Ivanisenko V.A. Computional Systems Biology. Novosibirsk: Publ. House SB RAS, 2008. (in Russian)

Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):833-850. (in Russian)

Kolchanov N.A., Matushkin Yu.G., Frolov A.S. Computer analysis of the evolution of genetic regulatory systems. In: Shumny V.K., Markel A.L. (Eds.). Current Concepts of Evolutionary Genetics. Novosibirsk: Institute of Cytology and Genetics SB RAS, 2000; 60-75. (in Russian)

Kraus M., Wolf Björn, Wolf Bernhard. Crosstalk between cellular morphology and calcium oscillation patterns. Insights from a stochastic computer model. *Cell Calcium*. 1996;19:461-472. DOI 10.1016/S0143-4160(96)90055-X.

Kuznetsov A.P., Seliverstova E.S., Trubetskov D.I., Turukina L.V. Phenomenon of the van der Pol equation. *Izvestiya Vysshikh Uchebnykh Zavedeniy. Prikladnaya Nelineynaya Dinamika = Izvestiya VUZ. Applied Nonlinear Dynamics*. 2014;22(4):3-42. DOI 10.18500/0869-6632-2014-22-4-3-42. (in Russian)

Lahav G., Rosenfeld N., Sigal A., Geva-Zatorsky N., Levine A.J., Elowitz M.B., Alon U. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet*. 2004;36:147-150. DOI 10.1038/ng1293.

Lavrentovich M., Hemkin S. A mathematical model of spontaneous calcium(II) oscillations in astrocytes. *J. Theor. Biol*. 2008;251:553-560. DOI 10.1016/j.jtbi.2007.12.011.

Le Novere N., Bornstein B., Broicher A., Courtot M., Donizelli M., Dharuri H., Li L., Sauro H., Schilstra M., Shapiro B., Snoep J.L., Hucka M. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*. 2006;34:D689-D691. DOI 10.1093/nar/gkj092.

Le Novère N., Hucka M., Mi H., Moodie S., Schreiber F., Sorokin A., Demir E., Wegner K., Aladjem M.I., Wimalaratne S.M., Bergman F.T., Gauges R., Ghazal P., Kawaji H., Li L., Matsuoka Y., Villéger A., Boyd S.E., Calzone L., Courtot M., Dogrusoz U., Freeman T.C., Funahashi A., Ghosh S., Jouraku A., Kim S., Kolpakov F.,

Luna A., Sahle S., Schmidt E., Watterson S., Wu G., Goryanin I., Kell D.B., Sander C., Sauro H., Snoep J.L., Kohn K., Kitano H. The Systems Biology Graphical Notation. *Nat. Biotechnol*. 2009;27: 735-741. DOI 10.1038/nbt.1558.

Likhoshvai V.A., Fadeev S.I., Matushkin Y.G. The global operation modes of gene networks determined by the structure of negative feedbacks. In: Bioinformatics of Genome Regulation and Structure. Boston, MA: Springer US, 2004;319-329. DOI 10.1007/978-1-4419-7152-4_34.

Likhoshvai V.A., Golubyatnikov V.P., Khlebodarova T.M. Limit cycles in models of circular gene networks regulated by negative feedback loops. *BMC Bioinform*. 2020;21:255. DOI 10.1186/s12859-020-03598-z.

Likhoshvai V.A., Matushkin Yu.G., Fadeev S.I. On the relationship of the gene network graph with the qualitative modes of its functioning. *Molekulyarnaya Biologiya = Molecular Biology.* 2001;35(6):1080-1087. (in Russian)

Likhoshvai V.A., Matushkin Yu.G., Fadeev S.I. Problems in the theory of the functioning of genetic networks. *Sibirskii Zhurnal Industrial'noi Matematiki = Journal of Applied and Industrial Mathematics*. 2003;6(2):64-80. (in Russian)

Lotka A.J. Contribution to the theory of periodic reactions. *J. Phys. Chem*. 1910;14:271-274. DOI 10.1021/j150111a004.

Lotka A.J. Undamped oscillations derived from the law of mass action. *J. Am. Chem. Soc*. 1920;42:1595-1599. DOI 10.1021/ja01453a010.

Malik-Sheriff R.S., Glont M., Nguyen T.V.N., Tiwari K., Roberts M.G., Xavier A., Vu M.T., Men J., Maire M., Kananathan S., Fairbanks E.L., Meyer J.P., Arankalle C., Varusai T.M., Knight-Schrijver V., Li L., Dueñas-Roca C., Dass G., Keating S.M., Park Y.M., Buso N., Rodriguez N., Hucka M., Hermjakob H. BioModels – 15 years of sharing computational models in life science. *Nucleic Acids Res*. 2019;48:407-415. DOI 10.1093/nar/gkz1055.

Meyer T. Calcium spiking. *Annu. Rev. Biophys. Biomol. Struct*. 1991; 20:153-174. DOI 10.1146/annurev.biophys.20.1.153.

Meyer T., Stryer L. Molecular model for receptor-stimulated calcium spiking. *Proc. Natl. Acad. Sci. USA*. 1988;85:5051-5055. DOI 10.1073/pnas.85.14.5051.

Michaelis L., Menten M.L. Die Kinetik der Invertinwirkung. *Biochem. Z.* 1913;49:333-369.

Mushtakova S.P. Oscillatory reactions in chemistry. *Sorosovskij Obrazovatelnyj Zhurnal = Soros Educational Journal.* 1997;7:31-36. (in Russian)

Novák B., Tyson J.J. Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol*. 2008;9:981-991. DOI 10.1038/nrm2530.

Ocone A., Millar A.J., Sanguinetti G. Hybrid regulatory models: a statistically tractable approach to model regulatory network dynamics. *Bioinformatics*. 2013;29:910-916. DOI 10.1093/bioinformatics/btt069.

Pasti L., Volterra A., Pozzan T., Carmignoto G. Intracellular calcium oscillations in astrocytes: a highly plastic, bidirectional form of communication between neurons and astrocytes *in situ*. *J. Neurosci.* 1997;17:7817-7830. DOI 10.1523/jneurosci.17-20-07817.1997.

Pertsev N.V., Loginv K.K. Stochastic model of dynamics of a biological community under conditions of consumption of harmful food resources by individuals. *Matematicheskaya Biologiya i Bioinformatika = Mathematical Biology and Bioinformatics.* 2011;6(1):1-13. (in Russian)

Podkolodnaya O.A., Tverdokhleb N.N., Podkolodnyy N.L. Computational modeling of the cell-autonomous mammalian circadian oscillator. *BMC Syst. Biol*. 2017;11:27-42. DOI 10.1186/s12918-016-0379-8.

Podkolodnyy N.L., Tverdokhleb N.N., Podkolodnaya O.A. Analysis of the circadian rhythm of biological processes in mouse liver and kidney. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2017;21(8):903-910. DOI 10.18699/VJ17.311 (in Russian)

Price J.L., Blau J., Rothenfluh A., Abodeely M., Kloss B., Young M.W. *double-time* is a novel *Drosophila* clock gene that regulates PERIOD protein accumulation. *Cell*. 1998;94:83-95. DOI 10.1016/S0092-8674(00)81224-6.

Prives C. Signaling to p53. *Cell*. 1998;95:5-8. DOI 10.1016/S0092-8674(00)81774-2.

Riznichenko G.Yu. Lectures on Mathematical Models in Biology. Moscow–Izhevsk, 2002. (in Russian)

Riznichenko G.Yu. Mathematical Modeling of Biological Processes. Models in Biophysics and Ecology. Moscow: Urait Publ., 2017. (in Russian)

Romanovsky Yu.M., Stepanova N.V., Chernavsky D.S. Mathematical Modeling in Biophysics. Moscow: Nauka Publ., 1975. (in Russian)

Rompala K., Rand R., Howland H. Dynamics of three coupled van der Pol oscillators with application to circadian rhythms. *Commun. Nonlinear Sci. Numer. Simul*. 2007;12:794-803. DOI 10.1016/j.cnsns.2005.08.002.

Ryashko L. Sensitivity analysis of the noise-induced oscillatory multistability in Higgins model of glycolysis. *Chaos*. 2018;28. DOI 10.1063/1.4989982.

Schaff J., Fink C.C., Slepchenko B., Carson J.H., Loew L.M. A general computational framework for modeling cellular structure and function. *Biophys. J*. 1997;73:1135-1146. DOI 10.1016/S0006-3495(97)78146-3.

Sel'kov E.E. Self-oscillations in glycolysis. 1. A simple kinetic model. *Eur. J. Biochem*. 1968;4:79-86. DOI 10.1111/j.1432-1033.1968.tb00175.x.

Shnol S.E. Kosmofysiska Faktorer i Slumpmassiga Processer. Stockholm (Sweden): Svenska fysikarkivet (the Swedish physics archive), 2009. (in Russian)

Shnol E.E. (Ed.). Studies in Mathematical Biology: Collection of scientific papers in memoriam A.D. Bazykin. Pushchino Scientific Center RAS, 1996. (in Russian)

Siwicki K.K., Eastman C., Petersen G., Rosbash M., Hall J.C. Antibodies to the period gene product of drosophila reveal diverse tissue distribution and rhythmic changes in the visual system. *Neuron*. 1988;1:141-150. DOI 10.1016/0896-6273(88)90198-5.

Smolen P., Baxter D.A., Byrne J.H. Mathematical modeling of gene networks. *Neuron*. 2000;26:567-580. DOI 10.1016/S0896-6273(00)81194-0.

Struyf E., Smis A., Van Damme S., Meire P., Conley D.J. The global biogeochemical silicon cycle. *Silicon*. 2009;1:207-213. DOI 10.1007/s12633-010-9035-x.

Tyson J.J., Chen K.C., Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol*. 2003;15:221-231. DOI 10.1016/S0955-0674(03)00017-6.

Tyson J.J., Laomettachit T., Kraikivski P. Modeling the dynamic behavior of biochemical regulatory networks. *J. Theor. Biol.* 2019; 462:514-527. DOI 10.1016/j.jtbi.2018.11.034.

Tyson J.J., Novák B. Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem*. 2010;61:219-240. DOI 10.1146/annurev.physchem.012809.103457.

Van Cappellen P. Biomineralization and global biogeochemical cycles. *Rev. Mineral. Geochem.* 2003;54:357-381. DOI 10.2113/0540357.

van der Pol B., van der Mark J. The heartbeat considered as a relaxation oscillation, and an electrical model of the heart. *Lond. Edinb. Dubl. Phil. Mag. J. Sci*. 1928;6:763-775. DOI 10.1080/14786441108564652.

Volterra V. Математическая теория борьбы за существование (пер. П.П. Лазарева). *Uspekhi Fizicheskih Nauk*. 1928;8:13-34. DOI 10.3367/UFNr.0008.192801b.0013.

Walter C. Oscillations in controlled biochemical systems. *Biophys. J.* 1969;9:863-872. DOI 10.1016/S0006-3495(69)86423-4.

Walter C.F. The occurrence and the significance of limit cycle behavior in controlled biochemical systems. *J. Theor. Biol*. 1970;27:259-272. DOI 10.1016/0022-5193(70)90141-4.

Wong A.S.Y., Huck W.T.S. Grip on complexity in chemical reaction networks. *Beilstein J. Organic Chem.* 2017;13:1486-1497. DOI 10.3762/bjoc.13.147.

Yasuma F., Hayano J.I. Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest*. 2004;125: 683-690. DOI 10.1378/chest.125.2.683.

Young M.W., Bargiello T.A., Jackson F.R. Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. *Nature.* 1984; 312:752-754.

Zavarzin G.A. Establishment of a system of biogeochemical cycles. *Paleontologicheskii Zhurnal = Paleontological Journal.* 2003;6:16-24. (in Russian)

Zavarzin G.A. Evolution of the Prokaryotic Biosphere: "Microbes in the cycle of life". 120 years afterwards: S.N. Winogradsky memorial lectures. Moscow: MAKS Press Publ., 2011. (in Russian)

Zhabotinsky A.M. Concentration Auto-oscillations. Moscow: Nauka Publ., 1974. (in Russian)

ORCID ID

T.N. Lakhova orcid.org/0000-0003-1729-7712
F.V. Kazantsev orcid.org/0000-0002-5711-7539
S.A. Lashin orcid.org/0000-0003-3138-381X
Yu.G. Matushkin orcid.org/0000-0001-7754-8611

Acknowledgements. This work was supported by the Budget Project No. 0259-2021-0009 "Systems biology and bioinformatics: reconstruction, analysis and modeling of the structural and functional organization and evolution of gene networks of humans, animals, plants and microorganisms" and the Russian Science Foundation project No. 18-14-00293 (visualization and structural analysis of graphs of metabolic networks).

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

Received November 13, 2020. Revised March 15, 2021. Accepted March 18, 2021.

Вавиловский журнал генетики и селекции. 2021;25(3):331-336
DOI 10.18699/VJ21.036

A hypomorphic mutation in the mouse *Csn1s1* gene generated by CRISPR/Cas9 pronuclear microinjection

A.V. Smirnov[1] ✉, T.A. Shnaider[1], A.N. Korablev[1], A.M. Yunusova[1], I.A. Serova[1], N.R. Battulin[1, 2] ✉

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
✉ hldn89@gmail.com; battulin@bionet.nsc.ru

Abstract. Caseins are major milk proteins that have an evolutionarily conserved role in nutrition. Sequence variations in the casein genes affect milk composition in livestock species. Regulatory elements of the casein genes could be used to direct the expression of desired transgenes into the milk of transgenic animals. Dozens of casein alleles have been identified for goats, cows, sheep, camels and horses, and these sequence variants are associated with altered gene expression and milk protein content. Most of the known mutations affecting casein genes' expression are located in the promoter and 3'-untranslated regions. We performed pronuclear microinjections with Cas9 mRNA and sgRNA against the first coding exon of the mouse *Csn1s1* gene to introduce random mutations in the *α-casein* (*Csn1s1*) signal peptide sequence at the beginning of the mouse gene. Sanger sequencing of the founder mice identified 40 mutations. As expected, mutations clustered around the sgRNA cut site (3 bp from PAM). Most of the mutations represented small deletions (1–10 bp), but we detected several larger deletions as well (100–300 bp). Functionally most mutations led to gene knockout due to a frameshift or a start codon loss. Some of the mutations represented in-frame indels in the first coding exon. Of these, we describe a novel hypomorphic *Csn1s1* (*Csn1s1*$^{c.4\text{-}5insTCC}$) allele. We measured Csn1s1 protein levels and confirmed that the mutation has a negative effect on milk composition, which shows a 50 % reduction in gene expression and a 40–80 % decrease in Csn1s1 protein amount, compared to the wild-type allele. We assumed that mutation affected transcript stability or splicing by an unknown mechanism. This mutation can potentially serve as a genetic marker for low *Csn1s1* expression.
Key words: casein; CRISPR; pronuclear microinjection; hypomorphic mutation.

For citation: Smirnov A.V., Shnaider T.A., Korablev A.N., Yunusova A.M., Serova I.A., Battulin N.R. A hypomorphic mutation in the mouse *Csn1s1* gene generated by CRISPR/Cas9 pronuclear microinjection. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):331-336. DOI 10.18699/VJ21.036

Гипоморфная мутация гена *Csn1s1* мыши, полученная пронуклеарной микроинъекцией CRISPR/Cas9

А.В. Смирнов[1] ✉, Т.А. Шнайдер[1], А.Н. Кораблев[1], А.М. Юнусова[1], И.А. Серова[1], Н.Р. Баттулин[1, 2] ✉

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ hldn89@gmail.com; battulin@bionet.nsc.ru

Аннотация. Казеины – это основные молочные белки, которые играют важную эволюционно консервативную роль в питании. Вариации последовательности казеиновых генов влияют на состав молока у животных. Регуляторные элементы казеиновых генов можно использовать для управления экспрессией желаемых трансгенов в молоке трансгенных животных. Десятки аллелей казеина идентифицированы у коз, коров, овец, верблюдов и лошадей, и эти варианты связаны с измененной экспрессией генов и содержанием молочного белка. Большинство известных мутаций, влияющих на экспрессию генов казеина, находится в промоторных и 3'-нетранслируемых областях. Мы выполнили пронуклеарные микроинъекции с мРНК Cas9 и sgRNA против первого кодирующего экзона гена *Csn1s1* мыши, чтобы ввести случайные мутации в последовательность сигнального пептида α-казеина (*Csn1s1*) в начале гена мыши. Секвенирование мышей-основателей по Сэнгеру выявило 40 мутаций. Как и ожидалось, мутации группировались вокруг сайта разреза sgRNA (3 п. н. от PAM). Большинство мутаций представляют собой небольшие делеции (1–10 п. н.), но мы также обнаружили несколько более крупных делеций (100–300 п. н.). Функционально большинство мутаций приводило к нокаутам генов из-за сдвига рамки считывания или потери стартовых кодонов. Некоторые из мутаций представлены инделами в рамке считывания в первом кодирующем экзоне. Из них мы описываем новый гипоморфный аллель *Csn1s1* (*Csn1s1*$^{c.4\text{-}5insTCC}$). Мы измерили уровни белка Csn1s1 и подтвердили, что мутация отрицательно влияет на состав молока, который показывает снижение экспрессии гена на 50 % и уменьшение количества белка Csn1s1 на 40–80 % по сравнению с аллелем дикого типа. Мы предположили, что мутация влияет на стабильность транскрипта или сплайсинг по неизвестному механизму. Эта мутация потенциально может служить генетическим маркером низкой экспрессии *Csn1s1*.
Ключевые слова: казеин; CRISPR; пронуклеарная микроинъекция; гипоморфные мутации.

Introduction

Caseins are major milk proteins that have an evolutionary conserved role in nutrition (Rijnkels et al., 2003). Casein locus has been studied for a long time to understand the principle of gene regulation and hormone-induced expression (Rijnkels et al., 2013; Dos Santos et al., 2015; Lee et al., 2017). At the same time, regulatory elements of the casein genes could be used to direct the expression of desired transgenes into milk of transgenic animals (Houdebine, 2009; Kim et al., 2015). This strategy is frequently employed to create "enriched" milk with improved composition (An et al., 2012; Wan et al., 2012; Yuan et al., 2014), or to achieve large scale production of recombinant human proteins in mouse models (Wu et al., 2012; Burkov et al., 2013; Qian et al., 2014) and in livestock species (Kalds et al., 2019). Milk-specific signal peptides are used in biotechnology to enhance recombinant protein secretion during lactation (Yu et al., 2006; Liu et al., 2014; Lu et al., 2019).

During breeding, casein genes acquired many sequence variations that can lead to altered gene expression and are characteristic of some goat and cow breeds (Yue et al., 2011; Fomichev et al., 2012; Guan et al., 2019). Many of these sequence variants could be used as markers for breeding. In dairy industry casein composition is an important milk trait directly influencing quality of dairy products (Sanchez et al., 2018; Cieslak et al., 2019). Hypomorphic casein mutation are also associated with other traits such as litter size (Wang et al., 2018). For example, K. Wang with colleagues showed that 11 bp del in the intron 8 of the goat *Csn1s1* negatively affects the expression of the gene (Wang et al., 2018). Other known hypomorphic mutations are located in the promoter and 3′-untranslated regions (UTRs) of casein genes (Huang et al., 2012; Noce et al., 2016).

Novel CRISPR methods greatly facilitate genome editing in farm animals (Kalds et al., 2019), including targeted transgene integration (Park et al., 2017) and mutation modeling (Li et al., 2017; Zhou et al., 2019). The latter approach has potential to explain the molecular mechanism of how hypomorphic mutations affect milk proteins. In this report, we used CRISPR/Cas9 to create a set of mutations within a signal peptide sequence of the α-casein (*Csn1s1*) gene in mice. One of the mutants was chosen to study effects of a small in-frame insertion on the *Csn1s1* expression during lactation.

