

Том 60, Номер 7

ISSN 0016-6758

Июль 2024



ГЕНЕТИКА



НАУКА
— 1727 —

СОДЕРЖАНИЕ

Том 60, номер 7, 2024

Обзорные и теоретические статьи

Искусственный интеллект и классические методы в генетике и селекции животных

А. Д. Солошенко, Э. А. Солошенко, М. Т. Семина, Н. Н. Спасская,
В. Н. Воронкова, Ю. А. Столповский

3

Генетика микроорганизмов

Роль различных субъединиц ремоделирующего комплекса INO80 в репарационной сборке хроматина у дрожжей *Saccharomyces cerevisiae*

Т. А. Евстюхина, Е. А. Алексеева, И. И. Скобелева, В. Т. Пешехонов, В. Г. Королев

17

Генетика животных

Генетическая структура популяций волка Северной Евразии: оценка влияния исключения из анализа родственных особей

П. А. Казимиров, Ю. С. Белоконь, М. М. Белоконь, А. Я. Бондарев, А. В. Давыдов,
Е. С. Захаров, С. В. Леонтьев, Д. В. Политов

31

«Эволюция» митохондриального генома пеночки-таловки (*Phylloscopus borealis sensu lato*) происходит в ее ядерном геноме

Л. Н. Спиридонова, О. П. Вальчук, Я. А. Редькин

45

Генетическое разнообразие малого суслика *Spermophilus pygmaeus* Pallas, 1779 (Sciuridae, Rodentia) на Северном Кавказе

Ф. А. Темботова, М. С. Гудова, А. Х. Амшокова, А. Х. Халидов

62

Генетика человека

Экспрессия гена $\beta 1$ -адренорецептора у пациентов с фибрилляцией предсердий до и после хирургического лечения

В. О. Попова, Э. Ф. Муслимова, Т. Ю. Реброва, Е. А. Арчаков,
Р. Е. Баталов, С. А. Афанасьев

75

Полиморфизм генов антиоксидантов и избыточный вес у детей

М. А. Шкурат, Е. В. Машкина, Н. П. Милютина, Е. Д. Теплякова, Т. П. Шкурат

82

Математические модели и методы

Реконструкция матрицы генотипических корреляций между вариантами внутри гена для совместного анализа импутированных и секвенированных данных

Г. Р. Свищёва, А. В. Кириченко, Н. М. Белоногова, Е. Е. Елгаева, Я. А. Цепилов,
И. В. Зоркольева, Т. И. Аксенович

91

Краткие сообщения

Разнонаправленное изменение уровня метилирования CpG-сайтов в 5' регионе гена *TBX20* в восходящей аорте при атеросклерозе и аневризме

Ю. А. Королёва, И. А. Гончарова, А. А. Зарубин, С. А. Шипулина,
А. А. Слепцов, Д. С. Панфилов, Б. Н. Козлов, М. С. Назаренко

100

Кариотип и молекулярно-генетическая дифференциация 24-хромосомной формы серого хомячка *Nothocricetulus migratorius* из Тянь-Шаня

О. В. Брандлер, А. В. Блехман

107

Профили экспрессии генов, вовлеченных в синтез лигнанов, в развивающихся семенах льна

Е. Н. Пушкова, Е. М. Дворянинова, Л. В. Повхова, Т. А. Рожмина, Р. О. Новаковский,
Е. А. Сигова, А. А. Дмитриев, Н. В. Мельникова

112

Секвенирование и аннотация хлоропластного генома *Triticum timonovum* Heslot et Ferrary

А. Р. Кулуев, Р. Т. Матниязов, Б. Р. Кулуев, Л. Ю. Привалов, А. В. Чемерис

118

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КЛАССИЧЕСКИЕ МЕТОДЫ В ГЕНЕТИКЕ И СЕЛЕКЦИИ ЖИВОТНЫХ

© 2024 г. А. Д. Солошенко^{1,2,*}, Э. А. Солошенко¹, М. Т. Семина¹,
Н. Н. Спасская³, В. Н. Воронкова¹, Ю. А. Столповский¹

¹Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

²Российский государственный аграрный университет — МСХА имени К.А. Тимирязева, Москва, 127434 Россия

³Московский государственный университет имени М.В. Ломоносова, Научно-исследовательский

Зоологический музей, Москва, 125009 Россия

*e-mail: alesol@rgau-msha.ru

Поступила в редакцию 29.11.2023 г.

После доработки 07.03.2024 г.

Принята к публикации 15.03.2024 г.

В настоящей статье проведены обзор и анализ основных методов популяционной генетики и селекции животных, а также математических методов машинного обучения, используемых в животноводстве. На примере двух одомашнированных видов — домашняя лошадь (*Equus caballus*) и северный олень (*Rangifer tarandus*) проведено обучение моделей библиотеки CatBoost. Для обучения модели на основе данных по одомашненным и диким северным оленям, европейским и российским породам лошадей использованы результаты, полученные с помощью микросателлитных панелей, соответственно локусов 16 и 17. Для определения успешности модели были рассчитаны стандартные показатели: Accuracy, Precision, Recall и F1, построены матрицы ошибок. Показаны новые возможности идентификации породной принадлежности животных.

Ключевые слова: машинное обучение, нейронные сети, животноводство, лошадь, *Equus caballus*, северный олень, *Rangifer tarandus*, микросателлитный анализ, генетическое разнообразие.

DOI: 10.31857/S0016675824070017 **EDN:** BIMRAP

Современное развитие индустриального, органического, традиционного животноводства связано с анализом большого спектра хозяйственно-полезных признаков (экстерьера и интерьера животных), качественных и количественных показателей продуктивности, генетических маркеров, генов-кандидатов, данных о секвенировании геномов, в совокупности различных критериев отбора в зоотехнии и ветеринарии. Необходимость работы с большими массивами данных, а также новейшие возможности получения “цифровых фенотипов” определили интенсивное развитие цифровых технологий, математических методов анализа и интегрирование машинного обучения и нейросетей в практику животноводства, в том числе в глобальную проблему сохранения генетических ресурсов, уникального адаптивного потенциала редких локальных пород животных [1].

Классические методы биоинформационного анализа в генетике и селекции животных включают в себя оценку частот, число эффективных и уникальных аллелей, расчеты генетических дистанций, определение филогенеза, определение различных коэффициентов оценки уровня генетического

разнообразия и т. д. Филогенетический анализ, изучение генетических и селекционных взаимоотношений между породными группами и популяциями проводятся с помощью построения дендрограмм. Наибольшее распространение получили методы UPGMA и Neighbor-joining, а также программа STRUCTURE, которая активно используется для исследования популяционной структуры методом байесовского анализа марковских цепей. Для оценки влияния генотипа на продуктивность животных применяются методы дисперсионного анализа, в частности BLUP (Best Linear Unbiased Prediction) и его подвиды, основанные на статистической модели, предложенной С.Р. Henderson [2].

Широкое распространение нейросетей и искусственного интеллекта позволяет проводить исследование в биологии, генетике и животноводстве. В настоящее время машинное обучение используется для мониторинга состояния и благополучия животных при их содержании, в идентификации отдельных особей и других направлениях, например для предсказания корреляций нуклеотидных замен и продуктивности животных. Очевидно, что использование современных методов мониторинга,

оценки родословных, идентификации животных может значительно улучшить качество управления стадом, разведения и селекции животных.

Одной из важных (насущных) проблем животноводства является идентификация породной принадлежности животных. В связи с этим в нашей работе на большом массиве данных по микросателлитным маркерам европейских и российских пород лошадей, а также домашних и диких северных оленей был использован один из методов машинного обучения (CatBoost) и проведено обучение модели с целью идентификации особей, популяций, пород.

В настоящей статье описаны классические методы, которые используются в генетике и селекции животных, проведен анализ современных методов машинного обучения и их перспектив в животноводстве.

ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКИЕ ПАРАМЕТРЫ

Параметры, применяемые для оценки “генетического благополучия” породы или популяции, рассчитываются на основе генетических профилей животных и частот аллелей. Под “генетическим благополучием” мы подразумеваем определенную степень инбридинга и генетического разнообразия в породе или популяции. Оценка генетического разнообразия и филогенетический анализ, основанные на молекулярных маркерах, позволяют идентифицировать породные группы с низким уровнем аллельного разнообразия, которое может снижаться ввиду длительной изолированности популяции, влияния внешних факторов, а также жесткой системы подбора пар животных в племенном животноводстве [3]. Как правило, в племенных хозяйствах используются несколько выдающихся по продуктивности производителей на большом количестве особей маточного поголовья. Интенсивные методы селекционной работы, а также сокращение поголовья сельскохозяйственных животных в России, особенно в коневодстве, приводят к необходимости постоянного контроля за уровнем генетического разнообразия с целью нивелирования негативных эффектов инбридинга [4].

В настоящее время во всем мире для контроля происхождения и определения статуса пород основных domesticiрованных видов животных применяются микросателлитные (short tandem repeats) маркеры, которые рекомендованы ISAG (международное общество генетики животных <https://www.isag.us/>). В то же время все интенсивнее используются подходы, связанные с детекцией однонуклеотидных замен (SNP, single nucleotide polymorphism), особенно ассоциированных с конкретными фенотипическими признаками. Микросателлитные маркеры отличаются высокой

вариабельностью и широкой представленностью в геноме. Панели, применяемые для идентификации животных, используют наиболее полиморфные локусы, которые при этом считаются условно нейтральными (не локализованы рядом с кодирующими участками ДНК, участвующими в отборе) [5].

Благодаря использованию SNP-маркеров в животноводстве возможно ускорение темпов селекции при привлечении таких смежных областей, как эмбриология, биоинформатика и математическая генетика [6].

Наиболее широко распространенными показателями при оценке генетического разнообразия являются: ожидаемая (H_e) и наблюдаемая (H_o) гетерозиготность. Данные параметры основаны на уравнении Харди — Вайнберга и позволяют выявить недостаток гетерозигот в популяции. Ожидаемая гетерозиготность (разнообразие по М. Nei) показывает вероятность гетерозиготности особи в популяции, рассчитывается по формуле:

$H_e = 1 - \sum_i p_i^2$, где p_i — частота i -го аллеля, n_i — общее число аллелей во всех локусах.

Значения для H_e и H_o варьируют от 0 (нет гетерозиготности) до практически 1.

H_o — наблюдаемая гетерозиготность, т. е. фактическая доля гетерозиготных образцов.

Так, при изучении генетического разнообразия пород лошадей России наибольший уровень гетерозиготности был идентифицирован для русской верховой породы лошадей ($H_o = 0.71$), в формировании которой участвуют несколько пород: чистокровная верховая, немецкие спортивные, ахалтекинская [7].

Также проводится расчет параметров инбридинга, в частности коэффициента F_{is} , с помощью которого измеряют внутривидовой инбридинг, т. е. снижение гетерозиготности индивида по причине близкородственных скрещиваний, F_{it} — межвидовой коэффициент инбридинга, где учтены поправки на дифференциацию между популяциями [8].

Для оценки различий между популяциями используют методы расчетов генетических расстояний F_{st} от 0 — различий нет до 1 — максимальное различие. F_{st} также является коэффициентом инбридинга в популяциях в сравнении с общей выборкой. Подобные методы помогают установить дифференциацию популяций [9]. Так, низкий параметр F_{st} может наблюдаться для двух близкородственных пород, например как донская и буденновская породы лошадей (0.02). И напротив, высокий показатель наблюдается у пород, имеющих различное историческое происхождение (тяжеловозные и верховые породы лошадей — 0.15).

Программная среда R, активно используемая для анализа генетических данных, позволяет

визуализировать различные показатели. При использовании пакета PopGenReport частоты аллелей представляются в виде тепловых карт (рис. 1), по которым выявляются общие тренды и индивидуальные для каждой отдельной популяции [10].

Так, при анализе генетического разнообразия ценного промыслового вида – соболя было выявлено среднее число аллелей на популяцию от 7.73 до 10.73 (табл. 1). Так как соболь обладает высокой миграционной активностью, это позволило сделать выводы о его миграциях. Наименьшее количество аллелей наблюдалось на Камчатке (7.73), а наибольшие показатели – для популяций, находящихся на пересечении миграционных потоков и обогащающихся за счет этого новыми аллельными вариантами [11].

Как указывает ряд авторов [12], важно учитывать тот факт, что для анализа разнообразия изначально исследователями выбираются наиболее полиморфные локусы. В исследовании сравнили оценку нуклеотидного разнообразия и анализ по микросателлитным маркерам. Была выявлена положительная корреляция между данными показателями, однако гетерозиготность была выше в 1.4 раза для нуклеотидных замен в сравнении

с панелью полиморфных микросателлитных ЛОКУСОВ.

Кроме того, изначально выбираются нейтральные маркеры, которые подчиняются закону Харди — Вайнберга. Это может являться причиной искажения результатов и смещения оценки ввиду утери части данных.

КЛАСТЕРИЗАЦИЯ ПОПУЛЯЦИЙ

Исследования филогенеза пород одомашненных и популяций диких животных позволяют оценить микроэволюционные процессы и их историческое формирование [10].

Для кластерного анализа (рис. 2) используются программы STRUCTURE и Geneland, с применением алгоритма Монте-Карло по схеме марковских цепей (MCMC) для байесовской статистики. В программах задают предполагаемое количество популяций и число итераций, далее рассчитываются графики для каждого числа популяций по каждой итерации [13].

Для построения непосредственно филогенетических деревьев широкое распространение получили методы UPGMA и Neighbor-joining. Эти

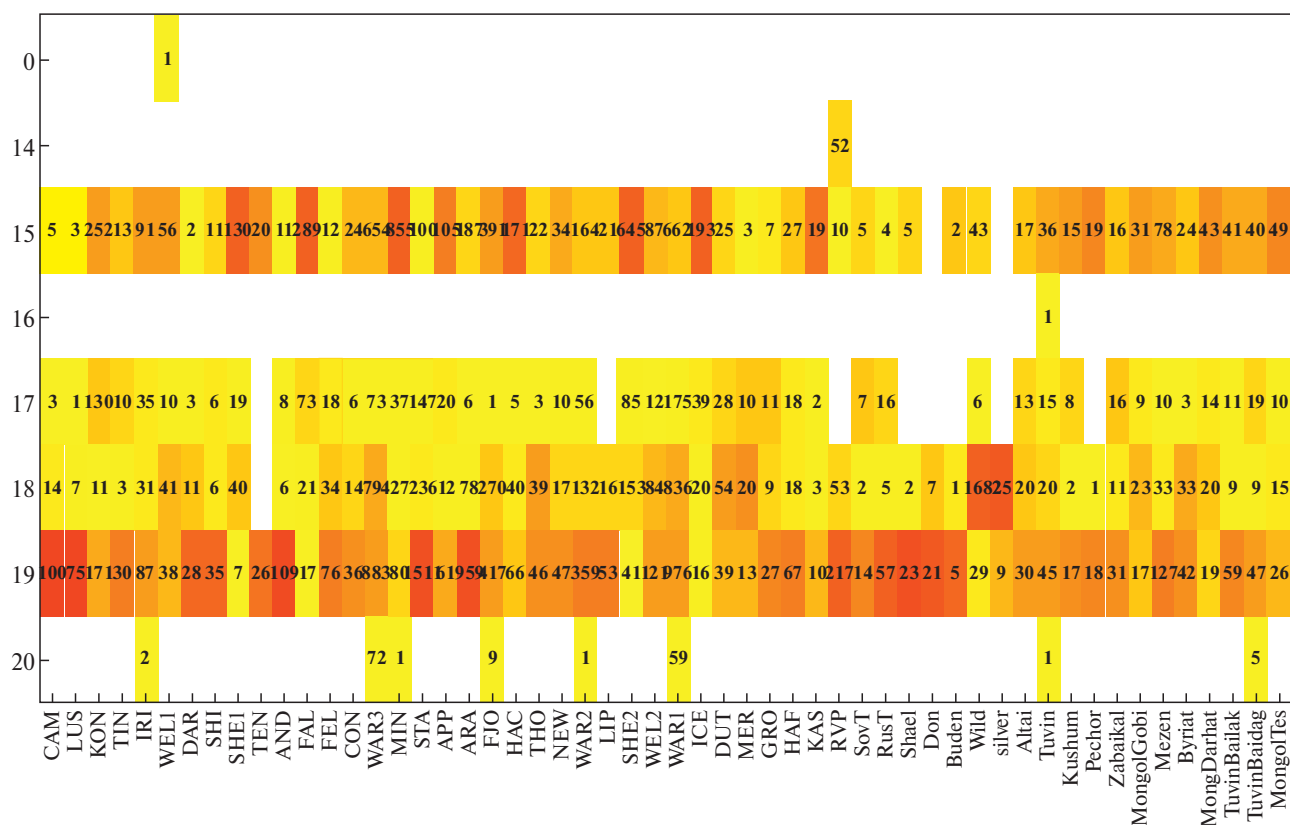


Рис. 1. Тепловая карта частот аллелей для 54 пород лошадей. Цветом от желтого к красному обозначена частота встречаемости аллеля в популяции. Идентифицирован приват-аллель 14-го локуса HTG7 для русской верховой породы лошадей (RVP).

Таблица 1. Показатели генетического разнообразия для различных популяций соболя

Популяция/регион	N	A	AR	H _o	H _e	P
Ванавара (Ц. Сибирь)	31	8.55	6.44	0.704	0.761	0.461
Ербогачен (Ц. Сибирь)	28	8.09	5.99	0.660	0.729	0.415
Братск (Ц. Сибирь)	23	8.18	6.43	0.675	0.770	0.277
Саяны	31	9.18	6.69	0.707	0.786	0.286
Сихотэ-Алинь (Приморье)	40	10.73	7.37	0.741	0.809	0.270
Северный Урал	71	9.36	6.17	0.695	0.733	0.471
Камчатка	37	7.73	5.29	0.713	0.706	0.926

Примечание. N – размер выборки; A – среднее число аллелей на локус; AR – обогаченность популяций аллелями (allelic richness), вычисленная как среднее число аллелей, нормированное на объем выборки N; H_o – наблюдаемая гетерозиготность; H_e – ожидаемая гетерозиготность; P – значение вероятности для теста Харди – Вайнберга с учетом всех локусов.

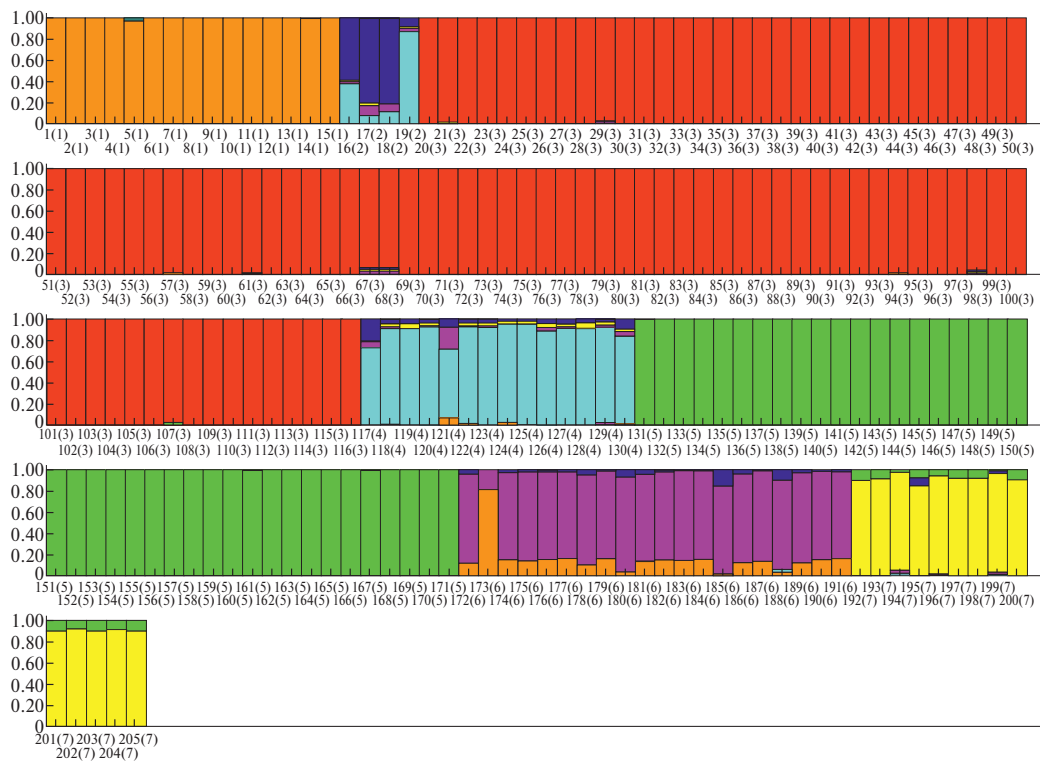


Рис. 2. Популяционная структура заводских пород лошадей. Оранжевый цвет - ахалтекинская, синий – буденновская, красный – одичавшие лошади о. Водный, голубой – донская, зеленый – русская тяжеловозная, розовый – русская верховая, желтый – советская тяжеловозная.

методы являются дистанционными, т. е. первоначальные данные рассчитываются в виде матрицы генетических расстояний, которая впоследствии преобразуется в дихотомичное дерево [14].

Основная идея бутстреп-процедуры по В. Efron состоит в том, чтобы методом статистических испытаний Монте-Карло многократно извлекать повторные выборки из эмпирического

распределения, возможно, при использовании метода главных компонент, позволяющего отойти от стандартной модели построения дендрограмм [15].

Метод главных компонент уменьшает размерность данных, преобразуя их в ковариационную матрицу – матрицу главных компонент. PCA (principal component analysis) отличается высокой степенью воспроизводимости результатов ввиду

отсутствия введения в расчеты различных поправок либо коэффициентов. Единственной мерой расчета является доля дисперсии данных [16].

Таким образом, при использовании метода возможно визуализировать данные в пространстве двух главных компонент (двухмерное пространство) и трех главных компонент (трехмерное пространство) [17].

Сотрудниками лаборатории сравнительной генетики животных был проведен анализ главных компонент для заводских пород лошадей (тяжеловозные и верховые породы) в сравнении с одичавшими лошадьми о. Водный, чей статус и происхождение остаются неуточненными [18]. Построенные методом UPGMA дендрограммы показали низкий уровень бутстреп-поддержки для верховых пород лошадей, что не позволяло достоверно разделить их на отдельные породы и сделать выводы о возникновении одичавшей популяции. Однако в пространстве двух главных компонент было обнаружено частичное перекрытие с буденновской и донской породами лошадей, что позволило сделать предположение о возникновении популяции одичавших лошадей от данных пород (рис. 3).

В большинстве случаев анализ данных происходит существующими пакетами и библиотеками для R или Python. Так, библиотека *poppr* для R позволяет строить UPGMA и NJ деревья. Расчет происходит на основе генетических дистанций Nei

(1972) [19]. Тем не менее авторами пакета указывается ряд моментов, которые необходимо учитывать при обсчетах: различные модели мутационных процессов (пошаговая мутация) либо отсутствие учета мутаций, а также проблемы с обсчетами для организмов с различной ploidyностью [20]. Кроме того, как уже указывалось выше, деревья ограничены дихотомией, что приводит к неоднозначным результатам. Л.А. Животовским в книге “Генетика природных популяций” [8] наглядно проиллюстрированы неоднозначность сжатия матрицы генетических дистанций и дальнейшее построение деревьев (рис. 4).

При построении дерева методом UPGMA наблюдается неоднозначность отнесения популяций 1 и 2 к различным кластерам. При включении обеих популяций или только одной из них наблюдаются три разных дерева с отнесением популяций 1 и 2 либо к отдельному кластеру, либо популяции 1 к кластеру А, а популяции 2 к кластеру В, что в дальнейшем объясняется распределением выборок в пространстве главных компонент. Тем не менее данные методы являются основными при наглядном представлении структуры выборок. Поэтому необходимо проводить расчеты различными методами с дальнейшим их глубоким анализом для понимания популяционных процессов в популяциях.

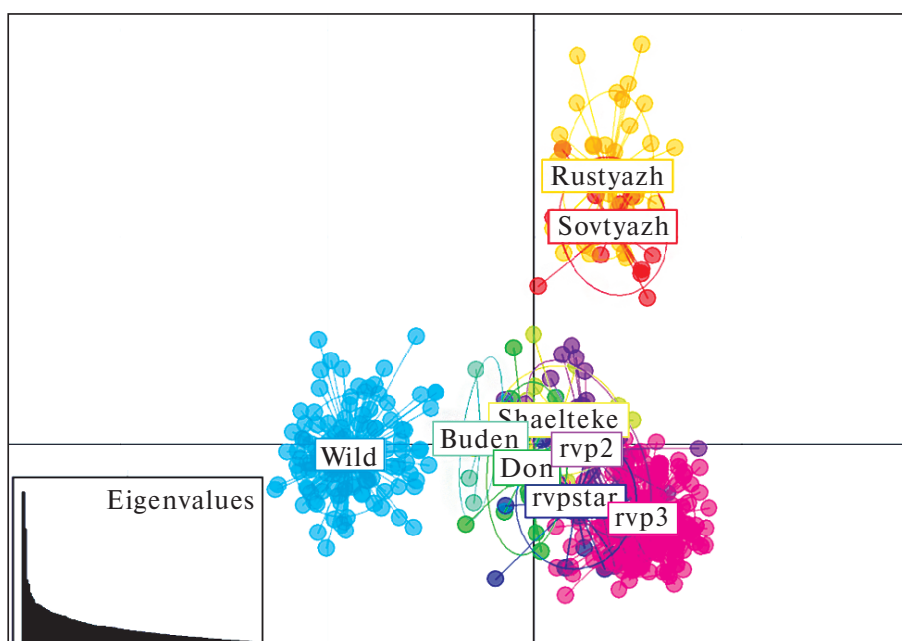


Рис. 3. Распределение верховых и тяжеловозных пород лошадей в пространстве двух главных компонент в сравнении с одичавшими лошадьми о. Водный для уточнения происхождения данной популяции. Wild — одичавшие лошади, Buden — буденновская порода; Shaelteke — ахалтекинские лошади завода “Шаэль”; Don — донская; rvp2, rvpstar, rvp3 — выборки русской верховой породы; Rustyazh — русская тяжеловозная; Sovtyazh — советская тяжеловозная порода.

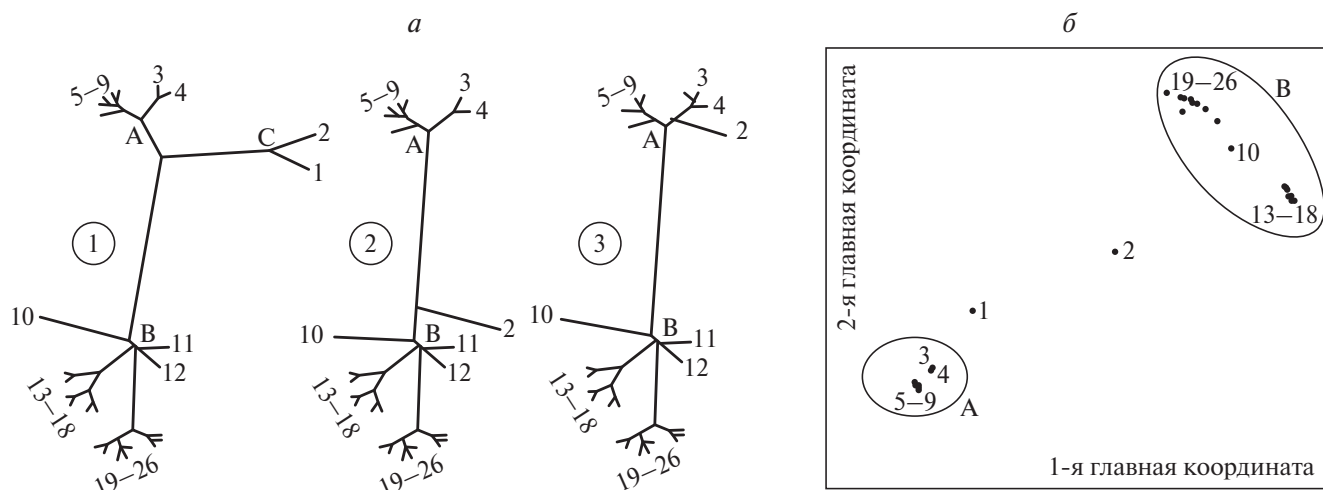


Рис. 4. Построение дерева методом VPGVA. а – дихотомическая кластеризация выборок кеты; б – метод главных компонент.

BLUP (BEST LINEAR UNBIASED PREDICTION)

Для оценки проводимых селекционных программ традиционно использовались методы “матери – дочери”, “дочери – сверстницы”, основанные на расчете разницы средних величин продуктивности животных. Различия в эколого-географических условиях, рационе, условиях содержания животных являлись факторами смещения данных показателей. Для повышения эффективности оценки селекционных процессов был введен метод оценки BLUP (наилучший линейный несмещенный прогноз), позволяющий учитывать влияние как генотипических, так и паратипических факторов, таких как возраст, возраст отела, год, дата постановки на откорм, среднесуточный удой и прочие. В зависимости от целей селекции, отбора и подбора животных, а также наличия тех или иных данных используют различные виды данного метода. Данный метод был предложен С.Р. Хендерсон в 1984 г. [21].

По данным П.И. Отраднова и соавт. приведено уравнение смешанной модели в матричном виде [22]:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Матрицы X и Z являются матрицами дизайна для фиксированных (X) и рандомизированных (Z) эффектов. Соответственно X' и Z' – транспонированные матрицы, h^2 – коэффициент наследуемости (его нет в данной формуле) оценивает отношение изменчивости, обусловленной генетическими факторами, к общей изменчивости (генетическая и паратипическая), из него рассчитывают λ обратная матрица родства).

При использовании методов GWAS (Genome-Wide Association Study) и наличии данных нуклеотидных замен возможно использование метода GBLUP (Genomic Best Linear Unbiased Prediction), при котором используется матрица геномного сходства G вместо матрицы родства.

Так, при исследовании голштинизированного черно-пестрого скота Московской области с использованием моделей BLUP Sire Model (BLUP SM), GBLUP был выявлен среднегодовой генетический тренд +37 кг молока, +1.7 и +0.8 кг продукции молочного жира и белка, а в Ленинградской области – +84, +3.3 и +2.3 кг. Причем было доказано, что оценка по нуклеотидным заменам (GBLUP) достовернее, чем по быкам-производителям (Sire Model) [23].

BLUP удобно использовать в современных условиях автоматизации процессов животноводства, например при использовании откормочных станций, которые позволяют учитывать нахождение на кормовой станции, среднесуточное потребление корма, скорость потребления корма, его конверсию и ряд других показателей, которые затем возможно использовать в будущей модели. Так, при исследовании свиней породы дюрок в модель включались год и месяц рождения животных, дата постановки на откормочную станцию, данные кормовой станции, количество недель выращивания на ней. Помимо этого в анализ была включена матрица родства [24].

Одним из плюсов BLUP следует отметить несмещенность прогноза и отсутствие необходимости в нормальности распределения данных, так как учитываются многие факторы, как генетические, так и средовые.

МАШИННОЕ ОБУЧЕНИЕ

Развитие цифровых технологий и непосредственно технической составляющей современных компьютерных систем позволили ввести анализ больших данных в современное животноводство, генетику и селекцию [25]. Обучение искусственному интеллекту позволяет отойти от классических стандартов применения математических моделей. По данным базы PubMed (<https://www.ncbi.nlm.nih.gov/>) за 2023-й год опубликовано 110 статей по использованию машинного обучения в животноводстве.

Машинное обучение (Machine Learning, ML) — класс методов искусственного интеллекта, основанный на статистических моделях и логических операциях, позволяющий автоматически улучшать вычислительные алгоритмы при отсутствии четких инструкций с использованием примеров данных или прошлого опыта [26].

Обучая программу (алгоритмы) на основе экспериментальных данных по генотипированию животных, мы получаем модель, которая может делать прогнозы (например, определить породу животного) на основе наблюдений (например, по генотипу и фенотипу). Если рассматривать информатику как предмет алгоритмов, то машинное обучение является предметом обучения алгоритмов.

Модель может быть прогностической — моделирует данные в будущем, описательной — получает знания из существующих данных или комбинированной. Машинное обучение использует теорию статистики при построении математических моделей, поскольку основная задача — делать выводы на основе выборки.

В классическом ML выделяют два основных способа:

- “Обучение с учителем” (supervised learning) — такой вид обучения подразумевает наличие размеченных данных (обучающей выборки), потенциально связанных некоторой закономерностью. Модель обучается по принципу “стимул — реакция” и позволяет решать задачи классификации или регрессии, а качество модели определяется по тестовой (иногда валидационной) выборке.
- “Обучение без учителя” или неконтролируемое обучение (unsupervised learning) — обучение на неразмеченных данных. В классических задачах unsupervised learning есть данные, но нет обучающей выборки (т. е. правильные ответы неизвестны). При таком обучении модель обучается выявлять скрытые взаимосвязи без контроля со стороны исследователя и позволяет решать задачи кластеризации, ассоциации и уменьшения размерности (обобщения).

Для определения точности, полученной в ходе обучения модели, используются метрики качества.

Для каждой задачи используются свои метрики. В контексте задач классификации выделим Accuracy, Precision, Recall и F_1 .

Для понимания концепции метрик качества необходимо сказать о матрице ошибок (confusion matrix). Матрица представляет таблицу, позволяющую проиллюстрировать качество обучения модели, как правило, контролируемого обучения (supervised learning). В случаях unsupervised learning ее называют матрицей соответствия (matching matrix) [27].

Разберем, как устроена матрица ошибок на примере задачи бинарной классификации. В таких задачах алгоритм учится предсказывать принадлежность объекта с определенным набором данных (features) к одному из двух классов. Матрица ошибок представлена на рис. 5.