Materials and methods

Generation and genotyping of the *Csn1s1* mutant mice. *In vitro* transcription and purification of the gRNA were performed with MEGAshortscript™ T7 Transcription Kit (Thermo Fisher Scientific) and MEGAclear™ Transcription Clean-Up Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Cas9 mRNA (GeneArt™ CRISPR Nuclease mRNA) was purchased from Thermo Fisher Scientific. 50 ng/μL Cas9 mRNA and 25 ng/μL gRNA (5′-GTGAG GATGAGGAGTTTCA-3′) were mixed in RNase-free water, backfilled into an injection needle with positive balancing pressure (Transjector 5246, Eppendorf) and injected into the cytoplasm of zygotes (C57BL/6 × CBA background). After injections, the embryos were cultured for 1 hour in drops of M16 medium at 37 °C and an atmosphere of 5 % CO_2. The viable microinjected zygotes were transplanted the same day into oviducts of pseudopregnant CD-1 females (0.5 days after coitus). Isoflurane inhalation anesthesia was applied in these experiments.

Mutations were detected using PCR and Sanger sequencing of the target region of the *Csn1s1* exon 2 (Supplementary 1)[1]. Primers for PCR were as follows: 5′-GCGCATAACTAAG CATCTTATGCT-3′ (forward primer), 5′-TGACTTGGAG TTTTAGATTTGGACA-3′ (reverse primer). Selected male mice founders were crossed with C57BL/6 females. For mutation c.4-5insTCC described in this paper, founder male was crossed with two F_1 heterozygous daughters. Offspring was genotyped and two sibling females were selected for each group (wt, heterozygous or homozygous mutation) for further analysis.

All experiments were conducted at the Centre for Genetic Resources of Laboratory Animals at the Institute of Cytology and Genetics, SB RAS (RFMEFI61914X0005 and RFMEFI61914X0010). All experiments were performed in accordance with protocols and guidelines approved by the Animal Care and Use Committee Federal Research Centre of the Institute of Cytology and Genetics, SB RAS operating under standards set by regulations documents Federal Health Ministry (2010/708n/RF), NRC and FELASA recommendations. Experimental protocols were approved by the Bioethics Review Committee of the Institute of Cytology and Genetics, SB RAS.

Droplet digital PCR analysis. Total cellular RNA was extracted from mouse mammary glands at day 8 of lactation using TRI Reagent (Sigma-Aldrich). 1 μg of total RNA was used to generate cDNA in a 20 μl reaction using RevertAid RT Kit (Thermo Fisher Scientific) with random hexamer primers according to the manufacturer's instructions. Droplet Digital PCR (ddPCR) was performed using a QX100 system (Bio-Rad) with primers and probes specific for the *Csn1s1* and *Csn2* mouse transcripts (Supplementary 2). The primers and probes sequences were as follows: 5′-TGTAGTGGAT CAGGCACTGG-3′ (*Csn1s1* forward primer), 5′-TCCTTG GAGACAATGGGCTT-3′ (*Csn1s1* reverse primer), 5′-HEX-CCAGTTCTCTGTTCAGCCCTTCCCACA-BHQ2–3′ (*Csn1s1* probe), 5′-AGGACTTGACAGCCATGAAGG-3′ (*Csn2* forward primer), 5′-ATGTTCAACAGATTCCTC ACTGGA-3′ (*Csn2* reverse primer), 5′-FAM-ATCCTCGCC TGCCTTGTGGCCCTTGC-BHQ1–3′ (*Csn2* probe). ddPCR reactions were set in 20 μl volumes containing 1× ddPCR™ Supermix for Probes (no dUTP), 900 nM primers and 250 nM probes, and 1 μl of 5000-fold diluted cDNA. ddPCR reactions for each sample were performed in duplicates. PCR was conducted according to the following program: 95 °C for 10 min, then 40 cycles of 95 °C for 30 s and 61 °C for 1 min, with a ramp rate of 2 °C per second, and a final step at 98 °C for 5 min. The results were analyzed using QuantaSoft software (Bio-Rad). Concentrations of cDNA copies of *Csn1s1* and *Csn2* were derived from ddPCR and relative expression of *Csn1s1* to *Csn2* was calculated for each animal.

Milk and mammary gland protein analysis. Milk was obtained from narcotized female mice at day 8 of lactation after oxytocin administration (Uusi-Oukari et al., 1997). The milk was collected with a pipette attached to an aspiration

[1] Supplementary materials 1 and 2 are available at:
http://www.bionet.nsc.ru/vogis/download/pict-2021-25/appx8.pdf

device, transferred into a microcentrifuge tube and stored at –80 °C. Inguinal mammary glands (MGs) were extracted from the same (euthanized) mice and stored at –80 °C. For protein extraction MGs were minced in Dounce homogenizers, resuspended in RIPA buffer (150 mM NaCl, 1 % Nonidet P-40, 0.5 % sodium deoxycholate, 0.1 % SDS, 15 mM Tris pH 7.4) with protease inhibitor cocktail (1x Complete ULTRA (Roche), 1x PhosSTOP (Roche), 5 mM NaF (Sigma)). The lysates were incubated on ice for 30 min and then centrifuged at 4 °C for 10 min at 10000 g. Supernatant was sonicated and stored at –80 °C. Total protein concentrations for milk and MG lysates were determined with Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific), according to the manufacturer's instructions. To prepare samples for SDS-PAGE, milk or MG protein samples were mixed with RIPA and SDS-PAGE loading buffer (Bio-Rad) to a final concentration of 1 μg/μl and heated at 65 °C for 20 minutes. The samples (25 μg) were separated on a 12 % polyacrylamide gel and stained with Coomassie Blue G-250. ThermoFisher PageRuler™ Prestained Protein Ladder (10 to 180 kDa) was used as a protein molecular weight marker. Csn1s1, albumin and total protein concentrations were evaluated using Quantity One (Bio-Rad) and ImageJ software.

Coomassie-stained gel was used for wet transfer of the proteins to PVDF membrane (0.45 μm Immobilon-P, Merck). The membrane was blocked with 5 % milk in TBST (20 mM Tris pH 7.5, 150 mM NaCl, 0.1 % Tween 20) for 2 hours and incubated with primary mouse anti-Csn1s1 antibodies (1:1000) (sc-373711, Santa Cruz Biotechnology) overnight at 4 °C. Next day membrane was repeatedly washed with TBST and incubated with secondary mouse HRP-antibodies (1:1000) (sc-516102, Santa Cruz Biotechnology) at 25 °C for 2 hours. Immunodetection was performed with ECL substrate solution (Millipore Corporation, Billerica, MA, USA), according to the manufacturer's instructions.

Results

Generation of the *Csn1s1* mutant mice

We performed pronuclear microinjections with Cas9 mRNA and sgRNA against the first coding exon of the mouse *Csn1s1* gene (Fig. 1, *a*). Cas9-induced mutations in this region could potentially affect signal peptide coding sequence and lead to altered milk composition. We screened founder mice by Sanger sequencing and identified multiple random mutations at the Cas9 cut site (presented in Supplementary 1). The sgRNA targeted the *Csn1s1* site with high efficiency as we detected 41 mutant alleles, 4–5 mosaic alleles (additional background signal) and only 4–5 wild-type alleles in 20 founder mice (~90 % allele mutation efficiency). As expected, mutations clustered around the sgRNA cut site (3 bp from PAM). Most of the mutations represented small deletions (1–10 bp), but we detected several larger deletions as well (100–300 bp). Of note, some of the unique mutation variants had increased incidence rate. For example, 12 bp deletion (GAAACTCCTCAT) arose independently four times and another 10 bp deletion (CCATGAAACT) – three times (see Supplementary 1). We suspect this bias towards some variants is caused by microhomology-mediated end-joining (MMEJ), since these two mutations are flanked with 3 and 2 similar nucleotides (CAT and CC, respectively) (see Supplementary 1). Although mutations were mostly deleterious for the gene expression and led to a *Csn1s1* knockout (KO) by frameshift, several in-frame mutation variants resulted in subtle changes in signal peptide coding sequence without KO (see Supplementary 1). We selected one of the mutants, tagged $Csn1s1^{c.4\text{-}5insTCC}$, which had a 3 bp insertion following the start codon (see Fig. 1, *b*). To study gene expression and milk composition we chose 6 female siblings (2 wild-type, 2 heterozygotes, 2 homozygotes) from the $Csn1s1^{c.4\text{-}5insTCC}$ line for milk and mammary glands collection.

Mutation c.4-5insTCC leads to reduced expression of the *Csn1s1* gene

We estimated the *Csn1s1* gene expression in mammary glands of wild-type and mutant mice at day 8 of lactation using droplet digital PCR (ddPCR). We selected *Csn2* (β-*casein*) as a reference gene as it has a similar expression profile in mammary gland (see Supplementary 2). ddPCR analysis revealed that *Csn1s1*:*Csn2* ratio was roughly 1:3 (0.338) in wild-type siblings (Fig. 2), which is in agreement with published data (Yamaji et al., 2013). Heterozygous and homozygous $Csn1s1^{c.4\text{-}5insTCC}$ siblings showed lower *Csn1s1* expression with ratios around 1:4 (0.248) and 1:6 (0.168), respectively (see Fig. 2), compared to wild-type siblings. We also used females from the parental strains C57BL/6 and CBA as controls for normal caseins level (see Fig. 2). In rare cases, Cas9 activity can provoke rearrangements near the target locus.

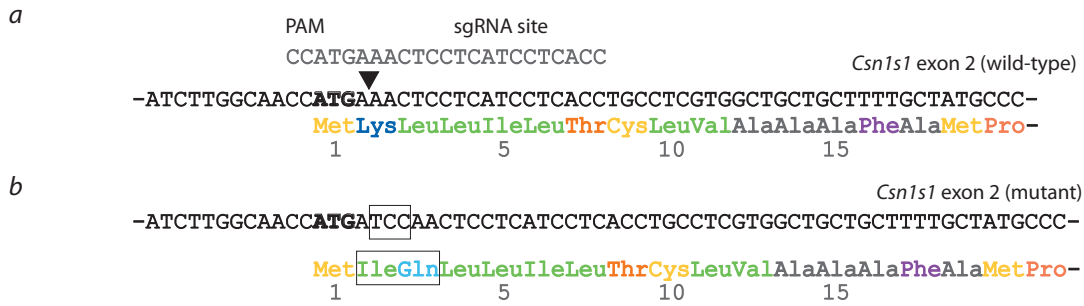


Fig. 1. Generation of the *Csn1s1* mutant mice by CRISPR/Cas9 pronuclear microinjection.

a – sequence of the first coding exon (63 bp) of the *Csn1s1* gene (exon 2, NM_007784.3). Black arrowhead indicates Cas9 cleavage site. Complementary 20 bp sgRNA sequence with PAM is shown above; *b* – sequence of the mutated *Csn1s1* exon with a 3 bp insertion in the second codon (framed). Start codon (ATG) is highlighted.

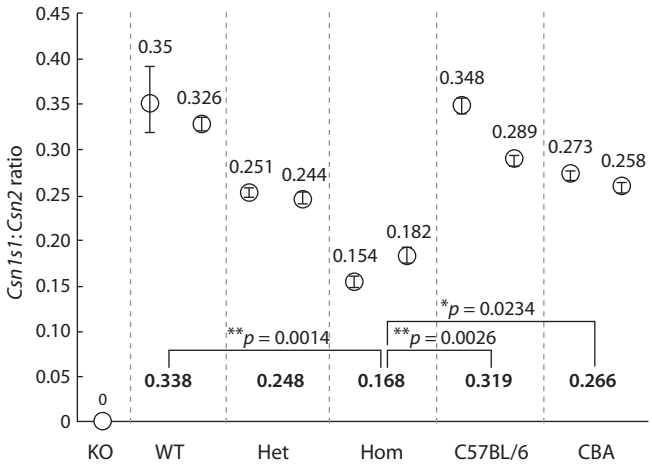


Fig. 2. Droplet digital PCR analysis of the *Csn1s1* and *Csn2* expression in mammary gland. Data presented as ratios of *Csn1s1*/*Csn2* transcripts.

Error bars – standard deviations. Bold numbers at the bottom represent mean ratios for two females of each group. KO – *Csn1s1* knockout female; WT, Het, Hom – siblings with corresponding mutation genotype; C57BL/6, CBA – females from the inbred strains. Statistics: one-way ANOVA.

We sequenced mutation-flanking regions, including the *Csn1s1* promoter and surrounding introns (2.3 kb+1.1 kb) from homozygous mice (data not shown). We also sequenced top two off-targets for the *Csn1s1* sgRNA in the F0 founder (Chr1:24388301; Chr4:140044179). No mutations were found in the examined sequences. Thus, we could confirm that the 3 bp in-frame insertion in the first coding exon led to a 30–50 % decrease in the *Csn1s1* gene expression.

Milk protein composition in c.4-5insTCC mice

We measured Csn1s1 protein levels and confirmed that mutation has a negative effect on milk composition. We collected milk and mammary glands from lactating females at the day 8 of lactation. *Csn1s1* knockout mouse was taken from another experiment ("KO") as additional control for the Csn1s1 levels. Separation of milk proteins on 12 % SDS-PAGE resulted in a typical band pattern for mouse milk (Fig. 3, *a*). In wild-type mice, Csn1s1 represents a major protein fraction and corresponds to approx. 30 % of total milk protein (Kolb et al., 2011). In homozygous mutants, loss of Csn1s1 could be observed both at the Coomassie-stained gel (see Fig. 3, *a*) and after western blotting (see Fig. 3, *b*). Csn1s1 levels fell down to 30 % in homozygotes, both for milk and for mammary gland lysates (see Fig. 3, *c*). However, exact ratios varied depending on the control protein band used for calculations. For instance, we performed the following calculations for the milk Csn1s1: Csn1s1 (gel) vs total protein (gel) – 40 % reduction; Csn1s1 (gel) vs albumin (gel) – 70 % reduction; Csn1s1 (western blot membrane) vs total protein (gel) – 80 % reduction. This effect was even more pronounced in mammary gland lysates (intracellular casein levels): Csn1s1 (western blot membrane) vs total gel – 80 % reduction.

Discussion

We report a novel in-frame *Csn1s1* hypomorphic mutation that leads to a 50 % gene expression decrease in mice. In most cases, mutation effect is tied to disruption of a regulatory element (enhancer, promoter, UTR, miRNA site) (Hogg, Harries, 2014). Frame-shifting indels in the coding sequence could initiate transcript surveillance pathway called non-

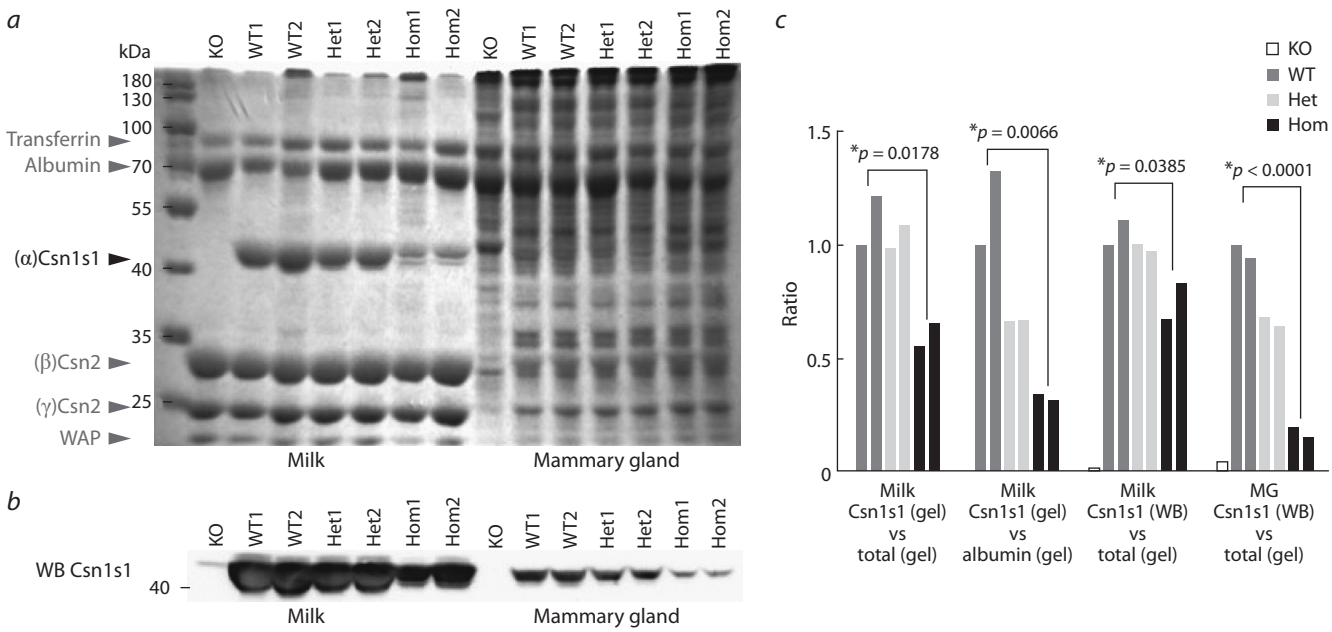


Fig. 3. Detection of the Csn1s1 protein in milk and mammary glands of the mutant mice.

a – coomassie-stained 12 % SDS-polyacrylamide gel analysis of the whole milk and mammary gland lysates from wild-type (WT1, WT2), heterozygous (Het1, Het2) and homozygous (Hom1, Hom2) *Csn1s1* mutant mice. KO – the *Csn1s1* knockout mice. ThermoFisher PageRuler™ Prestained Protein Ladder (10 to 180 kDa) was used as a protein molecular weight marker. Major milk proteins are indicated with arrows. Csn1s1 protein expected size – 43 kDa; *b* – Western blot of the same Coomassie-stained gel transferred to a PVDF membrane; *c* – quantitation of Csn1s1 protein in the milk of mutant mice using data from Fig. 3, *a* and *b*. Intensity of the Csn1s1 protein band was calculated in relation to the whole milk protein signal (total protein). MG – mammary gland. ImageJ software was used for analysis. Statistics: one-way ANOVA, *p*-values shown for WT vs homozygotes comparisons. One of the WT controls (WT1) was set to 1 (100 %).

sense-mediated mRNA decay (NMD) (Popp, Maquat, 2016; Lindeboom et al., 2019). Alternatively, in-frame mutations can lead to exon removal by alternative splicing (Mucaki et al., 2020; Thompson et al., 2020). Essentially, exon skipping could be promoted by internal exon splicing enhancers and suppressors (ESEs and ESSs) which are hard to predict (Sterne-Weiler, Sanford, 2014; Tuladhar et al., 2019), unlike typical splice site mutations (Cartegni et al., 2002). In our mutant mice, promoter had no alterations as the mutation happened in the coding sequence, quite far from a transcription start site. We assumed that it affected transcript stability or splicing by unknown mechanism. It should also be noted that mutated Csn1s1 protein was still secreted in milk, thus the function of N-terminal signal peptide was not critically affected by the mutation.

Conclusion

We demonstrated that CRISPR/Cas9 approach could be conveniently exploited to induce a spectrum of mutations in the *Csn1s1* gene either by random mutagenesis, or, ideally, by a set of single-stranded oligo DNA nucleotides (ssODNs). Our results warn that careful examination of the gene's expression is required in addition to protein analysis.

References

An L.Y., Yuan Y.G., Yu B.L., Yang T.J., Cheng Y. Generation of human lactoferrin transgenic cloned goats using donor cells with dual markers and a modified selection procedure. *Theriogenology.* 2012; 78:1303-1311. DOI 10.1016/j.theriogenology.2012.05.027.

Burkov I.A., Serova I.A., Battulin N.R., Smirnov A.V., Babkin I.V., Andreeva L.E., Dvoryanchikov G.A., Serov O.L. Expression of the human granulocyte-macrophage colony stimulating factor (hGM-CSF) gene under control of the 5′-regulatory sequence of the goat alpha-S1-casein gene with and without a MAR element in transgenic mice. *Transgenic Res.* 2013;22:949-964. DOI 10.1007/s11248-013-9697-4.

Cartegni L., Chew S.L., Krainer A.R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 2002;3(4):285-298. DOI 10.1038/nrg775. PMID 11967553.

Cieslak J., Wodas L., Borowska A., Pawlak P., Czyzak-Runowska G., Wojtowski J., Puppel K., Kuczynska B., Mackowski M. 5′-flanking variants of equine casein genes (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*) and their relationship with gene expression and milk composition. *J. Appl. Genet.* 2019;60(1):71-78. DOI 10.1007/s13353-018-0473-2.