Здесь True class — истинная метка класса на этом объекте, то есть истинные значения классов, изначально содержащиеся в данных. Predicted class — это ответ алгоритма на объекте, т. е. значения классов, которые предсказывает обученная модель для элементов выборки. Если истинная метка класса для объекта 1 и модель отнесла ее к соответствующему классу (1 → 1), то предсказание учитывается в поле True Positive (TP). Если для объекта с истинным классом 0 модель отнесла его к нулю (0 → 0), то предсказание учитывается в True Negative (TN). В случае если объект с меткой 1 был отнесен к классу 0 (1 → 0), предсказание учитывается в поле False Negative (FN), а для меток 0, отнесенных к 1 (0 → 1), — в False Positive (FP).

Таким образом, суммы правильных предсказаний классов записываются в True Positive и True Negative, а ошибки — в False Positive и False Negative.

Например, мы обучили модель на данных, содержащих 100 объектов, разделенных на два класса поровну (50 объектов класса 0 и 50 объектов класса 1), но по различным причинам алгоритм не обучился распознавать классы со 100% точностью. Представим, что проверка модели дала следующие результаты: объектов класса 1 с правильно предсказанными метками было 40 единиц, объектов класса 0 — 30 единиц соответственно, объектов класса 1, отнесенных к классу 0, — 10 единиц, а объектов класса 0, отнесенных к классу 1, — 20 единиц. Матрица ошибок представлена на рис. 6.

Метрика Accuracy (общая точность) — это метрика, которая характеризует долю правильных ответов алгоритма, т. е. то, насколько близок данный набор измерений (классов, наблюдений, показаний) к их истинному значению [28]:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

Predicted class	Positive 1	True Positive (TP) 1 → 1	False Positive (FP) 0 → 1
	Negative 0	False Negative (FN) 1 → 0	True Negative (TN) 0 → 0
		Positive 1	Negative 0
		True class	

Рис. 5. Матрица ошибок.

Predicted class	Positive 1	40 True Positive (TP) 1 → 1	20 False Positive (FP) 0 → 1
	Negative 0	10 False Negative (FN) 1 → 0	30 True Negative (TN) 0 → 0
		Positive 1	Negative 0
		True class	

Рис. 6. Матрица ошибок примера.

Используется, когда классы сбалансированы. В случае дисбаланса классов лучше воспользоваться другими метриками. Для оценки качества модели на каждом из классов по отдельности обычно используют метрики Precision и Recall, они не зависят, в отличие от Accuracy, от соотношения классов и потому применимы в условиях несбалансированных выборок.

Метрика Precision (положительная точность, positive predictive value) — это доля правильно предсказанных положительных объектов относительно всех объектов, отнесенных моделью к положительному классу [29]:

$$precision = \frac{TP}{TP + FP}.$$

Чем меньше ложноположительных срабатываний будет допускать модель, тем больше будет ее Precision.

Метрика Recall (полнота, sensitivity in diagnostic binary classification) характеризует долю правильно предсказанных положительных объектов среди всех объектов положительного класса [28]:

$$recall = \frac{TP}{TP + FN}.$$

Recall не зависит от True Negative и False Positive в отличие от Precision, то есть, если модель правильно прогнозирует все Positive объекты, Recall будет равен единице (100%) даже для случаев, когда все объекты класса Negative были ошибочно определены как Positive. Для объединения Precision и Recall в обобщенный критерий качества применяют F-меру.

Метрика F1-мера (F1-measure) — среднее гармоническое Precision и Recall, является мерой точности в задачах бинарной классификации [30]:

$$F_1 = 2 \frac{recall \times precision}{recall + precision} = \frac{TP}{TP + \frac{FP + FN}{2}}.$$

F1-мера достигает максимума при Precision и Recall, равных единице, и близка к нулю, если один из аргументов близок к нулю. F1-мера применяется, если метрики precision и recall одинаково важны.

Для случаев, когда одна из используемых метрик приоритетнее, используют F_β -меру:

$$F_\beta = (1 + \beta^2) \times \frac{precision \times recall}{\beta^2 \times precision + recall},$$

где β — весовой коэффициент Precision в F_β -мере.

F_β -мера применяет дополнительные веса, придавая одной из метрик Precision или Recall большее значение, чем другой.

Таким образом, качество обучения моделей определяется по соответствующим метрикам.

Машинное обучение применимо для решения задач в области генетики, например, когда в качестве исходных для обучения модели данных (features) используются нуклеотидные последовательности или генотипы. Предсказание уровня и направленности экспрессии генов на основе большого объема данных по SNP позволяет определять эффекты различных нуклеотидных замен, а также последствия соматических мутаций и изменения в хроматине. Существующая модель DeepSEA — система алгоритмов глубокого обучения, созданная для этих целей, предсказывает изменения в хроматине на основании данных секвенирования с чувствительностью до одного нуклеотида. Широкое развитие GWAS и NGS методов позволяет накапливать большое количество датасетов по rs (Reference SNP cluster ID), ассоциированных с

различными заболеваниями, что в совокупности с данной моделью может быть актуально для исследования в области медицинской генетики [31].

В области животноводства активно исследуются и разрабатываются различные технологии: компьютерное зрение для решения задач классификации и мониторинга состояния и поведения животных, электронные ошейники для наблюдений за состоянием здоровья и активностью носителя, модели машинного и глубокого обучения для ускорения и повышения качества селекционных процессов, экспертные системы в области оптимизации процессов учета животных, анализа родословных, точечных мутаций, их влияния на резистентность и продуктивность животных и другие. Использование подобных технологий позволяет снизить затраты труда при содержании животных и ветеринарном контроле [32].

Так как паспортизация и создание баз данных генетических профилей животных не всегда доступны для фермеров и селекционеров, в особенности в отдаленных регионах, все чаще начинают применяться методы обучения моделей на основе фотографического материала.

Например, модель VGG-16 (рис. 7), предложенная К. Simonyan и А. Zisserman в 2014 г., достигает

высокой точности при определении объектов на изображении (93%) [33]. В качестве DataSet используется база данных ImageNet (<https://image-net.org/>), в которой в настоящее время находится 1.2 млн изображений, относящихся к 1000 категориям.

На основе данной модели ряд исследователей обучили свои варианты нейросети для задач идентификации как отдельных животных, так и пород на основе их фотографий. Применение предобученных моделей для решения узконаправленных задач — распространенная практика на сегодняшний день.

S.A. Jwade с соавт. обучили модель определять одну из четырех пород овец: меринос, суффолк, белый суффолк, полл дорсет, используя данные 1642 животных [34]. Для стандартизации условий для фото был создан специальный загон. Таким образом удалось снизить вероятность ошибочного обучения ввиду различий в средовых условиях съемки. Показатель Аккуратности при этом достиг 95.8%.

Подобные исследования проведены и на других видах животных. В 2020 г. были проведены исследования для идентификации собак по коллекции фото [35]. Актуальность этой работы заключается в возможности внедрения предложенного метода для

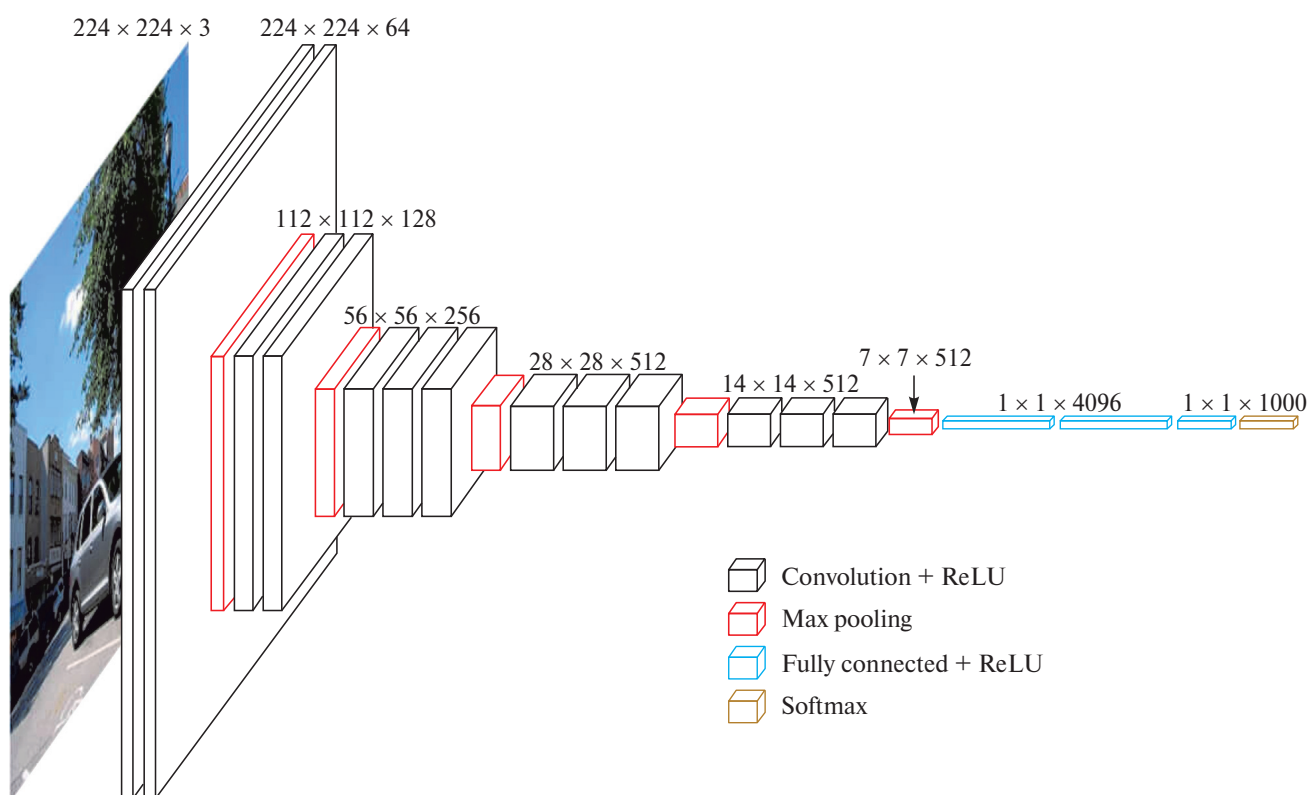


Рис. 7. Архитектура VGG-16.

поиска пропавших животных. Авторы ссылаются на ассоциацию American Pet Products Association (APPA), по данным которой из 78 млн домашних собак 15% теряются и 7% (819 000 особей) не удается вернуть владельцам. Несмотря на небольшой объем входных данных, включающих 21 собаку пород хаски и мопсов, а также данные из интернет-ресурсов по 10 породам, с пятью фотографиями по каждой породе, удалось обучить модель до 77.19% точности на кросс-валидации.

Сотрудники лаборатории сравнительной генетики животных ИОГен им. Н.И. Вавилова РАН провели исследование по оценке возможности использования моделей машинного обучения для определения классификации пород и популяций животных по данным анализа микросателлитных локусов. Исследования были проведены для двух видов: лошадь (*Equus caballus*) — классификация пород и северный олень (*Rangifer tarandus*) — идентификация диких и домашних популяций.

Для исследования лошадей использовалась стандартная панель из 17 микросателлитных локусов (AHT4, AHT5, ASB2, ASB17, ASB23, CA425, HMS1, HMS2, HMS3, HMS6, HMS7, HTG4, HTG6, HTG7, HTG10, LEX3, VHL20), одобренная для генотипирования и паспортизации ISAG (International Society for Animal Genetics). Датасет содержал информацию по генотипам микросателлитных локусов более чем 3400 особей 14 пород. Исследование северного оленя проводилось по 16 микросателлитным локусам (BMS1788, RT30, RT1, RT9, C143, RT7, ONEQ, FCB193, RT6, C217, RT24, C32, BMS745, NVHRT16, T40, C276) для популяций дикого и домашнего оленя с соответствующим разбиением на две популяции [36]. Выборки домашних оленей представлены четырьмя зарегистрированными в РФ породами: ненецкой, чукотской, эвенской и эвенкийской, а также двумя экотипами из Тофаларии и Тоджи. Популяции диких оленей были представлены образцами из Магаданской, Амурской, Таймырской, Якутской, Турской, Мурманской и Чукотской областей.

В качестве обучаемой модели была выбрана модель CatBoostClassifier из библиотеки CatBoost от Yandex. Функция потерь — MultiClass. [37] Соотношение обучающей и валидационной выборок составляло 80 к 20, с перетасовкой строк и выравниванием классов (стратификацией).

Наблюдаемые значения precision (табл. 2) находятся в диапазоне от 0.73 (чистокровная верховая порода) до 1 (советская тяжеловозная, одичавшие лошади о. Водный, русская тяжеловозная, нью-форест, донская, андалузская), минимальные значения свидетельствуют о высокой доле неправильно отнесенных к данному классу лошадей других пород. Метрика Recall, оценивающая способность улавливать необходимый класс, зависит

от доли неверно отрицательных (False Negative), т. е. упущенных объектов и варьирует от 0.5 (советская тяжеловозная) до 1 для большей части пород. Наглядное представление данных в виде матрицы ошибок представлено на рис. 7.

Цветом от синего к желтому показана точность идентификации. Одичавшие лошади о. Водный, ахалтекинская, русская тяжеловозная, фризская, фьорд, донская, арабская породы идентифицируются со 100% вероятностью.

Советская тяжеловозная порода в 50% случаев неверно определяется как ахалтекинская порода лошадей, что может быть вызвано низким уровнем генетического разнообразия и количества приват-аллелей в данной породе. Практический вопрос дифференциации советской тяжеловозной от ахалтекинской породы не был приоритетным, так как это абсолютно разные по фенотипическим показателям лошади; кроме того, ахалтекинская порода является чистокровной породой лошадей с “закрытым генофондом”. При идентификации советской тяжеловозной породы модель не совершает ошибок при сравнении с другими упряжными и тяжеловозными лошадьми и может использоваться при оценке уровня межпородной гибридизации. Порода аппалуза верно определяется в 63% случаев, потери (ошибки) связаны с чистокровной верховой породой лошадей, которая часто используется для улучшения скаковых и верховых качеств в селекции многих пород лошадей. Стоит отметить идентификацию в 100% случаев одичавших лошадей о. Водный и их дифференциацию от породы-основательницы — донской, так как эти породы схожи по экстерьерным показателям и масти, и в данном случае идентификация с использованием молекулярных маркеров является актуальной. В целом обучение модели можно считать успешным, средне-взвешенная метрика Ассигасу составляет 0.96.

При исследовании популяций северного оленя была обучена модель с целью идентификации породной принадлежности домашнего оленя и их дифференциации от диких популяций. Матрицы ошибок, полученные в ходе обучения модели, представлены на рис. 9 и 10.

Несмотря на то что в оленеводстве используется пастбищный тип содержания и нередко случаи для так называемого “освежения крови” или использования диких самцов в одомашненных стадах, модель успешно разделяет домашних и диких оленей, что предоставляет возможность исследовать процесс доместикации и решать практическую задачу, в частности контролировать браконьерство. При разделении оленей на две группы (дикие и домашние) ошибочно идентифицированы были только две особи из 770. Ошибка модели стремится к 0.

При разделении группы домашних оленей на различные породы и экотипы наблюдается

Таблица 2. Значения метрик качества модели для пород лошадей (*Equus caballus*)

Порода	Сокращенное название	Precision	Recall	F1-score
Андалузская	AND	1.00	1.00	1.00
Аппалуза	APP	0.86	0.63	0.73
Арабская	ARA	0.94	0.95	0.95
Донская	DON	1.00	1.00	1.00
Фьорд	FJO	0.99	1.00	1.00
Фризская	FRI	0.99	1.00	1.00
Нью форест	NEW	1.00	0.70	0.82
Русская тяжеловозная	RusT	1.00	1.00	1.00
Русская верховая	Rwp	0.88	0.88	0.88
Стандартбредная	STA	0.98	0.99	0.98
Ахалтекинская (к/з Шаэль)	Shael	0.67	1.00	0.80
Чистокровная верховая	THO	0.73	0.80	0.76
Одичавшие лошади о.Водный	Wild	1.00	1.00	1.00
Советская тяжеловозная	sovt	1.00	0.50	0.67
	Accuracy			0.96
	Macro avg	0.93	0.89	0.90
	Weighted avg	0.96	0.96	0.96

снижение вероятности верно идентифицировать отдельные породы.

Породная идентификация вызывает затруднения с эвенской породой, возможно это связано с отсутствием консолидированной структуры, геонгеографическими особенностями, случайными скрещиваниями с другими породами из-за разведения данной породы на огромных территориях: Саха-Якутии, Магаданской области и на севере Камчатского края. Именно эти ареалы разведения северного оленя являются центром миграционных путей северных оленей [38]. Чукотская порода определяется с вероятностью 90% (в 10% случаев идентифицируется как эвенская), эвенкийская — 75%, эвенская — 50%, ненецкая — 98%, экотипы из Тоджи — 89% и Тофаларии — 100%. При этом модель ошибочно относит в 10% случаев эвенскую породу к диким оленям. В остальных случаях ошибки в идентификации домашних оленей как диких отсутствуют, что позволяет использовать модель, а именно метод CatBoost, для решения вопросов идентификации индивидов и различий между домашними и дикими северными оленями.

ЗАКЛЮЧЕНИЕ

Одной из проблем, возникающих в процессе применения искусственного интеллекта в науке, является интерпретация результатов обучения моделей. Сложности возникают из-за того, что модели искусственного интеллекта часто работают на основе сложных алгоритмов и большого количества данных, которые трудно обработать человеку. Глубокое обучение и машинное обучение предоставляют разные подходы к созданию моделей искусственного интеллекта.

Машинное обучение зачастую использует простые модели, такие как деревья решений или линейные модели, которые легче интерпретировать, чем модели глубокого обучения. Обычно они имеют меньше параметров и используют более простой математический аппарат, что делает их понятнее для человека.

Модели глубокого обучения, такие как нейронные сети, могут быть более сложными и трудными для интерпретации, поскольку они имеют множество скрытых слоев и множество нейронов в каждом слое. Они способны аппроксимировать

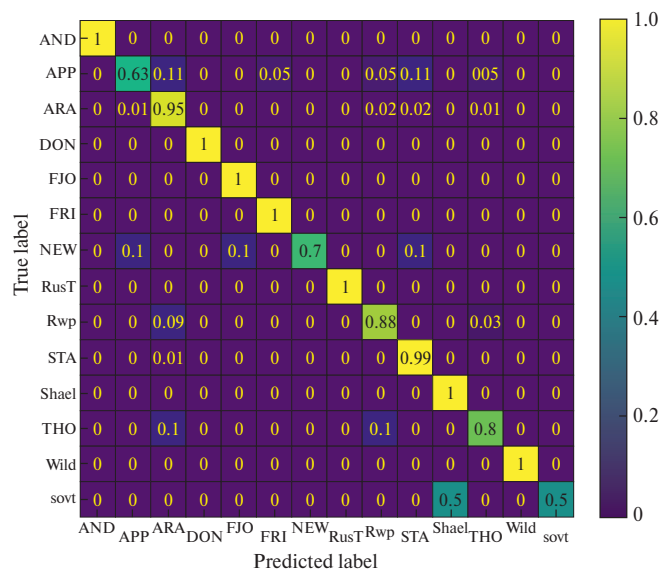


Рис. 8. Матрица ошибок для каждой пары исследуемых пород.

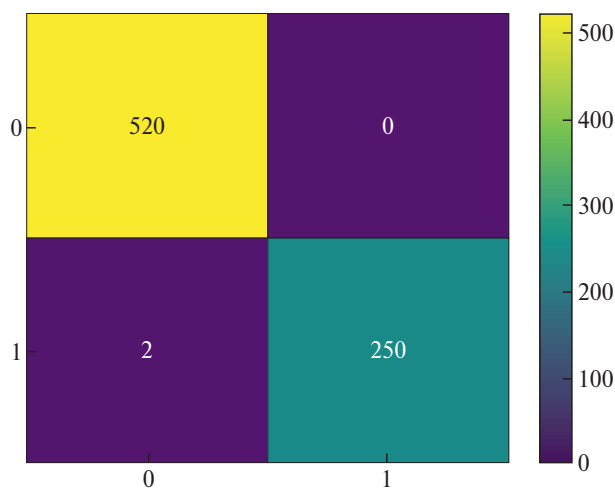


Рис. 9. Матрица ошибок модели бинарной классификации для домашних и диких оленей. 0 — домашние, 1 — дикие.

сложные функции, но это может затруднить понимание того, как они принимают решения.

Таким образом, интерпретируемость является важным фактором при выборе между глубоким и машинным обучением. Если требуется более интерпретируемая модель, то модель машинного обучения может стать лучшим выбором. Если же требуется более мощная модель, способная аппроксимировать сложные зависимости, возможно, стоит обратиться к глубокому обучению.

С точки зрения генетики и селекции животных, равно как и для многих других отраслей науки, существует проблема накопления и создания больших массивов данных. В области животноводства стоит отметить отсутствие достаточных объемов

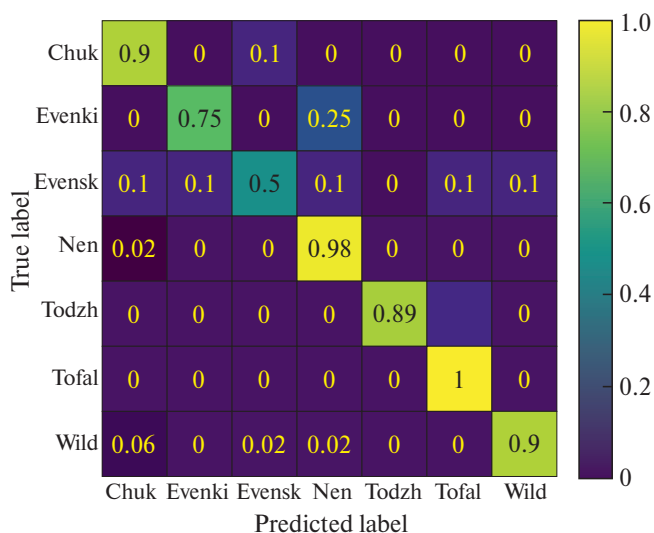


Рис. 10. Матрица ошибок модели для пород домашних оленей и их диких популяций.

фенотипических (зоотехнических и ветеринарных данных), генетических (различные типы маркеров, гены, секвенированные геномы), фотографических и видеоматериалов для создания баз данных, которые могут быть использованы или исследованы при помощи различных методов искусственного интеллекта. При этом достоверность полученных данных напрямую влияет на качество обучаемых моделей.

Методы машинного обучения в генетике и селекции в ближайшем будущем станут основой для решения широкого спектра научных и практических задач, таких как: оценка племенной и генофондной ценности животных, адаптивности, жизнеспособности, психотипа, генетического потенциала, прогноз по их использованию, созданию оптимальных условий содержания и кормления, подбор и отбор, меж- и внутрипородные варианты скрещивания, создания новых пород или селекционных достижений.

В настоящей работе использование микросателлитных баз данных по двум видам одомашненных животных и мультиклассовых моделей машинного обучения позволило с достаточной точностью идентифицировать породную принадлежность животных, а именно породы лошадей, и различить одомашненную и дику форму северного оленя.

Перспективы использования методов машинного обучения в традиционной, геномной, маркер-зависимой, эпигенетической селекции огромны. Оценка генетической ценности, здоровья животных, их адаптивности к различным агроклиматическим условиям, поиск новых генов, анализ их взаимодействия и т. д. выходят на новый уровень, где сложно переоценить возможности искусственного интеллекта.

Работа выполнена при поддержке гранта Российского научного фонда № 23-16-00059.

Все применимые международные, национальные и/или институциональные принципы ухода и использования животных были соблюдены.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Моисеева И.Г., Уханов С.В., Столповский Ю.А. и др. Генофонды сельскохозяйственных животных. Генетические ресурсы животноводства России. М.: Наука, 2006. 462 с.
2. Weigel K.A., VanRaden P.M., Norman H.D., Grosu H. A 100-year review: Methods and impact of genetic selection in dairy cattle—from daughter-dam comparisons to deep learning algorithms // *J. Dairy Sci.* 2017. V. 100. № 12. P. 10234–10250.
3. Храброва Л.А., Зайцев А.М., Суходольская И.В. и др. Проблемы учета и сохранения аборигенных пород лошадей // *Аборигенное коневодство России: история, современность, перспективы: Сб. науч. трудов по матер. II Всеросс. научно-практ. конф. с междунар. участием. Мезень, 2018. С. 170–176.*
4. Николаева Э.А., Спаская Н.Н., Столповский Ю.А., Воронкова В.Н. Структура популяций заводских и вторично одичавших лошадей // *Генетические процессы в популяциях: Материалы науч. Конф. с междунар. участием, посвященной 50-летию юбилею лаборатории популяционной генетики им. Ю.П. Алтухова ИОГен РАН и 85-летию со дня рождения академика Юрия Петровича Алтухова. 2022. С. 45.*
5. Ashley M.V., Dow B.D. The use of microsatellite analysis in population biology: background, methods and potential applications // *Mol. Ecol. Evol.: Approaches and Applications.* 1994. P. 185–201.
6. Столповский Ю.А., Пискунов А.К., Свищева Г.Р. Геномная селекция. I: Последние тенденции и возможные пути развития // *Генетика.* 2020. Т. 56. № 9. С. 1006–1017. <https://doi.org/10.31857/S0016675820090143>
7. Николаева Э.А., Воронкова В.Н., Политова М.А. и др. Генетическая структура русской верховой породы лошадей // *Генетика.* 2023. Т. 59. № 9. С. 1048–1058. <https://doi.org/10.31857/S0016675823090096>. EDN WUWYIE.
8. Животовский Л.А. Генетика природных популяций. Йошкар-Ола: Вертикаль, 2021. 600 с.
9. Meirmans P.G., Hedrick P.W. Assessing population structure: FST and related measures // *Mol. Ecol. Res.* 2011. V. 11. № 1. P. 5–18. <https://doi.org/10.1111/j.1755-0998.2010.02927.x>
10. Adamack A.T., Gruber B. Popgenreport: Simplifying basic population genetic analyses in R // *Methods Ecol. Evol.* 2014. V. 5. N 4. P. 384–387. <https://doi.org/10.1111/2041-210X.12158>
11. Кауштанов С.Н., Свищева Г.Р., Пищулина С.Л. и др. Географическая структура генофонда соболя (*Martes zibellina* L.): данные анализа микросателлитных локусов // *Генетика.* 2015. Т. 51. №. 1. С. 78–78. <https://doi.org/10.1134/S1022795415010044>
12. Väli Ü., Einarsson A., Waits L., Ellegren H. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? // *Mol. Ecol.* 2008. V. 17. № 17. P. 3808–3817.
13. Porras-Hurtado L., Ruiz Y., Santos C. et al. An overview of STRUCTURE: Applications, parameter settings, and supporting software // *Front. in Genet.* 2013. V. 4. P. 98. <https://doi.org/10.3389/fgene.2013.00098>
14. Gronau I., Moran S. Optimal implementations of UPGMA and other common clustering algorithms // *Inform. Proc. Letters.* 2007. V. 104. № 6. P. 205–210. <https://doi.org/10.1016/j.ipl.2007.07.002>
15. Efron B. Bootstrap methods: Another look at the jackknife // *Ann. Statist.* 1979. V. 7. P. 1–26. <https://doi.org/10.1214/aos/1176344552>
16. Reich D., Price A., Patterson N. Principal component analysis of genetic data // *Nat. Genet.* 2008. V. 40. P. 491–492. <https://doi.org/10.1038/ng0508-491>
17. Sievert C. Interactive Web-based Data Visualization With R, plotly, and shiny. CRC Press, 2020.
18. Spasskaya N.N., Voronkova V.N., Letarov A.V. et al. Features of reproduction in an isolated island population of the feral horses of the Lake Many-Gudilo (Rostov Region, Russia) // *App. An. Beh. Sci.* 2022. V. 254. <https://doi.org/10.1016/j.applanim.2022.105712>
19. Maloy S., Hughes K. Brenner's Encyclopedia of Genetics. MS, Cambridge: Academic Press., 2013.
20. Ruzica Bruvo, Nicolaas K. Michiels, Thomas G. D'Souza, Hinrich Shulenberg. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level // *Mol. Ecol.* 2004. V. 13(7). P. 2101–2106.
21. Henderson C.R. Applications of linear models in animal breeding. Guelph, Canada: Univ. Guelph Press. 1984. 462 p.
22. Отрадных П.И., Рудиянов Д.М., Белоус А.А. Валидация оценок племенной ценности свиней породы дюрок по признакам кормового поведения // *Свиноводство.* 2023. № 5. С. 22–26. <https://doi.org/10.37925/0039-713X-2023-5-22-26>
23. Сермягин А.А., Белоус А.А., Контэ А.Ф. и др. Валидация геномного прогноза племенной ценности быков-производителей по признакам молочной

- продуктивности дочерей на примере популяции черно-пестрого и голштинского скота // С.-х. биология. 2017. Т. 52. № 6. С. 1148–1156.
24. Контэ А.Ф., Белоус А.А., Отраднов П.И. Племенная ценность кормового поведения свиней // Аграрный вестник Урала. 2022. №. 10 (225). С. 44–53.
 25. Nayeri S., Sargolzaei M., Tulpan D. A review of traditional and machine learning methods applied to animal breeding // Animal Health Res. Rev. 2019. V. 20, P. 31–46. <https://doi.org/10.1017/S1466252319000148>
 26. Zhou Z.H. *Machine Learning*. London: Springer Nature, 2021. 460 p. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
 27. Stehman S.V. Selecting and interpreting measures of thematic classification accuracy // Remote Sensing of Environment. 1997. V. 62. № 1. P. 77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
 28. Erickson B.J., Kitamura F. Magician's corner: 9. Performance metrics for machine learning models // Radiology: Artificial Intelligence. 2021. V. 3. № 3. <https://doi.org/10.1148/ryai.2021200126>
 29. Powers D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation // arXiv preprint arXiv:2010.16061. 2020. <https://doi.org/10.48550/arXiv.2010.16061>
 30. Sasaki Y. The truth of the F-measure // Teach Tutor Mater. 2007. V. 1. № 5. P. 1–5.
 31. Penzar D.D., Zinkevich A.O., Vorontsov I.E. What do neighbors tell about you: The local context of cis-regulatory modules complicates prediction of regulatory variants // Front. Genet. 2019. V. 10. <https://doi.org/10.3389/fgene.2019.01078>
 32. Михальский А.И., Новосельцева Ж.А. Применение методов машинного обучения в задачах продуктивного животноводства // Пробл. биол. продуктивных животных. 2018. № 4. С. 98–109. <https://doi.org/10.25687/1996-6733.prodanimbiol.2018.3.98-109>
 33. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // arXiv preprint arXiv:1409.1556. 2014. doi 10.48550/arXiv.1409.1556
 34. Jwade S.A., Guzzomi A., Mian A. On farm automatic sheep breed classification using deep learning // Computers and Electronics in Agriculture. 2019. V. 167. <https://doi.org/10.1016/j.compag.2019.105055>
 35. Batic D., Culibrk D. Identifying individual dogs in social media images // arXiv:2003.06705. 2019.
 36. Столповский Ю.А., Бабаян О.В., Каишанов С.Н. и др. Генетическая оценка пород северного оленя (*Rangifer tarandus*) и их дикого предка с помощью новой панели STR-маркеров // Генетика. 2020. Т. 56. № 12. С. 1409–1425. <https://catboost.ai/en/docs/concepts/loss-functions-multiclassification#usage-information>
 37. Южаков А.А., Мухачев А.Д., Лайшев К.А. Породы и проблемы селекции северных оленей России. М.: Наука, 2023. 165 с.

Artificial Intelligence and Classical Methods in Animal Genetics and Breeding

A. D. Soloshenkov^{1,2,*}, E. A. Soloshenkova¹, M. T. Semina¹,
N. N. Spasskaya³, V. N. Voronkova¹, Y. A. Stolpovky¹

¹Vavilov Institute of General Genetic, Russian Academy of Sciences Moscow, 119991 Russia

²Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, 127434 Russia

³Zoo museum of Moscow State University, Moscow, 125009 Russia

*e-mail: alesol@rgau-msha.ru

The article analyses basic methods of population genetics and animal breeding, as well as mathematical methods of machine learning used in animal breeding. The training of cat boost library models was carried out on the example of two domesticated species – domestic horse (*Equus caballus*) and reindeer (*Rangifer tarandus*). Data from microsatellite panels of 16 and 17 loci, respectively, were used to train the model using data on domesticated and wild reindeer, European and Russian horse breeds. The standard indicators: accuracy, precision, recall and f1 were calculated to determine the success of the model. Confusion matrices were constructed. New possibilities of identification of animal breed affiliation were shown.

Keywords: machine learning, neural networks, animal husbandry, horse, *Equus caballus*, reindeer, *Rangifer tarandus*, microsatellite analysis, genetic diversity

РОЛЬ РАЗЛИЧНЫХ СУБЪЕДИНИЦ РЕМОДЕЛИРУЮЩЕГО КОМПЛЕКСА INO80 В РЕПАРАЦИОННОЙ СБОРКЕ ХРОМАТИНА У ДРОЖЖЕЙ *Saccharomyces cerevisiae*

© 2024 г. Т. А. Евстюхина^{1, 2}, Е. А. Алексеева^{1, 2, *}, И. И. Скобелева¹,
В. Т. Пешехонов^{1, 2}, В. Г. Королев^{1, 2}

¹Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра “Курчатовский институт”, Ленинградская область, Гатчина, 188300 Россия

²Курчатовский геномный центр — ПИЯФ, Ленинградская область, Гатчина, 188300 Россия

*e-mail: alekseeva_ea@npi.nrcki.ru

Поступила в редакцию 13.12.2023 г.

После доработки 02.02.2024 г.

Принята к публикации 12.02.2024 г.

Репаративная сборка хроматина является важным шагом в поддержании стабильности генома. Правильную сборку хроматина обеспечивают шапероны гистонов, нарушение функции которых может привести к развитию различных форм рака и к ряду наследственных заболеваний у человека. Действие факторов ремоделирования завершает репарацию хроматина. Ремоделирующий комплекс хроматина дрожжей INO80 играет важную роль в архитектуре хроматина. Мы использовали индуцированный мутагенез и ПЦР в реальном времени для изучения роли INO80 в репарационной сборке хроматина. У двойных мутантов *ies5Δ hsm3Δ (hif1Δ)* дефекты структуры нуклеосом, вызванные мутациями *hsm3Δ* и *hif1Δ*, приводят к гиперчувствительности клеток к УФ-облучению и исчезновению *hsm3-* и *hif1-*специфического мутагенеза. Двойные мутанты, несущие мутацию *nhp10Δ* и мутацию *hsm3Δ* или *hif1Δ*, были неотличимы от одиночного мутанта по летальному эффекту УФ-облучения, однако высокий УФ-индуцированный мутагенез, характерный для всех мутаций, исчезал. Таким образом, мы обнаружили, что мутации в генах, контролирующих субъединицы комплекса INO80, могут проявлять сильные взаимодействия с мутациями в генах гистоновых шаперонов. Была подтверждена гипотеза о том, что белок Nim1 выполняет шаперонную функцию в процессе репаративной сборки хроматина.

Ключевые слова: комплекс INO80, *Saccharomyces cerevisiae*, ген *IES5*, ген *NHP10*, УФ-индуцированный мутагенез.

DOI: 10.31857/S0016675824070023 **EDN:** BILQZP

Геном эукариот упакован в хроматин в виде ядерных нуклеосомных частиц, содержащих ядро октамера гистонов, обернутое примерно дважды 147 парами оснований ДНК [1]. Гистоны — это сильно заряженные основные белки, которые могут связываться с различными белками и нуклеиновыми кислотами. Для предотвращения неспецифического взаимодействия гистонов используются шапероны, которые помогают регулировать их правильную укладку в нуклеосомах [2]. Было показано, что посттрансляционное ацетилирование N-концевых хвостов гистонов является важным механизмом, с помощью которого клетки регулируют доступность хроматина [3]. Известно, что в ходе эксцизионной репарации нуклеотидов (НЭР) вновь репарированная ДНК собирается в хроматин

[4] при помощи гистоновых шаперонов Asf1, CAF-1 и комплекса NuB4 [4, 5].

Ремодуляторы хроматина — это специализированные мультибелковые машины, которые могут изменять структуру, состав и расположение нуклеосом в ДНК. Ремоделирующие комплексы (РМК) по-разному влияют на структуру хроматина. Эукариоты содержат как минимум пять семейств РМК: SWI/SNF, ISWI, NURD, INO80 и SWR1. Все РМК содержат АТФазные субъединицы, которые имеют очень похожие АТФазные домены. С-концы этих белков имеют уникальные домены, которые определяют их специфическое связывание или регуляцию. Ферментативная активность этих комплексов приводит к изменению положения нуклеосом вдоль ДНК, увеличению доступности ДНК на

поверхности нуклеосомы и/или обмену гистоновых белков внутри нуклеосом [6, 7].

Комплекс INO80 представляет собой эволюционно консервативный АТФ-зависимый комплекс ремоделирования хроматина, который первоначально был описан как участвующий в транскрипции и репарации ДНК [8–11]. В дрожжах *S. cerevisiae* INO80 необходим для адекватного процессинга репликативных вилок в клетках после обработки метилметансульфонатом вследствие его роли в толерантности к повреждениям ДНК [12]. Было показано, что INO80 необходим для эффективного привлечения Rad18, который инициирует безошибочный путь пострепликативной репарации. Таким образом, INO80 является регулятором толерантности к повреждениям ДНК во время репликации [13].

Комплекс INO80 у *S. cerevisiae* состоит из 15 различных субъединиц: Ino80, Rvb1, Rvb2, Arp4, Arp5, Arp8, Act1, Taf14, Nhp10, Ies1, Ies2, Ies3, Ies4, Ies5 и Ies6 [12]. В дополнение к каркасу Ino80–Ies2 были идентифицированы четыре топологических модуля: Nhp10–Ies1–Ies3–Ies5 (Nhp10-модуль), Arp4–Arp8–Act1–Taf14–Ies4 (Arp8-модуль), Rvb1–Rvb2 (Rvb1/2-модуль) и Arp5–Ies6 (Arp5-модуль). Сборка Arp5–Ies6 необходима для сопряжения гидролиза АТФ с продуктивным движением нуклеосом [14]. Функция модуля Arp5 важна для стимулированного нуклеосомами гидролиза АТФ и ремоделирования хроматина [15], что подтверждает идею о том, что модуль Arp5 функционирует как комплекс нуклеосома/гистон–шаперон и посредник ремоделирования, который связывается с ДНК. Модуль Nhp10 INO80 представляет собой модуль связывания нуклеосом с высоким сродством. Nhp10 представляет собой HMG2-box белок, который, как известно, связывает искаженную ДНК [16, 17]. Nhp10–Ies3–Ies5 тесно связан с ДНК и нуклеосомами, что позволяет предположить, что этот модуль способствует высокоаффинному нацеливанию INO80. Однако модуль Nhp10 не оказывал влияния на активность АТФазы и способность перемещать нуклеосомы, что позволяет предположить, что он не является частью основной ремоделирующей активности [12]. Таким образом, эти данные позволяют предположить, что модули Arp8 и Arp5 участвуют в реакции ремоделирования, тогда как модуль Nhp10 способствует высокоаффинному распознаванию ДНК или нуклеосом [12].

Комплекс NuB4 взаимодействует с шапероном Asf1p, который контролирует сборку хроматина. Это взаимодействие эволюционно консервативно и наблюдается у множества организмов [18–21]. Hat1 — каталитическая субъединица комплекса NuB4, в который входят Hat1p, Hat2p и Hif1p [22–23]. Hat2p обладает активностью шаперона гистонов и, как полагают, облегчает взаимодействие различных комплексов NuB4 с гистонами [24–27].

Hif1p является членом семейства гистоновых шаперонов N1 и специфически связывается с гистонами H3 и H4. Hif1p может играть роль в укладке гистонов на ДНК, что позволяет предположить, что Hat1p может быть непосредственно вовлечен в процесс сборки хроматина [22]. Комплекс NuB4 физически взаимодействует с Hsm3p, который мы ранее идентифицировали как функциональную субъединицу комплекса NuB4 [5, 28–30]. Ранее было показано, что инактивация субъединиц Hif1 и Hsm3 комплекса NuB4 приводит к снижению экспрессии гена *RNR3*, кодирующего субъединицу рибонуклеотидредуктазного комплекса RNR [5]. В настоящем исследовании мы сосредоточились на процессах, которые происходят во время репаративной сборки хроматина в дрожжевых клетках. Наши результаты впервые показали, что мутации в генах, контролирующих ремодуляторы хроматина, могут проявлять сильное взаимодействие с мутациями в генах гистоновых шаперонов.

МАТЕРИАЛЫ И МЕТОДЫ

Штаммы

Делеции генов получали путем ПЦР-опосредованного разрушения генов с использованием специфических маркеров [31]. Все штаммы, используемые в работе, происходят из коллекции группы генетических исследований хроматина и репарации Лаборатории экспериментальной генетики или были получены в ходе исследования. Генотипы штаммов дрожжей, использованных в данном исследовании, представлены в табл. 1.

Получение штаммов

Для разрушения открытой рамки считывания гена *IES5* (*YER092W*) клетки штамма дрожжей дикого типа (LMG-3031) трансформировали фрагментом ДНК, содержащим маркер *NatMX*, с фланкирующими нуклеотидными последовательностями, гомологичными флангам гена *IES5*. Фрагмент генерировали путем ПЦР-амплификации с праймерами IES5-1 (5'-TGACTGCGATGACTCTAGCG-3') и IES5-2 (5'-CGGTTTAAGCCACATGTTAGG-3') и ДНК, выделенной из штамма KFY-827 в качестве матрицы. Трансформанты с делецией гена *IES5* отбирали по их устойчивости к кланату с использованием полной среды, содержащей нурсеотрицин в концентрации 20 мкг/мл. Разрушение гена *IES5* было подтверждено с помощью ПЦР.

Для разрушения открытой рамки считывания гена *IES6* (*YEL044W*) клетки штамма дрожжей дикого типа (LMG-3031) трансформировали фрагментом ДНК, содержащим маркер *KanMX*, с фланкирующими нуклеотидными последовательностями,

Таблица 1. Штаммы дрожжей, использованные в работе

Штамм	Генотип	Источник
LMG-3031	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1</i>	Наша коллекция
TAE-200	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 ies5Δ</i>	Получены в ходе исследования
TAE-201	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 nhp10Δ</i>	»
TAE-203	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 ies5Δ hsm3Δ</i>	»
TAE-204	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 ies5Δ hif1Δ</i>	»
TAE-205	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 nhp10Δ hsm3Δ</i>	»
TAE-206	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 nhp10Δ rad2Δ</i>	»
LMG-352	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 ies5Δ rad2Δ</i>	»
EAA-1	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 rpd3Δ</i>	Наша коллекция
EAA-2	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 ies5Δ rpd3Δ</i>	Получены в ходе исследования
TAE-202	<i>MATα ade2Δ-248 ura3-160, 188 leu2-3, 112 trp1 ies6Δ</i>	»
TAE-207	<i>MATα leu2-3, 112 ade2-192 ies6Δ</i>	»
TAE-208	<i>MATα leu2-3, 112 ade2-192 ies6Δ/MATα leu2-3, 112 ade2-192 ies6Δ</i>	»
TAE-209	<i>MATα leu2-3, 112 ade2-192/MATα leu2-3, 112 ade2-192</i>	»

гомологичными флангам гена *IES6*. Фрагмент генерировали путем ПЦР-амплификации с праймерами *IES6-1* (5'-GGCTCTCTGTGCGAAGGAAG-3') и *IES6-2* (5'-GAGACTGGCATCCATGCATG-3') и ДНК, выделенной из штамма KFY-1437 в качестве матрицы. Трансформанты с делецией гена *IES6* отбирали по их устойчивости к генетицину с использованием полной среды, содержащей генетицин в концентрации 200 мкг/мл. Разрушение гена *IES6* было подтверждено с помощью ПЦР.

Для разрушения открытой рамки считывания гена *NHP10* (*YDL002C*) клетки штамма дрожжей дикого типа (LMG-3031) трансформировали фрагментом ДНК, содержащим маркер *NatMX*, с фланкирующими нуклеотидными последовательностями, гомологичными флангам гена *NHP10*. Фрагмент генерировали путем ПЦР-амплификации с праймерами *NHP10-1* (5'-GCTAAGAAATCTTGTGCACGA-3') и *NHP10-2* (5'-TAGGAATCGTCTGCCCCACGT-3') и ДНК, выделенной из штамма KFY-1803 в качестве матрицы. Трансформанты с делецией гена *NHP10* отбирали по их устойчивости к кланату с использованием полной среды, содержащей нурсеотрицин в концентрации 20 мкг/мл. Разрушение гена *NHP10* было подтверждено с помощью ПЦР.

Питательные среды

Состав минимальной среды, использованной в качестве селективной, а также среды D для выращивания культур и учета выживаемости приведены

в [31]. Специальную среду с добавлением 96%-ного этанола (15 мл/л) и 40%-ной глюкозы (4 мл/л), включающую рост мутантов с дыхательной недостаточностью, использовали для оценки индуцированного мутагенеза в локусах *ADE4–ADE8* [32]. Для определения частоты спонтанных мутаций устойчивости к канаванину методом упорядоченного посева использовали среду минимального состава с добавлением аминокислот и азотистых оснований, необходимых для роста тестируемых штаммов, и различных концентраций канаванина. Во флукуационном методе медиан использовали среду того же состава, но обогащенную полным набором аминокислот (за исключением аргинина) и азотистых оснований и содержащую более высокие концентрации канаванина [33].

Чувствительность к УФ-излучению

Тесты на выживаемость клеток проводили путем выращивания в течение ночи культуры соответствующего штамма в жидкой YPD при 30 °C. Клетки промывали и ресуспендировали в воде с плотностью 1×10^7 кл./мл. Суспензию клеток облучали УФ-лампой БУВ-30 (диапазон УФ-С) с мощностью дозы на верхнем уровне 1.4 Дж/м²с (для штаммов дикого типа, *ies5Δ*, *nhp10Δ*, *ies6Δ*, *rp3Δ*, *ies5Δ rp3Δ*, *ies5Δ him1Δ*, *ies5Δ hsm3Δ*, *ies5Δ hif1Δ*, *nhp10Δ him1Δ*, *nhp10Δ hsm3Δ* и *nhp10Δ hif1Δ*) и на нижнем уровне – 0.21 Дж/м²с (для штаммов *rad2Δ*, *rad2Δ ies5Δ* и *rad2Δ nhp10Δ*). Аликвоты отбирали после различных доз облучения, разбавляли и помещали на

чашки с YPD для определения количества выживших [5].

Мутационные тесты

Мутационные тесты проводили путем выращивания в течение ночи культуры соответствующего штамма в жидкой YPD при 30 °С. Клетки промывали и ресуспендировали в воде с плотностью 1×10^7 кл./мл. Клетки облучали УФ-лампой БУВ-30. Аликвоты отбирали после различных доз облучения, разбавляли и помещали на чашки с YPD для определения количества выживших. Для определения частоты мутаций неразведенные аликвоты высевали на среду YPD со спиртом вместо глюкозы, состав которой был описан ранее [32].

Спонтанный мутагенез определяли с использованием стандартного метода медиан (флуктуационный тест), который регистрирует в основном ошибки репликации [34]. После трехдневной инкубации отбирали 12 отдельных колоний, каждую из которых суспендировали в 1 мл воды и высевали на селективную среду с канавагином в концентрации, исключающей возможность роста чувствительных к канавагину клеток. При оценке количества высеянных клеток мы разбавляли суспензии и высевали их на полную среду. После инкубации в течение трех-четырех дней подсчитывали количество колоний, устойчивых к канавагину, и общее количество колоний, чувствительных к канавагину, на чашке Петри. Частоту возникновения спонтанных мутаций оценивали по специальной формуле [34].

В других экспериментах скорость спонтанных мутаций устойчивости к канавагину оценивали по методу упорядоченного посева Хромова-Борисова [35]. Этот метод позволяет измерять скорость репаративного мутагенеза. В этих экспериментах тестируемые дрожжевые культуры выращивали на чашках с полной средой в течение суток. Затем готовили 5 мл суспензии (1×10^6 кл./мл). Специальный репликатор на 150 штырей был погружен в эту суспензию и перенесен на чашку Петри со средой, содержащей канавагин. Репликатор переносил 150 равных капель дрожжевой суспензии (около 2 мкл каждая) на равных расстояниях друг от друга, каждая капля содержала приблизительно 2000 клеток. Концентрация канавагина для оценки скорости мутаций устойчивости к антибиотику была определена в специальных предварительных экспериментах для всех штаммов и составила 50 мг/л. Мутанты имеют более быстрый рост, который проявляется в виде “бородавок” на пятнах с ограниченным ростом исследуемой культуры. После 14–15 дней инкубации подсчитывали бородавки мутантов, устойчивых к канавагину, и общее количество клеток. Последнее было сделано после смыва клеток с ряда реплик капель, лишенных бородавок. Частоты

мутаций на одно деление клеток определяли путем деления числа бородавок на общее количество клеток на чашке. Результаты представлены как средние значения 3–5 независимых экспериментов с 95%-ным доверительным интервалом.

ПЦР в реальном времени

Для проведения ПЦР в реальном времени была использована система детекции RT-PCR CFX96 (Bio-Rad, Великобритания). Реакции проводили в 25 мкл объемах, состоящих из 10 мкл 2.5-кратной реакционной смеси для ОТ-ПЦР в присутствии красителя SYBR Green I и эталонного красителя Rox (Syntol, Россия), 13.8 мкл воды, 1 мкл кДНК и 0.1 (2 мМ) соответствующих праймеров (праймеры для гена *RNR3*: For*RNR3* 5'-ACACCTTTCATGGTTTATAAG-3' и Rev*RNR3* 5'-CGACGATTTCACAACATAA-3'; для гена *ACT1*: For*ACT1* 5'-GAAGGTCAAGATCATTGC-3' и Rev*ACT1* 5'-GTTGGAAGGTAGTCAAAG-3').

Условия амплификации ПЦР были следующими: 1 цикл по 5 мин при 95 °С, затем 39 циклов по 15 с при 95 °С и 20 с при 52 °С. Анализ кривой плавления показал 5-секундное увеличение на 1 °С с 55 до 95 °С.

Контрольные реакции с праймером и без матрицы реакционных смесей были включены. Два биологических и три технических повтора были выполнены для каждого образца. Результаты были обработаны с использованием программы CFX Manager.

Статистический анализ

Экспериментальные данные представлены как стандартная ошибка среднего (SEM) по крайней мере от трех повторов, а статистические различия определялись с помощью *t*-критерия Стьюдента. Значимость определяли на уровне $p < 0.05$ [5].

РЕЗУЛЬТАТЫ

Сборка хроматина накладывает ограничения на многие клеточные процессы, требующие доступа к хромосомной ДНК. Гистоновые шапероны играют ключевую роль в динамике хроматина, контролируя процессы его разборки и сборки [36–37]. В отличие от репликативной сборки хроматина процесс репарационной сборки менее изучен. Это связано с тем, что еще не все участники выявлены и не все связи между этапами этого процесса установлены. Показано, что мутанты по гену *INO80* влияют на процессы репарации у дрожжей [9]. Мутации некоторых компонентов комплекса INO80 проявляют гиперчувствительность к генотоксическим агентам [8, 38, 39].

Ранее мы показали, что анализ индуцированного мутагенеза у различных мутантов по репарации может служить тонким инструментом для изучения процесса репарационной сборки хроматина [5, 40, 41]. Как отмечалось ранее, *Ies5*, *Ies6* и *Nhp10* являются компонентами комплекса ремоделирования хроматина INO80. Роль генов *NHP10*, *IES5* и *IES6*, кодирующих субъединицы двух модулей комплекса INO80, в генетическом контроле мутационного процесса ранее не изучалась. Мы получили нулевые мутанты по этим генам и охарактеризовали их чувствительность к УФ-излучению, а также влияние мутаций в генах *NHP10*, *IES5* и *IES6* на частоту УФ-индуцированных и спонтанных мутаций.

IES6

Потеря каталитической субъединицы *Ies6* или *Ino80* приводит к быстрому увеличению плоидности. Одним из путей к полиплоидии является аномальная сегрегация хромосом из-за aberrантной структуры центромеры [42]. Поэтому мы попытались изучить эффекты мутации *ies6Δ*, когда мутация является гомозиготной в диплоидных клетках, несущих мутацию *ade2-192* также в гомозиготном состоянии. *ade2-192* представляет собой редко ревертирующую миссенс-мутацию. Частота спонтанных реверсий у диплоидного штамма дикого типа, гомозиготного по мутации *ade2-192*, составила $2.9 \pm 0.5 \times 10^{-8}$. У диплоида, гомозиготного по мутациям *ies6Δ* и *ade2-192*, эта частота составила $1.4 \pm 0.4 \times 10^{-8}$. Таким образом, мутация *ies6Δ* снижает частоту спонтанного мутагенеза примерно в 2 раза. Для дальнейшего тестирования выживаемости мутанта *ies6Δ ade2-192* мы провели эксперименты с УФ-светом. Было обнаружено, что этот мутант проявлял такую же чувствительность к УФ, как и штамм дикого типа (рис. 1).

Далее мы проверили, как мутация гена *IES6* влияет на УФ-индуцированный мутагенез. Как показано на рис. 1, мутант *ies6Δ ade2-192* продемонстрировал уровень УФ-индуцированного мутагенеза, свойственный штамму дикого типа. Таким образом, мутация *ies6Δ* не оказывает существенного влияния на выживаемость и мутагенез дрожжевых клеток после УФ-облучения.

IES5 и *NHP10*

В настоящей работе проведено сравнительное исследование генетических свойств мутаций *ies5Δ* и *nhp10Δ* и их взаимодействия с делеционными мутациями в генах *HSM3* и *HIF1*, которые кодируют субъединицы комплекса NuB4 и принимают участие в репаративной сборке хроматина. Сначала мы протестировали чувствительность к УФ-излучению штамма дикого типа и одиночных мутантов *ies5Δ* и *nhp10Δ*. Как видно из рис. 2, *a*, одиночный мутант

продemonстрировал чувствительность к УФ, сравнимую с чувствительностью штамма дикого типа. Мутация *ies5Δ* в сочетании с *rad2Δ* не приводила к изменению устойчивости к УФ-излучению клеток двойного мутанта *ies5Δ rad2Δ* по сравнению с клетками одиночного мутанта *rad2Δ* (рис. 3). На основании полученных данных можно сделать вывод, что мутации в гене *IES5*, кодирующей субъединицу модуля Nhp10 комплекса INO80, не изменяют устойчивости клеток к УФ-излучению и не влияют на НЭР. В отличие от *ies5Δ*, мутант *nhp10Δ* показал незначительное увеличение чувствительности к УФ-излучению (рис. 2, *б*). Двойной мутант *nhp10Δ rad2Δ* значительно отличался от одиночного мутанта *rad2Δ* (рис. 3). Таким образом, мутанты по генам, кодирующим субъединицы одного модуля

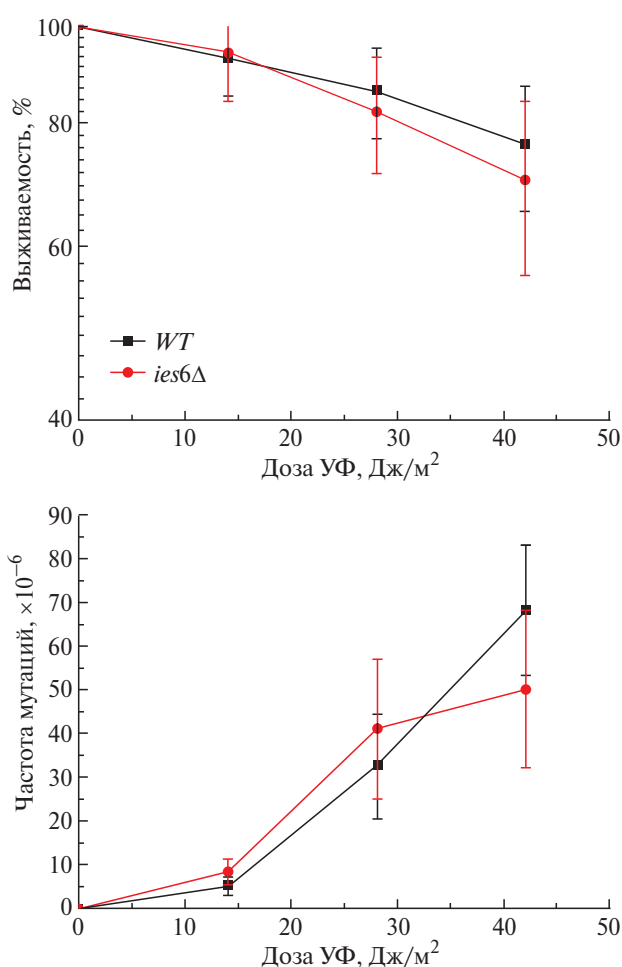


Рис. 1. Чувствительность к УФ-свету и частота УФ-индуцированного мутагенеза в локусе *CAN1* у штамма дикого типа и мутантного диплоидного штамма *ies6Δ* при обработке различными дозами ультрафиолетового излучения. На графиках показаны стандартные ошибки среднего значения ($\pm$ SEM), полученные в результате пяти независимых экспериментов.

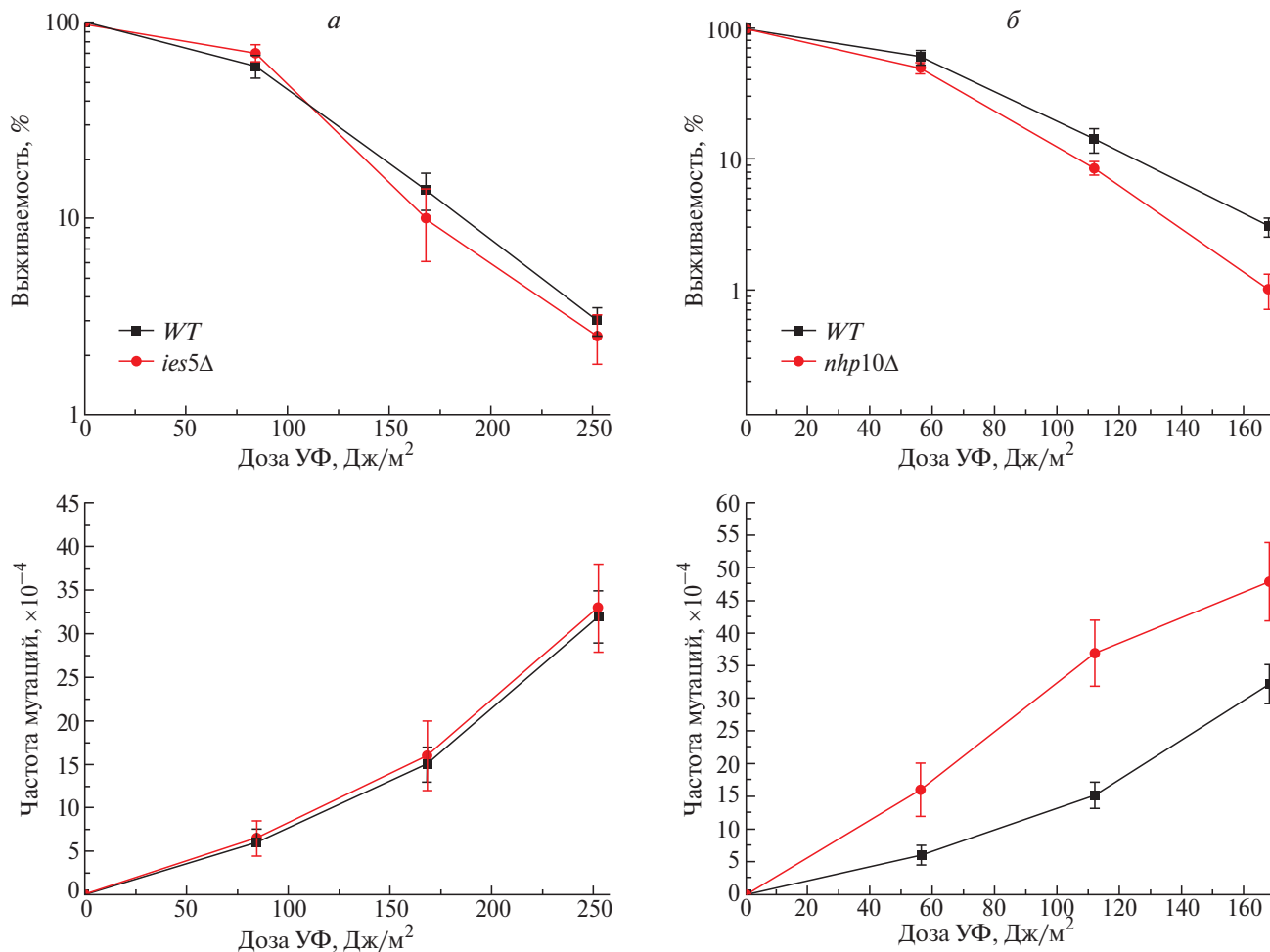


Рис. 2. Чувствительность к УФ-свету и частота УФ-индуцированного мутагенеза в локусах *ADE4–ADE8* у штамма дикого типа и мутантного штамма *ies5Δ* (а) и мутантного штамма *nhp10Δ* (б) при обработке различными дозами ультрафиолетового излучения.

комплекса INO80, по-разному влияют на УФ-чувствительность дрожжевых клеток.

Далее была исследована возможная роль *Ies5* и *Nhp10* в мутационном процессе. Мы решили сравнить спонтанный мутагенез у мутантов *ies5Δ* и *nhp10Δ* со штаммом дикого типа. При изучении частоты возникновения спонтанных мутаций использовали два разных метода: флуктуационный тест (метод “медианы Коулсона – Ли”), который в основном позволяет определить уровень спонтанного репликативного мутагенеза, и метод упорядоченного посева, по результатам которого можно судить об уровне репаративного мутагенеза [5]. В обоих случаях была рассчитана частота мутаций гена *CAN1*, приводящих к устойчивости к канаваину. Из табл. 2 видно, что наличие в клетках мутации *ies5Δ* приводит к снижению скорости спонтанного репликативного мутагенеза. Поскольку метод упорядоченного посева одновременно выявляет репликативный и репаративный спонтанный

мутагенез, вклад репаративного мутагенеза в общую скорость спонтанного мутагенеза у мутанта *ies5Δ* составит около 100%. Учитывая это обстоятельство, можно заключить, что мутация *ies5Δ* увеличивает скорость репаративного мутагенеза примерно в два раза. Мутация *nhp10Δ* не изменяла скорость репликативного спонтанного мутагенеза, но увеличивала репаративный мутагенез в 6–7 раз (табл. 2).

Для оценки влияния мутаций *ies5Δ* и *nhp10Δ* на УФ-индуцированный мутагенез мы измерили частоту прямых мутаций в локусах *ADE4–ADE8*, индуцированных УФ-лучами у штамма дикого типа и одиночных мутантов *ies5Δ* и *nhp10Δ*. Данные, представленные на рис. 2,б, позволяют предположить, что мутация *ies5Δ* не влияет на частоту УФ-индуцированного мутагенеза. Частота УФ-индуцированного мутагенеза у двойного мутанта *ies5Δ rad2Δ* практически не отличалась от одиночного *rad2Δ* (рис. 3). В отличие от *ies5Δ* мутант *nhp10Δ* усиливал

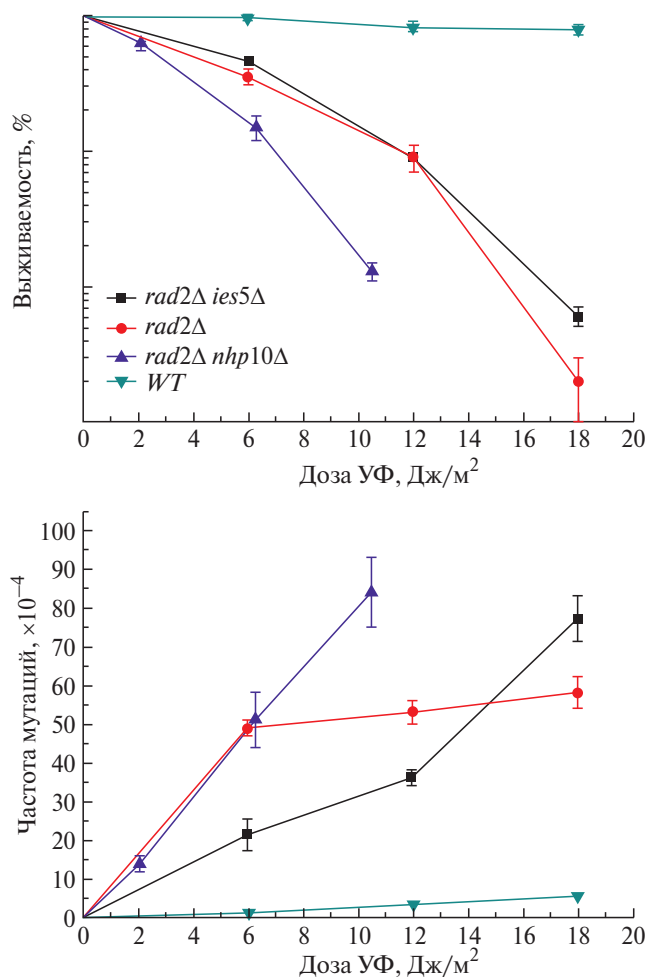


Рис. 3. Чувствительность к УФ-свету и частота УФ-индуцированного мутагенеза в локусах *ADE4–ADE8* у штамма дикого типа и мутантных штаммов *rad2Δ*, *rad2Δ ies5Δ* и *rad2Δ nhp10Δ* при обработке различными дозами ультрафиолетового излучения.

УФ-индуцированный мутагенез по сравнению со штаммом дикого типа (рис. 2,б). Двойной мутант *nhp10Δ rad2Δ* проявлял гиперчувствительность по сравнению с одиночным мутантом. Это взаимодействие было аддитивным, что указывает на разницу в механизмах чувствительности к УФ-излучению, опосредованную мутациями *nhp10Δ* и *rad2Δ* (рис. 3). Однако при равной выживаемости частота

мутаций у мутанта *nhp10Δ* и штамма дикого типа была одинаковой. Такая же картина наблюдалась для мутантов *rad2Δ* и *nhp10Δ rad2Δ*.

Ранее было показано, что комплекс NuB4 может принимать непосредственное участие в процессе сборки хроматина [5, 22]. На основании полученных данных мы предположили, что инактивация субъединиц Hif1 и Hsm3 комплекса NuB4 приводит к неправильной сборке нуклеосом, что влияет на активацию киназы Rad53 [5]. Модуль Nhp10 способствует высокоаффинному связыванию с ДНК или нуклеосомой [12]. Исходя из нашего предположения, если между белками Hsm3 и Hif1 существует генетическое взаимодействие с белками Ies5 и Nhp10, то можно ожидать изменения фенотипических проявлений их мутаций в двойных мутантах. С помощью этого анализа мы стремились подтвердить, что мутации *hsm3Δ* и *hif1Δ* действительно нарушают правильную сборку нуклеосом.