Dos Santos C.O., Dolzhenko E., Hodges E., Smith A.D., Hannon G.J. An epigenetic memory of pregnancy in the mouse mammary gland. *Cell Rep.* 2015;11:1102-1109. DOI 10.1016/j.celrep.2015.04.015.

Fomichev K.A., Sazanova A.L., Malewski T., Kaminski S., Sazanov A.A. Associations between two novel rSNPs in 5′-flanking region of the bovine casein gene cluster and milk performance traits. *Gene*. 2012;496:49-54. DOI 10.1016/j.gene.2011.12.038.

Guan D., Mármol-Sánchez E., Cardoso T.F., Such X., Landi V., Tawari N.R., Amills M. Genomic analysis of the origins of extant casein variation in goats. *J. Dairy Sci.* 2019;102:5230-5241. DOI 10.3168/jds.2018-15281.

Hogg D.R., Harries L.W. Human genetic variation and its effect on miRNA biogenesis, activity and function. *Biochem. Soc. Trans.* 2014; 42(4):1184-1189. DOI 10.1042/BST20140055. PMID 25110023.

Houdebine L.M. Production of pharmaceutical proteins by transgenic animals. *Comp. Immunol. Microbiol. Infect. Dis.* 2009;32:107-121. DOI 10.1016/j.cimid.2007.11.005.

Huang W., Peñagaricano F., Ahmad K.R., Lucey J.A., Weigel K.A., Khatib H. Association between milk protein gene variants and protein composition traits in dairy cattle. *J. Dairy Sci.* 2012;95:440-449. DOI 10.3168/jds.2011-4757.

Kalds P., Zhou S., Cai B., Liu J., Wang Y., Petersen B., Sonstegard T., Wang X., Chen Y. Sheep and goat genome engineering: from random transgenesis to the CRISPR Era. *Front. Genet*. 2019;10:750. DOI 10.3389/fgene.2019.00750. eCollection2019.

Kim J.J., Yu J., Bag J., Bakovic M., Cant J.P. Translation attenuation via 3′ terminal codon usage in bovine *csn1s2* is responsible for the difference in α_{s2}- and β-casein profile in milk. *RNA Biol.* 2015;12: 354-367. DOI 10.1080/15476286.2015.1017231.

Kolb A.F., Huber R.C., Lillico S.G., Carlisle A., Robinson C.J., Neil C., Petrie L., Sorensen D.B., Olsson I.A., Whitelaw C.B. Milk lacking α-casein leads to permanent reduction in body size in mice. *PLoS One.* 2011;6:e21775. DOI 10.1371/journal.pone.0021775.

Lee H.K., Willi M., Wang C., Yang C.M., Smith H.E., Liu C., Hennighausen L. Functional assessment of CTCF sites at cytokine-sensing mammary enhancers using CRISPR/Cas9 gene editing in mice. *Nucleic Acids Res.* 2017;45:4606-4618. DOI 10.1093/nar/gkx185.

Li W.R., Liu C.X., Zhang X.M., Chen L., Peng X.R., He S.G., Lin J.P., Han B., Wang L.Q., Huang J.C., Liu M.J. CRISPR/Cas9-mediated loss of FGF5 function increases wool staple length in sheep. *FEBS J.* 2017;284:2764-2773. DOI 10.1111/febs.14144.

Lindeboom R.G.H., Vermeulen M., Lehner B., Supek F. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat. Genet.* 2019;51(11):1645-1651. DOI 10.1038/s41588-019-0517-5. Epub 2019 Oct 28. PMID 31659324. PMCID PMC6858879.

Liu H.C., Pai S.Y., Chen H.L., Lai C.W., Tsai T.C., Cheng W.T., Yang S.H., Chen C.M. Recombinant Derp5 allergen with α_{S1}-casein signal peptide secreted in murine milk protects against dust mite allergen-induced airway inflammation. *J. Dairy Sci.* 2014;97:6792-6803. DOI 10.3168/jds.2014-8484.

Lu R., Zhang T., Song S., Zhou M., Jiang L., He Z., Yuan Y., Yuan T., Lu Y., Yan K., Cheng Y. Accurately cleavable goat β-lactoglobulin signal peptide efficiently guided translation of a recombinant human plasminogen activator in transgenic rabbit mammary gland. *Biosci. Rep.* 2019;39:6. DOI 10.1042/BSR20190596.

Mucaki E.J., Shirley B.C., Rogan P.K. Expression changes confirm genomic variants predicted to result in allele-specific, alternative mRNA splicing. *Front. Genet.* 2020;11:109. DOI 10.3389/fgene.2020.00109. PMID 32211018. PMCID PMC7066660.

Noce A., Pazzola M., Dettori M.L., Amills M., Castelló A., Cecchinato A., Bittante G., Vacca G.M. Variations at regulatory regions of the milk protein genes are associated with milk traits and coagulation properties in the Sarda sheep. *Anim. Genet.* 2016;47:717-726. DOI 10.1111/age.12474.

Park K.E., Powell A., Sandmaier S.E., Kim C.M., Mileham A., Donovan D.M., Telugu B.P. Targeted gene knock-in by CRISPR/Cas ribonucleoproteins in porcine zygotes. *Sci. Rep.* 2017;7:42458. DOI 10.1038/srep42458.

Popp M.W., Maquat L.E. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell.* 2016;165(6):1319-1322. DOI 10.1016/j.cell.2016.05.053. PMID 27259145. PMCID PMC4924582.

Qian X., Kraft J., Ni Y., Zhao F.Q. Production of recombinant human proinsulin in the milk of transgenic mice. *Sci. Rep*. 2014;4:6465. DOI 10.1038/srep06465.

Rijnkels M., Elnitski L., Miller W., Rosen J.M. Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics.* 2003;82:417-432. DOI 10.1016/s0888-7543(03)00114-9.

Rijnkels M., Freeman-Zadrowski C., Hernandez J., Potluri V., Wang L., Li W., Lemay D.G. Epigenetic modifications unlock the milk protein gene loci during mouse mammary gland development and differentiation. *PLoS One.* 2013;8:e53270. DOI 10.1371/journal.pone.0053270.

Sanchez M.P., Wolf V., El Jabri M., Beuvier E., Rolet-Répécaud O., Gaüzère Y., Minéry S., Brochard M., Michenet A., Taussat S.,

Barbat-Leterrier A., Delacroix-Buchet A., Laithier C., Fritz S., Boichard D. Short communication: Confirmation of candidate causative variants on milk composition and cheesemaking properties in Montbéliarde cows. *J. Dairy Sci*. 2018;101:10076-10081. DOI 10.3168/jds.2018-14986.

Sterne-Weiler T., Sanford J.R. Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol.* 2014;15(1):201. DOI 10.1186/gb4150. PMID 24456648. PMCID PMC4053859.

Thompson B.A., Walters R., Parsons M.T., Dumenil T., Drost M., Tiersma Y., Lindor N.M., Tavtigian S.V., de Wind N., Spurdle A.B., InSiGHT Variant Interpretation Committee. Contribution of mRNA splicing to mismatch repair gene sequence variant interpretation. *Front. Genet.* 2020;11:798. DOI 10.3389/fgene.2020.00798. PMID 32849802. PMCID PMC7398121.

Tuladhar R., Yeu Y., Tyler Piazza J., Tan Z., Rene Clemenceau J., Wu X., Barrett Q., Herbert J., Mathews D.H., Kim J., Hyun Hwang T., Lum L. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat. Commun.* 2019;10(1):4056. DOI 10.1038/s41467-019-12028-5. PMID 31492834. PMCID PMC6731291.

Uusi-Oukari M., Hyttinen J.M., Korhonen V.P., Västi A., Alhonen L., Jänne O.A., Jänne J. Bovine alpha s1-casein gene sequences direct high level expression of human granulocyte-macrophage colony-stimulating factor in the milk of transgenic mice. *Transgenic Res.* 1997;6:75-84. DOI 10.1023/a:1018461201385.

Wan Y.J., Zhang Y.L., Zhou Z.R., Jia R.X., Li M., Song H., Wang H.Z., Wang L.Z., Zhang G.M., You J.H., Wang F. Efficiency of donor cell preparation and recipient oocyte source for production of transgenic cloned dairy goats harboring human lactoferrin. *Theriogenology.* 2012;78:583-592. DOI 10.1016/j.theriogenology.2012.03.004.

Wang K., Yan H., Xu H., Yang Q., Zhang S., Pan C., Chen H., Zhu H., Liu J., Qu L., Lan X. A novel indel within goat *casein alpha S1* gene is significantly associated with litter size. *Gene.* 2018;671:161-169. DOI 10.1016/j.gene.2018.05.119.

Wu X., Lin Y., Xiong F., Zhou Y., Yu F., Deng J., Huang P., Chen H. The extremely high level expression of human serum albumin in the milk of transgenic mice. *Transgenic Res.* 2012;21:1359-1366. DOI 10.1007/s11248-012-9612-4.

Yamaji D., Kang K., Robinson G.W., Hennighausen L. Sequential activation of genetic programs in mouse mammary epithelium during pregnancy depends on STAT5A/B concentration. *Nucleic Acids Res.* 2013;41:1622-1636. DOI 10.1093/nar/gks1310.

Yu Z., Meng Q., Yu H., Fan B., Yu S., Fei J., Wang L., Dai Y., Li N. Expression and bioactivity of recombinant human lysozyme in the milk of transgenic mice. *J. Dairy Sci.* 2006;89:2911-2918. DOI 10.3168/jds.S0022-0302(06)72563-2.

Yuan Y.G., An L., Yu B., Song S., Zhou F., Zhang L., Gu Y., Yu M., Cheng Y. Expression of recombinant human alpha-lactalbumin in the milk of transgenic goats using a hybrid pomoter/enhancer. *J. Anal. Method. Chem.* 2014;281031. DOI 10.1155/2014/281031.

Yue X.P., Zhang X.M., Wang W., Ma R.N., Deng C.J., Lan X.Y., Chen H., Li F., Xu X.R., Ma Y., Lei C.Z. The *CSN1S1* N and F alleles identified by PCR-SSCP and their associations with milk yield and composition in Chinese dairy goats. *Mol. Biol. Rep.* 2011;38:2821-2825. DOI 10.1007/s11033-010-0428-0.

Zhou S., Cai B., He C., Wang Y., Ding Q., Liu J., Liu Y., Ding Y., Zhao X., Li G., Li C., Yu H., Kou Q., Niu W., Petersen B., Sonstegard T., Ma B., Chen Y., Wang X. Programmable base editing of the sheep genome revealed no genome-wide off-target mutations. *Front. Genet.* 2019;10:215. DOI 10.3389/fgene.2019.00215.

Acknowledgements. The authors are grateful to the Centre for Genetic Resources of Laboratory Animals at the Institute of Cytology and Genetics, SB RAS (RFMEFI61914X0005 and RFMEFI61914X0010) for help with some animal procedures. The authors were supported by Russian Science Foundation (16-14-00095). Data analysis was performed on computational nodes of the Institute of Cytology and Genetics (budget project No. 0259-2021-0016).

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

Received October 22, 2020. Revised January 28, 2021. Accepted February 8, 2021.

Вавиловский журнал генетики и селекции. 2021;25(3):337-343
DOI 10.18699/VJ21.037

Новые генетические технологии защиты растений от паразитических нематод

А.В. Кочетов[1], Т.А. Гавриленко[1, 2], О.С. Афанасенко[1, 3] ✉

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия
[3] Всероссийский научно-исследовательский институт защиты растений, Пушкин, Санкт-Петербург, Россия
✉ olga.s.afan@gmail.com

Аннотация. Нематоды относятся к числу значимых вредителей сельскохозяйственных растений. В обзоре рассмотрены последние данные о молекулярных механизмах устойчивости растений к цистообразующим и галловым нематодам, среди которых одни из наиболее вредоносных видов: *Globodera rostochiensis*, *G. pallida*, *Heterodera schachtii*, *Meloidogyne chitwoodi* и *M. incognita*. Например, золотистая картофельная нематода *G. rostochiensis*, зарегистрированная в 61 субъекте РФ на общей площади 1.8 млн га, способна приводить к потере от 19 до 90 % урожая картофеля. Биологические особенности нематод затрудняют разработку агротехнических способов борьбы с ними: цисты *G. rostochiensis* сохраняют жизнеспособность в почве в течение многих лет, нематициды токсичны или малоэффективны, поэтому предпочтительным методом борьбы с ними является интрогрессия генов устойчивости от родственных культурных и дикорастущих видов. Стратегия жизненного цикла цистообразующих и галловых нематод основана на способности личинок проникать в корни восприимчивых видов растений, репрограммировать клетки растения-хозяина, формирующие гигантские клетки или синцитии в качестве питающих структур, а также ингибировать иммунный ответ. Молекулярные механизмы, лежащие в основе такого взаимодействия в системе «патоген–хозяин», вызывают значительный интерес как с точки зрения управления морфогенезом растений, так и в аспекте разработки безопасных и эффективных способов борьбы с паразитическими нематодами. В обзоре рассмотрены данные об эффекторах, с помощью которых разные виды нематод контролируют иммунный ответ растения-хозяина, а также гены устойчивости (*R*-гены) и некоторые молекулярные механизмы, прерывающие формирование питающих структур и развитие паразита. Приведены новые данные о способах генетического контроля, основанных на одном из активно обсуждаемых в последнее время варианте механизма РНК-интерференции – HIGS (host induced gene silencing), представляющем собой адресное выключение экспрессии гена-мишени в клетках личинки нематоды с помощью специфических двуцепочечных РНК, синтезирующихся в клетках растения-хозяина. Индукция РНК-интерференции в клетках растений приводит к появлению молекул-медиаторов, способных инициировать аналогичный процесс в клетках фитофагов, взаимодействующих с растением, в том числе у личинок нематод. Описаны случаи, в которых такое адресное выключение экспрессии генов-мишеней приводило к нарушениям развития личинок и высокому уровню защиты сельскохозяйственных растений от наиболее опасных видов нематод.
Ключевые слова: цистообразующие нематоды; галловые нематоды; картофель; гены-эффекторы; *R*-гены; РНК-интерференция; хозяин-индуцированный генетический сайленсинг.

Для цитирования: Кочетов А.В., Гавриленко Т.А., Афанасенко О.С. Новые генетические технологии защиты растений от паразитических нематод. *Вавиловский журнал генетики и селекции*. 2021;25(3):337-343. DOI 10.18699/VJ21.037

New genetic tools for plant defense against parasitic nematodes

A.V. Kochetov[1], T.A. Gavrilenko[1, 2], O.S. Afanasenko[1, 3] ✉

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
[3] All-Russian Institute of Plant Protection, Pushkin, St. Petersburg, Russia
✉ olga.s.afan@gmail.com

Abstract. Nematodes belong to economically important pests. Here we reviewed the recent data on molecular mechanisms of plant resistance to cyst and gall nematodes including the most devastating *Globodera rostochiensis*, *G. pallida*, *Heterodera schachtii*, *Meloidogyne chitwoodi*, and *M. incognita*. The Golden Potato Cyst Nematode (*G. rostochiensis*, GPCN) may be taken as an example of an economically important pest: in Russia, it occurs in 61 regions with a total area of 1.8 million ha and may cause the yield loss from 19 to 90 %. The biological characteristics of sedentary nematodes makes their agrotechnical control problematic, i. e. the GPCN cysts remain dormant in soil for many years until a susceptible host appears, whereas nematicides are either toxic or inefficient. Introgression of resistance genes (*R*-genes) from related cultivated or wild species is likely to be the most appropriate way for their biocontrol. The life cycle of sedentary nematodes is based on juveniles' penetration into the host root where they reprogram plant cells into a syncytium or the so-called 'giant cells' and inhibit the plant defense response. Molecular mechanisms of plant-nematode interaction

are unusual and this phenomenon provides a very interesting model for the investigation of plant morphogenesis control as well as for the development of new genetic instruments of biocontrol. Here we reviewed recent publications on plant parasitic nematode effectors used for hijacking of the plant immune system, data on *R*-genes and molecular mechanisms of their activities. In addition, host-induced gene silencing (HIGS) is discussed as a perspective mechanism for nematode biocontrol. HIGS is based on the RNA interference in the cells of the host plant addressed against the nematode genes important for their development and productivity. Several recent investigations demonstrated efficiency of HIGS against sedentary nematodes.
Key words: cyst nematodes; gall nematodes; potato; effector genes; *R*-genes; RNA interference; host induced genetic silencing.

For citation: Kochetov A.V., Gavrilenko T.A., Afanasenko O.S. New genetic tools for plant defense against parasitic nematodes. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):337-343. DOI 10.18699/VJ21.037 (in Russian)

Введение

Фитопаразитические нематоды относятся к числу значимых вредителей сельского хозяйства и распространены по всему миру (около 4 тыс. видов, большинство из которых встречаются в странах с тропическим и субтропическим климатом) (Зиновьева и др., 2012). Из многочисленных видов фитопаразитических нематод в обзоре наибольшее внимание уделено цистообразующим и галловым, поскольку к ним относятся виды, наносящие значительный урон экономически важным сельскохозяйственным культурам растений, среди которых *Globodera rostochiensis* (Woll), *G. pallida* (Stone), *Heterodera schachtii* (Schmidt), *Meloidogyne chitwoodi* (Golden et al.), *Meloidogyne incognita* (Kofoid & White). Следует отметить, что *M. incognita* называют наиболее опасным и вредным организмом, поражающим культурные растения (Trudgill, Blok, 2001).

Разработка способов борьбы с паразитическими нематодами считается одной из актуальных биотехнологических задач, однако их биологические особенности затрудняют применение типовых агротехнических и химических методов защиты. Примером может служить золотистая картофельная нематода (ЗКН) *G. rostochiensis* – широко распространенный карантинный вредитель, приводящий к существенным потерям в картофелеводстве (Evans, Trudgill, 1992). В России ЗКН зарегистрирована в 61 субъекте, включающем 861 административный район на территории общей площадью 1.8 млн га (Справочник по карантинному фитосанитарному состоянию, 2017). В зависимости от сорта и сезонных условий потери урожая картофеля могут составлять от 19 до 90 % (Friedman, 1985), при этом покоящиеся в почве цисты ЗКН способны сохранять жизнеспособность более 30 лет (Winslow, Willis, 1972). Кроме того, следует учитывать высокую вероятность распространения ЗКН в новых, не характерных для этого вида регионах вследствие прогнозов по изменению климата (Jones et al., 2017), а также возможности появления на территории РФ других опасных для картофелеводства карантинных видов нематод.

Способы борьбы с ЗКН включают использование нематицидов (Kearn et al., 2017) и применение в севообороте родственных культур, способных инициировать выход личинок ЗКН, но не являющихся для них хозяевами, что приводит к гибели личинок и снижению инфицированности почвы (например, *Solanum sisymbriifolium*) (Dandurand et al., 2019; Kooliyottil et al., 2019). Однако эти методы могут быть недостаточно результативны или экологичны: так, эффективные нематициды токсичны и запрещены в европейских странах (Trudgill et al., 2003). Таким образом, предпочтительным подходом к контролю ЗКН в разных странах является возделывание устойчивых сортов с интрогрессированными генами устойчивости (*R*-генами) от родственных культурных (*S. tuberosum* ssp. *andigenum*) и дикорастущих (*S. canasense*, *S. oplocense*, *S. spegazzinii*, *S. vernei*) видов картофеля (Dalamu et al., 2012).