Одиночные мутанты *hif1Δ* и *hsm3Δ* показали такую же чувствительность к УФ, как и штамм дикого типа [5], а также одиночный *ies5Δ*. Однако двойной мутант *hsm3Δ ies5Δ* показал гиперчувствительность к УФ-облучению (рис. 4,а). Точно такую же природу взаимодействия показали мутации *hif1Δ* и *ies5Δ* (рис. 4,а). В отличие от мутанта *ies5Δ*, одиночный *nhp10Δ* проявлял гиперчувствительность к УФ-облучению, и оба двойных *hif1Δ nhp10Δ* и *hsm3Δ nhp10Δ* ничем не отличались от одиночного *nhp10Δ* (рис. 4,б). Таким образом, дефекты структуры нуклеосомы, вызванные мутациями *hif1Δ* и *hsm3Δ*, не приводят к дальнейшему увеличению УФ-чувствительности мутанта *nhp10Δ*.

Дальнейший генетический анализ изученных мутантов был проведен для понимания взаимодействия соответствующих белковых продуктов в УФ-индуцированном мутагенезе. Характерной особенностью мутантов *hif1Δ* и *hsm3Δ* является высокий УФ-индуцированный мутагенез [5], тогда как *ies5Δ* по этому показателю не отличается от штамма дикого типа. Как видно из рис. 4,а, мутация *ies5Δ* полностью подавляет характерный мутагенез *hif1Δ* и *hsm3Δ*, частота УФ-индуцированных мутаций у этих мутантов даже ниже по сравнению со штаммом дикого типа. Мутация *nhp10Δ* снижала характерный для *hif1Δ* и *hsm3Δ* мутагенез до уровня штамма дикого типа, в то же время подавлялся

Таблица 2. Спонтанный мутагенез устойчивости к канаванину

Штаммы	Упорядоченный посев, $\times 10^{-7}$	Флуктуационный тест, $\times 10^{-7}$
WT	3.0 ± 0.24	3.0 ± 0.16
<i>ies5Δ</i>	4.4 ± 0.7	>1
<i>ies6Δ / ies6Δ</i>	2.7 ± 0.7	>1
<i>nhp10Δ</i>	19.8 ± 3.4	–

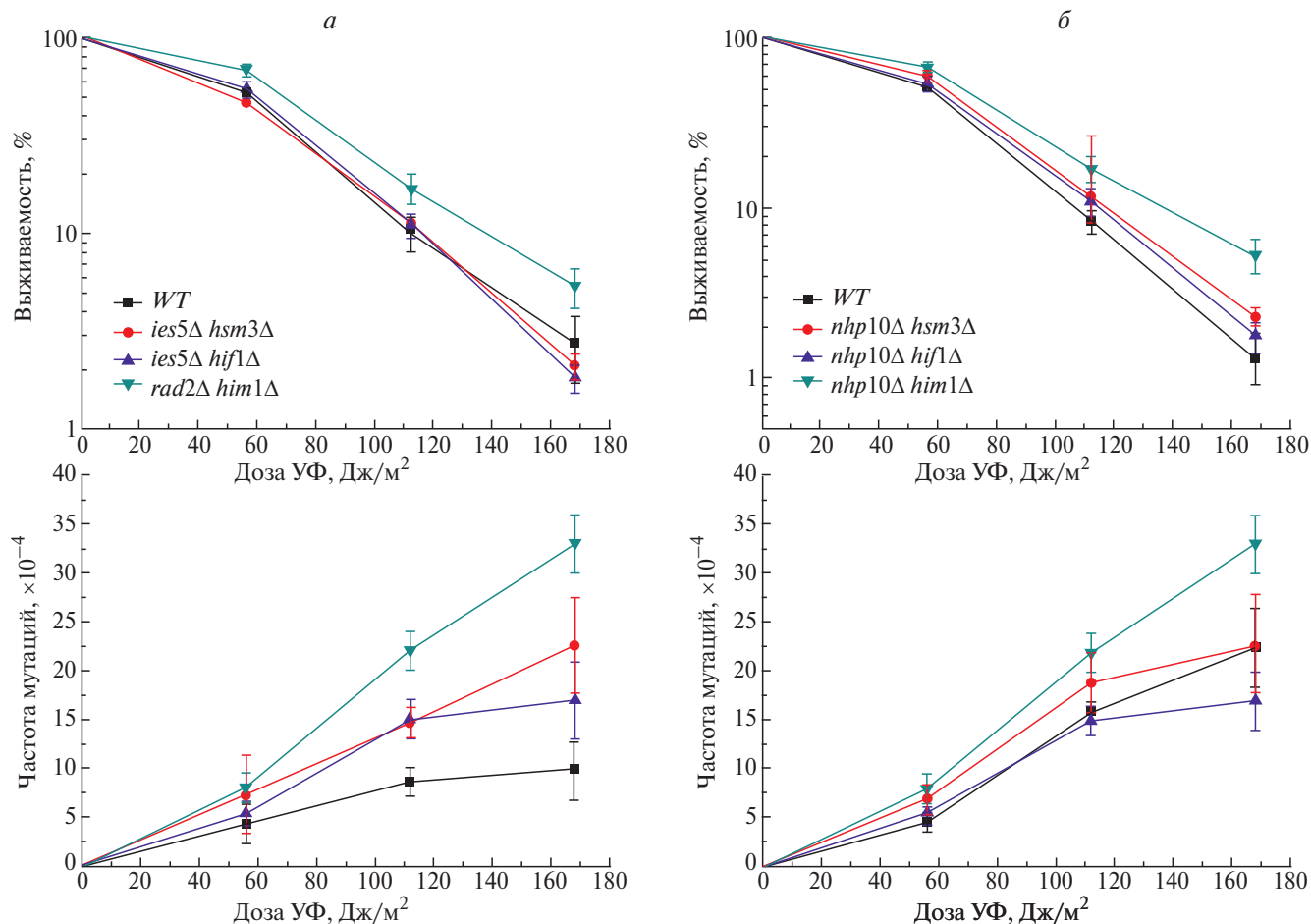


Рис. 4. Чувствительность к УФ-свету и частота УФ-индуцированного мутагенеза в локусах *ADE4–ADE8* у штамма дикого типа и мутантных штаммов *ies5Δ hsm3Δ*, *ies5Δ hif1Δ* и *ies5Δ him1Δ* (а) и *nhp10Δ hsm3Δ*, *nhp10Δ hif1Δ* и *nhp10Δ him1Δ* (б) при обработке различными дозами ультрафиолетового излучения.

nhp10-зависимый мутагенез (рис. 4, б). В совокупности эти результаты подтверждают нашу гипотезу о том, что мутации *hif1Δ* и *hsm3Δ* приводят к некорректной сборке нуклеосом и указывают на то, что ремоделирующий комплекс INO80 не может редактировать эти искажения структуры нуклеосом.

Для завершения сборки хроматина необходимо удалить ацетильные группы с N-концевых хвостов гистонов. Для этого клетка использует гистондеацетилазы (HDAC), одной из таких HDAC является комплекс RPD3. У *S. cerevisiae* этот комплекс представляет собой HDAC класса I и существует в виде комплексов Rpd3L или Rpd3S, оба из которых имеют общие субъединицы Rpd3, Sin3 и Ume1 [43, 44], которые могут проявлять свойства гистоновой шаперона во время репаративной сборки хроматина [45]. Было решено изучить взаимодействие мутаций *ies5Δ* и *rpд3Δ*. Чтобы оценить эффект мутации в гене, кодирующем субъединицу Rpd3 комплекса RPD3, мы разрушили ген *IES5* в одиночном мутанте *rpд3Δ*. Как показано на рис.

5, оба одиночных мутанта не отличались по выживаемости от двойного мутанта *ies5Δ rpд3Δ*. Однако УФ-индуцированный мутагенез у двойного мутанта был подавлен и находился на том же уровне, что и у мутанта *ies5Δ him1Δ* (рис. 4, 5). Эти результаты позволяют предположить, что отсутствие деацетилирования вновь собранных нуклеосом препятствует нормальному функционированию комплекса INO80.

Him1 и комплекс INO80

Одиночный мутант *him1Δ* показал большую устойчивость к УФ-излучению по сравнению со штаммом дикого типа [46]. На основании этих данных мы исследовали потенциальные эпистатические связи между мутациями генов *NHP10*, *IES5* и *HIM1*. Мутация *nhp10Δ* подавляет *him1*-зависимую устойчивость к УФ-излучению до уровня одиночного мутанта *nhp10Δ* (рис. 4). Как и в случае с мутантами *hsm3Δ* и *hif1Δ*, мутация *him1Δ* приводила

к гиперчувствительности двойного мутанта *him1Δ ies5Δ* (рис. 4,а).

УФ-индуцированный мутагенез у мутанта *him1Δ* высок, как и у мутантов *hsm3Δ* и *hif1Δ* [45]. Мы создали двойные мутанты *ies5Δ him1Δ* и *nhp10Δ him1Δ* и изучили УФ-индуцированный мутагенез у этих двойных мутантов. Характер взаимодействия мутации *him1Δ* с мутациями *ies5Δ* и *nhp10Δ* при УФ-индуцированном мутагенезе был аналогичен мутациям *hsm3Δ* и *hif1Δ* (рис. 4,б). Таким образом, на основании полученных данных можно заключить, что мутация *him1Δ*, как и мутации *hsm3Δ* и *hif1Δ*, приводит к некорректной сборке нуклеосом.

Экспрессия *RNR3* у мутантов *ies5Δ* и *nhp10Δ*

Ранее мы показали, что причиной усиленно-го УФ-мутагенеза мутантов *hsm3Δ* и *hif1Δ* является промежуточная УФ-индуцированная активация киназы Rad53 [5]. Уровень активации киназы

Rad53 можно определить при помощи уровня экспрессии гена *RNR3*, продукт которого является субъединицей рибонуклеотидредуктазного комплекса RNR, активность которого непосредственно регулируется киназой Rad53 [41]. При самом высоком и самом низком уровнях активации киназы *hsm3Δ*- и *hif1Δ*-специфический мутагенез подавляется [5]. Поскольку мутации *ies5Δ* и *nhp10Δ* подавляют мутагенез, специфичный для *him1Δ*, *hsm3Δ* и *hif1Δ*, было протестировано влияние мутаций *ies5Δ* и *nhp10Δ* на уровень УФ-индуцированной экспрессии гена *RNR3*. Мы облучали УФ-светом мутантные клетки дикого типа, *ies5Δ* и *nhp10Δ* и через 4 часа измеряли уровни мРНК гена *RNR3* в облученных и необлученных клетках. Как видно из рис. 6, уровень мРНК в клетках дикого типа увеличился почти в 2 раза.

У одиночного мутанта *ies5Δ* уровень экспрессии гена *RNR3* был немного снижен по сравнению с уровнем штамма дикого типа (рис. 6,а), что не

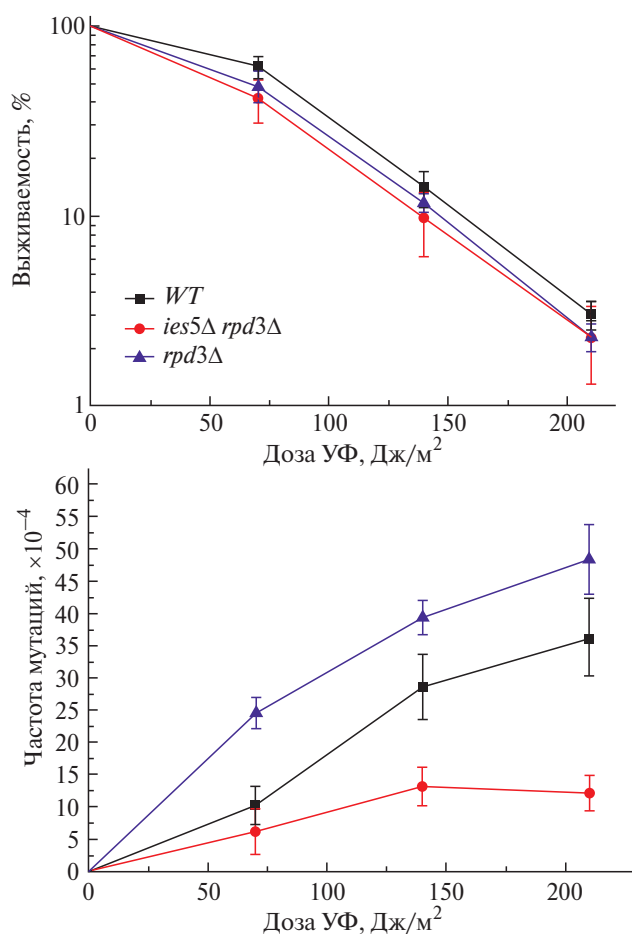


Рис. 5. Чувствительность к УФ-свету и частота УФ-индуцированного мутагенеза в локусах *ADE4–ADE8* штамма дикого типа и мутантных штаммов *ies5Δ rpd3Δ* и *rpd3Δ* при обработке различными дозами ультрафиолетового излучения.

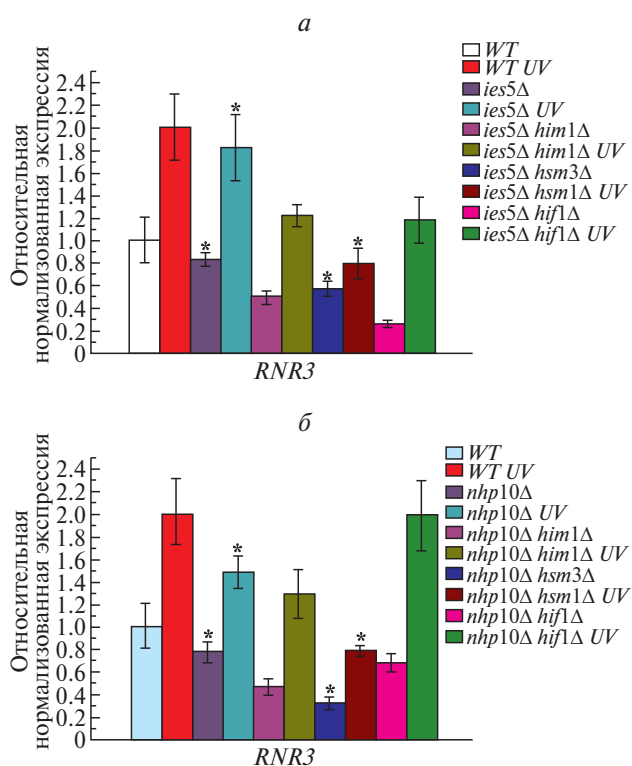


Рис. 6. Гистограммы показывают относительную нормализованную экспрессию гена *RNR3* в штамме дикого типа и мутантных штаммах: *ies5Δ*, *ies5Δ him1Δ*, *ies5Δ hsm3Δ* и *ies5Δ hif1Δ* (а); *nhp10Δ*, *nhp10Δ him1Δ*, *nhp10Δ hsm3Δ* и *nhp10Δ hif1Δ* (б) до и после облучения их ультрафиолетовым светом (после УФ-облучения клетки выдерживали в течение четырех часов при 30° С в термостате для индукции), доза УФ-излучения составляла 252 Дж/м²; * $p < 0.05$, t -критерий Стьюдента.

влияло на частоту УФ-индуцированного мутагенеза. С другой стороны, мутация *nhp10Δ* подавляет уровень экспрессии до промежуточного уровня, что объясняет усиление УФ-индуцированного мутагенеза у этого штамма. Возникает вопрос, почему мутации *ies5Δ* и *nhp10Δ* подавляют *him1*-, *hsm3*- и *hif1*-специфический мутагенез? Аналогичное подавление УФ-специфического мутагенеза мы наблюдали у двойных мутантов *hat1Δ hsm3Δ* и *hsm3Δ hif1Δ* [5]. Hat1 является ключевой субъединицей комплекса NuB4, и ее отсутствие полностью разрушает комплекс, поэтому у мутанта нарушается репаративная сборка нуклеосом. Именно это событие приводит к гиперактивации экспрессии гена *RNR3* у мутанта *hat1Δ*. У двойного мутанта *hsm3Δ hif1Δ* уровень УФ-индуцированной экспрессии гена *RNR3* подавлен до уровня клеток дикого типа без облучения. В этом случае подавление *RNR3* происходит за счет аддитивного эффекта обеих мутаций. В экспериментах с УФ-облучением двойные мутанты *ies5Δ him1Δ*, *ies5Δ hsm3Δ* и *ies5Δ hif1Δ* показали дальнейшее снижение уровней экспрессии *RNR3* по сравнению с одиночным мутантом *ies5Δ*. Согласно этому наблюдению, можно предположить, что у двойных мутантов *ies5Δ him1Δ*, *ies5Δ hsm3Δ* и *ies5Δ hif1Δ* наблюдается аддитивное взаимодействие мутаций. В экспериментах без УФ-облучения мутация *nhp10Δ* более эффективно снижает уровень экспрессии гена *RNR3* по сравнению с мутацией *ies5Δ*. Та же тенденция наблюдалась у двойных мутантов *nhp10Δ him1Δ* и *nhp10Δ hsm3Δ* (рис. 6,б). Однако мутант *nhp10Δ hif1Δ* имел более высокий уровень экспрессии *RNR3* по сравнению с мутантом *ies5Δ hif1Δ* как до, так и после УФ-облучения. При этом уровень экспрессии *RNR3* после облучения достиг уровня штамма дикого типа. Возможно, этот эффект связан, как и в случае с мутацией *hat1Δ*, с полной дисфункцией вновь собранных нуклеосом.

ОБСУЖДЕНИЕ

Было показано, что посттрансляционное ацетилирование N-концевых хвостов гистонов является важным механизмом, с помощью которого клетки регулируют доступность хроматина [3]. Процесс сборки репарационного хроматина имеет ряд отличий от процесса сборки при репликации и менее полно изучен, чем последний. В ряде недавних работ мы попытались расширить наши знания о механизме репарационной сборки хроматина [5, 40, 41]. Ассоциация гистоновой шапероны Hif1 с NuB4 стала первым прямым доказательством роли комплекса HAT1 в процессе сборки хроматина [22]. Показано, что Hat1, помимо Hif1 и Hat2, взаимодействует с шаперонами Asf1 и Hsm3 [5, 47]. Ранее мы показали, что инактивация обеих вспомогательных субъединиц NuB4 комплекса Hsm3 и

Hif1 частично подавляет УФ-индуцированную экспрессию генов комплекса RNR, тогда как инактивация обеих субъединиц приводит к ее полному подавлению [5]. Предполагается, что потеря вспомогательных субъединиц комплекса NuB4 при репарационной сборке хроматина приводит к дефектам нуклеосомы.

Эксперименты по взаимодействию мутаций *ies5Δ* и *rpд3Δ* показали, что отсутствие деацетилирования вновь встроенных нуклеосом препятствует нормальному функционированию комплекса INO80, что указывает на то, что ацетилированные нуклеосомы воспринимаются комплексом INO80 как поврежденные и подтверждают роль комплекса RPD3 в репаративной сборке хроматина. Nhp10–Ies3–Ies5 тесно связан с ДНК и нуклеосомы, что позволяет предположить, что этот модуль способствует высокоаффинному нацеливанию INO80 на ДНК. Мы предположили, что дефекты структуры нуклеосомы, возникающие в результате некорректной сборки хроматина, могут препятствовать эффективному связыванию комплекса INO80 с восстановленным хроматином. Для проверки этих гипотез были проведены эксперименты с двойными мутантами, включая мутации *ies5Δ*, *nhp10Δ*, с одной стороны, и мутации *hsm3Δ*, *hif1Δ*, *him1Δ*, с другой стороны. Мутация *nhp10Δ* вызывает гиперчувствительность к УФ-облучению, что указывает на нарушение связывания комплекса INO80 с восстановленным хроматином. У двойных мутантов дефекты структуры нуклеосом, вызванные мутациями *him1Δ*, *hsm3Δ* и *hif1Δ*, существенно не влияют на выживаемость мутанта *nhp10Δ* после УФ-облучения. Возможно, белок Nhp10 обеспечивает соединение комплекса INO80 с ДНК и не играет существенной роли в связывании этого комплекса с нуклеосомой. Одиночная мутация *ies5Δ* не вызывает гиперчувствительности к УФ-облучению, что указывает на нормальное связывание комплекса INO80 с репарированным хроматином. Однако у двойного мутанта дефекты структуры нуклеосомы, вызванные мутациями *him1Δ*, *hsm3Δ* и *hif1Δ*, приводят к гиперчувствительности клеток к УФ-облучению. Мы предполагаем, что белок Ies5, возможно, обеспечивает соединение комплекса INO80 с нуклеосомой и не играет существенной роли в связывании этого комплекса с ДНК. Наши данные о взаимодействии мутации *him1Δ* с мутациями *ies5Δ* и *nhp10Δ*, кодирующими субъединицы ремоделирующего комплекса INO80, подтверждают нашу гипотезу о том, что белок Him1 выполняет шаперонную функцию в процессе репарационной сборки хроматина.

Таким образом, мы впервые продемонстрировали, что мутации в генах, контролирующих ремодуляторы хроматина, могут проявлять сильные взаимодействия с мутациями в генах гистоновых шаперонов.

Ранее мы показали, что одиночные мутации в генах *ISW1*, *ISW2* и *CHD1* слабо влияют на репарационные и мутагенные процессы в дрожжевых клетках. В то же время существуют различия во влиянии этих мутаций на уровни спонтанных мутаций, что указывает на определенную специфичность действия белковых комплексов *ISW1*, *ISW2* и *CHD1* на экспрессию разных генов, контролируемых процессы репарации и мутации у дрожжей [48]. Функция модуля *Arg5* важна для стимулированного нуклеосомами гидролиза АТФ и ремоделирования хроматина [12], поэтому мы не наблюдали влияния инактивации гена *IES6* на индуцированный мутагенез, аналогичного мутациям в генах, кодирующих субъединицы других ремоделирующих комплексов *ISW1*, *ISW2* и *CHD1*. В то же время инактивация различных субъединиц *INO80* влияет на уровень спонтанного мутагенеза (табл. 2), как и мутации в генах, кодирующих упомянутые выше ремодуляторы.

Фосфорилированный *Rad9* взаимодействует с СООН-концевым доменом *Rad53*. Инактивация этого домена устраняет зависимое от повреждения ДНК фосфорилирование *Rad53*, остановку фазы клеточного цикла G2/M и увеличение транскрипции *RNR3* [49]. Ранее мы использовали аллель гена *RAD53* с поврежденной С-терминальной последовательностью. Мутация *rad53+HA-F* не увеличивала экспрессии *RNR3* по сравнению с клетками дикого типа в нормальных условиях роста и после УФ-облучения. В двойном мутанте *rad53+HA-F hsm3Δ* мутаторный фенотип мутанта *hsm3Δ* полностью подавлялся [41]. Таким образом, *Rad9* отвечает за максимальную гиперактивацию киназы *Rad53* при репарации повреждений ДНК. На основании этих данных мы предположили, что искажения структуры нуклеосомы, вызванные отсутствием вспомогательных субъединиц комплекса *NuB4*, ухудшают связывание белка *Rad9* с хроматином, что и приводит к супрессии *hsm3-*, *hif1-* и *him1-*специфического мутагенеза. Роль комплекса *INO80* в этом процессе, по-видимому, состоит в организации архитектуры хроматина, которая способствует эффективному связыванию *Rad9* с хроматином.

Действие факторов ремоделирования завершает репарацию хроматина. Наши результаты показывают, что комплекс *INO80* играет важную роль в этом процессе. Отсутствие белка *Nhp10*, входящего в состав модуля, отвечающего за связывание с ДНК, по-видимому, инактивирует основную функцию комплекса *INO80* и приводит к снижению эффективности систем репарации. Инактивация другой субъединицы этого модуля, *Ies5*, не приводит к заметным изменениям фенотипа клетки. Однако у двойных мутантов, несущих некорректно собранные нуклеосомы и мутации *nhp10Δ* или *ies5Δ*, мы наблюдали большие изменения в

функционировании пострепликативной репарации, выразившиеся в появлении гиперчувствительности к летальному действию УФ-облучения и подавлению мутаторного фенотипа.

Работа выполнена при финансовой поддержке “Курчатовского геномного центра – ПИЯФ” программой развития центров генетических исследований мирового уровня, соглашение № 075-15-2019-1663.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Luger K., Mader A.W., Richmond R.K. et al. Crystal structure of the nucleosome core particle at 2.8 Å resolution // *Nature*. 1997. V. 389. P. 251–260. <https://doi.org/10.1038/38444>
2. Ransom M., Dennehey B.K., Tyler J.K. Chaperoning histones during DNA replication and repair // *Cell*. 2010. V. 140, P. 183–195. <https://doi.org/10.1016/j.cell.2010.01.004>
3. Kouzarides T. Chromatin modifications and their function // *Cell*. 2007. V. 128. P. 693–705. <https://doi.org/10.1016/j.cell.2007.02.005>
4. Gaillard P.H., Martini E.M., Kaufman P.D. et al. Chromatin assembly coupled to DNA repair: A new role for chromatin assembly factor 1 // *Cell*. 1996. V. 86. P. 887–96. [https://doi.org/10.1016/s0092-8674\(00\)80164-6](https://doi.org/10.1016/s0092-8674(00)80164-6)
5. Evstyukhina T.A., Alekseeva E.A., Fedorov D.V. et al. Genetic analysis of the Hsm3 protein function in yeast *Saccharomyces cerevisiae* NuB4 complex // *Genes*. 2021. V. 12. <https://doi.org/10.3390/genes12071083>
6. Becker P.B., Horz W. ATP-dependent nucleosome remodeling // *Annu. Rev. Biochem.* 2002. V. 71. P. 247–273. <https://doi.org/10.1146/annurev.biochem.71.110601.135400>
7. Smith C.L., Peterson C.L. ATP-dependent chromatin remodeling // *Curr. Top. Dev. Biol.* 2005. V. 65. P. 115–148. [https://doi.org/10.1016/S0070-2153\(04\)65004-6](https://doi.org/10.1016/S0070-2153(04)65004-6)
8. Shen X., Mizuguchi G., Hamiche A., Wu C. A chromatin remodelling complex involved in transcription and DNA processing // *Nature*. 2000. V. 406. P. 541–544. <https://doi.org/10.1038/35020123>
9. Morrison A., Highland J., Krogan N. et al. INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair //

- Cell. 2004. V. 119. № 6. P. 767–775. <https://doi.org/10.1016/j.cell.2004.11.037>
10. Morrison A., Kim J.-A., Person M.D. et al. Mec1/Tel1 phosphorylation of the INO80 chromatin remodeling complex influences DNA damage checkpoint responses // Cell. 2007. V. 130. P. 499–511. <https://doi.org/10.1016/j.cell.2007.06.010>
 11. Van Attikum H., Fritsch O., Hohn B., Gasser S.M. Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand repair // Cell. 2004. V. 119. P. 777–788. <https://doi.org/10.1016/j.cell.2004.11.033>
 12. Tosi A., Haas C., Herzog F. et al. Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex // Cell. 2013. V. 154. P. 1207–1219. <https://doi.org/10.1016/j.cell.2013.08.016>
 13. Yao W., King D.A., Beckwith S.L. et al. The INO80 complex requires the Arp5-Ies6 subcomplex for chromatin remodeling and metabolic regulation // Mol. Cell Biol. 2016. V. 36. P. 979–991. <https://doi.org/10.1128/MCB.00801-15>
 14. Falbo K., Alabert C., Katou Y. et al. Involvement of a chromatin remodeling complex in damage tolerance during DNA replication // Nat. Struct. Mol. Biol. 2009. V. 16. P. 1167–1172. <https://doi.org/10.1038/nsmb.1686>
 15. Shen X., Choi E., Wu C. Involvement of actin-related proteins in ATP-dependent chromatin remodeling // Mol. Cell. 2003. V. 12. P. 147–155. [https://doi.org/10.1016/s1097-2765\(03\)00264-8](https://doi.org/10.1016/s1097-2765(03)00264-8)
 16. Ray S., Grove A. The yeast high mobility group protein HMO2 domains of the chromatin-remodeling complex INO80, binds DNA ends // Nucl. Acids Res. 2009. V. 37. P. 6389–6399. <https://doi.org/10.1093/nar/gkp695>
 17. Ray S., Grove A. Interaction of *Saccharomyces cerevisiae* HMO2 domains with distorted DNA // Biochemistry. 2012. V. 51. P. 1825–1835. <https://doi.org/10.1021/bi201700h>
 18. Fillingham J., Recht J., Silva A.C. et al. Chaperone control of the activity and specificity of the histone H3 acetyltransferase RTT109 // Mol. Cell Biol. 2008. V. 28. P. 4342–4353. <https://doi.org/10.1128/MCB.00182-08>
 19. Campos E.I., Fillingham J., Li J. et al. The program for processing newly synthesized histones H3.1 and H4 // Nat. Struct. Mol. Biol. 2010. V. 17. P. 1343–1351. PubMed ID: 20953179
 20. Barman H.K., Takami Y., Nashijima H. et al. Histone acetyltransferase-1 regulates integrity of cytosolic histone H3-H4 containing complex // Biophys. Res. Commun. 2008. V. 373. P. 624–630. <https://doi.org/10.1016/j.bbrc.2006.05.079>
 21. Jasencakova Z., Scharf A.N., Corpet A. et al. Replication stress interferes with histone recycling and predeposition marking of new histones // Mol. Cell. 2010. V. 37. P. 736–743. <https://doi.org/10.1016/j.molcel.2010.01.033>
 22. Ai X., Parthun M.R. The nuclear Hat1p/Hat2p complex: A molecular link between type B histone acetyltransferases and chromatin assembly // Mol. Cell. 2004. V. 14. P. 195–205. [https://doi.org/10.1016/s1097-2765\(04\)00184-4](https://doi.org/10.1016/s1097-2765(04)00184-4)
 23. Poveda A., Pamblanco M., Tafrov S. et al. Hif1 is a component of yeast histone acetyltransferase B, a complex mainly localized in the nucleus // J. Biol. Chem. 2004. V. 279. P. 16033–16043. <https://doi.org/10.1074/jbc.M314228200>
 24. Verreault A., Kaufman P.D., Kobayashi R., Stillman B. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase // Curr. Biol. 1998. V. 8. P. 96–108. [https://doi.org/10.1016/s0960-9822\(98\)70040-5](https://doi.org/10.1016/s0960-9822(98)70040-5)
 25. Saganuma T., Pattenden S.G., Workman J.L. Diverse functions of WD40 repeat proteins in histone recognition // Genes Dev. 2008. V. 22. P. 1265–1268. <https://doi.org/10.1101/gad.1676208>
 26. Song J.J., Garlick J.D., Kingston R.E. Structural basis of histone H4 recognition by p55 // Genes Dev. 2008. V. 22. P. 1313–1318. <https://doi.org/10.1101/gad.1653308>
 27. Furuyama T., Dalal Y., Henikoff S. Chaperone-mediated assembly of centromeric chromatin in vitro // Proc. Natl Acad. Sci. USA. 2006. V. 103. P. 6172–6177. <https://doi.org/10.1073/pnas.0601686103>
 28. Gavin A.C., M. Bösch M., Krause R. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes // Nature. 2002. V. 415. P. 141–147. <https://doi.org/10.1038/415141a>
 29. Blackwell J.S., Wilkinson S.T., Mosammaparast N., Pemberton L.F. Mutational analysis of H3 and H4 N termini reveals distinct roles in nuclear import // J. Biol. Chem. 2007. V. 282. P. 20142–20150. <https://doi.org/10.1074/jbc.M701989200>
 30. Ge Z., Wang H., Parthun M.R. Nuclear Hat1p complex (NuB4) components participate in DNA repair-linked chromatin reassembly // J. Biol. Chem. 2011. V. 286. P. 16790–16799. <https://doi.org/10.1074/jbc.M110.216846>
 31. Захаров И.А., Кожин С.А., Кожина Т.А. и др. Сборник методик по генетике дрожжей-сахаромицетов. Л.: Наука, 1984. изд. 2-е. 144 с.
 32. Ковальцова С.В., Королев В.Г. Штамм дрожжей *Saccharomyces cerevisiae* для тестирования мутагенов окружающей среды на основе взаимодействия мутаций *rad2* и *him1* // Генетика. 1996. Т. 32. № 3. С. 366–372.
 33. Fedorova I.V., Kovaltzova S.V., Gracheva L.M. et al. Requirement of *HSM3* gene for spontaneous mutagenesis in *Saccharomyces cerevisiae* // Mutat. Res. 2004. V. 554. P. 65–75.

34. Lea C.A., Coulson D.E. The distribution of the number of mutants in bacterial populations // *J. Genet.* 1949. V. 49. P. 264–285. <https://doi.org/10.1007/BF02986080>
35. Khromov-Borisov N.N., Saffi J., Henriques J.A.P. Perfect order plating: principal and applications // *TTO1*. 2002.TO2638
36. De Koning L., Corpet A., Haber J.E., Almouzni G. Histone chaperones: An escort network regulating histone traffic // *Nat. Struct. Mol. Biol.* 2007. V. 14. P. 997–1007. <https://doi.org/10.1038/nsmb1318>
37. Burgess R.J., Zhang Z. Histone chaperones in nucleosome assembly and human disease // *Nat. Struct. Mol. Biol.* 2013. V. 20. P. 14–22. <https://doi.org/10.1038/nsmb2461>
38. Chai B., Huang J., Cairns B., Laurent B. Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair // *Genes Dev.* 2005. V. 19. P. 1656–1661. <https://doi.org/10.1101/gad.1273105>
39. Shim E., Ma J., Oum J. et al. The yeast chromatin remodeler RSC complex facilitates end joining repair of DNA double-strand breaks // *Mol. Cell. Biol.* 2005. V. 25. P. 3934–3944. <https://doi.org/10.1128/MCB.25.10.3934-3944.2005>
40. Alekseeva E.A., Evstyukhina T.A., Peshekhonov V.T., Korolev V.G. Participation of the *HIM1* gene of yeast *Saccharomyces cerevisiae* in the error-free branch of post-replicative repair and role Pol η in *him1*-dependent mutagenesis // *Curr. Genet.* 2021.V. 67. P. 141–151. <https://doi.org/10.1007/s00294-020-01115-6>
41. Evstyukhina T.A., Alekseeva E.A., Peshekhonov V.T. et al. The role of chromatin assembly factors in induced mutagenesis at low levels of DNA damage // *Genes*. 2023. V. 14. <https://doi.org/10.3390/genes14061242>
42. Chambers A.L., Ormerod G., Durley S.C. et al. The INO80 chromatin remodeling complex prevents polyploidy and maintains normal chromatin structure at centromeres // *Genes Dev.* 2012. V. 26. P. 2590–2603. <https://doi.org/10.1101/gad.199976.112>
43. Carrozza M.J., Li B., Florens L. et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription // *Cell*. 2005. V. 123. P. 581–592. <https://doi.org/10.1016/j.cell.2005.10.023>
44. Keogh M.C., Kurdastani S.K., Morris S.A. et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex // *Cell*. 2005. V. 123. P. 593–605. <https://doi.org/10.1016/j.cell.2005.10.025>
45. Chen X.-F., Kuryan B., Kitada T. et al. The Rpd3 core complex is a chromatin stabilization module // *Curr. Biol.* 2012. V. 22. P. 56–63. <https://doi.org/10.1016/j.cub.2011.11.042>
46. Kelberg E.P., Kovaltsova S.V., Alekseev S.Yu. et al. *HIM1*, a new yeast *Saccharomyces cerevisiae* gene playing a role in control of spontaneous and induction mutagenesis // *Mutat. Res.* 2005. V. 578. P. 64–78. <https://doi.org/10.1016/j.mrfmmm.2005.03.003>
47. Keck K.M., Pemberton L.F. Histone chaperon link histone nuclear import and chromatin assembly // *Biochim. Biophys. Acta*. 2012. V. 1819. P. 277–289. <https://doi.org/10.1016/j.bbagr.2011.09.007>
48. Евстюхина Т.А., Алексеева Е.А., Федоров Д.В. и др. Роль ремодулирующих комплексов CHD1 и ISWI в контроле спонтанного и УФ-индуцированного мутагенеза у дрожжей *Saccharomyces cerevisiae* // *Генетика*. 2017. Т. 53. № 2. С. 173–180.
49. Sun Z., Hsiao J., Fay D.S., Stern D.F. Rad53 FHA domain associated with phosphorylated Rad9 in the DNA damage checkpoint // *Science*. 1998. V. 281. P. 272–274. <https://doi.org/10.1126/science.281.5374.272>

The Role of Different Subunits of INO80 Remodeling Complex in Repair Chromatin Assembly in Yeast *Saccharomyces cerevisiae*

T. A. Evstyukhina^{1,2}, E. A. Alekseeva^{1,2,*}, I. I. Skobeleva¹, V. T. Peshekhonov^{1,2}, V. G. Korolev^{1,2}

¹Petersburg Nuclear Physics Institute Named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", 188300 Gatchina, Russia;

²Kurchatov Genome Center—Petersburg Nuclear Physics Institute, 188300 Gatchina, Russia
*e-mail: alekseeva_ea@pnpi.nrcki.ru

Reparative chromatin assembly is an important step in maintaining genome stability. The correct assembly of chromatin is provided by histone chaperones, whose dysfunction can lead to the development of various forms of cancer and a number of hereditary diseases in humans. The effect of remodeling factors completes chromatin repair. The yeast chromatin remodeling complex INO80 plays an important role in chromatin architecture. We used induced mutagenesis and real-time PCR to study the role of INO80 in chromatin repair assembly. In double mutants *ies5Δ hsm3Δ(hif1Δ)*, defects in the structure of nucleosomes caused by mutations *hsm3Δ* and *hif1Δ* lead to hypersensitivity of cells to UV radiation and the disappearance of *hsm3*- and *hif1*-specific mutagenesis. Double mutants carrying the *nhp10Δ* mutation and *hsm3Δ* or *hif1Δ* mutations were indistinguishable from a single mutant in terms of the lethal effect of UV irradiation, however, the high UV-induced mutagenesis characteristic of all mutations disappeared. Thus, we found that mutations in the genes controlling the subunits of the INO80 complex can exhibit strong interactions with mutations in histone chaperone genes. We have confirmed the hypothesis that the Him1 protein performs a chaperone function in the process of reparative chromatin assembly.

Keywords: INO80 complex, *S. cerevisiae*, *IES5* gene, *NHP10* gene, UV-induced mutagenesis.

ГЕНЕТИЧЕСКАЯ СТРУКТУРА ПОПУЛЯЦИЙ ВОЛКА СЕВЕРНОЙ ЕВРАЗИИ: ОЦЕНКА ВЛИЯНИЯ ИСКЛЮЧЕНИЯ ИЗ АНАЛИЗА РОДСТВЕННЫХ ОСОБЕЙ

© 2024 г. П. А. Казимиров^{1, 2, *}, Ю. С. Белоконов¹, М. М. Белоконов¹, А. Я. Бондарев³,
А. В. Давыдов⁴, Е. С. Захаров⁵, С. В. Леонтьев⁶, Д. В. Политов^{1, 2, **}

¹Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

²Всероссийский научно-исследовательский институт охраны окружающей среды, Москва, 117628 Россия

³Алтайский государственный аграрный университет, Барнаул, 656049 Россия

⁴Федеральный научно-исследовательский центр развития охотничьего хозяйства, Москва, 105118 Россия

⁵Институт естественных наук Северо-Восточного федерального университета им. М.К. Аммосова,
Якутск, 677000 Россия

⁶ТОО “Национальный центр биотехнологии”, Астана, 010000 Казахстан

*e-mail: farenklaw@gmail.com

**e-mail: dmitri_p@inbox.ru

Поступила в редакцию 14.12.2023 г.

После доработки 12.2023 г.

Принята к публикации 12.2023 г.

По данным о полиморфизме 20 аутосомных микросателлитных локусов проведен анализ генетической структуры популяции и пространственной автокорреляции у волка, *Canis lupus* Linnaeus, 1758, Северной Евразии (территории Российской Федерации и Республики Казахстан). С помощью молекулярных маркеров выявлены скрытые генеалогические паттерны на различных дистанциях, наиболее сильно проявляющиеся на географических дистанциях до 150 км, но прослеживающиеся и на более широком пространственном масштабе (до 700–1600 км). Показано, что выявление и исключение родственных генотипов несущественно влияет на внутривидовые оценки генетического разнообразия. Однако данная процедура рекомендуется для более адекватной оценки популяционно-генетической структуры, позволяя оптимизировать ряд статистических процедур. Полученные данные, таким образом, демонстрируют интегрирующий эффект natalной миграции, действующей против дифференцирующего влияния филпатрии. Кроме того, мы продемонстрировали, что исключение подобных особей из выборок после анализа родства может приводить к снижению генетических дистанций между выборками.

Ключевые слова: волк, микросателлиты, генетическое родство, коэффициент родства, пространственная автокорреляция.

DOI: 10.31857/S0016675824070034 **EDN:** BIIKDR

Волк (*Canis lupus* Linnaeus, 1758) является крупным широко распространенным хищником, в целом характеризующимся выраженной территориальностью, но в то же время обладающим высокой способностью к расселению. Стая, состоящая в большинстве случаев из взрослой размножающейся пары и ее потомков, занимает участок, площадь которого может варьировать от нескольких десятков до нескольких тысяч квадратных километров. На размер участка влияет множество факторов: особенности ландшафта, доступность пищевых ресурсов и т. д. [1, 2]. В результате на определенной части ареала возникает семейная структура, что потенциально может приводить к инбридингу и негативно влиять на последующие поколения

вследствие инбредной депрессии. В то же время высокая мобильность большинства молодых особей обеспечивает перемешивание генетического материала и интеграцию видового генофонда в целом. Процессы расселения у волка представлены преимущественно natalной миграцией — молодыми потомками, покидающими родительскую территорию и составляющими основную массу нетерриториальных особей. При этом по данным GSM-трекинга мигрирующие волки могут преодолевать сотни километров [3]. В популяции, как правило, присутствуют нетерриториальные особи, способные перемещаться на еще большие расстояния [1, 2]. Таким образом, дифференцирующий эффект генетического дрейфа и локального

естественного отбора вследствие ограничения панмиксии семейной (демовой) структурой комбинируется с интегрирующим влиянием потока генов, и оба процесса уравнивают друг друга.

Генетические методы позволяют с высокой точностью реконструировать семейную структуру волка на конкретной территории. При этом чаще всего такие исследования родственных связей проводятся с использованием неинвазивных образцов [4–6]. В то же время при проведении популяционно-генетических исследований на более широкой географической шкале высокая территориальность вида может создать препятствия к интерпретации результатов, так как при отборе нескольких проб на территории одной стаи значительно повышается вероятность того, что пробы окажутся от близкородственных особей, которые впоследствии, по мере взросления, имеют все более низкие шансы остаться на этой территории вследствие конкуренции с родителями и другими членами стаи. Не менее сильно подобный эффект может быть выражен при использовании инвазивных проб, в частности материала, полученного от охотников, так как при охоте на волков часто добывают семейную группу целиком или большую ее часть (например, волчицу с волчатами). Чтобы более четко выявить географические тренды и снизить эффект демовой организации на пространственную структуру популяций, ряд исследователей исключает из анализа близкородственных особей (по одной особи из родственных пар или исключают из анализа отпрысков, оставляя родителей) [7, 8]. Однако, с другой стороны, подобная элиминация может привести к потере информации о мигрантах, так как образцы от близкородственных особей могут быть собраны на значительных расстояниях в случае, если часть волков из такой группы мигрировали.

Имеющиеся данные о волках Европы, юга азиатской части ареала и большей части Северной Америки относятся к популяциям с антропогенно нарушенной генетической структурой, находящимся в состоянии депрессии, восстанавливающимся или недавно восстановившимся после прохождения “бутылочного горлышка”. В Новом Свете широкомасштабные исследования на пространствах в миллионы квадратных километров возможны только в Канаде и на Аляске. В Северной Евразии обширные географически связанные территории с относительно ненарушенной популяционно-генетической структурой волка остались только в лесной зоне России, где несмотря на все усилия по регулированию численности хищника, обитают наиболее интактные популяционные группировки. Имеющаяся в нашем распоряжении коллекция образцов волка с территории России, дополненная выборками с территории Республики Казахстан, позволяет впервые исследовать вопрос о дальности миграции родственных особей

на материале, собранном с большей части ареала бореальной макропопуляции волка.

В настоящей работе мы поставили своей целью описать паттерны родственной структуры волка на широком спектре географических дистанций с использованием бипарентально наследуемых молекулярно-генетических маркеров (аутосомных микросателлитных локусов), а также определить степень влияния присутствующих в выборке близкородственных особей на результаты популяционно-генетического анализа. Обсуждаются также полученные данные о генетической дифференциации волка на исследованной территории.

МАТЕРИАЛЫ И МЕТОДЫ

Характеристика материала. В данном исследовании в качестве материала использовано 878 образцов от животных, добытых в 48 регионах России и Республики Казахстан. Данный набор включает 326 генотипов особей, вошедших в ранее опубликованную статью [9]. Высушенные или заспиртованные образцы шкур из большинства регионов РФ были собраны в рамках договора с ФГБУ “Федеральный научно-исследовательский центр развития охотничьего хозяйства” (П.М. Павлов, а также авторы статьи А.Я. Бондарев и А.В. Давыдов), сборы из Республики Саха (Якутия) осуществлены одним из авторов (Е.С. Захаров), с территории Республики Казахстан – сбор организован С.В. Леонтьевым [9]. Все особи на территории Российской Федерации добыты в рамках программ регулирования численности волка по официально выданным разрешениям, образцы биоматериала получены по запросам Министерства природных ресурсов и экологии РФ от региональных уполномоченных органов в области охоты и сохранения охотничьих ресурсов. На территории Казахстана исследуемые особи были добыты легально, по специальным разрешениям на изъятие видов животных, численность которых подлежит регулированию, выданным уполномоченным государственным органом. Для целей первичной обработки образцы были отнесены к географическим популяциям. Важно отметить, что если информация о тесных родственных связях уже была в сопроводительной информации к той или иной серии образцов, такие очевидные родственники исключались нами уже на этом этапе. Таким образом, мы исследовали скрытые генеалогические связи в матрице генотипов, представляющей огромную территорию – большую часть ареала бореальной макропопуляции волка.

Выделение ДНК и проведение полимеразной цепной реакции (ПЦР). Выделение ДНК проводили наборами “ДНК-Экстран-2” производства компании “Синтол” (Москва, Россия) в соответствии с протоколом производителя. Дополнительно после лизиса проводили осаждение жиров с

использованием тетрахлорметана — равный объем CCl_4 добавляли к лизату, встряхивали на вортексе 5–10 с, центрифугировали при 15000 об./мин 3 мин, после чего отбирали супернатант — водную фазу, содержащую лизат, — в чистые пробирки.

Генетический анализ изначально проводили с использованием 27 микросателлитных локусов ядерной локализации. Состав мультиплексных ПЦР-реакций (8 панелей) и режимы амплификации соответствовали протоколу, приведенному в [4]. Фрагментный анализ проводили на капиллярном секвенаторе с использованием коммерческого сервиса в компании “Синтол”. Расшифровку генотипов по хроматограммам проводили при помощи программы STRand v2.4.110 [10]. После формирования матрицы многолокусных генотипов из нее удалялись локусы и особи с более чем 15 и 10% отсутствующих данных соответственно.

Статистическая обработка матрицы генотипов. Оценка минимального количества локусов, необходимых для дифференциации генотипов, проводилась в пакете poppr [11, 12] для среды R [13] с числом итераций для каждого количества локусов 50000. Оценка полиморфности локусов (Polymorphism Information Content, PIC) проводилась в пакете polysat [14, 15], отклонение от равновесия Харди — Вайнберга в пакете regas [16] с использованием 100000 МСМС (Монте-Карло с марковскими цепями) итераций. Оценка частоты присутствия нуль-аллеля для каждой пары популяция — локус проводилась в R-пакете popgenreport [17, 18]. Значимость отличия индекса фиксации (F) от нуля рассчитывали через χ^2 -тест при помощи программы GenAlEx [19, 20] и базовых функций среды R.

Локусы, для которых наблюдалось одновременное значимое отклонение от равновесия Харди — Вайнберга и индекса фиксации от нуля и высокая (> 0.15) частота нуль-аллеля в более чем половине географических популяций, исключались из анализа.

Оценка индексов родства. Сравнение методов оценки родственных связей проводилось при помощи функции “comparestimator” пакета Related [21] с количеством симулированных пар для каждого уровня родства (неродственные; полусибсы; полные сибсы; родитель — потомок) — 1000. Функция симулирует заданное количество пар на основе частот аллелей, рассчитанных по таблице генотипов, после чего определяет корреляцию между индексом родства, полученным различными методами оценки, с ожидаемыми значениями индексов для каждой из выбранных степеней родства (0; 0.25; 0.5). Дальнейшую оценку индексов родства методом, выбранным на основе предыдущего шага, проводили функцией “coancestry” пакета Related.

Расчет координат сбора образцов. Так как точное место сбора (населенный пункт) было доступно

лишь для части образцов, в качестве координат сбора для образцов без точной географической локализации использовали координаты географического центра субъекта РФ, в котором была собрана проба, или района субъекта, если таковой был указан в сопроводительных документах. Для некоторых видов анализа в географические данные вносили случайный шум. Для этого при помощи базовых инструментов среды R к описанным выше координатам добавляли случайное значение вплоть до: 1 градуса широты/долготы если место сбора было известно до субъекта РФ; 0.7 градуса если место сбора было известно до района; 0.1 градуса если место сбора было известно до населенного пункта/точной географической точки.

Работа с датами сбора. Работу с данными о датах сбора для тех образцов, для которых эти данные были доступны, проводили с использованием пакета lubridate [22] для среды R. Даты сбора были стандартизированы в формате год — месяц. Для образцов с неизвестным месяцем сбора он был указан как январь. Расчет временных промежутков между сбором образцов проводили с точностью до года: 0 — добыты в один год, 1 — добыты с разницей в один год и т. д.

Корреляции между уровнем родства и географическим расстоянием и датами сбора. Расчет корреляций между переменными проводился в среде R (пакет rstatix [23]) и в программе STATISTICA [24]. Корреляции рассчитывали между следующими парами переменных:

индекс родства — расстояние между местами добычи в километрах;

индекс родства — разница во времени добычи в годах.

Расчет моделей двухступенчатой пороговой регрессии (threshold regression) проводили в пакете chngpt [25] с параметрами bootstrap = 1000, model = “segmented”. Значение p для модели рассчитывалось при помощи пакета lmtest [26] функцией “lmtest”, выполняющей тест отношения правдоподобия (asymmetric likelihood ratio test) в сравнении с моделью, предусматривающей отсутствие корреляции (данный тест рекомендуется авторами пакета chngpt).

Пространственная автокорреляция и тест Мантелла. Расчет пространственной автокорреляции и тест Мантелла проводили в программе GenAlEx. Расчет географических дистанций для этого этапа анализа проводили как на реальных значениях координат, так и на координатах с внесенным случайным шумом. И расчет пространственной автокорреляции, и тест Мантелла проводили с 999 итерациями.

Исключение близкородственных особей и анализ популяционной структуры. Для определения влияния близкородственных особей на выявленную популяционную структуру было проведено два

анализа: один с использованием данных по всем особям, добытым на территории РФ, второй с использованием сокращенного набора данных, из которого были исключены близкородственные особи. Для исключения близкородственных особей были идентифицированы пары особей с индексом родства более или равным 0.5 и добытые на расстоянии менее 100 км. Данная отсечка по расстоянию была выбрана на основе полученных данных по зависимости индекса родства от расстояния, а также в связи с описанными выше особенностями данных о месте добычи волков — при отсечке в 100 км набор особей, отобранный с использованием координат с внесенным случайным шумом и без него, не различался. Из отобранных пар исключалась либо особь с большим процентом отсутствующих данных, либо случайная особь (при равном проценте отсутствующих данных). Если одна из особей в паре уже была исключена ранее, вторая особь по умолчанию включалась в анализ.

Анализ популяционной структуры методами байесовского анализа проводили в программе STRUCTURE v2.3.4 [27–29] со следующими параметрами: длина периода “разогрева” (“burn-in”) — 25000; количество итераций MCMC повторов после “разогрева” — 250000; модель наследования — “admixture”; число кластеров $K = 1-8$, с 10 повторами для каждого K , флаг UsePopInfo = 0. Так как в рамках анализа предусматривалось сравнение полученных результатов для двух наборов данных, была использована опция “use sequential integer random seed”, которая позволяет задать “зерно”, используемое для расчетов случайных параметров. Таким образом, случайные аспекты для соответствующих итераций по обоим наборам данных будут одинаковыми, что позволяет более точно оценить вклад исключения родственных особей в результаты. В качестве стартового “зерна” использовалось число 1758.

Для работы с результатами программы STRUCTURE использовался пакет rphelper [30] для среды R, за исключением процесса “выравнивания” кластеров между итерациями (присвоения одинаковым кластерам одинакового номера), для которого использовался скрипт, специально написанный с использованием базовых функций среды R. На основе анализа генетической структуры проводилось определение наиболее вероятного числа исходных кластеров (K) при помощи метода Эванно [31].

Для выделенных популяций в программе GenAlEx были рассчитаны основные популяционно-генетические параметры: среднее число аллелей на locus — N_a ; число эффективных аллелей — N_e ; индекс Шеннона — I ; средняя наблюдаемая гетерозиготность — H_o ; средняя ожидаемая гетерозиготность — H_e ; внутрипопуляционный индекс фиксации — F ; несмещенная оценка ожидаемой гетерозиготности — uH_e ; показатель межпопуляционной

дифференциации — F_{ST} ; теоретическое число мигрантов между популяциями на поколение при наблюдаемом значении F_{ST} — N_m ; среднее число приватных (специфичных для популяции) аллелей — N_p .

РЕЗУЛЬТАТЫ

Характеристика матриц генотипов

После исключения особей и локусов с долей отсутствующих данных выше принятых пороговых значений для дальнейшего анализа оказались доступны данные по 738 особям и 22 локусам. По результатам оценки отклонения от равновесия Харди — Вайнберга, индекса фиксации и частоты нуль-аллелей из дальнейшего анализа были исключены также локусы ANTI26 и 2006, так как у них наблюдались отклонения по всем трем параметрам в половине или более географических популяций (9 и 6 соответственно). Таким образом, итоговый набор данных составил 738 особей, генотипированных по 20 локусам (2137, 2010, CXX253, vWf, 2079, PEZ03, ANTI19, 2201, 2096, CXX250, CXX225, 2140, 2054, 2168, 2159, ANTI38, CXX123, CXX204, ANTI06, 2001).

По результатам анализа минимального количества локусов, необходимых для дифференциации генотипов, максимальная информативность возможна уже при шести локусах и стабильно достигается (медиана уникальных генотипов = максимальному числу генотипов) при 17 локусах (рис. 1). По результатам анализа PIC используемые локусы показали средний уровень информативности — медиана значения PIC по всем локусам по общему набору данных составила 0.55, максимум 0.83 (locus vWf), минимум 0.23 (locus 2054).

Оценка коэффициентов родства

По результатам сравнения методов оценки коэффициентов родства (r) мы обнаружили наиболее сильную корреляцию между ожидаемыми и наблюдаемыми показателями (0.892) по индексу Wang [32]. Индексы Li [33] и Queller–Goodnight [34] показали практически идентичные значения корреляции ~0.89, а индекс Lynch–Ritland [35] — наименьшее значение $r = 0.819$ (рис. 2). На основании этих результатов для дальнейших расчетов мы использовали индекс Wang [32].

При рассмотрении всего набора данных целиком ожидаемо наблюдалась низкая (близкая к нулю) степень родства между особями с медианой 0.043.

Корреляция коэффициента родства с расстоянием

При рассмотрении простой линейной модели зависимости r от расстояния наблюдается значимая корреляция ($p < 0.001$; индекс корреляции Спирмена -0.159). Однако, подобная модель объясняет лишь малую часть дисперсии в коэффициентах родства — $R^2 = 0.028$. Практически идентичные результаты (с отклонением ~ 0.0001 для R^2 и индекса корреляции) были получены при использовании матриц расстояний, рассчитанных по координатам с внесенным случайным шумом.

Основываясь на данных о семейной и территориальной структуре популяций волка, мы предположили, что зависимость r от расстояния может иметь ступенчатый характер — на малых расстояниях уровень родства будет меняться больше на единицу расстояния, чем на больших дистанциях.

При рассмотрении ступенчатой (пороговой) регрессии были идентифицированы две точки преломления на отметках в 633.63 и 124.49 км (ст. ошибка 17.60 и 12.14 км соответственно). В данном анализе наблюдались различия в результатах при использовании координат с внесенным случайным шумом. Для них точки преломления составили 652.83 и 153.11 км (ст. ошибка 16.19 и 6.47 км соответственно) (рис. 3).

На основе полученных точек преломления были рассчитаны линейные модели для расстояний менее 120 км; от 120 до 633 км и более 633 км. Во всех трех случаях модели имели высокий уровень статистической значимости ($p < 0.001$). Наибольшее значение R^2 наблюдалось при расстояниях до 120 км — 0.031. Для двух других моделей значения R^2 составили 0.004 и 0.008 соответственно. При построении линейных моделей на основе дистанций с внесенным случайным шумом и соответствующих точек преломления статистическая значимость сохранялась ($p < 0.001$). Значения R^2 для данных моделей составили 0.027, 0.005 и 0.008.

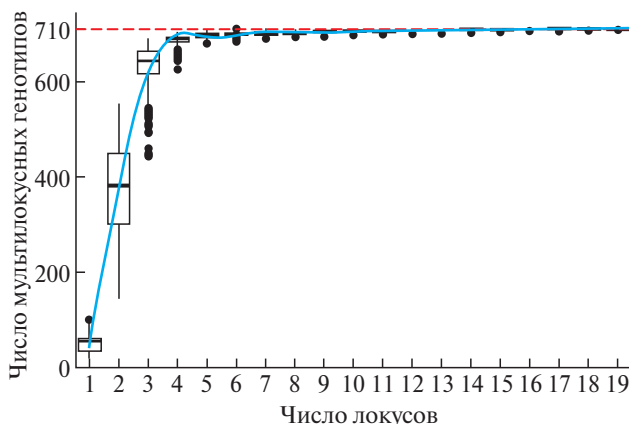


Рис. 1. Зависимость числа уникальных многолокусных генотипов от числа использованных локусов.

Коэффициенты корреляции Спирмена составили -0.180 , -0.065 , -0.087 соответственно для моделей на основе оригинальных координат и -0.145 , -0.070 , -0.086 для координат с внесенным случайным шумом.

Значимой корреляции между r и разницей во времени сбора не наблюдалось.

Пространственная автокорреляция и тест Мантелла

При проведении анализа пространственной автокорреляции с использованием размера классов в 100 км (рис. 4) наблюдается значение автокорреляции $r = 0.120$ на расстояниях < 100 км, к отметке в 200 км значение r снижается до 0.045 и сохраняется на этом уровне до отметки в 500 км, после чего резко снижается до 0.014 на отметке в 700 км. Далее значение r плавно снижается и теряет статистическую значимость на отметке в 1600 км, после чего колеблется в районе 0 вплоть до 2800 км. После этой отметки значения r приобретают значимо отрицательные значения (вплоть до -0.034) и сохраняют их вплоть до отметки в 6700 км, где происходит повторная потеря статистической значимости отклонения r от 0. Схожая картина с более сглаженными изменениями значений r наблюдается при использовании координат с внесенным случайным шумом.

При использовании размера классов в 25 км (рис. 5) выявляется более плавное снижение r с 0.139 на расстояниях до 25 км до 0.051 на отметке в 125 км. Понижение значений r после 500 км также идет более плавно. При использовании координат с внесенным случайным шумом динамика снижения r дополнительно снижается. В этом случае первичное снижение происходит с 0.131 на расстояниях до 25 км до 0.047 на отметке 150 км. Следующий этап снижения также начинается после отметки в 500 км. Стоит отметить близость

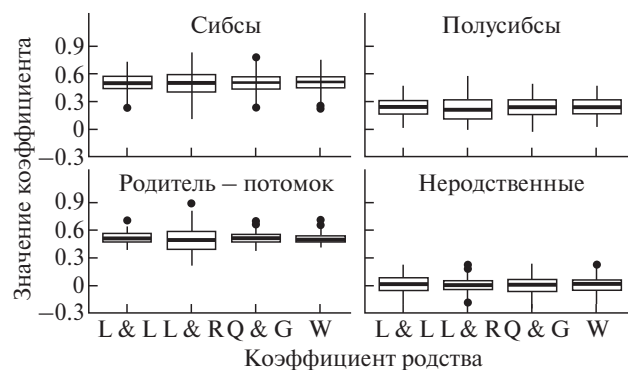


Рис. 2. Сравнение различных коэффициентов родства на основе 1000 симулированных родственных пар. Коэффициенты родства: L&L — *Li*; L&R — *Lynch-Ritland*; Q&G — *Queller-Goodnight*; W — *Wang*.

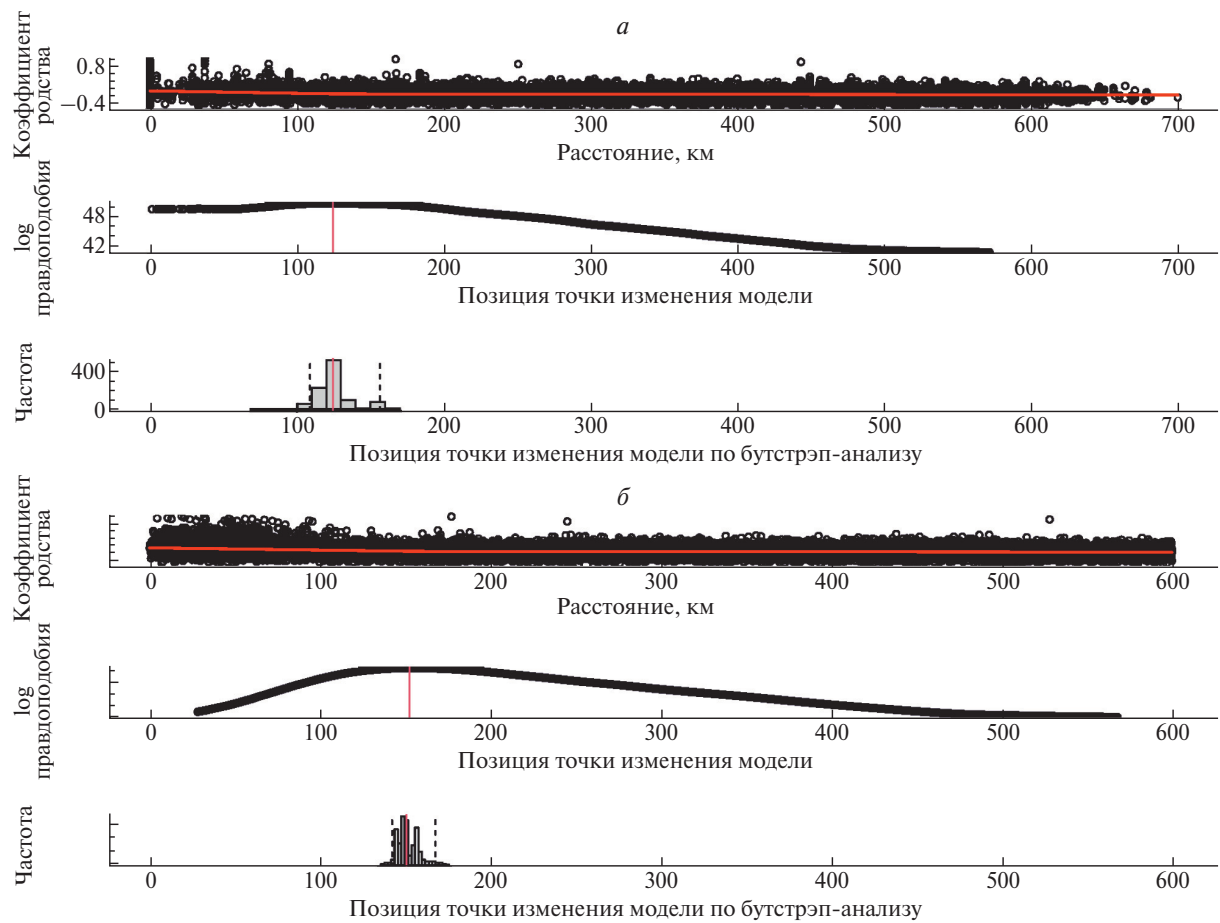


Рис. 3. График определения точки изменения линейной модели путем расчета по координатам без внесенного случайного шума (а) и с ним (б). Верхний график — зависимость индекса родства от расстояния в км; средний график — логарифм вероятности положения точки изменения модели; нижний график — результаты бутстрэп-анализа точки изменения модели.

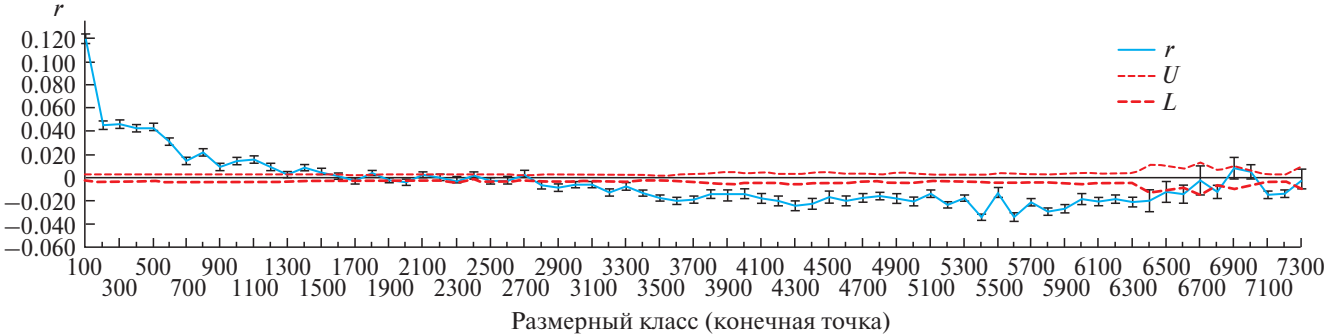


Рис. 4. Результаты анализа пространственной автокорреляции с использованием размерных классов в 200 км. r — индекс пространственной автокорреляции; U , L — верхний и нижний доверительные интервалы для отсутствия автокорреляции.

границ этапов снижения значения r с точками преломления, выявленными при анализе ступенчатой регрессии.

Тест Мантел показал значимую ($p < 0.001$) корреляцию между генетическими и географическими дистанциями со значением $R^2 = 0.016$ вне

зависимости от того, использовались координаты с внесенным случайным шумом или без него. При использовании логарифмированных географических дистанций значение R^2 составило 0.031 и 0.028 для координат с внесением случайного шума и без него соответственно.

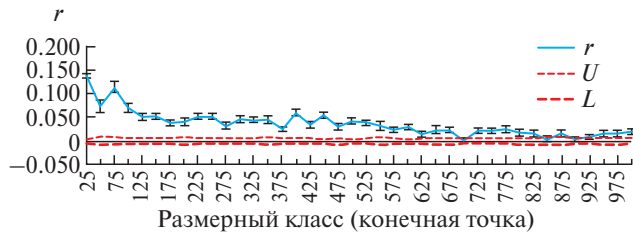


Рис. 5. Пространственная автокорреляция с использованием размерных классов в 25 км. r — индекс пространственной автокорреляции; U , L — верхний и нижний доверительные интервалы для отсутствия автокорреляции.