Молекулярные механизмы патогенеза также превращают ЗКН в непростую мишень для существующих способов защиты растений. Личинки ЗКН после выхода из яиц в почве проникают в корни восприимчивых растений, движутся до проводящих тканей и инициируют формирование питающего синцития. Для этого личинка ЗКН впрыскивает в клетки растения секрет из пищеводных желез, инициирующий репрограммирование и слияние клеток в синцитий (описаны синцитии, образовавшиеся в результате последовательного слияния более чем 200 клеток), а также ингибирующий иммунный ответ у восприимчивых сортов (Mejias et al., 2019). Сходный механизм формирования системы питания выработали галлообразующие нематоды семейства *Meloidogynidae* (мелойдогинид), также способные репрограммировать клетки растения-хозяина, инициировать процесс митоза без цитокинеза с формированием так называемых гигантских клеток, размер которых может более чем в 300 раз превышать исходную клетку. Личинка нематоды в данном случае формирует специализированный орган – галл, в котором может быть несколько гигантских клеток (Зиновьева и др., 2012; Mejias et al., 2019). Оба типа питающих структур (синцитий и гигантские клетки) характеризуются общими структурно-функциональными особенностями: развитый эндоплазматический ретикулум, утолщенные клеточные стенки, большое количество митохондрий, фрагментированная вакуоль и т. п. (Rodiuc et al., 2014; Palomares-Rius et al., 2017; Mejias et al., 2019). Механизмы индуцированного нематодами репрограммирования активно исследуют, так как этот биологический феномен представляет собой перспективную модель для изучения генетического контроля морфогенеза растений.

Известно, что личинки южной галловой нематоды *Meloidogyne incognita* секретируют сложный набор веществ различной природы, содержащий не менее 500 белков, функции большинства из которых неизвестны (Wang et al., 2012). В настоящее время можно выделить несколько PPN-эффекторов (plant parasitic nematode effectors), о функциях которых есть экспериментальные данные.

PPN-эффекторы, способные связывать активные формы кислорода (АФК), которые образуются в результате сверхчувствительной реакции. Эффективные способы борьбы с патогенами, выработанные растениями в ходе эволюции, включают распознавание специфических молекулярных паттернов (pathogen-associated molecular patterns, PAMP), индукцию системного защитного ответа и локально – программируемой клеточной смерти в месте инвазии. В случае личинок нематод зона мертвых клеток блокирует поступление питательных веществ и останавливает жизненный цикл. Известно, что АФК являются одним из ключевых медиаторов в регуляторных контурах, контролирующих этот вид иммунного ответа (effector-triggered immunity, ETI). Показано, что один из PPN-эффекторов яванской галловой нематоды *Meloidogyne javanica* (Treub, 1885) взаимодействует с каталитической субъединицей тиоредоксинредуктазы и связывает АФК, что блокирует передачу сигнала и ингибирует развитие иммунного ответа (Lin et al., 2016). Другую клеточную систему используют для этой же цели личинки свекловичной цистообразующей нематоды *Heterodera schachtii*: эффектор Hs10A06 связывается со спермидинсинтазой, что увеличивает синтез спермидина, также связывающего АФК (Hewezi et al., 2015).

PPN-эффекторы, ингибирующие локальный и системный иммунный ответ. Помимо связывания АФК PPN-эффекторы способны инактивировать PR-белки и ингибировать системный иммунный ответ, в частности в качестве мишеней могут быть использованы ферменты бета-1,3-эндоглюканазы, 1,3-бета-глюксинтазы и др. (Mejias et al., 2019). Для ингибирования специфического иммунитета можно применять инактивацию специфического белка-рецептора: например, PPN-эффектор *G. pallida* RHA1B является убиктвитинлигазой, способной как активировать убиквитинилирование и протеолиз белка-рецептора семейства NBS-LRR, так и ингибировать индукцию специфического иммунитета в целом, например в ответ на молекулярные паттерны других патогенов (Kud et al., 2019).

PPN-эффекторы, способные ингибировать иммунный ответ на уровне молекулярно-генетических регуляторных контуров. В качестве примера можно привести эффектор *M. incognita*, взаимодействующий с одной из субъединиц сигналосомы COP9 (Bournaud et al., 2018). Характерно, что этот белок, участвующий в пути передачи сигнала салициловой кислоты, также является мишенью для ряда эффекторов грибов и вирусов. Аналогичным образом PPN-эффекторы сразу нескольких видов нематод (*M. chitwoodi*, *G. rostochiensis*, *H. schachtii*) используют в качестве мишени белок PLCP (papain-like cysteine protein), который также служт одной из общих мишеней для различных патогенов (Mejias et al., 2019)

PPN-эффекторы, задействованные в индукции формирования синцития или гигантских клеток. Про функции этих белков известно немного. Цистообразующая бледная картофельная нематода *G. pallida* продуцирует эффектор GpSPRY-414-2, который связывается с белком CLASP, ассоциированным с микротрубочками и участвующим в делении и росте клеток (Mei et al., 2018); вероятно, эту мишень используют для реорганизации цитоскелета в синцитии. Можно отметить PPN-эффекторы цистообразующих нематод, структура которых сходна с CLAVATA3 (CLV3)/ESR (CLE) – регуляторными пептидами, задействованными в контроле деления и дифференцировки клеток растений, а также эффекторы *H. schachtii*, взаимодействующие с транспортером ауксина LAX3 (Gheysen, Mitchum, 2019). Тема репрограммирования клеток растения-хозяина тесно связана и с вопросом специфичности нематод: помимо преодоления иммунной системы важным элементом жизненного цикла патогена является адресное воздействие на системы молекулярно-генетического управления морфогенезом, что также относится к актуальным направлениям исследований (Sabeh et al., 2019).

Гены устойчивости (*R*-гены)

Гены устойчивости к *G. rostochiensis* свидетельствуют о том, что в настоящее время большинство сортов картофеля содержат по одному *R*-гену (Whitworth et al., 2018), обеспечивающему высокую устойчивость к патотипу ЗКН Ro1. К их числу относятся *H1*, интрогрессированный от *S. tuberosum* ssp. *andigenum* (Bakker et al., 2004), и *Gro1-4* от *S. spegazzinii* (Barone et al., 1990; Ballvora et al., 1995). Защита, контролируемая идентифицированными генами устойчивости, основана на механизме реакции сверхчувствительности: индукция программируемой клеточной смерти формирует зону мертвых клеток, отсекающих личинку ЗКН от поступления питательных веществ, что не позволяет ей завершить жизненный цикл (Kaloshian et al., 2011; Зиновьева и др., 2012; Kochetov et al., 2017). Несмотря на то что гены *H1* и *Gro1-4* сохраняют эффективность в Европе и России, следует учитывать, что нематоды способны к быстрой эволюции, поэтому поиск новых генов устойчивости к ЗКН считается одной из приоритетных задач генетики растений (Whitworth et al., 2018; Strachan et al., 2019). Одним из классических способов расширения набора генов устойчивости с перспективами их дальнейшего пирамидирования является поиск новых генов в природных популяциях растений. В качестве примера можно привести систематический скрининг биоресурсных коллекций ВИР, показавший наличие нового *R*-гена для *G. rostochiensis* у некоторых генотипов *S. phureja* (Limantseva et al., 2014). Сравнительный анализ транскриптомов устойчивых и восприимчивых генотипов позволил определить механизм устойчивости, основанный на индукции реакции сверхчувствительности и локальной программируемой клеточной смерти в зоне инвазии личинок. Особенностями взаимодействия нематод с корнями растений служат существенное повреждение тканей при миграции личинок и индукция неспецифического ответа на этот вид стресса (wounding), который у устойчивых линий дополняется специфическим PAMP-опосредованным иммунным ответом в виде реакции сверхчувствительности и системного синтеза защитных белков (Kochetov et al., 2017, 2020).

Подобный подход использован для изучения генетических механизмов устойчивости к галловым нематодам различных экономически важных видов пасленовых, на-

пример картофеля (Bali et al., 2019), томата (Schaff et al., 2007) и табака (Li et al., 2018). В работе S. Bali и коллег исследован селекционный клон с интрогрессированным из дикого диплоидного мексиканского вида картофеля *S. bulbocastanum* геном *RMC1*(*bulb*), обеспечивающим устойчивость к *Meloidogyne chitwoodi*. Аналогично данным, представленным A.V. Kochetov и сотрудниками (2017, 2020), личинки нематоды проникают в корни как восприимчивых, так и устойчивых сортов растений, однако PAMP-опосредованная реакция сверхчувствительности ограничивает формирование питающих гигантских клеток, и жизненный цикл нематоды прерывается. Развитие иммунного ответа приводит к масштабным изменениям в транскриптоме, что позволяет выявлять дифференциально экспрессирующиеся гены у восприимчивых и устойчивых генотипов, реконструировать системы взаимодействующих генов (генные сети), пути передачи сигнала и молекулярные механизмы устойчивости. Показано, что индукция устойчивости связана с накоплением АФК, увеличением активности генов сигнальных путей жасмоновой и салициловой кислот, синтеза компонентов клеточной стенки, полиаминов, PR-белков (Bali et al., 2019). X. Li и коллеги исследовали взаимодействие устойчивых и восприимчивых генотипов *Nicotiana tabacum* с *M. incognita*. На этой модели также продемонстрировано, что личинки нематод проникают в корни как устойчивых, так и восприимчивых растений, но у первых на седьмой день инициируются реакция сверхчувствительности и локальный некроз в районе гигантских питающих клеток. В результате сравнительного анализа транскриптома и дифференциально экспрессирующихся генов выявлены путь передачи сигнала салициловой кислоты, гены антиоксидантов, а также метаболические пути синтеза фенилпропаноидов, алкалоидов и терпеноидов, что может быть характерно для табака (Li et al., 2018).

Несмотря на то что защитные механизмы часто связаны со специфическими рецепторами и индукцией реакции сверхчувствительности, в основе иммунного ответа могут быть разные молекулярные события. Можно отметить, что разнообразие защитных механизмов в природных популяциях растений изучено недостаточно и новые варианты генов устойчивости могут отличаться от классических рецепторов семейства NBS-LRR, активирующих реакцию сверхчувствительности при появлении нематод-специфических молекулярных паттернов. Например, ген *Rhg1* сои, кодирующий специфический вариант белка везикулярного транспорта α-SNAP (Bayless et al., 2018), способен интерферировать с молекулярными механизмами патогенеза цистообразующих нематод, однако при этом его экспрессия в трансгенных растениях других видов (картофеля, арабидопсиса, других разновидностей сои) приводила как к увеличенной устойчивости к *H. schachtii, G. rostochiensis* и *G. pallida*, так и негативно влияла на рост и развитие растений (Butler et al., 2019). По-видимому, в процессе эволюции у устойчивых форм сои сформировался баланс в экспрессии генов, кодирующих белки α-SNAP, обеспечивающий устойчивость к патогену и компенсацию негативного эффекта такого необычного *R*-гена.

Новые способы генетического контроля устойчивости растений к нематодам

В последнее время значительный интерес вызывает один из вариантов технологии РНК-интерференции, а именно HIGS (host induced gene silencing), представляющий собой адресное выключение экспрессии гена-мишени в клетках личинки нематоды с помощью специфических двуцепочечных РНК (дцРНК), синтезирующихся в клетках растения-хозяина. дцРНК способны индуцировать систему защитного ответа на основе РНК-интерференции, приводящую к появлению коротких интерферирующих РНК (short interfering RNA, siRNA) и комплексов RISC, способных распознавать и разрушать мишени – молекулы РНК, содержащие участки, идентичные или высокогомологичные такой дцРНК-затравке. Этот механизм является не только способом борьбы с вирусами, но и одним из фундаментальных молекулярных механизмов контроля экспрессии генов у эукариот. Однако дцРНК, siRNA или другие промежуточные формы процесса РНК-интерференции могут проникать в клетки организмов, взаимодействующих с растением, например в клетки пищеварительной системы насекомых-фитофагов, гифы паразитических грибов, клетки нематод и др. При этом если дцРНК сконструирована из сегментов мРНК, соответствующих не гену растения-хозяина, а гену фитофага, то в ряде случаев при взаимодействии этого фитофага с растением отмечены индукция РНК-интерференции и специфическое ингибирование экспрессии гена-мишени в клетках фитофага. Если данный ген выполнял жизненно важные функции, то растения, продуцирующие такие дцРНК, становились токсичными для соответствующего вредителя.

Указанный биологический феномен демонстрирует возможность обмена регуляторной генетической информацией между организмами различной таксономической принадлежности в природных и искусственных экосистемах, что требует дальнейшего всестороннего исследования. Тем не менее очевидно и прикладное значение феномена HIGS, так как дцРНК, специфические для мРНК конкретного гена-мишени, не оказывают воздействия на организмы, в транскриптомах которых нет мРНК с протяженными участками сходства. В структуре мРНК эукариот помимо белок-кодирующей части (открытой рамки считывания, CDS) выделяют 5′- и 3′-концевые нетранслируемые последовательности (НТП). 5′-НТП играет важную роль в контроле инициации трансляции, в то время как функции 3′-НТП могут быть связаны с управлением цитоплазматической стабильности индивидуальных мРНК (Кочетов и др., 2002; Kochetov, Sarai, 2004; Ventoso et al., 2012). В качестве адресной нуклеотидной последовательности для индукции РНК-интерференции при конструировании дцРНК можно использовать более протяженные 3′-НТП мРНК гена-мишени, которые, в отличие от белок-кодирующих участков, в большинстве случаев не являются эволюционно консервативными, что расширяет диапазон возможностей селективного видоспецифического выключения отдельных генов.

Следует отметить, что эффективность воздействия дцРНК может зависеть от морфологических и биохимиче-

ских особенностей организма-мишени, в частности включающих барьеры на пути проникновения дцРНК внутрь клеток и содержание в тканях дцРНК-специфических рибонуклеаз. При этом не все гены организма-реципиента могут быть супрессированы с помощью РНК-интерференции. Так, в работе S. Iqbal и коллег (2020) приведены результаты систематического исследования генов домашнего хозяйства *M. incognita*, которые в перспективе могут быть использованы как мишени для HIGS. Выбор подходящего гена-мишени – одна из важных задач, так как некоторые гены домашнего хозяйства могут быть малочувствительны к РНК-интерференции либо эффект от их супрессии может быть недостаточно выражен. Авторы использовали методику, адаптированную из ранних экспериментов на *Caenorhabditis elegans*, – вымачивание личинок нематоды в растворе, содержащем дцРНК, и последующий анализ характеристик, включая инфекционную способность личинок и характеристики взрослых нематод после инфицирования личинками растений томата. Проанализировано 20 генов: экспрессия восьми генов не репрессировалась дцРНК и не вызывала морфологических проявлений, в то время как супрессия десяти генов приводила к аберрациям в морфологических характеристиках нематод и снижению их способности к инфицированию растений. Трансгенные растения, экспрессирующие дцРНК против шести генов нематоды, в экспериментах проявляли устойчивость, сопоставимую с присутствием у растений *R*-генов: снижение инфицированности растений доходило до 89 % (Iqbal et al., 2020).

Далее приведены данные успешного применения технологии HIGS как потенциального инструмента борьбы с паразитическими нематодами. Интересные результаты получены при попытке использования HIGS в качестве мишени гена *M. incognita*, кодирующего PPN-эффектор Mi-MSP2, участвующий в супрессии защитного ответа растения-хозяина. Трансгенные растения, экспрессирующие такие дцРНК, характеризовались сниженной на 88 % экспрессией гена-мишени у самок нематоды, а также уменьшенной на 80 % продукцией яиц (Joshi et al., 2019). Применение этой технологии для защиты растений баклажана показало, что дцРНК стабильно экспрессировалась в трансгенных формах вплоть до третьего поколения от самоопыления и обеспечивала высокий уровень защиты от нематоды (Chaudhary et al., 2019). Для *M. incognita*, одного из наиболее опасных вредителей растений, обладающего широким кругом хозяев, экспериментально доказана эффективность HIGS для нескольких генов, включая PPN-эффекторы (Shivakumara et al., 2017; снижение продуктивности самок нематоды составило 40–70 %), ген стерол-связывающего белка (Shivakumara et al., 2019; снижение продуктивности нематод до 50 %), ген L-цистеиновой протеазы (Dutta et al., 2015; снижение продуктивности на 60–80 %), гены кутикулярного коллагена (Banerjee et al., 2018; снижение продуктивности нематод до 80 %, структурные аберрации у личинок).

Развитие других видов нематод также может быть супрессировано с помощью HIGS. Отмечено снижение продуктивности самок *H. schachtii* на трансгенных растениях, экспрессирующих дцРНК сложной структуры, содержащую сегменты генов эндоглюканазы и белка MSP (major sperm protein) (Amin et al., 2018); кроме того, показана возможность применения HIGS для *Heterodera glycines* (Hu et al., 2019), *Bursaphelenchus xylophilus* (Qiu et al., 2019) и других видов нематод. Свекловичная цистообразующая нематода *H. schachtii* также служит примером актуальности разработки новых генетических технологий биологического контроля. Этот паразит поражает более 200 видов растений, принадлежащих 98 родам, и является одним из основных вредителей сахарной свеклы. Яйца в цистах сохраняют жизнеспособность в почве в течение нескольких лет. Стандартные способы борьбы с помощью нематицидов на основе фосфорорганических соединений и карбаматов затруднены из-за их высокой токсичности; применение в севообороте других видов растений, способных инициировать выход личинок из яиц, но не являющихся хозяевами для *H. schachtii*, экономически неэффективно для высокоинтенсивных технологий возделывания сахарной свеклы. Использование устойчивых форм растений также проблематично из-за их повышенной чувствительности к некоторым грибным патогенам (Amin et al., 2018), что в совокупности демонстрирует отсутствие эффективных способов борьбы и необходимость разработки новых технологий, к числу которых относится HIGS.

Заключение

Применение технологии HIGS основано на получении генетически-модифицированных форм растений, которые не экспрессируют чужеродные белки, но производят некодирующую двуцепочечную РНК, содержащую участки мРНК гена-мишени, в данном случае гена нематоды. РНК-интерференцию, как и белок-кодирующие гены, активно применяют для супрессии генов растений (Кочетов и др., 2004; Trifonova et al., 2007; Sugawara et al., 2016). Поскольку при продукции некодирующей дцРНК не производятся новые для растения-хозяина белки, отсутствует вероятность развития у потребителей аллергических реакций или специфических изменений в метаболоме растений, связанных с новыми ферментативными или регуляторными активностями. К рискам применения таких растений в практике сельского хозяйства следует отнести потенциальную возможность неспецифического действия дцРНК на другие организмы, взаимодействующие с растениями, если их мРНК каких-либо генов содержат протяженные участки сходства. Дальнейшее систематическое исследование структуры геномов организмов разной принадлежности в различных природных и агроэкосистемах, вероятно, позволит с точностью прогнозировать такие риски и откроет путь к инженерии устойчивости сельскохозяйственных растений к разнообразным вредителям и патогенам.

Список литературы / References

Зиновьева С.В., Чижов В.Н., Приданников М.В., Субботин С.А., Рысс А.Ю., Хусаинов Р.В. Фитопаразитические нематоды России. Изд-во: Товарищество научных изданий КМК, 2012;374. [Zinovieva S.V., Chizhov V.N., Pridannikov M.V., Subbotin S.A., Ryss A.Yu., Khusainov R.V. Phytoparasitic nematodes of Russia. KMK Scientific Publishing Association, 2012;374. (in Russian)]

Кочетов А.В., Сырник О.А., Рогозин И.Б., Глазко Г.В., Комарова М.Л., Шумный В.К. Контекстная организация 5′-нетранслируемых районов генов высших растений. *Молекулярная биология*. 2002;36(4):649-656.
[Kochetov A.V., Sirnik O.A., Rogosin I.B., Glazko G.V., Komarova M.L., Shumny V.K. Contextual features of higher plant mRNA 5′-untranslated regions. *Molekulyarnaya Biologiya = Molecular Biology*. 2002;36(4):649-656. (in Russian)]

Кочетов А.В., Титов С.Е., Колодяжная Я.С., Комарова М.Л., Коваль В.С., Макарова Н.Н., Илинский Ю.Ю., Трифонова Е.А., Шумный В.К. Повышение содержания пролина и осмотического давления клеточного сока у трансформантов табака, несущих антисмысловой супрессор гена пролиндегидрогеназы. *Генетика*. 2004;40(2):282-285.
[Kochetov A.V., Titov S.E., Kolodyazhnaia Ia.S., Komarova M., Koval' V.S., Makarova N.N., Ilinskyi Iu.Iu., Trifonova E.A., Shumny V.K. Tobacco transformants bearing antisense suppressor of proline dehydrogenase gene are characterized by higher proline content and cytoplasm osmotic pressure. *Genetika = Genetics*. 2004;40(2):282-285. (in Russian)]

Справочник по карантинному фитосанитарному состоянию территорий государств – участников СНГ на 01.01.2017 г. М.: ФГБУ «ВНИИКР», 2017;106.
[Handbook on the plant quarantine conditions in CIS member states as of January 1, 2017. Moscow: All-Russia Plant Quarantine Center Publ., 2017;106. (in Russian)]

Amin R.B., Karegar A., Afsharifar A., Niazi A., Karimi M. Disruption of the pathogenicity and sex ratio of the beet cyst nematode *Heterodera schachtii* by host-delivered RNA interference. *Mol. Plant. Microbe Interact*. 2018;31(12):1337-1346. DOI 10.1094/MPMI-05-18-0141-R.