Влияние близкородственных особей на результаты анализа популяционно-генетической структуры

Для получения сокращенного набора генотипов из анализа были исключены 124 особи, таким образом, размеры сокращенного набора данных составили 614 особей, не являющихся близкими родственниками.

При анализе обоих наборов генотипов наблюдался консенсус между всеми 10 итерациями STRUCTURE вплоть до $K = 3$, т. е. во всех итерациях особи относились к одному и тому же кластеру примерно с одинаковой вероятностью (рис. 6). В обоих случаях наблюдалось отклонение одной итерации от консенсуса для $K = 4$ и $K = 6$ (в последнем случае при анализе полного набора генотипов также наблюдались легкие отклонения для четырех итераций). Для $K = 5$ наблюдалось отклонение от консенсуса двух итераций при анализе полного набора генотипов, но не при анализе сокращенного.

Для $K = 7$ число итераций с отклонением от консенсуса составило пять и три для полного и сокращенного наборов генотипов соответственно.

При анализе обоих наборов данных наибольшее значение ΔK наблюдалось для $K = 2$ (рис. 7), что соответствует разделению популяции на европейский и сибирский кластеры. Второе наибольшее значение ΔK наблюдалось для $K = 3$, при котором обособляется кластер, соответствующий популяции Чукотки. Начиная с $K = 4$ значения ΔK выходили на плато. При $K = 4$ в отдельный кластер выделялись популяции юга России, а также некоторые западносибирские популяции. При расчете ΔK без учета $K = 1$ максимальное значение наблюдалось для $K = 3$.

При сравнении результатов распределения особей по генетическим кластерам на основе полного и сокращенного наборов генотипов также наблюдались различия, не ассоциированные с расхождениями между результатами итераций STRUCTURE. Так, при использовании сокращенного набора генотипов при $K = 6$ популяции Волго-Вятского региона объединялись в один, отдельный от популяций севера и северо-запада европейской части России, кластер, в который также вошли приуральские популяции и некоторые популяции Западной Сибири. Этого выделения не наблюдалось при использовании полного набора генотипов. По причине наличия таких значительных различий именно $K = 6$ было выбрано для сравнения значений популяционно-генетических параметров (рис. 8).

Результаты анализа внутривидовой изменчивости для шести генетических кластеров приведены в табл. 1. Примечательно, что

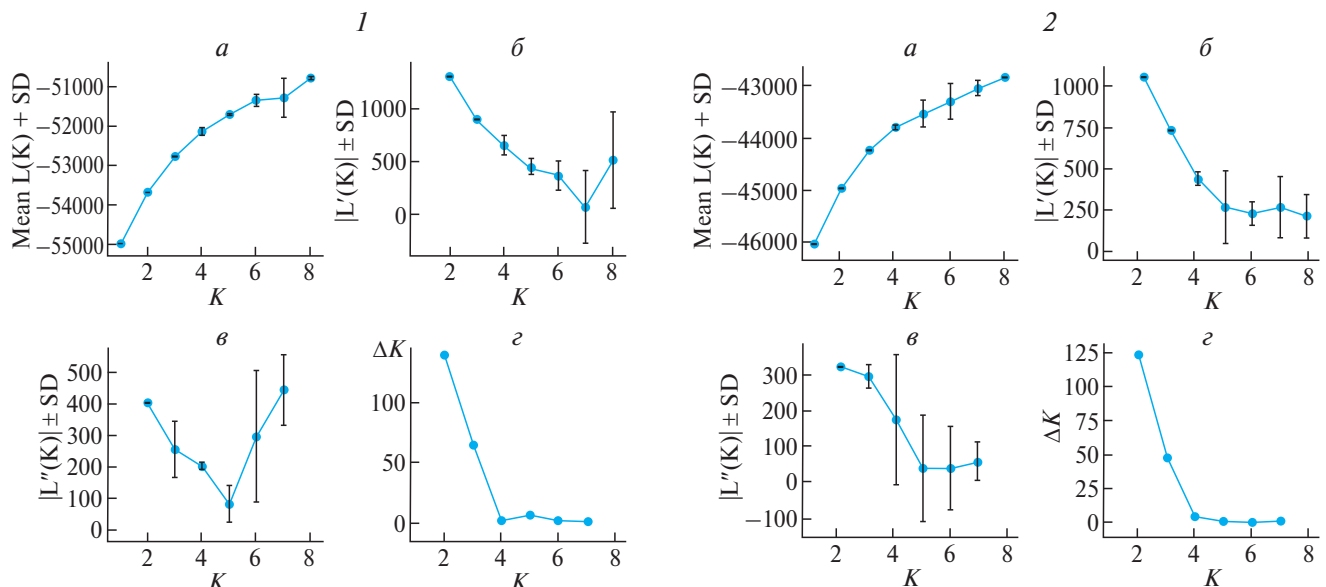


Рис. 6. Результаты анализа методом Эванно для полного (1) и сокращенного (2) набора генотипов.

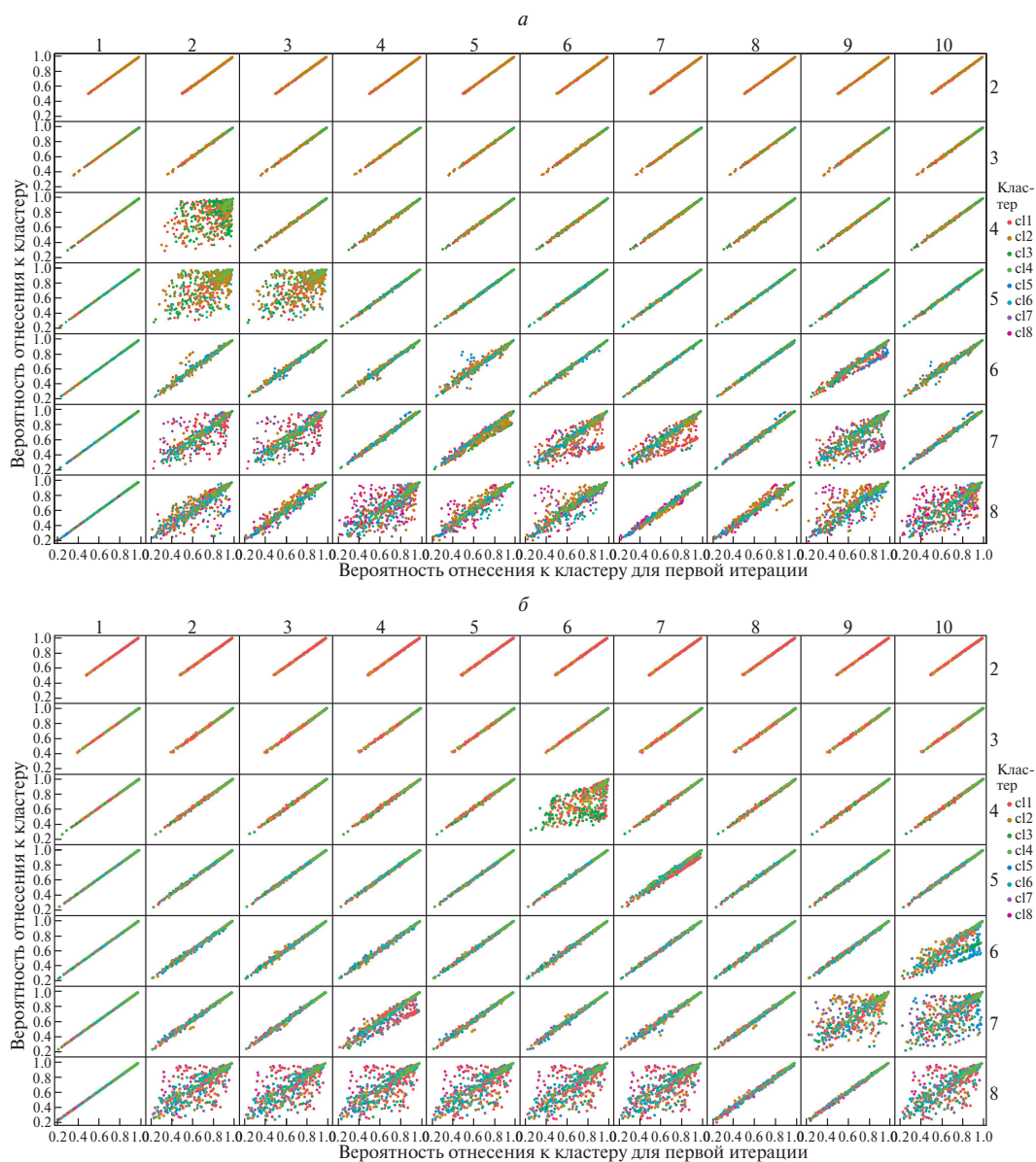


Рис. 7. График консенсуса между итерациями STRUcTURE для полного (а) и сокращенного (б) набора генотипов. В каждой ячейке по вертикали — вероятность отнесения особи к номинальному кластеру для данной итерации, по горизонтали — вероятность отнесения особи к номинальному кластеру для первой итерации. Организация ячеек по вертикали — значение K , по горизонтали — порядковый номер итерации STRUcTURE.

наблюдались минимальные различия в значениях показателей внутрипопуляционного генетического разнообразия для полного и сокращенного наборов генотипов. Наибольшие изменения наблюдались по общему и эффективному числу аллелей для первого и шестого кластеров, связанного с перераспределением особей из одного в другой. Для большинства других параметров разница между значениями для полного и сокращенного наборов генотипов находилась в пределах двух стандартных ошибок.

Оценка генетических дистанций между популяциями показала, что в большинстве пар при анализе полного набора генотипов полученные значения больше, чем при анализе сокращенного набора (табл. 2). Наибольшая разница в полученных значениях наблюдалась между 5 и 6, 1 и 6 кластерами (0.08), что вероятно связано с перераспределением особей между кластерами. Для двух пар наблюдалось повышение значений генетических дистанций при использовании сокращенного набора генотипов (1 и 2, 1 и 5 кластеры).



Рис. 8. Результаты байесовского анализа популяционной структуры для полного (а) и сокращенного (б) набора генотипов STRUCTURE.

Таблица 1. Средние значения показателей внутрипопуляционной изменчивости для каждого кластера, рассчитанные по полному и сокращенному наборам генотипов

Кластер	N	N_a	N_e	I	H_O	H_E	uH_E	F
1 полный	212.75 ± 1.36	11.30 ± 0.91	5.71 ± 0.44	1.88 ± 0.08	0.68 ± 0.02	0.80 ± 0.02	0.81 ± 0.02	0.16 ± 0.02
1 сокращенный	96.30 ± 1.03	9.90 ± 0.71	5.43 ± 0.41	1.83 ± 0.07	0.71 ± 0.02	0.80 ± 0.01	0.80 ± 0.01	0.11 ± 0.02
Разность	116.45	1.40	0.28	0.04	-0.03	0.01	0.00	0.05
2 полный	186.65 ± 2.19	10.70 ± 0.77	5.37 ± 0.44	1.81 ± 0.08	0.67 ± 0.02	0.79 ± 0.02	0.79 ± 0.02	0.15 ± 0.02
2 сокращенный	161.25 ± 1.87	10.70 ± 0.81	5.46 ± 0.45	1.83 ± 0.08	0.67 ± 0.02	0.79 ± 0.02	0.79 ± 0.02	0.15 ± 0.02
Разность	25.40	0.00	-0.08	-0.02	0.00	0.00	0.00	0.00
3 полный	91.40 ± 1.26	8.75 ± 0.66	4.56 ± 0.30	1.68 ± 0.07	0.68 ± 0.02	0.76 ± 0.02	0.77 ± 0.02	0.10 ± 0.02
3 сокращенный	80.05 ± 0.94	8.65 ± 0.63	4.63 ± 0.31	1.69 ± 0.07	0.68 ± 0.02	0.76 ± 0.02	0.77 ± 0.02	0.11 ± 0.02
Разность	11.35	0.10	-0.07	-0.01	0.00	0.00	0.00	-0.01
4 полный	108.70 ± 1.26	9.55 ± 0.65	5.20 ± 0.35	1.80 ± 0.07	0.67 ± 0.03	0.79 ± 0.02	0.79 ± 0.02	0.15 ± 0.04
4 сокращенный	106.60 ± 1.03	9.75 ± 0.63	5.36 ± 0.36	1.82 ± 0.07	0.66 ± 0.03	0.80 ± 0.01	0.80 ± 0.01	0.17 ± 0.04
Разность	2.10	-0.20	-0.16	-0.02	0.01	-0.01	-0.01	-0.02
5 полный	68.30 ± 0.42	8.60 ± 0.54	4.59 ± 0.35	1.66 ± 0.07	0.69 ± 0.02	0.76 ± 0.02	0.76 ± 0.02	0.08 ± 0.03
5 сокращенный	61.85 ± 0.28	8.55 ± 0.54	4.63 ± 0.35	1.67 ± 0.07	0.70 ± 0.02	0.76 ± 0.02	0.77 ± 0.02	0.08 ± 0.03
Разность	6.45	0.05	-0.04	0.00	0.00	0.00	0.00	0.00
6 полный	41.40 ± 0.22	7.20 ± 0.55	4.51 ± 0.34	1.60 ± 0.08	0.64 ± 0.02	0.75 ± 0.02	0.76 ± 0.02	0.14 ± 0.03
6 сокращенный	85.80 ± 0.51	10.00 ± 0.74	5.66 ± 0.48	1.84 ± 0.09	0.65 ± 0.02	0.79 ± 0.02	0.80 ± 0.02	0.17 ± 0.02
Разность	-44.40	-2.80	-1.15	-0.24	-0.02	-0.05	-0.04	-0.03

Таблица 2. Генетические дистанции между кластерами, рассчитанные по полному (ниже диагонали) и сокращенному (выше диагонали) наборам генотипов

Кластеры	1	2	3	4	5	6
1	0.000	0.151	0.243	0.123	0.214	0.104
2	0.151	0.000	0.216	0.124	0.110	0.133
3	0.243	0.216	0.000	0.243	0.187	0.292
4	0.123	0.124	0.243	0.000	0.207	0.115
5	0.214	0.110	0.187	0.207	0.000	0.214
6	0.104	0.133	0.292	0.115	0.214	0.000

Среднее число мигрантов по всем популяциям составило 5.71 ± 0.43 для полного набора генотипов и 7.16 ± 0.51 для сокращенного набора.

ОБСУЖДЕНИЕ

В большинстве регионов мира популяции волка находятся или в состоянии депрессии после длительного истребления, или в фазе роста после депрессии. Исследования родства проводились традиционно в основном на малых популяциях волка, подвергшихся недавнему “бутылочному горлышку”, например, скандинавской, где вследствие реколонизации (после полного истребления) несколькими мигрантами с территории СССР с 1980-х гг. [36] обитает высокоинбредная популяция волка, в которой наблюдаются выраженные признаки инбредной депрессии [37]. Подобные результаты были получены и в отношении небольших фрагментированных популяций мексиканского подвида волка, *Canis lupus baileyi* [38]. Однако ситуация на большей части России иная — на огромных пространствах лесной зоны, горно-лесных экотопов и, в меньшей степени, тундры и степи волк никогда не был доведен до депрессии. Если локально это и удавалось (в степях, некоторых территориях зоны тундры и в антропогенных ландшафтах), то восстановление происходило естественным образом из окружающих группировок с высокой численностью и нативной генетической структурой. В целом волк с территории России является уникальной моделью относительно ненарушенной (квазинативной) макропопуляции бореального волка Северной Евразии.

Исходя из полученных нами данных, можно заключить, что на изученной территории влияние семейной и территориальной структуры наиболее выражено на расстояниях до 125–150 км. При этом даже на таких дистанциях пространственное распределение особей объясняет лишь небольшую долю вариации в генетической близости. На расстояниях ~150–500 км все еще заметно влияние филопатрии и вызванной ею пространственной структуры, однако в значительно меньшей степени. Предположительно это является следствием процессов расселения и миграции на средне-дальние расстояния. На расстояниях 500–600 км и более влияние пространственной структуры популяции практически сходит на нет, но все еще остается достоверно положительным вплоть до ~1600 км. Очевидно, это обеспечивается дальними миграциями, особенно “пассионарными” в плане миграционной активности особей, а также, вероятно, несколькими поколениями средне-дальних миграций, если мы предположим, что способность к ним в какой-то значительной степени наследуется. Таким образом, можно заключить, что генетическая структура каждой популяции может быть подвержена влиянию со стороны популяций,

находящихся на расстояниях вплоть до 1600 км, при этом сила эффекта падает с расстоянием.

Стоит отметить, что в исследованиях, проводившихся на небольшом географическом масштабе, значения автокорреляции достигали нуля и негативных значений на значительно меньших расстояниях в ~10–20 км [5, 6, 39]. В то же время метаанализ популяционно-генетических исследований европейских популяций волка показал, что значения автокорреляции остаются достоверно положительными вплоть до 850 км [40], что больше соотносится с нашими данными, однако наши оценки примерно вдвое выше. Можно предположить, что анализ на малых масштабах не захватывает миграционных процессов внутри большей популяционной группировки — так, при рассмотрении всего нескольких стай стоит ожидать, что значения автокорреляции будут отрицательными на расстояниях, выходящих за границы территории каждой из стай. При рассмотрении больших территорий возможно захватить большее количество расселившихся и мигрировавших особей, что приводит к положительным значениям автокорреляции на больших расстояниях.

Тем не менее полученные нами наиболее высокие значения автокорреляции, которые можно отнести к родственным животным из одной стаи, все еще наблюдаются вплоть до расстояний (125 км), значительно превышающих упомянутые выше. Это может быть связано с тем, что наши данные включают в себя волков, обитающих в степных и тундровых экотопах, для которых описан значительно больший размер участков [1, 2]. Таким образом, на данных территориях особи, принадлежащие к одной семейной группировке, могут быть добыты на больших расстояниях.

Полученные нами результаты указывают на то, что удаление из выборки близкородственных особей позволяет более однозначно определить генетическую структуру исследуемых популяций, по крайней мере оценить ту часть изменчивости, которая относится к автохтонной компоненте и отражает локальные адаптации особей. При элиминации из матрицы для расчетов родственников не происходит существенного изменения в получаемых данных о внутривидовой изменчивости. Наблюдаемое некоторое снижение генетических дистанций между популяциями было ожидаемо (нивелируется эффект межвыборочного смещения частот аллелей из-за разной представленности семейных групп), однако за исключением кластеров, претерпевших перераспределение особей, даже эти изменения находятся на достаточно низком уровне, лишь для двух пар превышая 0.01.

Таким образом, можно сделать вывод о целесообразности удаления близкородственных особей, добытых в непосредственной близости друг от друга, для повышения информативности кластерного

анализа популяционной структуры. Тем не менее расчеты с родственными генотипами вносят свой вклад, отражают естественную ситуацию, позволяют, в частности, оценить эффекты дальней миграции. Кроме того, нами продемонстрировано, что основные черты проявляющейся популяционно-генетической структуры при анализе с родственниками в матрице и с их исключением из расчетов остаются неизменными. Так, в нашем случае на пространствах Северной Евразии (Россия и Казахстан) в обоих вариантах анализа выявлялись одни и те же территориальные группировки: 1) юг европейской части России (ЕЧР); 2) центр, север, и северо-запад ЕЧР; 3) северо-восток ЕЧР и Урал; 4) часть Западной Сибири; 5) Западная и Средняя Сибирь; 6) Восточная Сибирь и 7) Чукотка. При расчетах байесовской кластеризации с исключенными родственниками мы наблюдали лишь некоторое перераспределение доли исходных кластеров между североазиатскими, уральскими и частью западносибирских группировок.

Оценки генетического родства являются необходимой компонентой знаний о популяционной структуре не только потому, что помогают получить новые данные о комплексе поведенческих особенностей особей, составляющих биологические виды, и их территориальных и экологических группировок [41], но и поскольку пространственный автокорреляционный анализ косвенно, но объективно оценивает паттерны миграции, в первую очередь natalной (от рождения до первой репродукции), на исследуемой части видовой ареала. На небольшой географической шкале анализ родства незаменим в реконструкции семейной структуры и социальной организации [42]. На взятой нами в качестве полигона для исследований макрогеографической шкале на первый план выходят паттерны миграции. В нашей матрице данных исходно не было заведомо родственных генотипов волка, т. е. большая часть близких родственников элиминировалась на этапе выбора образцов для включения в анализ, и лишь скрытые, неочевидные родственные связи были объектом исследования. Несмотря на такой подход, наличие шлейфа особей с достоверно отличающимися от нуля коэффициентами родства прослеживалось на масштабе в сотни километров от места происхождения подобных особей.

Полученные данные, таким образом, демонстрируют интегрирующий эффект natalной миграции, действующей против дифференцирующего влияния филпатрии. Кроме того, мы продемонстрировали, что исключение подобных особей из выборки после анализа родства может приводить к занижению оценок генетических дистанций между выборками. При этом исключение родственных генотипов помогает более адекватно оценить некоторые параметры популяционно-генетической структуры, но в то же время наличие семейных

кластеров является в определенных условиях частью реальной картины, которая остается за кадром, если родственники вообще исключаются из анализа. Так, для волков Центрально-Лесного государственного биосферного заповедника было показано, что в резком увеличении численности волка в Тверской области в исследованный период времени большую роль играли “резидентные” особи, natalная миграция которых предполагалась лишь на небольшие дистанции [43]. Однако лишь применение технологий молекулярно-генетического анализа дает возможности изучить эти процессы более детально с получением количественных оценок параметров пространственной генетической структуры.

В целом мы можем заключить, что оба вида анализа (по полной матрице и с исключением родственных генотипов) информативны и дополняют друг друга, позволяя получить и дополнительные данные для характеристики генетической структуры вида.

Исследование выполнено за счет гранта Российского научного фонда № 23-24-00635, <https://rscf.ru/project/23-24-00635/>.

Этические нормы обращения с животными не нарушались, ни одно животное не было добыто специально для данного исследования. Все образцы взяты от законно добытых особей. Все применимые международные, национальные и/или институциональные принципы использования животных были соблюдены.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Mech L.D., Boitani L. Wolves: behavior, ecology, and conservation. Chicago: Univ. Chicago Press, 2006. 472 p.
2. Бибиков Д.И. Волк. Происхождение, систематика, морфология, экология. М.: Наука, 1985. 608 с.
3. Mystajek R.W., Tracz M., Tracz M. et al. Spatial organization in wolves *Canis lupus* recolonizing north-west Poland: Large territories at low population density // Mamm. Biol. 2018. V. 92. P. 37–44. <https://doi.org/10.1016/j.mambio.2018.01.006>
4. Åkesson M., Liberg O., Sand H. et al. Genetic rescue in a severely inbred wolf population // Mol. Ecol. 2016. V. 25. № 19. P. 4745–4756. <https://doi.org/10.1111/mec.13797>
5. Caniglia R., Fabbri E., Galaverni M. et al. Noninvasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population // J. Mamm. 2014. V. 95. № 1. P. 41–59. <https://doi.org/10.1644/13-MAMM-A-039>
6. Randall D.A., Pollinger J.P., Argaw K. et al. Fine-scale genetic structure in Ethiopian wolves imposed by sociality, migration, and population bottlenecks //

- Cons. Genet. 2010. V. 11. № 1. P. 89–101.
<https://doi.org/10.1007/s10592-009-0005-z>
7. *Szewczyk M., Nowak S., Niedźwiecka N. et al.* Dynamic range expansion leads to establishment of a new, genetically distinct wolf population in Central Europe // *Sci. Rep.* 2019. V. 9. № 1. P. 1–16.
<https://doi.org/10.1038/s41598-019-55273-w>
 8. *Pilot M., Dabrowski M.J., Hayrapetyan V. et al.* Genetic variability of the grey wolf *Canis lupus* in the Caucasus in comparison with Europe and the Middle East: distinct or intermediary population? // *PLoS One.* 2014. V. 9. № 4.
<https://doi.org/10.1371/journal.pone.0093828>
 9. *Казимиров П.А., Леонтьев С.В., Нечаева А.В. и др.* Популяционно-генетическая структура степного волка России и Казахстана по микросателлитным локусам // *Генетика.* 2022. Т. 58. № 11. С. 1261–1272. <https://doi.org/10.31857/S0016675822110042>
 10. *Toonen R.J., Hughes S.* Increased throughput for fragment analysis on ABI Prism 377 automated sequencer using a membrane comb and STRand software // *Biotechniques.* 2001 V. 31 P. 1320–1324.
<http://www.vgl.ucdavis.edu/STRand>
 11. *Kamvar Z.N., Brooks J.C., Grünwald N.J.* Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality // *Front. Genet.* 2015. V. 6.
<https://doi.org/10.3389/fgene.2015.00208>
 12. *Kamvar Z.N., Tabima J.F., Grünwald N.J.* Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction // *PeerJ.* 2014. V. 2.
<https://doi.org/10.7717/peerj.281>
 13. *R Core Team* R: A language and environment for statistical computing. 2022. Available online at <https://www.R-project.org/>
 14. *Clark L.V., Drauch Schreier A.* Resolving microsatellite genotype ambiguity in populations of allopolyploid and diploidized autopolyploid organisms using negative correlations between allelic variables // *Mol. Ecol. Res.* 2017. V. 17. № 5. P. 1090–1103.
<https://doi.org/10.1111/1755-0998.12639>
 15. *Clark L.V., Jasieniuk M.* Polysat: An R package for polyploid microsatellite analysis // *Mol. Ecol. Res.* 2011. V. 11. № 3. P. 562–566.
<https://doi.org/10.1111/j.1755-0998.2011.02985.x>
 16. *Paradis E.* pegas: An R package for population genetics with an integrated–modular approach // *Bioinformatics.* 2010. V. 26. P. 419–420.
<https://doi.org/10.1093/bioinformatics/btp696>
 17. *Adamack A.T., Gruber B.* PopGenReport: Simplifying basic population genetic analyses in R // *Meth. Ecol. Evol.* 2014. V. 5(4). 384–387.
<https://doi.org/10.1111/2041-210X.12158>
 18. *Gruber B., Adamack A.T.* Landgenreport: A new R function to simplify landscape genetic analysis using resistance surface layers // *Mol. Ecol. Res.* 2015. V. 15(5). 1172–1178. doi:10.1111/1755-0998.12381
 19. *Peakall R., Smouse P.E.* GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research // *Mol. Ecol. Notes.* 2006. V. 6. № 1. P. 288–295.
<https://doi.org/10.1111/j.1471-8286.2005.01155.x>
 20. *Peakall R., Smouse P.E.* GenALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update // *Bioinformatics.* 2012. V. 28. № 19. P. 2537–2539.
<https://doi.org/10.1093/bioinformatics/bts460>
 21. *Pew J., Wang J., Muir P., Frasier T.* Related: an R package for analyzing pairwise relatedness data based on codominant molecular markers // *Mol. Ecol. Res.* 2015. V. 15. P. 557–561. <https://doi.org/10.1111/1755-0998.12323>
 22. *Grolemund G., Wickham H.* Dates and times made easy with lubridate // *J. Stat. Software.* 2011. V. 40. № 3. P. 1–25.
<https://doi.org/10.18637/jss.v040.i03>
 23. *Kassambara A.* rstatix: Pipe-Friendly Framework for Basic Statistical Tests // 2023.
<https://cran.r-project.org/web/packages/rstatix/index.html>
 24. StatSoft, Inc. STATISTICA (data analysis software system), version 8. 2007. www.statsoft.com
 25. *Fong Y., Huang Y., Gilbert P., Permar S.* chngpt: Threshold regression model estimation and inference // *BMC Bioinformatics.* 2017. V. 18. № 454.
<https://doi.org/10.1186/s12859-017-1863-x>
 26. *Zeileis A., Hothorn T.* Diagnostic checking in regression relationships // *R News.* 2002. V. 2. № 3. P. 7–10.
 27. *Falush D., Stephens M., Pritchard J.K.* Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies // *Genetics.* 2003. V. 16. № 4. P. 1567–1587.
<https://doi.org/10.1093/genetics/164.4.1567>
 28. *Hubisz M.J., Falush D., Stephens M., Pritchard J. K.* Inferring weak population structure with the assistance of sample group information // *Mol. Ecol. Res.* 2009. № 9. P. 1322–1332.
<https://doi.org/10.1111/j.1755-0998.2009.02591.x>
 29. *Pritchard J.K., Stephens M., Donnelly P.* Inference of population structure using multilocus genotype data // *Genetics.* 2000. V. 155. № 2. P. 945–959.
<https://doi.org/10.1093/genetics/155.2.945>
 30. *Francis R.M.* pophelper: An R package and web app to analyse and visualize population structure // *Mol. Ecol. Res.* 2017 V. 17. P. 27–32.
<https://doi.org/10.1111/1755-0998.12509>
 31. *Evanno G., Regnaut S., Goudet J.* Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study // *Mol. Ecol.* 2005. V. 14. № 8. P. 2611–2620.
<https://doi.org/10.1111/j.1365-294X.2005.02553.x>
 32. *Wang J.* An estimator for pairwise relatedness using molecular markers // *Genetics.* 2002. V. 160. № 3. P. 1203–1215.
<https://doi.org/10.1093/genetics/160.3.1203>

33. Li C.C., Weeks D.E., Chakravarti A. Similarity of DNA fingerprints due to chance and relatedness // Human Heredity. 1993. V. 43. № 1. P. 45–52. <https://doi.org/10.1159/000154113>
34. Queller D.C., Goodnight K.F. Estimating relatedness using genetic markers // Evolution. 1989. V. 43. № 2. P. 258–275. <https://doi.org/10.2307/2409206>
35. Lynch M., Ritland K. Estimation of pairwise relatedness with molecular markers // Genetics. 1999. V. 152. № 4. P. 1753–1766. <https://doi.org/10.1093/genetics/152.4.1753>
36. Ellegren H. Inbreeding and relatedness in Scandinavian grey wolves *Canis lupus* // Hereditas. 1999. V. 130. № 3. P. 239–244. <https://doi.org/10.1111/j.1601-5223.1999.00239.x>
37. Liberg O., Andren H., Pedersen H.C. et al. Severe inbreeding depression in a wild wolf (*Canis lupus*) population // Biol. Letters. 2005. V. 1. № 1. P. 17–20. <https://doi.org/10.1098/rsbl.2004.0266>
38. Fredrickson R.J., Siminski P., Woolf M., Hedrick P.W. Genetic rescue and inbreeding depression in Mexican wolves // Proc. Royal Soc. B-Biol. Sci. 2007. V. 274. № 1623. P. 2365–2371. <https://doi.org/10.1098/rspb.2007.0785>
39. Korablev M.P., Korablev N.P., Korablev P.N. Genetic diversity and population structure of the grey wolf (*Canis lupus* Linnaeus, 1758) and evidence of wolf× dog hybridisation in the centre of European Russia // Mamm. Biology. 2020. P. 1–14. <https://doi.org/10.1007/s42991-020-00074-2>
40. Hindrikson M., Remm J., Pilot M. et al. Wolf population genetics in Europe: A systematic review, meta-analysis and suggestions for conservation and management // Biol. Reviews. 2017. V. 92. № 3. P. 1601–1629. <https://doi.org/10.1111/brv.12298>
41. Gompper M.E., Wayne R.K. Genetic relatedness among individuals within carnivore societies // Carnivore Behavior, Ecol. and Evol. V. 2. Ithaca and London: Cornell Univ. Press, 1996. P. 429–452. <https://doi.org/10.7591/9781501745829-020>
42. Jędrzejewski W., Branicki W., Veit C. et al. Genetic diversity and relatedness within packs in an intensely hunted population of wolves *Canis lupus* // Acta Theriologica. 2005. V. 50. № 1. P. 3–22. <https://doi.org/10.1007/BF03192614>
43. Кочетков В.В. Филопатрия и дисперсия в популяции волка (*Canis lupus* L.) // Сиб. экол. журн. 2015. Т. 22. № 3. С. 388–397. <https://doi.org/10.15372/SEJ20150306>

Genetic Structure of Wolf Populations in North Eurasia: the Effect of Exclusion of Closely Related Individuals from Analysis

P. A. Kazimirov^{a, b, *}, Yu. S. Belokon^a, M. M. Belokon^a, A. Ya. Bondarev^c,
A. V. Davydov^d, E. S. Zakharov^e, S. V. Leontyev^{f, b, **}

^aVavilov Institute of General Genetics Russian Academy of Sciences, Moscow, 119991 Russia

^bAll-Russian Research Institute for Environmental Protection, Moscow, 117628 Russia

^cAltai State Agricultural University, Barnaul, 656049 Russia

^dFederal center of development of the hunting economy, Moscow, 105118 Russia

^eAmnosov North-Eastern Federal University, Yakutsk, 677000 Russia

^fLimited Liability Partnership “National Center of Biotechnology”, Astana, 010000 Kazakhstan

*e-mail: farenklaw@gmail.com

**e-mail: dmitri_p@inbox.ru

We describe the results of analysis of genetic structure and spatial autocorrelation in the populations of grey wolf (*Canis lupus* Linnaeus, 1758) on the territory of the Russian Federation and the Republic of Kazakhstan, based on 20 autosomal microsatellite markers. With the use of molecular markers, we uncovered hidden genealogical patterns reaching as far as 700–1600 km and having the most pronounced effect on distances up to 150 km. Our research has shown that identification and exclusion of closely related genotypes has limited effect on the results of analysis of intrapopulation genetic diversity. Meanwhile, such procedure is recommended for researching population structure, as it allows for streamlining some statistical approaches. Results of our work demonstrate integral effect of natal migration, working against the differentiation effect of philopatry. Finally, we also show that the exclusion of closely related individuals can lead to underestimation of values of genetic distances between populations.