Bakker E., Achenbach U., Bakker J., van Vliet J., Peleman J., Segers B., van der Heijden S., van der Linde P., Graveland R., Hutten R., van Eck H., Coppoolse E., An der Vossen E., Bakker J., Goverse A. A high-resolution map of the H1 locus harboring resistance to the potato cyst nematode Globodera rostochiensis. *Theor. Appl. Genet*. 2004;109:146-152. DOI 10.1007/s00122-004-1606-z.

Bali S., Vining K., Gleason C., Majtahedi H., Brown C.R., Sathuvalli V. Transcriptome profiling of resistance response to *Meloidogyne chitwoodi* introgressed from wild species *Solanum bulbocastanum* into cultivated potato. *BMC Genomics*. 2019;20(1):907. DOI 10.1186/s12864-019-6257-1.

Ballvora A., Hesselbach J., Niewöhner J., Leiste D., Salamini F., Gebhardt C. Marker enrichment and high-resolution map of the segment of potato chromosome VII harbouring the nematode resistance gene *Gro1*. *Mol. Gen. Genet*. 1995;249:82-90.

Banerjee S., Gill S.S., Gawade B.H., Jain P.K., Subramaniam K., Sirohi A. Host Delivered RNAi of Two Cuticle Collagen Genes, *Mi-col-1* and *Lemmi-5* Hampers Structure and Fecundity in *Meloidogyne incognita*. *Front. Plant Sci.* 2018;8:2266. DOI 10.3389/fpls.2017.02266.

Barone A., Ritter E., Schachtschabel U., Debener T., Salamini F., Gebhardt C. Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Mol. Gen. Genet*. 1990;224:177-182.

Bayless A.M., Zapotocny R.W., Grunwald D.J., Amundson K.K., Diers B.W., Bent A.F. An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. *Proc. Natl. Acad. Sci USA*. 2018;115(19):E4512-E4521.

Bournaud C., Gillet F.-X., Murad A.M., Bresso E., Albuquerque E.V.S., Grossi-de-Sá M.F. *Meloidogyne incognita PASSE-MURAILLE* (*MiPM*) gene encodes a cell-penetrating protein that interacts with the CSN5 subunit of the COP9 signalosome. *Front. Plant Sci.* 2018;9:904. DOI 10.3389/fpls.2018.00904.

Butler K.J., Chen S., Smith J.M., Wang X., Bent A.F. Soybean Resistance Locus Rhg1 Confers Resistance to Multiple Cyst Nematodes in Diverse Plant Species. *Phytopathology*. 2019;109(12):2107-2115. DOI 10.1094/PHYTO-07-19-0225-R.

Chaudhary S., Dutta T.K., Tyagi N., Shivakumara T.N., Papolu P.K., Chobhe K.A., Rao U. Host-induced silencing of Mi-msp-1 confers resistance to root-knot nematode *Meloidogyne incognita* in eggplant. Transgenic. Res. 2019;28(3-4):327-340. DOI 10.1007/s11248-019-00126-5.

Dalamu B.V., Umamaheshwari R., Shrama R., Kaushik S.K., Joseph T.A., Singh B.P., Gebhardt C. Potato cyst nematode (PCN) resistance: genes, genotypes and markers – an update. *SABRAO J. Breed. Genet.* 2012;44:202-228.

Dandurand L.M., Zasada I.A., LaMondia J.A. Effect of the trap crop, *Solanum sisymbriifolium*, on *Globodera pallida*, *Globodera tabacum*, and *Globodera ellingtonae*. *J. Nematol.* 2019;51:1-11. DOI 10.21307/jofnem-2019-030.

Dutta T.K., Papolu P.K., Banakar P., Choudhary D., Sirohi A., Rao U. Tomato transgenic plants expressing hairpin construct of a nematode protease gene conferred enhanced resistance to root-knot nematodes. *Front. Microbiol.* 2015;6:260. DOI 10.3389/fmicb.2015.00260.

Evans K., Trudgill D.L. Pest aspects of potato production. Part 1. The nematode pests of potatoes. In: Harris P. (Ed.). The potato crop. London: Chapman & Hall, 1992.

Friedman W. Pests not known to occur in the United States or of limited distribution, No. 68: Golden Nematode. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, 1985;10.

Gheysen G., Mitchum M.G. Phytoparasitic nematode control of plant hormone pathways. *Plant Physiol.* 2019;179:1212-1226. DOI 10.1104/pp.18.01067.

Hewezi T., Juvale P.S., Piya S., Maier T.R., Rambani A., Rice J.H., Mitchum M.G., Davis E.L., Hussey R.S., Baum T.J. The cyst nematode effector protein 10A07 targets and recruits host posttranslational machinery to mediate its nuclear trafficking and to promote parasitism in *Arabidopsis*. *Plant Cell*. 2015;27:891-907. DOI 10.1105/tpc.114.135327.

Hu Y., You J., Li C., Pan F., Wang C. The *Heterodera glycines* effector Hg16B09 is required for nematode parasitism and suppresses plant defense response. *Plant Sci.* 2019;289:110271. DOI 10.1016/j.plantsci.2019.110271.

Iqbal S., Fosu-Nyarko J., Jones M.G.K. Attempt to silence genes of the RNAi pathways of the root-knot nematode, *Meloidogyne incognita* results in diverse responses including increase and no change in expression of some genes. *Front. Plant Sci.* 2020;11:328. DOI 10.3389/fpls.2020.00328.

Jones L.M., Koehler A.K., Trnka M., Balek J., Challinor A.J., Atkinson H.J., Urwin P.E. Climate change is predicted to alter the current pest status of *Globodera pallida* and *G. rostochiensis* in the United Kingdom. *Glob. Chang Biol.* 2017;23(11):4497-4507. DOI 10.1111/gcb.13676.

Joshi I., Kumar A., Singh A.K., Kohli D., Raman K.V., Sirohi A., Chaudhury A., Jain P.K. Development of nematode resistance in *Arabidopsis* by HD-RNAi-mediated silencing of the effector gene *Mi-msp2. Sci. Rep.* 2019;9(1):17404. DOI 10.1038/s41598-019-53485-8.

Kaloshian I., Desmond O.J., Atamian H.S. Disease resistance-genes and defense responses during incompatible interactions. In: Jones J., Gheysen G., Fenoll C. (Eds.). Genomics and molecular genetics of plant–nematode interactions. New York: Springer, 2011;309-324.

Kearn J., Lilley C., Urwin P., O'Connor V., Holden-Dye L. Progressive metabolic impairment underlies the novel nematicidal action of fluensulfone on the potato cyst nematode *Globodera pallida*. *Pestic. Biochem. Physiol.* 2017;142:83-90. DOI 10.1016/j.pestbp.2017.01.009.

Kochetov A.V., Egorova A.A., Glagoleva A.Y., Strygina K.V., Khlestkina E.K., Gerasimova S.V., Shatskaya N.V., Vasilyev G.V., Afonnikov D.A., Shmakov N.A., Antonova O.Y., Alpatyeva N.V., Khiutti A., Afanasenko O.S., Gavrilenko T.A. The mechanism of potato

resistance to *Globodera rostochiensis*: comparison of root transcriptomes of resistant and susceptible *Solanum phureja* genotypes. *BMC Plant Biol.* 2020;20(Suppl 1):350. DOI 10.1186/s12870-020-02334-2.

Kochetov A.V., Glagoleva A.Y., Strygina K.V., Khlestkina E.K., Gerasimova S.V., Ibragimova S.M., Shatskaya N.V., Vasilyev G.V., Afonnikov D.A., Shmakov N.A., Antonova O.Y., Gavrilenko T.A., Alpatyeva N.V., Khiutti A., Afanasenko O.S. Differential expression of NBS-LRR-encoding genes in the root transcriptomes of two *Solanum phureja* genotypes with contrasting resistance to *Globodera rostochiensis*. *BMC Plant Biol.* 2017;17(Suppl 2):251. DOI 10.1186/s12870-017-1193-1.

Kochetov A.V., Sarai A. Translational polymorphism as a potential source of plant proteins variety in *Arabidopsis thaliana*. *Bioinformatics.* 2004;20:445-447.

Kooliyottil R., Dandurand L.-M., Kuhl J.C., Caplan A., Xiao F., Mimee B., Lafond-Lapalme J. Transcriptome analysis of *Globodera pallida* from the susceptible host Solanum tuberosum or the resistant plant *Solanum sisymbriifolium*. *Sci. Rep.* 2019;9(1):13256. DOI 10.1038/s41598-019-49725-6.

Kud J., Wang W., Gross R., Fan Y., Huang L., Yuan Y., Gray A., Duarte A., Kuhl J.C., Caplan A., Goverse A., Liu Y., Dandurand L.M., Xiao F. The potato cyst nematode effector RHA1B is a ubiquitin ligase and uses two distinct mechanisms to suppress plant immune signaling. *PLoS Pathog.* 2019;15(4):e1007720. DOI 10.1371/journal.ppat.1007720.

Li X., Xing X., Tian P., Zhang M., Huo Z., Zhao K., Liu C., Duan D., He W., Yang T. Comparative transcriptome profiling reveals defense-related genes against *Meloidogyne incognita* invasion in tobacco. *Molecules.* 2018;23(8):2081. DOI 10.3390/molecules23082081.

Limantseva L., Mironenko N., Shuvalov O., Antonova O., Khiutti A., Novikova L., Afanasenko O., Spooner D., Gavrilenko T. Characterization of resistance to *Globodera rostochiensis* pathotype Ro1 in cultivated and wild potato species accessions from the Vavilov Institute of Plant Industry. *Plant Breed.* 2014;133:660-665.

Lin B., Zhuo K., Hu L., Sun L., Liao J., Zhang L.H., Chen S., Wang X. A novel nematode effector suppresses plant immunity by activating host reactive oxygen species scavenging system. *New Phytol*. 2016; 209:1159-1173. DOI 10.1111/nph.13701.

Mei Y., Wright K.M., Haegeman A., Bauters L., Diaz-Granados A., Goverse A., Gheysen G., T Jones J., Mantelin S. The *Globodera pallida* SPRYSEC effector *Gp*SPRY-414-2 that suppresses plant defenses targets a regulatory component of the dynamic microtubule network. *Front. Plant Sci.* 2018;9:1019. DOI 10.3389/fpls.2018.01019.

Mejias J., Truong N.M., Abad P., Favery B., Quentin M. Plant proteins and processes targeted by parasitic nematode effectors. *Front. Plant Sci.* 2019;10:970. DOI 10.3389/fpls.2019.00970.

Palomares-Rius J.E., Escobar C., Cabrera J., Vovlas A., Castillo P. Anatomical alterations in plant tissues induced by plant-parasitic nematodes. *Front. Plant Sci.* 2017;8:1987. DOI 10.3389/fpls.2017.01987.

Qiu X., Yang L., Ye J., Wang W., Zhao T., Hu H., Zhou G. Silencing of *cyp-33C9* gene affects the reproduction and pathogenicity of the pine wood nematode, *Bursaphelenchus xylophilus. Int. J. Mol. Sci.* 2019;20(18):4520. DOI 10.3390/ijms20184520.

Rodiuc N., Vieira P., Banora M.Y., de Almeida Engler J. On the track of transfer cell formation by specialized plant-parasitic nematodes. *Front. Plant Sci*. 2014;5:160. DOI 10.3389/fpls.2014.00160.

Sabeh M., Lord E., Grenier É., St-Arnaud M., Mimee B. What determines host specificity in hyperspecialized plant parasitic nematodes? *BMC Genomics.* 2019;20(1):457. DOI 10.1186/s12864-019-5853-4.

Schaff J.E., Nielsen D.M., Smith C.P., Scholl E.H., Bird D.M. Comprehensive transcriptome profiling in tomato reveals a role for glycosyltransferase in Mi-mediated nematode resistance. *Plant Physiol*. 2007;144(2):1079-1092. DOI 10.1104/pp.106.090241.

Shivakumara T.N., Chaudhary S., Kamaraju D., Dutta T.K., Papolu P.K., Banakar P., Sreevathsa R., Singh B., Manjaiah K.M., Rao U. Host-induced silencing of two pharyngeal gland genes conferred transcriptional alteration of cell wall-modifying enzymes of *Meloidogyne incognita* vis-à-vis perturbed nematode infectivity in eggplant. *Front. Plant Sci.* 2017;8:473. DOI 10.3389/fpls.2017.00473.

Shivakumara T.N., Somvanshi V.S., Phani V., Chaudhary S., Hada A., Budhwar R., Shukla R.N., Rao U. *Meloidogyne incognita* (Nematoda: Meloidogynidae) sterol-binding protein Mi-SBP-1 as a target for its management. *Int. J. Parasitol.* 2019;49(13-14):1061-1073. DOI 10.1016/j.ijpara.2019.09.002.

Strachan S.M., Armstrong M.R., Kaur A., Wright K.M., Lim T.Y., Baker K., Jones J., Bryan G., Blok V., Hein I. Mapping the H2 resistance effective against *Globodera pallida* pathotype Pa1 in tetraploid potato. *Theor. Appl. Genet*. 2019;132(4):1283-1294. DOI 10.1007/s00122-019-03278-4.

Sugawara T., Trifonova E.A., Kochetov A.V., Kanayama Y. Expression of extracellular ribonuclease gene increases resistance to cucumber mosaic virus in tobacco. *BMC Plant Biol*. 2016;16(Suppl 3):246. DOI 10.1186/s12870-016-0928-8.

Trifonova E.A., Sapotsky M.V., Komarova M.L., Scherban A.B., Shumny V.K., Polyakova A.M., Lapshina L.A., Kochetov A.V., Malinovsky V.I. Protection of transgenic tobacco plants expressing bovine pancreatic ribonuclease against tobacco mosaic virus. *Plant Cell Rep.* 2007;26:1121-1126. DOI 10.1007/s00299-006-0298-z.

Trudgill D.L., Blok V.C. Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic rootpathogens. *Annu. Rev. Phytopathol*. 2001;39:53-77. DOI 10.1146/annurev.phyto.39.1.53.

Trudgill D.L., Elliot M.J., Evans K., Phillips M.S. The white potato cyst nematode (*Globodera pallida*) – a critical analysis of the threat in Britain. *Ann. Appl. Biol*. 2003;143:73-80.

Ventoso I., Kochetov A., Montaner D., Dopazo J., Santoyo J. Extensive translatome remodeling during ER stress response in mammalian cells. *PLoS One*. 2012;7(5):e35915. DOI 10.1371/journal.pone.0035915.

Wang X.R., Moreno Y.A., Wu H.R., Ma C., Li Y.F., Zhang J.A., Yang C., Sun S., Ma W.-J., Geary T.G. Proteomic profiles of soluble proteins from the esophageal gland in female *Meloidogyne incognita*. *Int. J. Parasitol.* 2012;42:1177-1183. DOI 10.1016/j.ijpara.2012.10.008.

Whitworth J.L., Novy R.G., Zasada I.A., Wang X., Dandurand L.M., Kuhl J.C. Resistance of potato breeding clones and cultivars to three species of potato cyst nematode. *Plant Dis.* 2018;102(11):2120-2128. DOI 10.1094/PDIS-12-17-1978-RE.

Winslow R.D., Willis R.J. Nematode diseases of potatoes. II. Potato cyst nematode, Heterodera rostochiensis. In: Webster J. (Ed.). Economic Nematology. New York: Acad. Press, 1972;18-34.

ORCID ID

A.V. Kochetov orcid.org/0000-0003-3151-5181
T.A. Gavrilenko orcid.org/0000-0002-2605-6569
O.S. Afanasenko orcid.org/0000-0001-7368-0797

Acknowledgements. The review preparation was supported by the Russian Science Foundation, project 16-16-04073P.

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

Received September 10, 2020. Revised November 23, 2020. Accepted November 25, 2020.

Вавиловский журнал генетики и селекции. 2021;25(3):344-355
DOI 10.18699/VJ21.038

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

Methods of massive parallel reporter assays for investigation of enhancers

S.E. Romanov[1, 2], D.A. Kalashnikova[1, 2], P.P. Laktionov[1, 2]

[1] Novosibirsk State University, Epigenetics Laboratory, Department of Natural Sciences, Novosibirsk, Russia
[2] Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Genomics Laboratory, Novosibirsk, Russia
laktionov@mcb.nsc.ru

Abstract. The correct deployment of genetic programs for development and differentiation relies on finely coordinated regulation of specific gene sets. Genomic regulatory elements play an exceptional role in this process. There are few types of gene regulatory elements, including promoters, enhancers, insulators and silencers. Alterations of gene regulatory elements may cause various pathologies, including cancer, congenital disorders and autoimmune diseases. The development of high-throughput genomic assays has made it possible to significantly accelerate the accumulation of information about the characteristic epigenetic properties of regulatory elements. In combination with high-throughput studies focused on the genome-wide distribution of epigenetic marks, regulatory proteins and the spatial structure of chromatin, this significantly expands the understanding of the principles of epigenetic regulation of genes and allows potential regulatory elements to be searched for *in silico*. However, common experimental approaches used to study the local characteristics of chromatin have a number of technical limitations that may reduce the reliability of computational identification of genomic regulatory sequences. Taking into account the variability of the functions of epigenetic determinants and complex multicomponent regulation of genomic elements activity, their functional verification is often required. A plethora of methods have been developed to study the functional role of regulatory elements on the genome scale. Common experimental approaches for *in silico* identification of regulatory elements and their inherent technical limitations will be described. The present review is focused on original high-throughput methods of enhancer activity reporter analysis that are currently used to validate predicted regulatory elements and to perform *de novo* searches. The methods described allow assessing the functional role of the nucleotide sequence of a regulatory element, to determine its exact boundaries and to assess the influence of the local state of chromatin on the activity of enhancers and gene expression. These approaches have contributed substantially to the understanding of the fundamental principles of gene regulation.
Key words: gene regulatory elements; enhancers; massive parallel assays.

For citation: Romanov S.E., Kalashnikova D.A., Laktionov P.P. Methods of massive parallel reporter assays for investigation of enhancers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(3):344-355. DOI 10.18699/VJ21.038

Методы высокопроизводительного репортерного анализа энхансеров

С.Е. Романов[1, 2], Д.А. Калашникова[1, 2], П.П. Лактионов[1, 2]

[1] Новосибирский национальный исследовательский государственный университет, лаборатория эпигенетики факультета естественных наук, Новосибирск, Россия
[2] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, лаборатория геномики, Новосибирск, Россия
laktionov@mcb.nsc.ru

Аннотация. Корректное развертывание генетических программ развития и дифференцировки опирается на тонко координированную регуляцию экспрессии специфических наборов генов. Исключительную роль в управлении этим процессом играют регуляторные элементы генома, к которым относятся промоторы, энхансеры, инсуляторы и сайленсеры. Нарушения в их работе могут приводить к развитию различных патологий, включая онкологические заболевания, пороки развития и аутоиммунные заболевания. Развитие технологий высокопроизводительного геномного анализа позволило значительно ускорить накопление информации о специфичных эпигенетических характеристиках регуляторных элементов. В совокупности с полногеномными исследованиями распределения эпигенетических меток, регуляторных белков и пространственной структуры хроматина такие данные значительно расширяют представления о принципах эпигенетической регуляции генов и позволяют осуществлять поиск потенциальных регуляторных элементов *in silico*. Вместе с тем основные экспериментальные подходы, используемые для исследования локальных характеристик хроматина, имеют ряд технических ограничений, которые снижают достоверность биоинформатической идентификации

регуляторных областей генома. В связи с этим, а также с учетом вариабельности функций эпигенетических детерминант и многокомпонентной регуляции работы элементов генома определение их регуляторной роли часто требует функциональной проверки. Разработано множество методов, позволяющих провести исследование функциональной роли регуляторных элементов в масштабе генома. В настоящем обзоре кратко описаны основные экспериментальные подходы для проведения идентификации регуляторных элементов *in silico* и присущие им технические ограничения. Рассмотрены оригинальные методы высокопроизводительного репортерного анализа активности энхансеров, которые используют для валидации предсказанных регуляторных элементов и *de novo* поиска. Описанные методы анализа дают возможность оценить функциональную роль нуклеотидной последовательности регуляторного элемента, определить его точные границы, а также оценить влияние локального состояния хроматина на активность энхансеров и экспрессию генов. Применение таких методологических подходов обеспечило значительный вклад в понимание фундаментальных принципов регуляции генной экспрессии.
Ключевые слова: регуляторные элементы генома; энхансеры; высокопроизводительные методы анализа.

Introduction

The progress of programs for the development and maintenance of body functions is based on the expression of gene sets specific to cells and tissues. The gene expression is coordinated by a multilevel regulatory system that includes genetic and epigenetic mechanisms based on the interaction of genomic sequences, epigenetic modifications, regulatory proteins, and specific transcription factors. Certain genomic regions associated with the specific epigenetic determinants, as well as serving as a site for attracting regulatory proteins, are capable of modifying gene expression. Such regulatory elements in the genome play a key role in the implementation of genetic programs for development, differentiation, and maintenance of cellular and tissue homeostasis (Phillips-Cremins, Corces, 2013; Andersson et al., 2014; Kundaje et al., 2015).

Dysfunction of genomic regulatory elements may lead to the development of various pathologies, including cancer, developmental defects and autoimmune diseases (Maurano et al., 2012; Corradin et al., 2014; Miguel-Escalada et al., 2015; Bradner et al., 2017; Chatterjee, Ahituv, 2017). The genome wide association studies show that more than 90 % of disease-associated single nucleotide polymorphisms are located in non-coding genomic regions (Manolio et al., 2009; Maurano et al., 2012). The significant part of the genomic variants are located in regions that show epigenetic characteristics of enhancers, as well as affect enhancers, specific for the cell lines involved in the disease pathogenesis (Ernst et al., 2011; Akhtar-Zaidi et al., 2012; Trynka et al., 2013). The genetic variants associated with the development of type 2 diabetes (T2D), which were located in regions of putative enhancers in pancreatic islets, can be a good example (Stitzel et al., 2010; Pasquali et al., 2014).

Today, a lot of information is available regarding the specific properties of the epigenetic regulatory elements that alleviate identification of potential regulatory genomic regions *in silico* (Ernst et al., 2011). However, the validation and functional characterization of the regulatory elements often requires direct experimental verification. The classic methods are different modifications of reporter assays and functional mutagenesis. With the development of massively parallel sequencing methods, methodologies that allow studying the activity of the regulatory elements in genome-scale have been developed.

This review will describe the existing methodological solutions in the high-throughput analysis of enhancers that have significantly contributed to the understanding of the fundamental principles of their functions.

Types of regulatory elements

Several types of genomic regulatory elements, including promoters, enhancers, insulators, and silencers are distinguished.

Promoters are located near the transcription start site and serve as a DNA site where the transcription complex is assembled. In eukaryotes, such transcription complexes consist of the main transcription factors, RNA polymerase, and other regulatory proteins, including those which mediate the interaction with enhancers (Andersson, Sandelin, 2020).

Enhancers are nucleotide sequences in genomic DNA that contain binding sites for transcription factors and co-factors. As part of a protein complex, enhancers can physically interact with the promoter to activate gene expression (Shlyueva et al., 2014). Enhancers are able to regulate target promoters from a long distance, and regardless of mutual spatial orientation (Pennacchio et al., 2013). For example, the ZRS enhancer, the dominant mutation of which leads to familial forms of polydactyly, is located approximately 1 Mb from the controlled *Sonic hedgehog* (*Shh*) gene in the mouse genome (Lettice et al., 2014). On average, the enhancers are mapped at 20–50 Kb from the target gene in vertebrate genomes, and at 4–10 Kb in the genome of the fruit fly (Furlong, Levine, 2018).

The regulatory interactions network of promoters and enhancers can be quite complex. A separate gene can share enhancers with other genes, might be regulated either by several enhancers or specific enhancers in different types of cells. The *Arx* gene expression, for example, is controlled by four enhancers in mouse brain tissue (Dickel et al., 2018). Regulation of a gene by specific enhancers is also observed during the development of pathologies. For example, the *Myc* proto-oncogene enhancer is located in transcription termination sites in case of pancreatic cancer.

In case of rectal cancer, it is detected in the 5′-region of the gene, and in case of T-cell acute lymphoblastic leukemia, it can be found downstream of the 3′-region of the gene (Sur, Taipale, 2016).

Studies conducted in *Drosophila melanogaster* have shown that up to 30 percent of enhancers can act as remote regulatory elements without affecting the expression of genes located between them and target genes (Ghavi-Helm et al., 2014; Kvon et al., 2014). This means that there must be fine-tuned regulatory mechanisms that address interaction between the target gene promoter and specific enhancer. Today, there are several functionally intersecting concepts describing mechanics of the promoter-enhancer interactions, the main of which are contact formation via protein homo-oligomers and chromatin looping, caused by the action of motor proteins, such as RNA polymerase II and cohesin.

The regulatory elements – insulators – play an important role in regulation of the chromatin spatial structure. Interacting with specific proteins, insulators are able to block enhancer-promoter interaction and prevent the distribution of repressive chromatin marks acting as barrier elements (Kellum, Schedl, 1991, 1992; Geyer, Corces, 1992; Cai, Levine, 1995). With the development of modern methods of the nuclear architecture analysis, it became apparent that the functional impact of insulators is largely determined by their participation in the regulation of intra- and inter-chromosomal contacts (Yang, Corces, 2011). The insulator proteins play a key role in the formation of topologically associated domain (TAD) (Dixon et al., 2012). Such fragments are characterized by a high frequency of internal DNA contacts and are often flanked by the binding sites of insulator proteins and actively transcribed genes (Phillips-Cremins et al., 2013; Rao et al., 2014). Along with the regulation of the nucleus spatial structure, insulators are involved in many regulatory processes, including activation and repression of the gene expression, alternative splicing, and RNA polymerase pausing (Shukla et al., 2011; Paredes et al., 2013; Phillips-Cremins, Corces, 2013).

The silencers function is to suppress the gene expression, and such repression is mainly implemented by establishing repressive chromatin state and competition with activating proteins (Li et al., 2004; Srinivasan, Atchison, 2004; Harris et al., 2005; Lanzuolo et al., 2007; Tiwari et al., 2008).

Identification of regulatory genomic elements

The development of modern methods of high-throughput analysis has significantly accelerated and simplified the search for potential regulatory elements. The assumptions about the possible regulatory role of a genomic region are usually based on several types of data, including: (1) DNA accessibility for regulatory proteins, (2) presence of characteristic epigenetic determinants, (3) evaluation of gene expression and (4) analysis of DNA contacts.

Active regulatory elements are associated with specific proteins, and, hence, are free from nucleosomes. The treatment of genomic DNA with DNase I (DNase-seq), micrococcal nuclease (MNase-seq) and Tn5 transposase (assay for transposase-accessible chromatin, ATAC-seq), followed by high throughput sequencing and FAIRE-seq method, is used to identify such nucleosome-free loci (Nagy et al., 2003; Gaulton et al., 2010; Song, Crawford, 2010; Buenrostro et al., 2013). The listed methods are used for identification of putative enhancers, insulators, and silencers; however, to determine functional class of detected regulatory element, data on DNA accessibility should be combined with other descriptive data, e. g. chromatin properties (Song et al., 2011; Murtha et al., 2014; Huang et al., 2019).

The genomic mapping of the chromatin characteristic factors and histone modifications is also used to identify individual classes of regulatory elements. The basic method for assessing the representation of such epigenetic determinants in a particular genomic region is the chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-seq). Promoters are enriched in H3K4me3 histone mark (Bernstein et al., 2005). Monomethylation at the same position of the H3 histone (H3K4me1) is associated with enhancers, and the simultaneous presence of the H3K27me3 modification indicates that the enhancer might be poised for activation, while the H3K27ac modification indicates that the enhancer is active (Heintzman et al., 2007; Creyghton et al., 2010; Rada-Iglesias et al., 2011; Bonn et al., 2012; Arnold et al., 2013). Enrichment in the p300 histone acetyltransferase is characteristic of the enhancers (Visel et al., 2009). Mapping of specific transcription factors is also used to identify enhancers. For example, DNA regions enriched by the active enhancer histone marks, the Mediator complex proteins and the Oct4, Sox2, Nanog, Klf4, Esrrb master regulators are called super-enhancers and control the expression of tissue-specific sets of genes in embryonic stem cells (Whyte et al., 2013). To identify insulators in vertebrates, the genomic distribution of the CTCF protein and cofactors involved in the formation of loops, such as Rad21 and YY, are analyzed (Dixon et al., 2012, 2015; Nora et al., 2017; Rao et al., 2017). Silencers are enriched by the H3K27me3 histone modification associated with the effect of repressive Polycomb group proteins, as well as the H3K9me2/3 modifications related to heterochromatin (Barski et al., 2007).

The spatial organization of the nucleus mediates the interactions between target genomic loci and distal regulatory elements. The spatial chromatin structure is studied by methods that allow to fix and analyze DNA-DNA contacts, which originate from the 3C method (chromosome conformation capture) (Dekker et al., 2002; Tolhuis et al., 2002). The most widely used HiC method allows to build DNA genome-wide contacts map, in contrast to earlier methods (Gavrilov et al., 2009; Lieberman-Aiden et al., 2009). Combinations of chromatin spatial structure analysis and chromatin immunoprecipitation-based methods (ChIA-PET, HiChIP, and PLAC-ChIP) make it possible to establish DNA contacts in genomic regions which are specifically enriched in specific chromatin proteins or histone modifications (Fullwood, Ruan, 2009; Fang et al., 2016; Mumbach et al.,

2016). Analysis of the DNA-DNA contacts allows identifying promoter-enhancer interactions, defining borders of the topologically associating domain and larger chromatin compartments.

The data on the epigenetic characteristics and spatial genomic organization of model objects are available to a wide range of researchers in ENCODE, The Epigenome Roadmap, FANTOM and other databases (Birney et al., 2007; Bernstein et al., 2010; Andersson et al., 2014; Forrest et al., 2014; Kellis et al., 2014; Kundaje et al., 2015). These data are widely used for the prediction and research of potential regulatory elements.

However, it must be noted that the methods of analysis of protein-DNA and DNA-DNA interactions are capable of detecting non-functional interaction that can result in a false positive result. Local enrichment with characteristic epigenetic determinants detected by ChIP-seq does not necessarily indicate the presence of regulatory elements in a specific genomic region (Kvon et al., 2012). This can be due to the fact that implementation of the regulatory element function might require the coordinated binding of several transcription factors, and binding of only one of them is simply not enough (Halfon et al., 2000; Sandmann et al., 2007).

Nonfunctional transcription factor binding events can be transient, and caused by the general DNA-binding activity (Hammar et al., 2012). The chromatin immunoprecipitation method detects such transient interactions since it is based on the fixation of chromatin with formaldehyde with the formation of covalent cross-links between DNA and associated proteins. Modifications of the ChIP method that eliminate the need for chromatin fixation and potentially improve the accuracy of the method have been proposed (Skene, Henikoff, 2017; Kaya-Okur et al., 2019). A micrococcal nuclease fused with protein A is used in the variation of the CUT&RUN method (Skene, Henikoff, 2017). Protein A binds with the specific antibodies to the target protein, and micrococcal nuclease makes DNA breaks in the region of its binding. This allows selecting short genomic fragments, which are rich in proteins of interest, and identifying them with high-throughput sequencing. The Tn5 transposase is used in the CUT&TAG method instead of nuclease, which makes it possible to simultaneously introduce DNA adapters for massive parallel sequencing, flanking the recognition site of the protein of interest (Kaya-Okur et al., 2019). However, these methods have been developed recently and have not been widely used yet.

The false positive results in the ChIP-seq experiments may also be due to experimental variations, such as chromatin fragmentation mode, sequencing depth, and the threshold values for the identification of binding sites (Rye et al., 2011; Gomes et al., 2014; Jung et al., 2014). It is also important to note that in the presence of high- and low-affinity protein binding sites, the ChIP-seq method predominantly detects high-affinity ones (Nettling et al., 2016). This feature is also a limitation of the method, since it was shown that suboptimal binding sites for transcription factors in enhancers are needed for fine regulation of gene activity during development (Crocker et al., 2015, 2016; Farley et al., 2015).

In addition to the technical limitations of experimental methods, it is important to note that often functional regulatory elements demonstrate the presence of epigenetic determinants, which is generally uncharacteristic for their class. Functionally tested silencers in the K562 and HepG2B cell cultures, according to the ENCODE database, in addition to being enriched with the H3K9me3 and H3K27me3 repressive histone modifications, also contained H3K36me3 and H3K79me2 active chromatin histone marks (Pang, Snyder, 2020). Due to the experimental limitations of the methods, the variability of the functions of epigenetic determinants, and the participation of many components in the implementation of the functions of genomic elements, the determination of their regulatory role often requires functional verification.

Enhancer research methods

The functional role of genomic regulatory elements is commonly assessed with different modifications of reporter analysis. Pioneer work where the functional role of the genomic regulatory elements was demonstrated was devoted to the study of the enhancer of the early gene of the SV40 virus (Banerji et al., 1981). This work showed that a DNA fragment from the 5′-end of the early gene of the SV40 virus, consisting of two 72 bp repeats, can cause 200-fold activation of rabbit β-globin reporter gene expression in the HeLa cells (Banerji et al., 1981).

The standard genetic constructs used for the analysis of enhancer activity contain a reporter gene under the control of a minimal promoter, which confers minimal or no expression without additional activation. The genomic sequence of the enhancer is cloned into the construct, either upstream of the promoter or downstream of the coding sequence of the reporter gene. The obtained construct is used to transform cells and the change in the expression of the reporter gene comparing to a control construct that does not contain a potential enhancer is analyzed.

One of the first works aimed at *in vivo* functional testing of enhancers in genome-wide scale was based on the principles of classical reporter analysis (Kvon et al., 2014). Around 8,000 of the *D. melanogaster* lines were used, which contained a transgenic construct consisting of potential enhancer, minimal promoter, and Gal4 protein gene integrated in the same genomic region. The Gal4 expression was assessed at different stages of embryogenesis by *in situ* hybridization, and 400 embryos at different stages of development were analyzed for each potential regulatory element. As a result, more than three thousand enhancers have been identified. About a quarter were located in the vicinity of regulated genes, and a little more than a quarter were located at a distance of 20–100 Kb. On average, they were mapped around 10 Kb from the target genes (Kvon et al., 2014). About one third of the detected enhancers were located in the intergenic regions of regulated genes. Subse-

quently, it was also functionally confirmed that enhancers are able to regulate not only nearby genes, but also ones located through one or two genes (Kvon et al., 2014). The data obtained have significantly expanded the understanding of the fundamental principles of the operation of enhancers; however, the implementation of such projects requires a colossal amount of time and resources.

High-throughput enhancer reporter assays

The methods of high-throughput reporter analysis have evolved from classical approaches and allow simultaneous interrogation of thousands of regulatory sequences. There are two principal approaches in high-throughput reporter assays (Fig. 1). Within the first, a reporter gene contains a DNA barcode before the polyadenylation signal, and is placed under the control of a genomic fragment – a potential enhancer and a minimal promoter (see Fig. 1, *a*). In the case of activation of the reporter gene expression, such DNA barcode will be contained at the 3′-end of its transcript. After the pooling of such constructs, high-throughput sequencing is carried out, and unique DNA barcodes corresponding to each of the studied genomic fragments are determined (Fig. 2). After transformation using such constructs, the presence of DNA barcodes is analyzed by the transcriptome sequencing (RNA-seq). The expression level of a particular DNA barcode allows to assess activating ability of the corresponding specific regulatory element. This approach is used in the methods of quantitative assessment of the activity of genomic fragments, called MPRA (massive parallel reporter assay), different variations of which will be covered in this review (Kwasnieski et al., 2012, 2014; Melnikov et al., 2012; Kheradpour et al., 2013; Maricque et al., 2017).

The second approach allows evaluating the qualitative ability of the genomic fragment to exhibit the enhancers' properties. At first, a pool of genetic constructs that contain the genomic fragment of interest, a minimal promoter, and a reporter gene encoding a fluorescent protein or luciferase is prepared. At the next stage, the obtained pool of constructs is used for transformation and cells expressing the fluorescent protein are sorted using flow cytometry. The activation of the reporter gene expression means that the genomic fragment of interest demonstrates enhancer activities. The sorted cells are subjected to DNA isolation, fragments of constructs corresponding to the studied genomic fragments are amplified, and massive parallel sequencing is carried out, thus allowing to identify specific genomic fragments exhibiting the properties of enhancers. Examples of such methods include FIREWACh and SIF-seq (Dickel et al., 2014; Murtha et al., 2014) (see Fig. 1, *b*, *c*).

Combinations of the two approaches described above are also used. In this case, at the first stage cells carrying constructs containing potential enhancers are selected by flow cytometry. Then, the activating ability of specific genomic fragments is quantified by analyzing the representation of DNA barcodes by the RNA-seq method (Maricque et al., 2018).

MPRA methods are successfully used to study the activating properties of the nucleotide sequence of enhancers, the functional influence of regulatory protein binding motifs, and to search for and validate enhancers. Using this methodology, the effect of single mutations in the composition of three enhancers – ALDOB, ECR11, and LTV1, active in liver cells, was studied (Patwardhan et al., 2012). In the course of this research, a DNA library containing more than 100,000 mutated variants of the enhancers was synthesized. Such DNA fragments were cloned into constructs containing the minimal promoter, luciferase gene and transcribed DNA barcodes. The resulting DNA libraries were injected into the liver of mice, and a day later the transcriptome of liver cells was analyzed by RNA-seq (Kim, Ahituv, 2013).

It was found that the majority of single mutations had a weak effect on the activity of the studied enhancers. In addition, it was shown that mutations disrupting the enhancer function affect the predicted binding sites of the HNF4 and HNF1 transcription factors, which are active in liver cells (Kel et al., 2003). It is important to note that the experiment also showed significant discrepancies in theory and practice. Thus, within the ECR11 enhancer, mutations causing functional disorders were concentrated in the region that did not contain predicted binding sites for transcription factors, while mutations in the region containing most of these predicted sites did not change the enhancer activity. On the one hand, this clearly demonstrates that MPRA are applicable to clarify the boundaries of enhancers, and on the other hand, it emphasizes the importance of experimental verification of predictive data.

MPRA are also used for *de novo* search and validation of predicted enhancers. An elegant approach to search for enhancers was implemented in the STARR-seq method (Arnold et al., 2013) (see Fig. 1, *d*). The authors used the ability of enhancers to activate expression regardless of the position relative to the gene and promoter, and developed reporter constructs, containing studied genomic fragments cloned into open reading frame downstream from minimal promoter. In case a genomic region exhibits an enhancer function, this will lead to its transcription in cells. The expression level of this fragment in the cell transcriptome makes it possible to assess its activating function. This approach completely eliminates the need to use DNA barcodes, since the fragments play that role themselves.