Keywords: wolf, microsatellites, genetic relatedness, relatedness estimator, spatial autocorrelation

УДК 575.11:598.288.6

“ЭВОЛЮЦИЯ” МИТОХОНДРИАЛЬНОГО ГЕНОМА ПЕНОЧКИ-ТАЛОВКИ (*Phylloscopus borealis* SENSU LATO) ПРОИСХОДИТ В ЕЕ ЯДЕРНОМ ГЕНОМЕ

© 2024 г. Л. Н. Спиридонова¹, О. П. Вальчук¹, Я. А. Редькин^{2,3}

¹Федеральный научный центр Биоразнообразия наземной биоты Восточной
Азии Дальневосточного отделения Российской академии наук, Владивосток, 690022 Россия
²Научноисследовательский Зоологический музей Московского государственного университета
им. М.В. Ломоносова, Москва, 125009 Россия

³Институт биологических проблем криолитозоны Сибирского отделения Российской академии наук,
г. Якутск, 677000 Россия
e-mail: spiridonova@biosoil.ru

Поступила в редакцию 20.12.2023 г.

После доработки 11.01.2024 г.

Принята к публикации 23.01.2024 г.

Проведено секвенирование частичного фрагмента генов *ND5–cytb* митохондриального генома (1553 пн) и его ядерных копий *Phylloscopus borealis sensu lato* (s. l.), принадлежащих к отдельным таксономическим группам из разных частей ареала. Показана идентичность большинства таксон-специфичных и уникальных митохондриальных замен у форм *examinandus* и *xanthodryas* с таковыми в ядерных копиях мтДНК *borealis*. Различия митохондриальных гаплотипов *examinandus* с ядерными копиями мтДНК *borealis* имели низкие значения ($p = 0.02$), а генетическая дивергенция собственно мтДНК *borealis–examinandus*, *borealis–xanthodryas* и *examinandus–xanthodryas* значительно превышала предыдущие значения ($p = 0.035$, 0.044 и 0.046 соответственно). Впервые в ядерном геноме одной особи *borealis* в западной части гнездового ареала (Республика Коми) обнаружена ядерная копия митохондриального гаплотипа самой восточной формы *xanthodryas*, а ядерные копии мтДНК *xanthodryas* из префектуры Тояма (Япония) оказались, в свою очередь, близки митохондриальным гаплотипам *borealis* из Кытлыма (Свердловская обл.) ($p = 0.018$). Таким образом, источником большинства замен в митохондриальных ДНК исследованных форм являются мутации, возникшие в ядерных копиях митохондриальных генов. Происхождение митохондриальных гаплотипов *examinandus* и *xanthodryas* из ядерных копий мтДНК *borealis* и близкое сходство их ядерных геномов дают основание рассматривать митогеномы этих форм в качестве вариантов гаплотипа одного вида *Ph. borealis* s. l. С большой долей вероятности можно утверждать, что время дивергенции гаплотипов анализируемых форм значительно меньше, чем 2.5–3 млн лет, как то предполагалось рядом авторов ранее, а “молекулярные часы”, не учитывающие рекомбинационные события между ядерным и митохондриальным геномами, в данном случае не могут быть использованы.

Ключевые слова: митохондриальный геном, ядерные копии мтДНК, межгеномная рекомбинация, *Phylloscopus borealis*.

DOI: 10.31857/S0016675824070045 **EDN:** VIJXE

Митохондриальные маркеры широко используются для изучения генетического разнообразия и построения молекулярных филогений. Однако согласованность классической и молекулярной систематики часто нарушается. Анализ хранящихся в Генбанке последовательностей митохондриальных (мт) генов из работ по филогенетике разных таксонов убедительно показал, что одной из причин несоответствия молекулярных данных

классическим морфологическим признакам, несомненно, является ложное использование ядерных копий мт-генов (NUMT) вместо подлинных мтДНК [1]. Ядерные копии митохондриальных генов, с одной стороны, могут вносить существенные ошибки в филогенетический анализ [2], а с другой — могут помочь разрешению вопросов о происхождении новых вариантов гаплотипов мтДНК [3–5]. В настоящее время NUMT обнаружены во

всех исследованных в этом отношении растений и животных. Научный интерес к ядерным копиям митохондриальной ДНК, ранее считавшимся “молекулярным мусором” в филогенетических исследованиях, в последнее время возрастает [1, 6–9 и мн. др.]. Особый интерес для исследования молекулярных причин глубокой дивергенции митохондриальных линий у таксонов низкого ранга со слабыми морфологическими отличиями представляют группы видов и подвидов со значительно дифференцированными митохондриальными гаплотипами. В настоящее время имеется много примеров глубокой дивергенции митохондриальных гаплотипов, особенно у птиц, на так называемые восточные и западные [10–13]. В этом отношении интересным оказался комплекс широко распространенной пеночки-таловки, представленный тремя сильно различающимися филогруппами мт-гаплотипов.

Систематика форм, входящих в состав группировки пеночек-таловок *Phylloscopus borealis* (s. l.), представляет собой один из сложных случаев среди пеночек Северной Палеарктики. В настоящее время существуют два противоположных взгляда. По мнению российских орнитологов, совпадающему с критериями классической систематики, это по-прежнему один политипический вид со сложной внутривидовой структурой, включающий семь подвидов [14–15]. По особенностям вокализации эти формы образуют три хорошо обособленные подвидовые группы. Первая — “*borealis*”, объединяет северо-европейские, сибирские и аляскинские популяции, отличающиеся так называемой “быстрой” песней, включает расы: *Ph. b. talovka* Portenko, 1938; *Ph. b. transbaicalica* Portenko, 1938; *Ph. b. borealis* (Blasius, 1858) и *Ph. b. kennicotti* (S.F. Baird, 1869). Вторая — “*hylebata*”, включает популяции южной части Сихотэ-Алиня, Сахалина, Хоккайдо, Курильских островов и Камчатки, объединяет два подвида с “медленным” типом песни: *Ph. b. hylebata* Swinhoe, 1860 (южный, более мелкий) и *Ph. b. examinandus* Stresemann, 1913 (северный, более крупный). И третья — “японская таловка”, гнездящаяся на островах Хонсю и Кюсю, представленная единственной формой *Ph. b. xanthodryas* (Swinhoe, 1863) и обладающая также очень специфичным характером песни. Согласно другой точке зрения, более распространенной в мировой литературе [16–21], основанной на результатах молекулярных исследований (цитохром *b*) в сочетании с указанными данными по вокализации, выделяются три разных вида: “*Ph. borealis*”, “*Ph. examinandus*” и “*Ph. xanthodryas*”.

Следует отметить, что области распространения “медленного” и “быстрого” типа песни у таловок заметно перекрываются. Самые западные районы, где среди птиц с “быстрой” песней изредка отмечались песни “медленного” типа, установлены в

Хэнтэй-Чикойском нагорье [22] и нижнем течении р. Олёкмы в Южной Якутии [15]. При этом А.А. Назаренко [22] подчеркивает, что в Забайкалье песни обоих типов воспроизводят одни и те же самцы. Среди особей с “медленной” песней на Сахалине изредка можно слышать и “быстрый” вариант песни или сочетание элементов обоих вариантов у одних и тех же птиц (наши данные). Анализ окраски коллекционных экземпляров, собранных в гнездовой период на Сахалине и в Приамурье, показывает, что в этих регионах часто встречаются таловки со смешанными признаками. Е.Г. Лобков [23] прямо указывает, что весь север Камчатского полуострова, граничащий с Корьякским нагорьем, где обитают *Ph. b. borealis*, населяют таловки с переходными признаками. Все это указывает на то, что в зонах вторичного контакта таловок с быстрым и медленным типами песни в Приамурье, Приморье и сопредельных регионах, а также на севере Камчатки и на Сахалине сформировались переходные (гибридные) популяции этих птиц со смешанными морфологическими признаками и смешанной вокализацией. Выполненные к нынешнему моменту генетические исследования таловок с разным типом песни из разных частей ареала [16, 18, 20] основаны на малом количестве особей из такого же соответственно малого числа локалитетов, что без проведения должного анализа фенотипической принадлежности птиц не позволяет выявить характер репродуктивных отношений между ними. В связи со сказанным выше нам представляется более оправданным считать эти группы конспецифичными. Кроме того, использование названия “*examinandus* Stresemann, 1913” для отделения таловок с медленным типом песни неверно, поскольку птицам этой же группы должно принадлежать старшее имя “*hylebata* Swinhoe, 1860” [15, 24].

Секвенирование полного митогенома трех особей с о. Сахалин выявило у одной из них два значительно дифференцированных гаплотипа: собственно митохондриальный гаплотип *examinandus* и ядерную копию мтДНК (NUMT), сходную с мт-гаплотипом *borealis* [5]. В связи с этим мы выдвинули гипотезу происхождения гаплотипа *examinandus* из ядерного генома *borealis* посредством межгеномной рекомбинации. Принимая во внимание результаты наших предыдущих исследований [3–5], цель настоящей работы заключалась в поиске причин глубокой дифференциации митохондриальных гаплотипов комплекса форм *Ph. borealis* s. l. на фоне слабой морфологической изменчивости. Для этого необходимо было на обширном материале, по возможности со всего видового ареала, исследовать наиболее дифференцированные участки митохондриального генома и их ядерные копии и сравнить их у всех представителей комплекса.

МАТЕРИАЛЫ И МЕТОДЫ

Сбор материала и амплификация фрагмента *ND5-cytb* мтДНК

В дальнейшем изложении для упрощения наименования трех таксономических группировок комплекса мы будем использовать названия “*Ph. borealis*”, “*Ph. examinandus*” и “*Ph. xanthodryas*” в соответствии с их обозначением в современных Списках птиц мира [21, 25]. Материал для исследования представлен 80 образцами из 28 локалитетов, которые были собраны в разное время и охватывают большую часть гнездового ареала форм этого комплекса в России и в Японии (рис. 1, табл. 1). Для упрощения исследований образцам, имеющим оригинальные коды (номера колец живых птиц или уникальные номера и коды экземпляров зоологических коллекций), были присвоены также порядковые номера (табл. 1) Образцы мышечных тканей и крови были фиксированы в 96%-ном этаноле. ДНК выделяли набором ISILATE II Genomic DNA Kit (Bioline, Inc.). Кроме того, использован материал в виде перьев от образцов, хранящихся в зоологических музеях МГУ и ДВФУ.

Удачным фрагментом митохондриального генома для проведения данной работы оказался участок длиной 1.5 тпн., включающий смежные части генов NADH-дегидрогеназы единицы 5 (*ND5*) и цитохрома *b* (*cytb*). Ядерная копия мтДНК (NUMT) этого фрагмента была обнаружена в результате полногеномного секвенирования у одной из птиц о. Сахалин [5]. Для синтеза фрагмента из митохондриального и ядерного геномов использованы специфичные праймеры Ph NUMT-Fw-12890 5'-CAGGCTCTATCATCCACAGCCTT-3' и Ph NUMT-Rv-14417 5'-CACCTAGTAGGTTA GGTGAAAATAGG-3'. По причине плохой сохранности ДНК в музейном материале для амплификации исследуемого фрагмента у музейных образцов использовали дополнительные внутренние праймеры (Ph NUMT-Fw-13788 5'-AAGTTTCATCAAGCAGAGATGTTTG-3' и Ph NUMT-Rv-13764 5'-CAAACATCTCTGCTTGATGA AACTT-3'). В сравнительный анализ мы взяли следующие последовательности генов *ND5* и *cytb*, хранящиеся в Генбанке NCBI: *Ph. xanthodryas* (о. Сикоку AB362465; AB362466), *Ph. examinandus* (о. Сахалин AB362464), *Ph. borealis* (Магадан AB362462). Выборка *Ph. xanthodryas* была дополнена десятью последовательностями части гена *cytb* из Генбанка (MH079391, AB362458, AB362457, AB362456, AB362455, AB362452, AB362451, AB362449, AB362447, AB362448). Эти последовательности были использованы только в филогенетической реконструкции, за исключением сети гаплотипов.

Амплификацию фрагмента проводили в 20 мкл объема, содержащего 20 нг матрицы ДНК, 5

пмолей каждого праймера, 1 × готовой смеси для ПЦР с Taq AB ДНК-полимеразой (Компания “Алкор Био”, Россия). Цикл амплификации включал преденатурацию 3 мин/95 °С, затем 35 циклов: денатурация 30 с/95 °С, отжиг праймера 30 с/58 °С, элонгация в зависимости от длины фрагмента 30 или 90 с/72 °С, и финальная элонгация 72 °С в течение 5 мин.

Секвенирование, клонирование и компьютерная обработка последовательностей

Продукт амплификации был использован для циклического секвенирования с набором реагентов ABI PRISM® BigDye™ Terminator v. 3.1. Реакцию присоединения метки проводили с теми же праймерами. Секвенирование проводили на автоматическом лазерном секвенаторе ABI PRISM 3130 (Applied Biosystems, США/Hitachi Япония).

Для образцов, имеющих в амплифицированном фрагменте две наиболее дивергированные последовательности, проведено клонирование с помощью набора InsTAclone™ PCR cloning Kit (Fermentas, Литва) согласно инструкции производителя. Более подробная методика клонирования описана ранее [3]. Прямые и обратные последовательности были собраны в пакете программы Staden 1.53 [26]. Выравнивание проводилось в программе ClustalW, предложенной в MEGA v 7.0.21 [27].

Изменчивость гаплотипического (*h*) и нуклеотидного (π) разнообразия [28] форм анализировали в программе DnaSP v5.10.01 [29]. Дивергенцию между гаплотипами (*p*-дистанция) определяли с использованием модели замен GTR+G+I [28] в MEGA v 7.0.21. Выбор филогенетической модели проводился с помощью modeltest [30] в той же программе. Филогенетические реконструкции гаплотипов осуществлялись методом максимального правдоподобия (ML) и байесовским подходом (BI). Построение филогенетических деревьев ML проводили в программе MEGA, а BI был выполнен в программе MrBayes 3.1.2 [31]. Для оценки достоверности ветвлений филогенетического дерева, построенного по методу ML, проводилось 500 репликаций [32]. При BI-анализе создавали три миллиона генераций цепей Маркова, отбирая пробы каждые 100 генераций, т. е. 30000 проб. Первые 25% проб исключались из анализа. В программе Tracer 1.7.1 визуально оценивали сближение марковских цепей Монте-Карло (MCMC) к стационарному распределению [33]. Все значения эффективного размера выборки (ESS) были больше 200.

Сеть гаплотипов и поиск деревьев максимальной экономии (MP) анализировали методом усредненного объединения Median-Joining (MJ) в программе Network 10.2.0.0 [34]. Первичные последовательности клонов фрагмента *ND5-cytb* и его ядерных копий депонированы в Генбанк под

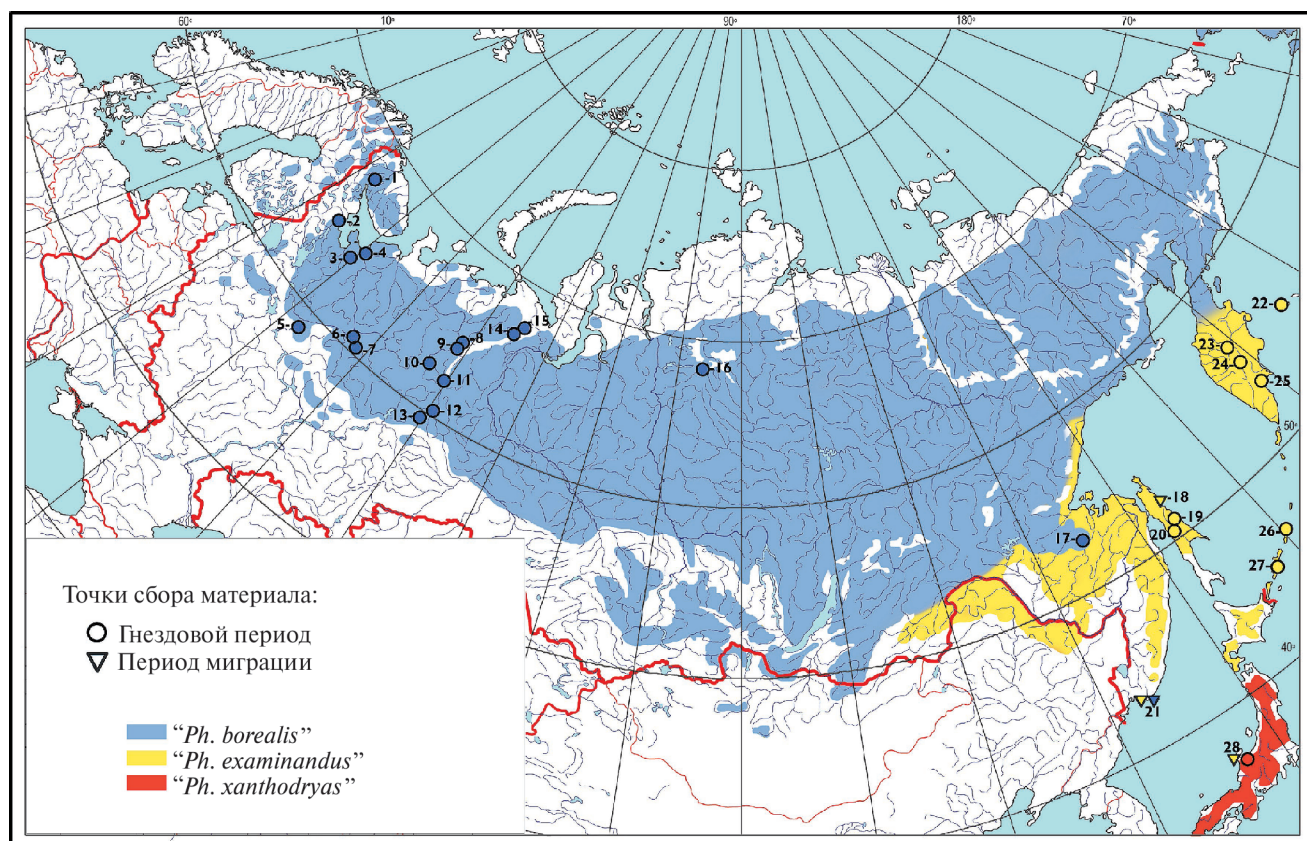


Рис. 1. Ареал и места сбора образцов *Ph. borealis* s. l. Номера точек соответствуют материалам, представленным в табл. 1.

номерами доступа для митохондриального фрагмента OR890288–OR890367 и для его ядерных копий OR890368–OR890425.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Изменчивость последовательностей фрагмента генов *ND5*–*cytb*

В результате анализа для 80 особей пеночки-таловки получены первичные последовательности митохондриального генома фрагмента генов *ND5*–*cytb* длиной около 1.5 тпн. У 58 образцов секвенирование фрагментов показало наличие двух значительно дифференцированных гаплотипов, один из которых являлся таксон-специфичным митохондриальным гаплотипом, а второй ядерной копией мтДНК (рис. 2). Одна птица (IN17880), отловленная и окольцованная в Японии в период осенней миграции, по результатам секвенирования имела отличающийся от всех остальных набор двойных пиков. Клонирование ее фрагмента обнаружило присутствие двух вариантов последовательностей, один из которых оказался гаплотипом *xanthodryas*, а второй – копией мтДНК, аналогичной таковым

borealis. В свою очередь, гнездящаяся особь (R-34870) из окрестностей пос. Порог (Респ. Коми, Троицко-Печорский р-н) в секвенированной последовательности имела сходные двойные пики, представленные на рис. 2.

Изменчивость митохондриальных гаплотипов у всех форм оказалась одинаково высокой. Число переменных сайтов для всех особей составило 181 (общее с NUMT – 277), из них 109 (общее с NUMT – 154) парсимони-информативных и 72 (общее с NUMT – 123) – единичные замены. Скорости замен транзиций и трансверсий, выраженные в процентном соотношении, приведены в табл. 2. Транзиции G/A и T/C внесли максимальный вклад в нуклеотидное разнообразие митохондриальной ДНК и ее ядерных копий. В несколько раз чаще, по сравнению с другими трансверсиями, встречались G/C-замены. Нуклеотидное разнообразие имело низкие значения и значительно варьировало, снижаясь в ряду *examinandus*–*borealis*–*xanthodryas* (табл. 3)

Полиморфные митохондриальные сайты *borealis*, *examinandus* и NUMT представлены на рис. 3 (а, б, в соответственно). В выборке *borealis* отсутствовала четко выраженная популяционная

Таблица 1. Образцы *Ph. borealis* s. l., использованные в анализе

Номер экзем- пляра	Форма	Место сбора	Номер точки на карте	Номер доступа Генбанка (мтДНК)	Номер доступа Генбанка (NUMT)
R-41117	borealis	РФ, Мурманская обл., Мончегорский р-н, дол. р. Нявки	1	OR890317	OR890391
R-41119	«	»	1	OR890318	OR890392
R-84460	«	Карелия, Кемский р-н, окр. пос. Кузема	2	OR890322	OR890396
R-97928	«	РФ, Архангельская обл., Плесецкий р-н, 40 км к югу от Архангельска	3	OR890319	OR890393
R-115722	«	РФ, Архангельская обл., Приморский р-н, Двинская губа, о. Мудьюг	4	OR890320	OR890394
R-115723	«	»	4	OR890321	OR890395
R-98720	«	РФ, Костромская обл., Чухломский р-н, окр. д. Рагозино	5	OR890315	OR890389
R-98721	«	»	5	OR890316	OR890390
ZMMU 612	«	РФ, Кировская обл., Подосиновский р-н	6	OR890290	
ZMMU 611	«	РФ, Кировская обл., Опаринский р-н	7	OR890289	OR890368
ZMMU 174	«	»	7	OR890288	OR890369
R-126924	«	РФ, Кировская обл., Опаринский р-н, окр. пос. Латышский	7	OR890291	OR890370
4204-1	«	»	7	OR890293	
5533-1	«	»	7	OR890292	OR890371
R-13109	«	РФ, Респ. Коми, гор. окр. Вуктыл, р. Шугор, окр. пос. Усть-Паток	8	OR890302	OR890379
R-13106	«	РФ, Респ. Коми, гор. окр. Вуктыл, р. Печора, дер. Мича-Бичевник	9	OR890301	OR890378
R-34868	«	РФ, Респ. Коми, Троицко-Печорский р-н, окр. пос. Порог	10	OR890303	
R-34870	«	»	10	OR890304	OR890380; OR890381
R-34871	«	»	10	OR890305	
R-34874	«	»	10	OR890306	
R-34875	«	»	10	OR890307	
R-34876	«	»	10	OR890308	OR890382
R-66106	«	РФ, Респ. Коми, Троицко-Печорский р-н, окр. г. Медвежий Камень	11	OR890309	OR890383

Таблица 1. (продолжение)

Номер экзем- пляра	Форма	Место сбора	Номер точки на карте	Номер доступа Генбанка (MT ДНК)	Номер доступа (NU MT)
R-131414	«	РФ, Свердловская обл., Ивдельский р-н, окр. г. Ивдель	12	OR890294	OR890372
R-131419	«	»	12	OR890295	OR890373
R-131421	«	»	12	OR890296	
2013-1	«	РФ, Свердловская обл., гор. окр. Карпинск, окр. пос. Кытлым	13	OR890298	OR890375
2012-1	«	»	13	OR890297	OR890374
5531-1	«	»	13	OR890299	OR890376
5532-1	«	»	13	OR890300	OR890377
R-116350	«	РФ, Респ. Коми, гор. окр. Воркута, окр. пос. Елецкий, р. Лёк-Елец	14	OR890310	OR890384
R-116351	«	»	14	OR890311	OR890385
R-116352	«	РФ, Респ. Коми, гор. окр. Воркута, хр. Енганэпэ, гора Южная	15	OR890312	OR890386
R-116353	«	»	15	OR890313	OR890387
R-115724	«	РФ, Респ. Коми, гор. окр. Воркута, хр. Енганэпэ, ручей Изъявож	15	OR890314	OR890388
R-123732	«	РФ, Эвенкийский авт. окр., Илимпиский р-н, устье р. Тайменья	16	OR890323	
R-123733	«	»	16	OR890324	
MIA 327	«	РФ, Хабаровский край, Верхнебуреинский р-н, хр. Эзоп	17	OR890325	
VF17218	<i>examinandus</i>	РФ, о. Сахалин, Ногликский р-н, залив Чайво	18	OR890337	OR890422
VF17223	«	»	18	OR890336	OR890423
K48-13	«	РФ, о. Сахалин, окр. Луньского зали- ва, гора Пик (480 м)	19	OR890338	OR890424
RYA-3202	«	РФ, о. Сахалин, Ногликский р-н, р. Плелярна	19	OR890343	OR890401
RYA-2888	«	РФ, о. Сахалин, Ногликский р-н, долина р. Набиль	19	OR890341	OR890399
RYA-2892	«	РФ, о. Сахалин, Тымовский р-н, пойма р. Малой Тыми (правый берег)	19	OR890342	OR890400
RYA-2878	«	РФ, о. Сахалин, Тымовский р-н, лев. берег р. Пиленги, ур. Пастушье	19	OR890340	OR890398
RYA-2874	«	РФ, о. Сахалин, Тымовский р-н, лев. бер. р. Тыми, окр. пос. Тымовское	19	OR890339	OR890397
VF57191	«	РФ, о. Сахалин, Тымовский р-н, пос. Зональное	20	OR890346	OR890404
VF57186	«	»	20	OR890344	OR890402
VF57190	«	»	20	OR890345	OR890403
VN85096	мигрант	Приморский край, р. Средняя Литовка	21	OR890326	OR890408; OR890407

Таблица 1. (окончание)

Номер экзем- пляра	Форма	Место сбора	Номер точки на карте	Номер доступа Генбанка (MT ДНК)	Номер доступа (NU MT)
VN85047	«	РФ, Приморский край, р. Литовка, приток Новорудная	21	OR890327	
VN85077	«	»	21	OR890329	
VN85138	«	»	21	OR890328	OR890409
M72	«	»	21	OR890330	
VN85359	«	»	21	OR890333	
VN85342	«	»	21	OR890334	OR890405
VN85382	«	»	21	OR890335	OR890406
VF17419	«	»	21	OR890331	
VF58089	«	»	21	OR890332	
CBH5193	<i>examinandus</i>	РФ, Командорские о-ва, о. Беринга	22	OR890353	OR890415
CBH5233	«	»	22	OR890352	OR890414
RYA-3543	«	РФ, Камчатский край, Быстринский р-н, окр. пос. Эссо	23	OR890348	OR890413
RYA-3372	«	РФ, Камчатский край, Милюковский р-н, пос. Милюково	24	OR890351	OR890410
RYA-3429	«	»	24	OR890349	OR890411
RYA 3453	«	»	24	OR890347	OR890412
RYA-3334	«	РФ, Камчатский край, Елизовский р-н, пойма р. Авачи	25	OR890350	
RYA-4084	«	РФ, о. Уруп, ур. Компанейское у бухты Новокурильская	26	OR890358	
RYA-4106	«	»	26	OR890359	
RYA-3948	«	РФ, о. Итуруп, берег бухты Касатка	27	OR890355	OR890417
RYA-4016	«	»	27	OR890357	OR890418
RYA-3940	«	»	27	OR890354	OR890416
RYA-3959	«	»	27	OR890356	OR890425
5-xanth	<i>xanthodryas</i>	Я., о. Хонсю, преф. Тояма, ст. Фучу	28	OR890360	OR890419
6-xanth	«	»	28	OR890361	OR890420
IH17880	«	»	28	OR890362	OR890421
IH17886	<i>examinandus</i>	»	28	OR890365	
IH17878	«	»	28	OR890363	
IH17892	«	»	28	OR890366	
1G86030	«	»	28	OR890364	
1G83034	«	»	28	OR890367	

Примечание. РФ – Российская Федерация, Я. – Япония.

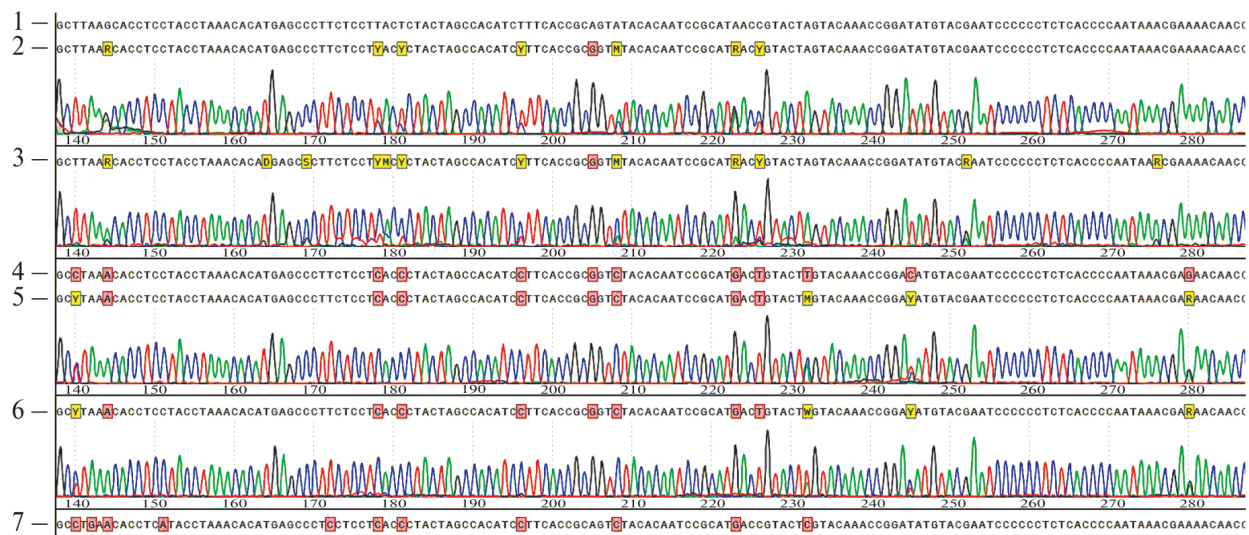


Рис. 2. Участок гена ND5 мтДНК форм *Ph. borealis* s. l.: 1 – *xanthodryas* (Генбанк AB362465), 2 – *xanthodryas* (IH17880), 3 – *borealis* (R-34870), 4 – *examinandus* (Генбанк AB362424), 5 – *examinandus* (VF57191), 6 – *examinandus* (RYA-3202), 7 – *borealis* (Генбанк AB362462). Красным цветом отмечены сайты, отличающиеся от последовательности *xanthodryas*. Сайты с двойными пиками показаны желтым цветом, с буквенным обозначением транзиций (R, Y) и трансверсий (D, M, S, W).

генетическая структура. Исключением явились мигрирующие птицы, отловленные в Приморье, гаплотипы которых отличались от западных тремя

Таблица 2. Значения нуклеотидных замен (в %), посчитанные по модели максимального правдоподобия (Maximum Composite Likelihood)

с\на	A	T	C	G
A	—	0.80	1.24	17.05
T	1.07	—	21.25	0.65
C	1.01	20.6	—	1.55
G	37.30	1.05	4.19	—

Примечание. Полужирным шрифтом обозначены транзиции, курсивом – трансверсии.

Таблица 3. Гаплотипическое (*h*), нуклеотидное (π) и внутривидовое (*p*-dist) разнообразие митохондриального фрагмента ND5–cytb и его ядерных копий *Ph. borealis* s. l

Формы (N/H)	h	π	<i>p</i> -dist
<i>borealis</i> (47/42)	0.99	0.00898	0.004
<i>examinandus</i> (29/29)	1.00	0.01310	0.010
<i>xanthodryas</i> (4/4)	1.00	0.00548	0.005
NUMT (58/58)	1.00	0.01929	0.020

Примечание. N – количество образцов, H – количество гаплотипов.

фиксированными заменами. Близким к ним оказался мт-гаплотип магаданской особи, взятой из Генбанка (AB362462).

Наиболее дивергированными митохондриальными гаплотипами обладали птицы из Кытлыма (Свердловская обл.) (2012-1, 5531-1 и 5532-1), Карелии (R-84460) и Костромской области (R-98721). Только два гаплотипа *borealis* встречены повторно, один у птиц из Респ. Коми (R-13106, R-13109, R-34876) и второй из Респ. Коми (R-34871), Мурманской обл. (R-41117 и R-41119) и Эвенкии (R-123732). Ядерные копии мтДНК обладали максимальным нуклеотидным разнообразием, у них выявлен 221 полиморфный сайт: для *examinandus* и *borealis* – 83 и 84 соответственно, для *xanthodryas* – 19. Из 58 последовательностей NUMT только у восьми выявлены стоп-кодоны митохондриальной ДНК (AGA, AGG).

Нуклеотидные последовательности *examinandus*, в отличие от *borealis*, разделились на две филогенетические линии, различающиеся между собой по более чем 30 фиксированным сайтам, однако не имеющие строгой географической привязки (рис. 3,б). Возможно, данный факт отражает существование в составе группировки таловок с “медленной” песней двух географических рас: *Ph. b. examinandus* и упомянутой выше *Ph. b. hylebata*. Выборка *xanthodryas* из-за малого количества исследованных образцов на рисунке не представлена.