To seek for enhancers, a plasmid library containing millions of random fragments of the fruit fly genome was prepared. After transfection of the S2 cell culture, a high-throughput RNA-seq transcription profile analysis was performed. As a result, thousands of genomic fragments were identified, which demonstrated the properties of enhancers. The most active were located close to housekeeping genes and developmental transcription factors. About a third of the fragments demonstrating pronounced activating properties in the S2 cell genome were located in areas of closed chromatin, lacking H3K27ac mark of active enhancers. Thus, it seems unlikely that such fragments are capable of performing the

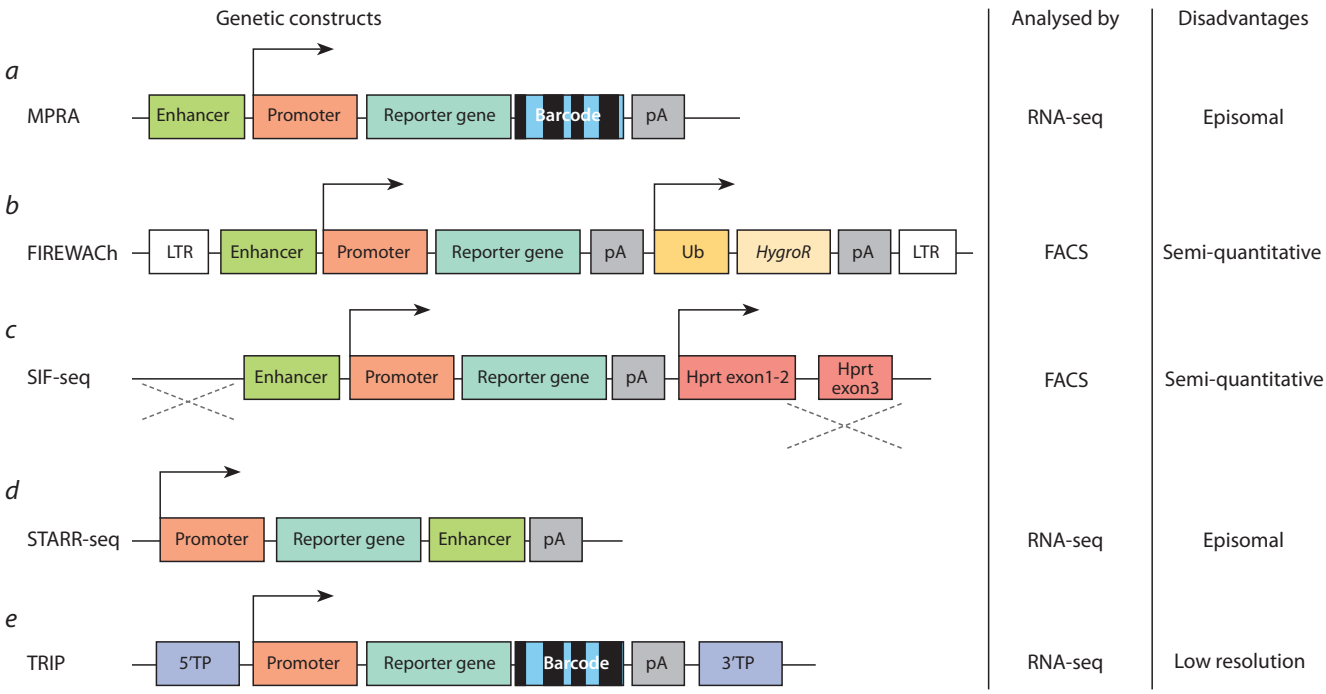


Fig. 1. Genetic constructs used for MPRA (*a*–*d*) and TRIP method (*e*).

pA – the polyadenylation signal; LTR – long terminal repeat; Ub – ubiquitin promoter; *HygroR* – hygromycin resistance gene; Hprt exon – the *Hprt* gene exon.

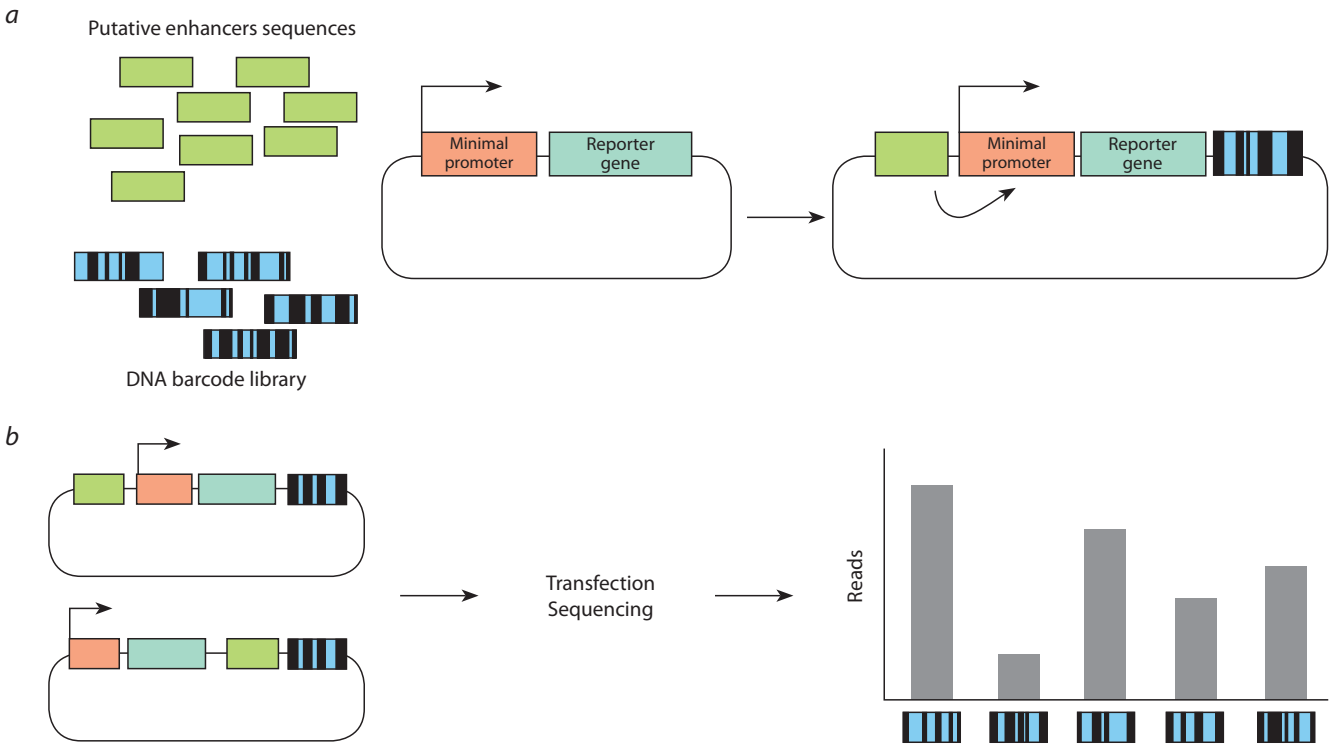


Fig. 2. Enhancer testing with MPRA.

a – at the first stage of MPRA, a pool of potential regulatory sequences is prepared. To obtain such sequences, synthesis technologies are used, or enrichment by chromatin immunoprecipitation methods, etc. Then, a pool of genetic constructs containing a minimal promoter and a reporter gene is created under the control of the putative regulatory elements. Each regulatory element in such constructs is associated with unique DNA barcode located at the end of the coding sequence of the reporter gene; *b* – upon transfection of cells, the expression of the reporter gene is activated in case putative regulatory element exhibits enhancer properties. The RNA-seq method is used to assess the expression level of unique DNA barcodes in the cell transcriptome. Normalization for barcode representation in initial pool and defining DNA fragments corresponding to each unique barcode allows to determine enhancer activity of tested genome region.

role of enhancers in the genome of the studied cells, and this finding highlights some of the limitations of episomal MPRA, which will be discussed below.

An interesting modification of the STARR-seq method was used in a subsequent work on the study of enhancers in human embryonic stem cells (Barakat et al., 2018). In the original work, DNA libraries were obtained by ultrasonic fragmentation of the *D. melanogaster* genomic DNA and subsequent mass cloning of the obtained fragments (Arnold et al., 2013). However, this approach is poorly applicable to larger genomes, since a sufficient representation of regulatory elements in the resulting DNA libraries is an extremely difficult task to achieve. Indeed, the use of the original STARR-seq method to study the regulatory elements of the mouse genome will require the creation of more than 200 million unique constructs (Murtha et al., 2014). Experimental verification showed that the use of a plasmid library containing 1.3 million unique fragments of the human genome made it possible to identify only six enhancers (Murtha et al., 2014).

To overcome this limitation, the ChIP-STARR-seq method was proposed. In the original paper, the chromatin immunoprecipitation was used to isolate genomic fragments enriched with the OCT4, NANOG transcription factors, as well as the H3K4me1 and H3K27ac histone modifications (Barakat et al., 2018). Obtained DNA fragments were then cloned into DNA libraries similar to those used in the original method. It was found that only a part of the genomic fragments that demonstrate enrichment by these factors in genome exhibited enhancer activity. Only about 25 % of the fragments bound by OCT4 showed enhancer properties. For the fragments enriched in NANOG and the H3K4me1 and H3K27 histone modifications, the results were 15, 9, and 10 %, respectively. It has been shown that neither individual factors nor their combinations are capable of unambiguously predicting enhancers. In addition, a group of enhancers associated with the regulation of general cellular processes, which had not previously been found in ESCs, were found. It turned out that such enhancers demonstrate a rather weak enrichment in TF OCT4 and NANOG, as well as in the histone modification H3K4me1, and, most likely, for this reason they were not previously detected in prospecting studies based on the chromatin immunoprecipitation method.

Data on chromatin accessibility and the genomic distribution of histone modifications and regulatory proteins deposited in open repositories allow to predict regulatory elements. Using MPRA, the activity of regulatory elements in the K562 cells and the E1 human embryonic stem cells, identified on the basis of chromatin structure analysis and annotated in ENCODE, was studied (Kwasnieski et al., 2014). It turned out that only about a quarter of them had an effect on gene expression, which underlines the importance of such experimental verification (Kwasnieski et al., 2014). At the same time, this effect may be due to the experimental limitations of the described MPRA. Indeed, the methods described above are episomal, which means that reporter constructs are not integrated into the genome, hence the activity of enhancers is assessed outside the chromatin context. Significant differences in the activity of enhancers analyzed in episomal manner and upon integration into the genome were also confirmed experimentally (Inoue et al., 2017).

This experimental discrepancy looks logical, because the observation of the effect of chromatin structure on gene regulation was demonstrated in classical genetic experiments long ago (Muller, 1930). The use of an original high-throughput reporter analysis method, combined with MPRA-approaches, made it possible to characterize the local effects of chromatin on gene expression in mouse embryonic stem cells (Akhtar et al., 2013) (see Fig. 1, *e*). Within the framework of this study, using the PiggyBac transposase-based genomic integration system, reporter constructs containing unique DNA barcodes at the 3′-end of the reporter gene were randomly inserted into the cell genome.

In the next step, such insertions were mapped and each DNA barcode was associated with a specific genomic locus. In total, more than 17 thousand of such inserts were received. Then, the expression level of DNA barcodes was analyzed using the RNA-seq method, which made it possible to assess the transcriptional activity of each insertion as well as the effect of the local chromatin structure on it. Reporter constructs integrated into regions of compacted chromatin and regions of domains associated with the nuclear lamina, as expected, showed a reduced level of expression. Reporter constructions located within 200 Kb from active genes were more actively transcribed. It is interesting to mention that an increased frequency of enhancers was observed within approximately the same range. Enhancers had an activating effect on the expression of reporter constructs at a distance of up to 20 Kb. It is important to note that in this case a linear distance is considered, and the spatial structure of chromatin is not taken into account. It was assumed that the formation of extended, actively transcribed regions is based on the action of several enhancers. This emphasizes the need to study regulatory elements in conditions close to native ones.

The effect of chromatin on the function of regulatory elements is to some extent taken into account in MPRA, which are based on the genomic integration of the reporter construct (Dickel et al., 2014; Murtha et al., 2014; Maricque et al., 2017, 2018). These FIREWACh and SIF-seq methods were used to identify enhancers in mouse ESCs, but did not allow quantitative assessment of the activity of regulatory elements (see the general description of approaches above) (Dickel et al., 2014; Murtha et al., 2014).

The FIREWACh method is based on genomic integration of reporter constructs using lentiviral transduction (see Fig. 1, *b*). This method of genomic integration ensures the insertion of the construct into random regions of the genome (Yang et al., 2008). Thus, an adequate comparison of the activity of various regulatory elements seems to be difficult, because it is highly likely that reporter constructs

will be integrated into different genomic regions, with an unpredictable effect of the local chromatin environment.

The SIF-seq method avoided such a drawback, since the integration of reporter constructs is carried out in the same region of the genome located in the region of the *Hprt* gene (Dickel et al., 2014) (see Fig. 1, *b*). However, this might serve as a disadvantage, since the correct operation of an enhancer is determined by a specific set of chromatin factors, and it is highly likely that it will become non-functional when transferred to a non-identical chromatin environment.

The approaches described above did not allow answering one of the fundamental questions of understanding the principles of enhancers' activity, namely, to what extent is it determined by the DNA sequence, and to what extent – by the properties of the surrounding chromatin? A systematic study of this issue was carried out in the research on the influence of different chromatin environments on the comparative activity of enhancers (Maricque et al., 2018). Within the framework of this study, 15 lines of the K562 cells were prepared, containing single insertions of reporter constructs located in different chromatin environments and containing the Cre-recombinase (*loxP*) recognition sites, allowing targeted insertion of transgenes. Such insertions contained a DNA barcode and a polyadenylation signal outside of the fragment flanked by *loxP*-sites, with a single unique DNA barcode corresponding to each line.

The described lines were pooled together, and Cre-mediated integration was used to integrate reporter constructs, that contained a reporter gene ending with a DNA barcode under the control of the minimal promoter and the genomic fragment of interest. As such fragments, 300 synthesized regulatory elements were used, which were previously studied by episomal MPRA and ranked according to the level of activity (Kwasnieski et al., 2014). For each genomic fragment, the corresponding unique DNA barcodes had been previously established. In case of successful integration, the original loxP cassette was replaced with a reporter construct containing the putative enhancers. Moreover, in the case of activation of the reporter gene, two DNA barcodes will be transcribed in its composition. Deciphering barcodes allows to identify which fragment was analyzed in which cell line.

The analysis of the representation of combinations of DNA barcodes in the transcriptome of cells made it possible to assess the level of activity of the studied regulatory elements in different chromatin environments. It was found that the chromatin environment has pronounced effect on the activity of cis-elements. However, being placed in the same chromatin environment, regulatory elements save their relative activity. It was also demonstrated that the activity of the promoter affects the expression of reporter constructs, but at the same time does not affect the comparative activity of regulatory elements. The results obtained support the model according to which the nucleotide sequence of the enhancer determines its overall activity, which is already modulated by the structure of the chromatin environment.

Conclusion

MPRA methods allow to perform detailed study of the regulatory potential of the genomic fragments, and it is a convenient tool for studying the effect of variations in the nucleotide sequence on their function. However, it is necessary to note the limitations of the methods, which should be taken into account when interpreting the results obtained. The common drawback of MPRA is the need to use a minimal promoter that is unable to activate the expression of the reporter gene in the absence of an enhancer, since the presence of basal activity can significantly distort the results. At the same time, the selected promoter can significantly influence the activity of a particular enhancer (Zabidi et al., 2015; Maricque et al., 2018). In this sense, the analysis of the activity of enhancers in combination with various promoters seems to be an ideal experiment. However, such work seems to be extremely difficult and time-consuming.

The synthesis of DNA fragments used as studied regulatory elements imposes restrictions on the total length of such a fragment. Usually, the length of the studied fragments is limited to about 200 bp, which often complicates the analysis of the influence of the rest of the enhancer regions falling outside these limits (Kwasnieski et al., 2014). MPRA methods based on episomal constructs do not take into account the possible influence of the chromatin environment on the regulatory element; therefore, they can be used to study the direct activating ability of a DNA sequence. MPRA based on the genomic integration of reporter constructs make it possible to overcome this limitation to some extent. However, random or site-specific integration still does not allow the analysis of the activity of a regulatory element in native genomic environment. The impossibility of studying the enhancers function in native environment is a serious MPRA limitation, since the function of the regulatory genomic element depends on the structure of the surrounding chromatin and the spatial organization of the locus.

Modern methods of high-throughput CRISPR/Cas9 mutagenesis, as well as methods of directed expression modulation based on the use of an inactivated form of the Cas9 endonuclease (dCas9) fused with activator or repressor proteins, make it possible to study regulatory elements in native genomic environment (Chavez et al., 2015; Sanjana et al., 2016; Canver et al., 2017; Li et al., 2020). While there are obvious advantages, such methods also have potential drawbacks. For example, point mutations produced by the targeted mutagenesis may not be sufficient to disrupt enhancer function. In addition, directed mutagenesis is associated with errors in the recognition of target genome regions (off-targets), which can lead to the generation of experimental noise.

It is important to note that the KRAB repressor protein, which is widely used for the targeted inactivation of enhancers, is capable of initiating the formation of heterochromatin regions of 1–2 Kb in length (Gasperini et al., 2019). This feature can reduce the resolution of the method and complicate the identification of specific functional fragments

of the enhancer, as well as increase unwanted side effects in the case of the presence of erroneous dCas9 recognition sites. In addition to possible technical difficulties, in the case of a successful disruption of the enhancer function, phenotypic manifestations can be restored rather quickly due to the presumable existence of duplicate enhancers (Diao et al., 2016).

Thus, MPRA and high-performance methods based on the CRISPR/Cas9 system are quite complementary and make it possible to characterize in detail the regulatory functions of the studied genomic fragments. Coupled with vast amounts of accumulated data on the chromatin structure and spatial organization in various cells and tissues, the use of such methods makes it possible to significantly advance in the understanding of the mechanisms of precise regulation of gene expression during development and in various pathologies. Altogether, this allows hoping that in the near future modern genomics will be able to move from a detailed functional description of regulatory elements to the creation of quantitative biological models for the regulation of gene expression.

References

Akhtar-Zaidi B., Cowper-Sal-lari R., Corradin O., Saiakhova A., Bartels C.F., Balasubramanian D., Myeroff L., Lutterbaugh J., Jarrar A., Kalady M.F., Willis J., Moore J.H., Tesar P.J., Laframboise T., Markowitz S., Lupien M., Scacheri P.C. Epigenomic enhancer profiling defines a signature of colon cancer. *Science.* 2012;336:736-739.

Akhtar W., de Jong J., Pindyurin A.V., Pagie L., Meuleman W., de Ridder J., Berns A., Wessels L.F., van Lohuizen M., van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell.* 2013;154:914-927.

Andersson R., Gebhard C., Miguel-Escalada I., Hoof I., Bornholdt J., Boyd M., Chen Y., …, Suzuki H., Hayashizaki Y., Muller F., Forrest A.R.R., Carninci P., Rehli M., Sandelin A. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507: 455-461.

Andersson R., Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet*. 2020;21:71-87.

Arnold C.D., Gerlach D., Stelzer C., Boryn L.M., Rath M., Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science.* 2013;339:1074-1077.

Banerji J., Rusconi S., Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell.* 1981;27:299-308.

Barakat T.S., Halbritter F., Zhang M., Rendeiro A.F., Perenthaler E., Bock C., Chambers I. Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell.* 2018;23:276-288 e278.

Barski A., Cuddapah S., Cui K., Roh T.Y., Schones D.E., Wang Z., Wei G., Chepelev I., Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell.* 2007;129:823-837.

Bernstein B.E., Kamal M., Lindblad-Toh K., Bekiranov S., Bailey D.K., Huebert D.J., McMahon S., Karlsson E.K., Kulbokas E.J. 3rd, Gingeras T.R., Schreiber S.L., Lander E.S. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell.* 2005;120:169-181.

Bernstein B.E., Stamatoyannopoulos J.A., Costello J.F., Ren B., Milosavljevic A., Meissner A., Kellis M., Marra M.A., Beaudet A.L., Ecker J.R., Farnham P.J., Hirst M., Lander E.S., Mikkelsen T.S., Thomson J.A. The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* 2010;28:1045-1048.

Birney E., Stamatoyannopoulos J.A., Dutta A., Guigo R., Gingeras T.R., Margulies E.H., Weng Z., …, Lander E.S., Koriabine M., Nefedov M., Osoegawa K., Yoshinaga Y., Zhu B., de Jong P.J. Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. *Nature.* 2007;447: 799-816.

Bonn S., Zinzen R.P., Girardot C., Gustafson E.H., Perez-Gonzalez A., Delhomme N., Ghavi-Helm Y., Wilczynski B., Riddell A., Furlong E.E. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* 2012;44:148-156.

Bradner J.E., Hnisz D., Young R.A. Transcriptional addiction in cancer. *Cell.* 2017;168:629-643.

Buenrostro J.D., Giresi P.G., Zaba L.C., Chang H.Y., Greenleaf W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods.* 2013;10:1213-1218.

Cai H., Levine M. Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature.* 1995;376:533-536.

Canver M.C., Bauer D.E., Orkin S.H. Functional interrogation of noncoding DNA through CRISPR genome editing. *Methods*. 2017; (121-122):118-129.

Chatterjee S., Ahituv N. Gene regulatory elements, major drivers of human disease. *Annu. Rev. Genomics Hum. Genet.* 2017;18:45-63.

Chavez A., Scheiman J., Vora S., Pruitt B.W., Tuttle M., Eswar P.R.I., Lin S., Kiani S., Guzman C.D., Wiegand D.J., Ter-Ovanesyan D., Braff J.L., Davidsohn N., Housden B.E., Perrimon N., Weiss R., Aach J., Collins J.J., Church G.M. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods.* 2015;12:326-328.

Corradin O., Saiakhova A., Akhtar-Zaidi B., Myeroff L., Willis J., Cowper-Sallari R., Lupien M., Markowitz S., Scacheri P.C. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* 2014;24:1-13.

Creyghton M.P., Cheng A.W., Welstead G.G., Kooistra T., Carey B.W., Steine E.J., Hanna J., Lodato M.A., Frampton G.M., Sharp P.A., Boyer L.A., Young R.A., Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA.* 2010;107:21931-21936.

Crocker J., Abe N., Rinaldi L., McGregor A.P., Frankel N., Wang S., Alsawadi A., Valenti P., Plaza S., Payre F., Mann R.S., Stern D.L. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*. 2015;160:191-203.

Crocker J., Noon E.P., Stern D.L. The soft touch: low-affinity transcription factor binding sites in development and evolution. *Curr. Top. Dev. Biol.* 2016;117:455-469.

Dekker J., Rippe K., Dekker M., Kleckner N. Capturing chromosome conformation. *Science*. 2002;295:1306-1311.

Diao Y., Li B., Meng Z., Jung I., Lee A.Y., Dixon J., Maliskova L., Guan K.L., Shen Y., Ren B. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res*. 2016;26:397-405.

Dickel D.E., Ypsilanti A.R., Pla R., Zhu Y., Barozzi I., Mannion B.J., Khin Y.S., Fukuda-Yuzawa Y., Plajzer-Frick I., Pickle C.S., Lee E.A., Harrington A.N., Pham Q.T., Garvin T.H., Kato M., Osterwalder M., Akiyama J.A., Afzal V., Rubenstein J.L.R., Pennacchio L.A., Visel A. Ultraconserved enhancers are required for normal development. *Cell*. 2018;172:491-499.e415.

Dickel D.E., Zhu Y., Nord A.S., Wylie J.N., Akiyama J.A., Afzal V., Plajzer-Frick I., Kirkpatrick A., Gottgens B., Bruneau B.G., Visel A., Pennacchio L.A. Function-based identification of mammalian enhancers using site-specific integration. *Nat. Meth*. 2014;11:566-571.

Dixon J.R., Jung I., Selvaraj S., Shen Y., Antosiewicz-Bourget J.E., Lee A.Y., Ye Z., Kim A., Rajagopal N., Xie W., Diao Y., Liang J., Zhao H., Lobanenkov V.V., Ecker J.R., Thomson J.A., Ren B. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518:331-336.

Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J.S., Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376-380.

Ernst J., Kheradpour P., Mikkelsen T.S., Shoresh N., Ward L.D., Epstein C.B., Zhang X., Wang L., Issner R., Coyne M., Ku M., Durham T., Kellis M., Bernstein B.E. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473:43-49.

Fang R., Yu M., Li G., Chee S., Liu T., Schmitt A.D., Ren B. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* 2016;26:1345-1348.

Farley E.K., Olson K.M., Zhang W., Brandt A.J., Rokhsar D.S., Levine M.S. Suboptimization of developmental enhancers. *Science*. 2015;350:325-328.

Forrest A.R., Kawaji H., Rehli M., Baillie J.K., de Hoon M.J., Haberle V., Lassmann T., …, Bajic V.B., Taylor M.S., Makeev V.J., Sandelin A., Hume D.A., Carninci P., Hayashizaki Y. A promoter-level mammalian expression atlas. *Nature*. 2014;507:462-470.

Fullwood M.J., Ruan Y. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem*. 2009;107: 30-39.

Furlong E.E.M., Levine M. Developmental enhancers and chromosome topology. *Science*. 2018;361:1341-1345.

Gasperini M., Hill A.J., McFaline-Figueroa J.L., Martin B., Kim S., Zhang M.D., Jackson D., Leith A., Schreiber J., Noble W.S., Trapnell C., Ahituv N., Shendure J. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell*. 2019;176: 1516.

Gaulton K.J., Nammo T., Pasquali L., Simon J.M., Giresi P.G., Fogarty M.P., Panhuis T.M., Mieczkowski P., Secchi A., Bosco D., Berney T., Montanya E., Mohlke K.L., Lieb J.D., Ferrer J. A map of open chromatin in human pancreatic islets. *Nat. Genet*. 2010;42: 255-259.

Gavrilov A., Eivazova E., Priozhkova I., Lipinski M., Razin S., Vassetzky Y. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol. Biol.* 2009;567:171-188.

Geyer P.K., Corces V.G. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes Dev.* 1992;6: 1865-1873.

Ghavi-Helm Y., Klein F.A., Pakozdi T., Ciglar L., Noordermeer D., Huber W., Furlong E.E. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*. 2014; 512: 96-100.

Gomes A.L., Abeel T., Peterson M., Azizi E., Lyubetskaya A., Carvalho L., Galagan J. Decoding ChIP-seq with a double-binding signal refines binding peaks to single-nucleotides and predicts cooperative interaction. *Genome Res*. 2014;24:1686-1697.

Halfon M.S., Carmena A., Gisselbrecht S., Sackerson C.M., Jimenez F., Baylies M.K., Michelson A.M. Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell*. 2000;103:63-74.

Hammar P., Leroy P., Mahmutovic A., Marklund E.G., Berg O.G., Elf J. The lac repressor displays facilitated diffusion in living cells. *Science*. 2012;336:1595-1598.

Harris M.B., Mostecki J., Rothman P.B. Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *J. Biol. Chem.* 2005;280:13114-13121.

Heintzman N.D., Stuart R.K., Hon G., Fu Y., Ching C.W., Hawkins R.D., Barrera L.O., Van Calcar S., Qu C., Ching K.A., Wang W., Weng Z., Green R.D., Crawford G.E., Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 2007;39:311-318.

Huang D., Petrykowska H.M., Miller B.F., Elnitski L., Ovcharenko I. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Res*. 2019;29:657-667.

Inoue F., Kircher M., Martin B., Cooper G.M., Witten D.M., McManus M.T., Ahituv N., Shendure J. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res*. 2017;27:38-52.

Jung Y.L., Luquette L.J., Ho J.W., Ferrari F., Tolstorukov M., Minoda A., Issner R., Epstein C.B., Karpen G.H., Kuroda M.I., Park P.J. Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res.* 2014;42:e74.

Kaya-Okur H.S., Wu S.J., Codomo C.A., Pledger E.S., Bryson T.D., Henikoff J.G., Ahmad K., Henikoff S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun*. 2019;10:1930.

Kel A.E., Gossling E., Reuter I., Cheremushkin E., Kel-Margoulis O.V., Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* 2003;31:3576-3579.

Kellis M., Wold B., Snyder M.P., Bernstein B.E., Kundaje A., Marinov G.K., Ward L.D., Birney E., Crawford G.E., Dekker J., Dunham I., Elnitski L.L., Farnham P.J., Feingold E.A., Gerstein M., Giddings M.C., Gilbert D.M., Gingeras T.R., Green E.D., Guigo R., Hubbard T., Kent J., Lieb J.D., Myers R.M., Pazin M.J., Ren B., Stamatoyannopoulos J.A., Weng Z., White K.P., Hardison R.C. Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. USA*. 2014;111:6131-6138.

Kellum R., Schedl P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell*. 1991;64:941-950.

Kellum R., Schedl P. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell. Biol.* 1992;12: 2424-2431.

Kheradpour P., Ernst J., Melnikov A., Rogov P., Wang L., Zhang X., Alston J., Mikkelsen T.S., Kellis M. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res*. 2013;23:800-811.

Kim M.J., Ahituv N. The hydrodynamic tail vein assay as a tool for the study of liver promoters and enhancers. *Methods Mol. Biol.* 2013; 1015:279-289.

Kundaje A., Meuleman W., Ernst J., Bilenky M., Yen A., Heravi-Moussavi A., Kheradpour P., …, Hirst M., Meissner A., Milosavljevic A., Ren B., Stamatoyannopoulos J.A., Wang T., Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518: 317-330.

Kvon E.Z., Kazmar T., Stampfel G., Yanez-Cuna J.O., Pagani M., Schernhuber K., Dickson B.J., Stark A. Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature*. 2014;512:91-95.

Kvon E.Z., Stampfel G., Yanez-Cuna J.O., Dickson B.J., Stark A. HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. *Genes Dev.* 2012;26:908-913.

Kwasnieski J.C., Fiore C., Chaudhari H.G., Cohen B.A. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res*. 2014;24:1595-1602.

Kwasnieski J.C., Mogno I., Myers C.A., Corbo J.C., Cohen B.A. Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc. Natl. Acad. Sci. USA*. 2012;109:19498-19503.

Lanzuolo C., Roure V., Dekker J., Bantignies F., Orlando V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 2007;9: 1167-1174.

Lettice L.A., Williamson I., Devenney P.S., Kilanowski F., Dorin J., Hill R.E. Development of five digits is controlled by a bipartite long-range *cis*-regulator. *Development*. 2014;141:1715-1725.

Li K., Liu Y., Cao H., Zhang Y., Gu Z., Liu X., Yu A., Kaphle P., Dickerson K.E., Ni M., Xu J. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.* 2020; 11:485.

Li L., He S., Sun J.M., Davie J.R. Gene regulation by Sp1 and Sp3. *Biochem. Cell Biol.* 2004;82:460-471.

Lieberman-Aiden E., van Berkum N.L., Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O., Sandstrom R., Bernstein B., Bender M.A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L.A., Lander E.S., Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289-293.

Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorff L.A., Hunter D.J., McCarthy M.I., …, Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F., McCarroll S.A., Visscher P.M. Finding the missing heritability of complex diseases. *Nature*. 2009;461: 747-753.

Maricque B.B., Chaudhari H.G., Cohen B.A. A massively parallel reporter assay dissects the influence of chromatin structure on *cis*-regulatory activity. *Nat*. *Biotechnol*. 2018. DOI 10.1038/nbt.4285.

Maricque B.B., Dougherty J.D., Cohen B.A. A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of *cis*-regulatory activity in neural cells. *Nucleic Acids Res.* 2017; 45:e16.

Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E., Wang H., Reynolds A.P., …, Ziegler S., Cotsapas C., Sotoodehnia N., Glass I., Sunyaev S.R., Kaul R., Stamatoyannopoulos J.A. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190-1195.

Melnikov A., Murugan A., Zhang X., Tesileanu T., Wang L., Rogov P., Feizi S., Gnirke A., Callan C.G. Jr., Kinney J.B., Kellis M., Lander E.S., Mikkelsen T.S. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 2012;30:271-277.

Miguel-Escalada I., Pasquali L., Ferrer J. Transcriptional enhancers: functional insights and role in human disease. *Curr. Opin. Genet. Dev.* 2015;33:71-76.

Muller H.J. Types of visible variations induced by X-rays in *Drosophila*. *J. Genet*. 1930;299-334.

Mumbach M.R., Rubin A.J., Flynn R.A., Dai C., Khavari P.A., Greenleaf W.J., Chang H.Y. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods.* 2016;13:919-922.

Murtha M., Tokcaer-Keskin Z., Tang Z., Strino F., Chen X., Wang Y., Xi X., Basilico C., Brown S., Bonneau R., Kluger Y., Dailey L. FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat. Methods.* 2014;11: 559-565.

Nagy P.L., Cleary M.L., Brown P.O., Lieb J.D. Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl. Acad. Sci. USA.* 2003;100: 6364-6369.

Nettling M., Treutler H., Cerquides J., Grosse I. Detecting and correcting the binding-affinity bias in ChIP-seq data using inter-species information. *BMC Genomics.* 2016;17:347.

Nora E.P., Goloborodko A., Valton A.L., Gibcus J.H., Uebersohn A., Abdennur N., Dekker J., Mirny L.A., Bruneau B.G. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*. 2017;169:930-944.e922.

Pang B., Snyder M.P. Systematic identification of silencers in human cells. *Nat. Genet.* 2020;52:254-263.

Paredes S.H., Melgar M.F., Sethupathy P. Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. *Bioinformatics*. 2013;29:1485-1487.

Pasquali L., Gaulton K.J., Rodriguez-Segui S.A., Mularoni L., Miguel-Escalada I., Akerman I., Tena J.J., …, Berney T., Gloyn A.L., Ravassard P., Skarmeta J.L.G., Muller F., McCarthy M.I., Ferrer J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 2014;46:136-143.

Patwardhan R.P., Hiatt J.B., Witten D.M., Kim M.J., Smith R.P., May D., Lee C., Andrie J.M., Lee S.I., Cooper G.M., Ahituv N., Pennacchio L.A., Shendure J. Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat. Biotechnol.* 2012;30:265-270.

Pennacchio L.A., Bickmore W., Dean A., Nobrega M.A., Bejerano G. Enhancers: five essential questions. *Nat. Rev. Genet.* 2013;14:288-295.

Phillips-Cremins J.E., Corces V.G. Chromatin insulators: linking genome organization to cellular function. *Mol. Cell.* 2013;50:461-474.

Phillips-Cremins J.E., Sauria M.E., Sanyal A., Gerasimova T.I., Lajoie B.R., Bell J.S., Ong C.T., Hookway T.A., Guo C., Sun Y., Bland M.J., Wagstaff W., Dalton S., McDevitt T.C., Sen R., Dekker J., Taylor J., Corces V.G. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013;153:1281-1295.

Rada-Iglesias A., Bajpai R., Swigut T., Brugmann S.A., Flynn R.A., Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011;470:279-283.

Rao S.S., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., Aiden E.L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-1680.

Rao S.S.P., Huang S.C., Glenn St Hilaire B., Engreitz J.M., Perez E.M., Kieffer-Kwon K.R., Sanborn A.L., Johnstone S.E., Bascom G.D., Bochkov I.D., Huang X., Shamim M.S., Shin J., Turner D., Ye Z., Omer A.D., Robinson J.T., Schlick T., Bernstein B.E., Casellas R., Lander E.S., Aiden E.L. Cohesin loss eliminates all loop domains. *Cell*. 2017;171:305-320 e324.

Rye M.B., Saetrom P., Drablos F. A manually curated ChIP-seq benchmark demonstrates room for improvement in current peak-finder programs. *Nucleic Acids Res.* 2011;39:e25. DOI 10.1093/nar/gkq1187.

Sandmann T., Girardot C., Brehme M., Tongprasit W., Stolc V., Furlong E.E. A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* 2007;21: 436-449.

Sanjana N.E., Wright J., Zheng K., Shalem O., Fontanillas P., Joung J., Cheng C., Regev A., Zhang F. High-resolution interrogation of functional elements in the noncoding genome. *Science*. 2016;353:1545-1549.

Shlyueva D., Stampfel G., Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 2014;15: 272-286.

Shukla S., Kavak E., Gregory M., Imashimizu M., Shutinoski B., Kashlev M., Oberdoerffer P., Sandberg R., Oberdoerffer S. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*. 2011;479:74-79.

Skene P.J., Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*. 2017;6:e21856. DOI 10.7554/eLife.21856.

Song L., Crawford G.E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* 2010;2010(2):pdbprot 5384.

Song L., Zhang Z., Grasfeder L.L., Boyle A.P., Giresi P.G., Lee B.K., Sheffield N.C., …, Winter D., Clarke N.D., Birney E., Iyer V.R.,

Crawford G.E., Lieb J.D., Furey T.S. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 2011;21:1757-1767.

Srinivasan L., Atchison M.L. YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev.* 2004;18:2596-2601.

Stitzel M.L., Sethupathy P., Pearson D.S., Chines P.S., Song L., Erdos M.R., Welch R., Parker S.C., Boyle A.P., Scott L.J., Margulies E.H., Boehnke M., Furey T.S., Crawford G.E., Collins F.S. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab.* 2010;12:443-455.

Sur I., Taipale J. The role of enhancers in cancer. *Nat. Rev. Cancer.* 2016;16:483-493.

Tiwari V.K., McGarvey K.M., Licchesi J.D., Ohm J.E., Herman J.G., Schubeler D., Baylin S.B. PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* 2008;6:2911-2927.

Tolhuis B., Palstra R.J., Splinter E., Grosveld F., de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell.* 2002;10:1453-1465.

Trynka G., Sandor C., Han B., Xu H., Stranger B.E., Liu X.S., Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* 2013;45:124-130.

Visel A., Blow M.J., Li Z., Zhang T., Akiyama J.A., Holt A., Plajzer-Frick I., Shoukry M., Wright C., Chen F., Afzal V., Ren B., Rubin E.M., Pennacchio L.A. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009;457:854-858.

Whyte W.A., Orlando D.A., Hnisz D., Abraham B.J., Lin C.Y., Kagey M.H., Rahl P.B., Lee T.I., Young R.A. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307-319.

Yang J., Corces V.G. Chromatin insulators: a role in nuclear organization and gene expression. *Adv. Cancer Res.* 2011;110:43-76.

Yang S.H., Cheng P.H., Sullivan R.T., Thomas J.W., Chan A.W. Lentiviral integration preferences in transgenic mice. *Genesis*. 2008;46:711-718.

Zabidi M.A., Arnold C.D., Schernhuber K., Pagani M., Rath M., Frank O., Stark A. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*. 2015;518:556-559.

ORCID ID

S.E. Romanov orcid.org/0000-0002-5989-5756
D.A. Kalashnikova orcid.org/0000-0002-0209-4616
P.P. Laktionov orcid.org/0000-0003-2174-6496

Acknowledgements. The reported study was funded by Russian Foundation for Basic Research, project No. 20-34-70141.

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

Received November 23, 2020. Revised March 28, 2021. Accepted March 28, 2021.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции»/"Vavilov Journal of Genetics and Breeding"
до 2011 г. выходил под названием «Информационный вестник ВОГиС»/
"The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Регистрационное свидетельство ПИ № ФС77-45870 выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 20 июля 2011 г.

«Вавиловский журнал генетики и селекции» включен ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Российский индекс научного цитирования, ВИНИТИ, базы данных Emerging Sources Citation Index (Web of Science), Zoological Record (Web of Science), Scopus, PubMed Central, Ebsco, DOAJ, Ulrich's Periodicals Directory, Google Scholar, Russian Science Citation Index на платформе Web of Science, каталог научных ресурсов открытого доступа ROAD.

Открытый доступ к полным текстам:
на сайте ИЦиГ СО РАН – bionet.nsc.ru/vogis/
платформе Elpub – vavilov.elpub.ru/index.php/jour
платформе Научной электронной библиотеки – elibrary.ru/title_about.asp?id=32440
PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

Подписку на «Вавиловский журнал генетики и селекции» можно оформить в любом почтовом отделении России. Индекс издания 42153 по каталогу «Пресса России».

✉ e-mail: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Подписано в печать 20.05.2021. Выход в свет 31.05.2021. Формат 60 × 84 $^{1}/_{8}$. Усл. печ. л. 12.79.
Уч.-изд. л. 15.3. Тираж 150 экз. (1-й завод 1–50 экз.). Заказ № 103. Цена свободная.

Отпечатано в типографии ФГУП «Издательство СО РАН», Морской проспект, 2, Новосибирск, 630090.