Последовательности ядерных копий мтДНК, несмотря на их высокую гетерогенность, разделились на две филогруппы, которые мы обозначили как NUMT I (западная) и NUMT II (восточная)

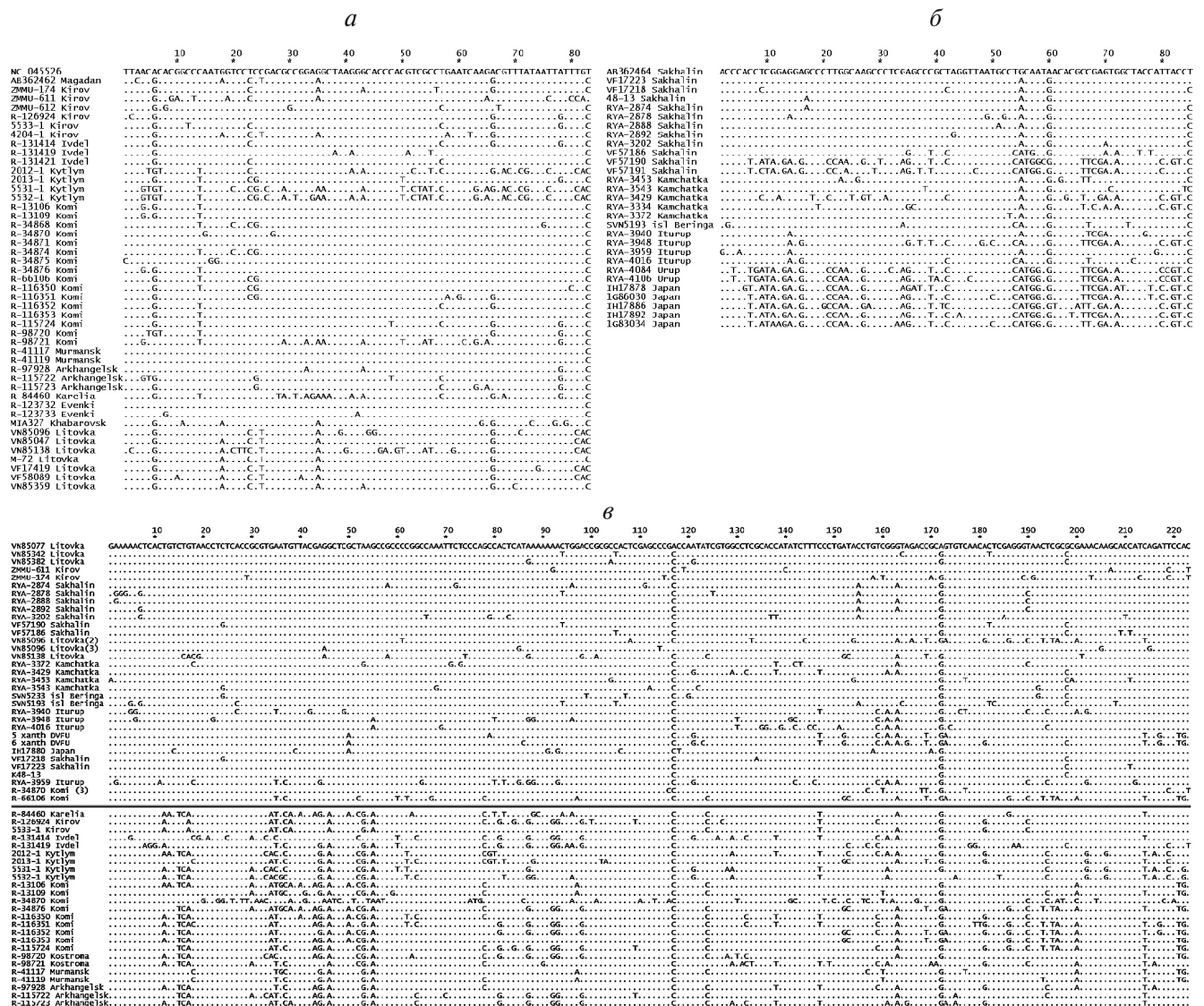


Рис. 3. Полиморфные сайты фрагмента *ND5-cytb* и его ядерных копий *Ph. borealis* s. l.: *borealis* (a), *examinandus* (б) и NUMT (чертой разделены филогруппы NUMT II (вверху) и NUMT I (внизу)) (в).

(рис. 3, в). В первую вошли в основном *borealis* из западной части ареала и часть мигрирующих птиц из Приморья, а также, что особенно важно, две японские *xanthodryas* из музея ДВФУ. Ядерные копии мтДНК *xanthodryas* оказались близки митохондриальным гаплотипам птиц из Кытлыма ($p = 0.018$). Такие же значения ($p = 0.02$) получены между митохондриальными гаплотипами *examinandus* и ядерными копиями мтДНК западных *borealis*. В линии NUMT II объединились в основном ядерные копии мтДНК представителей восточной части ареала и некоторых особей из западных территорий. Кроме того, в нее вошли последовательности неясной локализованности (ядерной или

митохондриальной), поскольку митохондриальные гаплотипы, недавно возникшие из ядерных копий, имеют с ними одинаковые последовательности и не могут быть адекватно идентифицированы.

В ядерных копиях мтДНК исследуемого фрагмента двух образцов с Камчатки (RYA-3453(2)) и Сахалина (VF17223(2)) обнаружена делеция триплета TCA, кодирующего аминокислоту Ser₍₁₄₀₎. Такая же делеция присутствует в гене *cytb* мтДНК у образцов с островов Сахалин (VF57186) и Уруп (RYA-4084, RYA-4106) и птиц, окольцованных в Японии в период осенней миграции (IH17878, 1G86030).

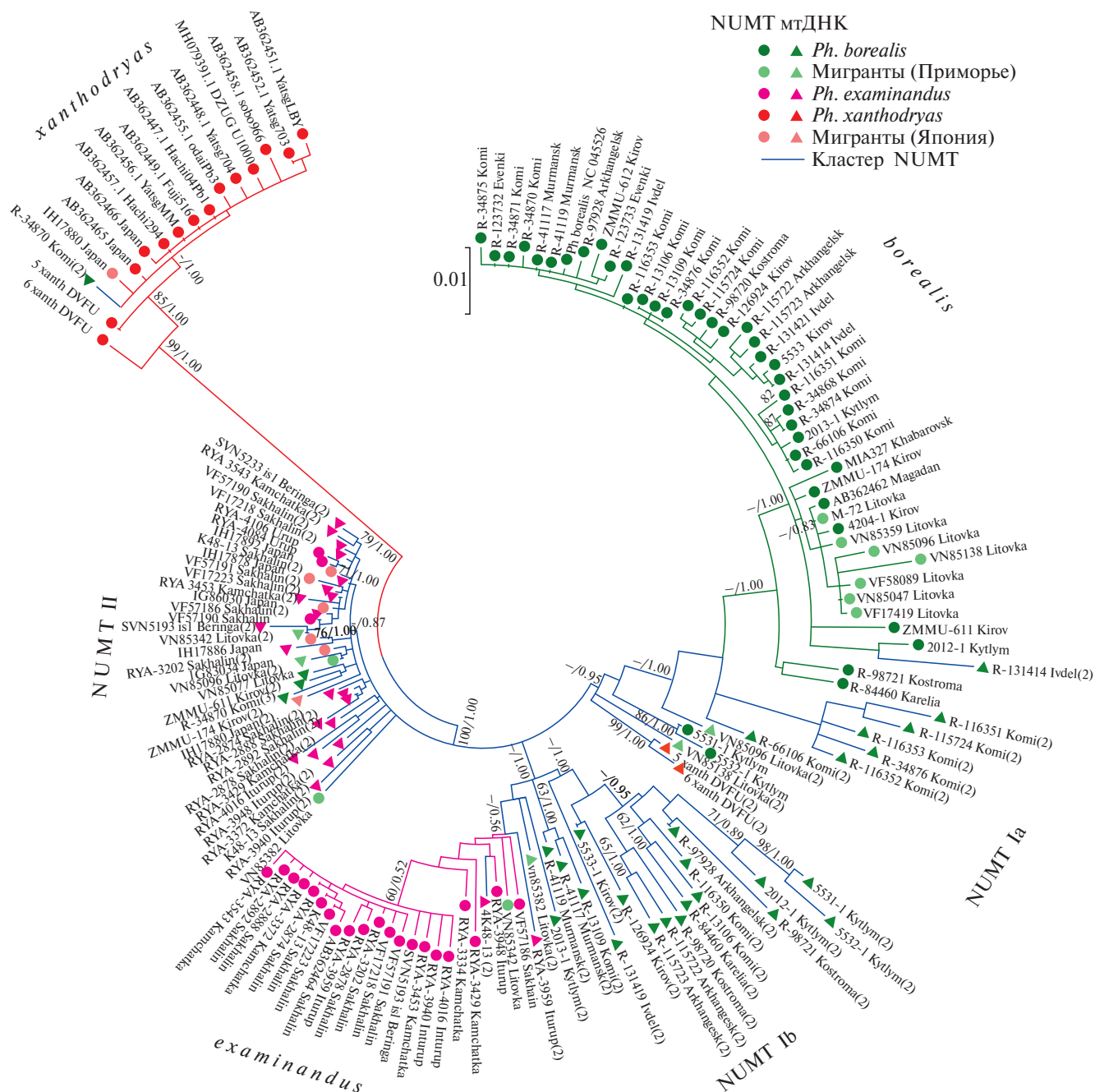


Рис. 4. Филогенетическое дерево последовательностей мтДНК фрагмента *ND5-cytb* и его ядерных копий *Ph. borealis* s. l., построенное по методу ML. В узлах ветвлений указаны значения поддержки (ML/BI $\geq 50\%$). Вверху дерева указан масштаб с длиной ветви, измеряемой количеством замен на сайт.

Дивергенция мтДНК в парах *borealis-examinandus*, *borealis-xanthodryas* и *examinandus-xanthodryas* значительно превышала вышеописанные значения ($p = 0.035$, 0.044 и 0.046 соответственно), что согласуется с литературными данными по этим формам (3.8–5.1%) [16].

Филогенетические связи митохондриальных гаплотипов и NUMT

Для построения реконструкции связей мт-гаплотипов и их ядерных копий использованы метод максимального правдоподобия (модель замен GTR+G+I, $bic = 22168.443$) и байесовский

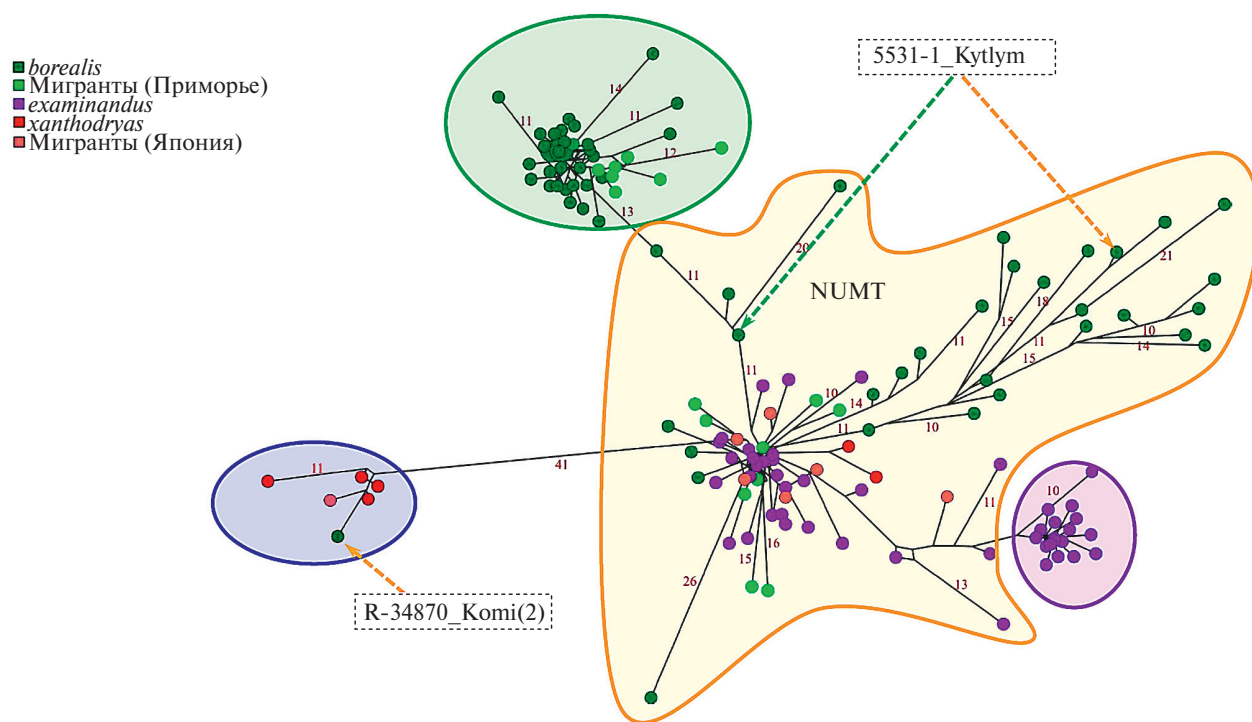


Рис. 5. Филогенетическая сеть последовательностей мтДНК фрагмента *ND5–cytb* *Ph. borealis* s. l. и его ядерных копий, построенная по методу Median Joining. На ветвях красным цветом указано количество замен (≥ 10). Зеленая пунктирная стрелка указывает мт-гаплотип, желтая — ядерную копию мтДНК.

подход. На обеих реконструкциях представлены четыре кластера, три из которых были существенно дифференцированы и соответствовали митохондриальным гаплотипам *borealis*, *examinandus* и *xanthodryas* (рис. 4). Четвертый кластер оказался наиболее гетерогенным, мы называем его супер-кластером по причине большого количества объединенных в нем последовательностей. Он включил не только все ядерные копии мтДНК, но и митохондриальные гаплотипы *examinandus*, которые обособились отдельной филогруппой. Условно супер-кластер разделился на три кластера, два из которых (NUMT Ia и NUMT Ib) соответствовали западной линии NUMT I и третий — восточной NUMT II. Часть NUMT отдельными длинными ветвями заняла промежуточное положение между митохондриальным и NUMT I кластерами.

Кластер *borealis* имел характерную для митохондриальных генов гребенчатую структуру с короткими внутренними ветвями и объединил большую часть западных таловок. В его состав обособленно вошла часть птиц, мигрирующих через Приморье вместе с одной особью из Магадана (AB362462). Ядерные копии мтДНК этих птиц объединились в западном кластере NUMT I. Интересным оказался факт присутствия у мигрирующей особи (K-25 VN85096) с митохондриальным гаплотипом *borealis* разных ядерных копий мтДНК, отнесенных

к кластерам NUMT I и NUMT II, что может указывать на ее гибридное происхождение *borealis* × *examinandus*. Значительно дивергированными оказались некоторые митохондриальные последовательности нескольких птиц — одной из Кытлыма (2012-1), объединившейся с ядерной копией мтДНК особи из Ивделя (R-131414) (обе из Свердловской обл.), и двух птиц из Костромской области (R-98721) и Карелии (R-84460). Их ядерные копии мтДНК вошли в удаленную ветвь кластера NUMT I. Заслуживает внимания тот факт, что митохондриальные гаплотипы двух других птиц из Кытлыма (5531-1 и 5532-1) также принадлежат к кластеру ядерных копий NUMT I. При этом их собственные ядерные копии мтДНК сгруппированы в другой ветви этого же кластера, однако на значительном удалении от собственных митохондриальных гаплотипов. Такое смешение митохондриальных гаплотипов с ядерными копиями на филогенетической реконструкции указывает на близкое родство этих последовательностей и позволяет сделать предположение об источнике возникновения первых.

Структура митохондриального кластера *examinandus*, объединившего птиц Камчатки, Командорских островов (о. Беринга), Сахалина и Итурупа, была сходна с предыдущим и имела монофилетическое происхождение. Он вошел

в состав кластера с ядерными копиями мтДНК *borealis* NUMT I. Интересно, что митохондриальные гаплотипы двух птиц с о. Уруп объединились с ядерными копиями более гетерогенного по таксономическому составу восточного кластера NUMT II. У одной птицы (VN85342), мигрировавшей в Приморье, обнаружен митохондриальный гаплотип *examinandus*, что подтверждает миграционные пути части *examinandus* через материк. Все ядерные копии мт-гаплотипов этих особей оказались в восточном кластере NUMT II.

Наиболее дифференцированной оказалась форма *xanthodryas*, которая отделилась от суперкластера ядерных копий, с максимальным количеством мутаций (более 40 замен). Два образца *xanthodryas* из музея ДВФУ нам удалось проанализировать на предмет ядерных копий мтДНК, которые оказались близки к митохондриальным последовательностям образцов из Кытлыма, из западной части ареала *borealis*. В свою очередь, одна из ядерных копий мтДНК образца (R-34870) из Республики Коми оказалась очень сходной с митохондриальными последовательностями *xanthodryas*.

Структура суперкластера NUMT, разделенного на три филогруппы NUMT Ia, NUMT Ib и NUMT II, значительно отличается от структуры кластеров, включающих митохондриальные гены. В нем преобладают длинные внутренние ветви, характерные для некодирующих ядерных генов. Кластер NUMT I представлен несколькими гетерогенными группами и занимает промежуточное положение между кластерами *borealis* и NUMT II. Он объединил не только ядерные копии мтДНК западных и восточных *borealis* и двух японских *xanthodryas*, но, что самое неожиданное, в него вошли некоторые мт-гаплотипы западных *borealis* и восточных *examinandus*.

Кластер NUMT II располагается между кластерами NUMT I и *xanthodryas*. Структура этого кластера характеризуется смешанным типом ветвления, представленным длинными ветвями, относящимися к ядерным копиям мтДНК, и короткими ветвями, типичными для митохондриальных генов (см. рис. 4). Наряду с ядерными копиями мтДНК в нем хаотично объединились также некоторые мт-гаплотипы *examinandus*. В этот же кластер вошли ядерные копии двух птиц из Кировской области (ZMMU-611 и ZMMU-174). Необходимо особо отметить базовое положение суперкластера NUMT для кластеров митохондриальных гаплотипов. Его центральное расположение подтверждает построенная в программе NetWork сеть гаплотипов (рис. 5). Из сети гаплотипов были исключены взятые из Генбанка десять последовательностей гена *cytb* *xanthodryas* по причине недостающей части фрагмента гена *ND5*. Количество замен между суперкластером NUMT и остальными филогруппами указано на ветвях сети. Важно отметить, что

некоторые ядерные копии мтДНК имели более 40 замен с центральным кластером, что указывает на скрытый в ядерном геноме огромный резерв разнообразия митохондриальных копий.

Интересным явился факт присутствия в кластере *xanthodryas* одной из ядерных копий, принадлежавшей особи *borealis* из западной части ареала (Респ. Коми, R-34870). В свою очередь, ядерные копии мтДНК двух *xanthodryas* (5-xanth, 6-xanth) сгруппировались в кластере NUMT I с мигрирующим *borealis* (VN85138) из Приморья, митохондриальный гаплотип которого сходен с магаданским гаплотипом из Генбанка (AB362462). Такое перекрестное расположение мт-гаплотипов и ядерных копий косвенно указывает на произошедший обмен между ядерным и митохондриальным геномами. Несмотря на обособленность каждого митохондриального кластера, базовым для всех форм оказался суперкластер с ядерными копиями мтДНК.

ОБСУЖДЕНИЕ

Выбор фрагмента митохондриального генома *ND5-cytb* для анализа и сравнения его ядерных копий был определен нашим предыдущим исследованием [5]. Данные гены пользуются широкой популярностью в филогенетических исследованиях, и их последовательности в большом количестве представлены в разных генетических банках. Например, для гена *cytb* птиц прочитано около 56000 последовательностей, а для гена *ND5* — около 8000. Необходимо отметить, что около 30% этих последовательностей имеют ошибки разного характера, включая ядерные копии гена *cytb* мтДНК, на которые приходится до 1% ошибок [35]. Это свидетельствует о том, что не все ядерные копии митохондриальных генов по характерным признакам (отсутствие/присутствие терминирующих кодонов, наличие инделей, приводящих к сдвигу рамки считывания, и повышенный полиморфизм последовательностей за счет несинонимичных замен) могут быть отнесены к псевдогенам. Результаты настоящего исследования митохондриальных генов и их ядерных копий и ранее опубликованных работ [3–5] показывают в большинстве случаев полную функциональность ядерных копий, и при переходе из ядерного генома в митохондриальный они становятся полноценными кодирующими генами. Из 58 ядерных копий, представленных в данной работе, только у восьми обнаружены преждевременные терминирующие кодоны митохондриальной ДНК. Напомним, что стоп-кодоном копии митохондриального гена в ядерном геноме является TGA-триплет, кодирующий в митохондриальном геноме аминокислоту триптофан [3]. Поэтому не требуется каких-либо случайных мутаций и длительного

времени, чтобы ген стал “молчащим” при переходе в ядерный геном и наоборот.

Все таксон-специфичные и большинство уникальных замен митогенома исследованных образцов были обнаружены в его ядерных копиях, включая уникальную делецию ТСА-триплета, кодирующего аминокислоту серин Ser₍₁₄₀₎. Данный факт дополнительно подтверждает родство последовательностей, полученных из ядерного и митохондриального геномов. Кроме того, по таким маркерным сайтам можно также устанавливать родственные отношения и соответственно популяционную принадлежность птиц в периоды миграций и зимовки. Еще раз обращаем внимание на то, что митохондриальные последовательности *examinandus* объединились в одном кластере с ядерными копиями мтДНК *borealis*. У японской *xanthodryas* (IN17880) и западной *borealis* (Респ. Коми, R-34870) выявлены перекрестные ядерные копии мтДНК. Сходная ситуация была описана у других видов воробьиных *Calliope calliope* (Pallas, 1776) [3] и *C. pectoralis* Gould, 1837 [5].

Полученные данные указывают на то, что источником митохондриальных замен, т. е. “эволюции” таксон-специфичных гаплотипов мтДНК, в данном случае являются мутации в ядерных копиях мт-генов. Причем межгеномное рекомбинационное событие, происходящее в гоноцитах [3], предшественниках половых клеток птиц [36], предоставляет новый вариант гаплотипа с уже готовыми многими специфичными сайтами, передающимися по наследству сразу многим потомкам. Обнаруженные факты значительно сокращают время дивергенции мт-гаплотипов для анализируемых форм, чем это предполагалось ранее [18, 20], а “молекулярные часы”, не учитывающие рекомбинационных событий между ядерным и митохондриальным геномами, в этом случае не могут быть использованы. Появление новых филогрупп мтДНК обычно связывают с событиями Ледникового периода, климатическими колебаниями плейстоцена и рефугиумами, прохождением популяциями “бутылочного горлышка”. Принимая во внимание как описанные ранее факты по другим видам, так и представляемую работу, мы полагаем, что гомологичная рекомбинация между ядерным и митохондриальным геномами, являющаяся уникальным событием, раскрывает причину монофилии и глубокой дивергенции современных митохондриальных филогрупп *Ph. borealis* s. l. без всех выше описанных условий.

Сравнение митохондриальных генов и их ядерных копий может прояснить происхождение таксон-специфичных митохондриальных гаплотипов, как показано нами ранее [5, 8]. Поскольку все модели филогенетических реконструкций для мт-гаплотипов основаны на постепенном накоплении мутаций в самом митогеноме, рекомбинационные

события, случающиеся между ядерным и митохондриальным геномами с резкой сменой “старого” гаплотипа на более дифференцированный “новый”, не укладываются в эти модели. NUMTs практически не используются для построения филогенетических реконструкций, однако в ряде случаев анализ их изменчивости позволил провести реконструкцию филогенетических отношений, как, например, у полевок рода *Microtus* [37] и шимпанзе [38]. NUMTs являются быстроэволюционирующими видоспецифичными последовательностями и полезны для филогенетического анализа прежде всего в группах близкородственных видов и видов-двойников [39, 40].

Одним из примеров глубокой митохондриальной дивергенции при отсутствии ядерных отличий является пара *Ph. affinis* (Tickell, 1833) — *Ph. occisinensis* Martens et al., 2008. Выделение *occisinensis* в отдельный вид [41] основано главным образом на значительной митохондриальной дивергенции. Анализ ядерных геномов (SNP), наряду с биоакустическими и морфологическими признаками, показал высокое сходство этих таксонов [13]. В то же время результаты секвенирования митохондриальных геномов продемонстрировали их глубокую дивергенцию, которую объясняют экспансией формы *affinis* на запад, где она заменила ныне “вымершую” родственную форму, считавшуюся донором ее мтДНК и незначительной части ее ядерного генома. Причина замены собственного митохондриального гаплотипа *affinis* на чужой осталась загадкой. В связи с этим Д. Чжан с соавт. предлагают новый термин — “призрачная интрогрессия” и предлагают ее в качестве причины глубокой дивергенции митохондриальных гаплотипов *Ph. affinis* и *Ph. occisinensis* [13]. Время расхождения *occisinensis* и *affinis* по митохондриальному гену *цитохрома b* П. Альстрем с соавт. датируют более 4 млн лет, а для форм *borealis*, *examinandus* и *xanthodryas* — периодом в 2.5–3 млн лет [42]. Однако, как показывают наши исследования, все таксон-специфичные митохондриальные гаплотипы существующих в настоящее время форм *Ph. borealis* представлены в его ядерных копиях мтДНК.

Скрытые межгеномные рекомбинационные процессы, случающиеся в организме, можно увидеть по косвенным признакам, анализируя значительный материал не только мтДНК, но и ее ядерных копий. Если в анализе по митохондриальным маркерам присутствуют длинные ветви с особями, отличающимися большим количеством мутаций от базового гаплотипа, есть большая вероятность обнаружения либо рекомбинационного события, либо ядерной копии мтДНК. Как показывает наше исследование, митохондриальные копии в ядерном геноме не всегда имеют признаки, отличающие их от ядерных псевдогенов. К примеру, мт-гаплотипы особей *borealis* из Кытлыма,

Костромской области и Карелии по их кластеризации на филогенетической реконструкции (см. рис. 4) можно было бы рассматривать в качестве ядерных копий мтДНК, если бы мы не обнаружили их собственные NUMT, которые оказались значительно дивергировавшими.

Разделение суперкластера NUMT на три филогруппы (NUMT Ia, NUMT Ib и NUMT II) можно объяснить несколькими, одновременными межгеномными рекомбинационными событиями, в результате которых новая ядерная копия мтДНК изменялась независимо от предыдущих. Высокая скорость мутаций, характерная для ядерных псевдогенов, обеспечила глубокую дифференциацию обнаруженных филогрупп. Ядерные копии мтДНК, пребывающие длительное время в ядерном геноме, могут существенно изменяться от собственных митохондриальных геномов. Объединение в одном кластере с ядерными копиями *borealis* гаплотипов *examinandus* указывает на источник их происхождения, причем собственные ядерные копии *examinandus*, продолжая “эволюционировать” в ядерном геноме, значительно дивергировали от первоначальной последовательности ядерной копии *borealis*. Структура кластеров на филогенетической реконструкции, построенной с использованием митохондриальных генов и их ядерных копий, поразительно различается. Кластеры, включающие митохондриальные последовательности, имеют характерную картину распределения, соответствующую так называемому принципу “эффекта основателя”, длинную монофилетическую ветвь с короткими внутренними гребенками.

Суперкластер NUMT в основном образован длинными ветвями за счет высокого нуклеотидного разнообразия ядерных копий мтДНК, что согласуется с литературными данными, указывающими на высокую степень изменчивости псевдогенов митохондриального происхождения, и противоречит представлению о них как о “молекулярных ископаемых” [7]. Принимая во внимание эти отличия, можно предположить, что в целом ряде исследований митохондриальных маркеров, результатом которых на филогенетических реконструкциях стало выявление высокой гетерогенности либо полифилетического происхождения, анализировались не собственно митохондриальные гены, а их ядерные копии.

Таким образом, все таксон-специфичные митохондриальные гаплотипы *borealis*, *examinandus* и *xanthodryas*, по нашему мнению, произошли из ядерных копий мтДНК одной формы, обитающей в настоящее время, в частности, на юге Предуралья (соответствующей подвиду *Ph. b. talovka*). Ядерный геном служит местом, где происходят “эволюция” митохондриального генома и накопление его генетической изменчивости. Обнаруженные варианты ядерных копий являются потенциальным

источником новых гаплотипов мтДНК и представляют собой скрытый молекулярный резерв мтДНК, который может быть реализован через межгеномную рекомбинацию. В то же время ядерный геном, кодирующий морфологические признаки, остается прежним, что объясняет слабую морфологическую дивергенцию форм при глубоких отличиях митохондриальных гаплотипов. Основываясь на наших выводах о происхождении митохондриальных гаплотипов форм *examinandus* и *xanthodryas* из ядерных копий мтДНК *borealis* и учитывая высокое сходство их ядерных геномов, мы рассматриваем митогеномы этих форм в качестве вариантов гаплотипа одного вида *Ph. borealis* s. l.

Авторы чрезвычайно признательны В.Н. Сотникову, предоставившему уникальные материалы собственных сборов из различных частей гнездового ареала исследуемого вида.

Работа выполнена в рамках Государственного задания Министерства науки и высшего образования Российской Федерации (тема № 121031500274-4 “Эволюционные аспекты формирования наземной биоты востока Азии”), а также гостемы: “Таксономический и биохорологический анализ животного мира как основа изучения и сохранения структуры биологического разнообразия” номер ЦИТИС-121032300105-0.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных. Был использован музейный материал, а не живые особи.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Все применимые международные, национальные и/или институциональные принципы ухода и использования животных были соблюдены.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Гребельный С.Д., Иванова Н.Ю., Нефедова Е.А. Обмен между ядерным и митохондриальными геномами (по результатам анализа ядерных и цитоплазматических копий митохондриальных генов – *numts* и *cymts*) // Цитология. 2018. Т. 60. № 11. С. 899–902. <https://doi.org/10.1134/S004137711811007X>
2. Sorenson M.D., Quinn T.W. Numts: A challenge for avian systematics and population biology. Auk. 1998. V. 115. P. 214–221.

3. Спиридонова Л.Н., Вальчук О.П., Редькин Я.А., Крюков А.П. Ядерные копии митохондриальных генов – источник новых гаплотипов гена цитохрома *b* мтДНК *Luscinia calliope* (Muscicapidae, Aves) // Генетика. 2016. Т. 52. № 9. С. 1069–1080. <https://doi.org/10.7868/S0016675816090137>
4. Спиридонова Л.Н., Вальчук О.П., Редькин Я.А. Новый случай рекомбинации между ядерным и митохондриальным геномами в роде *Calliope* Gould, 1836 (Muscicapidae, Aves): гипотеза происхождения *Calliope pectoralis* Gould, 1837 // Генетика. 2019. Т. 55. № 1. С. 81–93. <https://doi.org/10.1134/S0016675819010132>.
5. Спиридонова Л.Н., Вальчук О.П. Полный митохондриальный геном *Phylloscopus eximius* и гипотеза его происхождения // Генетика. 2022. Т. 58. № 3. С. 362–366. <https://doi.org/10.31857/S0016675822030146>
6. Triant D.A., DeWoody J.A. Demography and phylogenetic utility of numt pseudogenes in the southern red-backed // J. Mammalogy. 2009. V. 90. № 3. P. 561–570. <https://doi.org/10.1644/08-MAMM-A-149R1.1>
7. Романов Д.А., Андрианов Б.В. Митохондриальные последовательности в ядерном геноме животных // Успехи соврем. биологии. 2013. Т. 133. № 3. С. 254–268.
8. Спиридонова Л.Н., Вальчук О.П., Редькин Я.А. и др. Филогеография и демографическая история соловья-красношейки *Luscinia calliope* // Генетика. 2017. Т. 53. № 8. С. 933–951. <https://doi.org/10.7868/S0016675817080100>
9. Волков В.А., Григорьева Е.А., Лебедева М.В., Потомкина Е.К. Гетероплазмия и ядерные копии митохондриальных генов (NUMTs), выявленные в зоне интрогрессивной гибридизации ели европейской и ели сибирской // Тр. Санкт-Петербургского науч.-исслед. ин-та лесного хозяйства. 2022. № 1. С. 4–19. DOI: 10.21178/2079-6080.2022.1.4
10. Fok K.W., Wade C.M., Parkin D.T. Inferring the phylogeny of disjunct populations of the azurewinged magpie *Cyanopica cyanus* from mitochondrial control region sequences // Proc. R. Soc. Lond. B. 2002. V. 269. P. 1671–1679. <https://doi.org/10.1098/rspb.2002.2057>
11. Kryukov A., Iwasa M.A., Kakizawa R. et al. Synchronic east-west divergence in azure-winged magpies (*Cyanopica cyanus*) and magpies (*Pica pica*) // J. Zool. Syst. Evol. Res. 2004. V. 42. P. 342–351. <https://doi.org/10.1111/j.1439-0469.2004.00287.x>
12. Kryukov A., Spiridonova L., Mori S. et al. Deep phylogeographic breaks in magpie *Pica pica* across the Holarctic: concordance with bioacoustics and phenotypes // Zool. Sci. 2017. V. 34. P. 185–200. <https://doi.org/10.2108/zs160119>
13. Zhang D., Tang L., Cheng Y., Hao Y., et al. “Ghost introgression” as a cause of deep mitochondrial divergence in a bird species complex // Mol. Biol. Evol. 2019. V. 36. № 11. P. 2375–2386. <https://doi.org/10.1093/molbev/msz170>
14. Коблик Е.А., Редькин Я.А., Архипов В.Ю. Список птиц Российской Федерации. М.: Тов-во науч. изданий КМК, 2006. 281 с.
15. Глуценко Ю.Н., Нечаев В.А., Редькин Я.А. Птицы Приморского края: краткий фаунистический обзор. М.: Тов-во науч. изданий КМК, 2016. 523 с.
16. Reeves A.B., Drovetski S.V., Fadeev I.V. Mitochondrial DNA data imply a stepping-stone colonization of Beringia by arctic warbler *Phylloscopus borealis* // J. Avian Biol. 2008. V. 39. P. 567–575. <https://doi.org/10.1111/j.0908-8857.2008.04421.x>
17. Saitoh T., Shigeta Y., Ueda K. Morphological differences among populations of the Arctic Warbler with some intraspecific taxonomic notes // Ornith. Sci. 2008. V. 7 (2). P. 135–142. <https://doi.org/10.2326/1347-0558-7.2.135>
18. Saitoh T., Alström P., Nishiumi I. et al. Old divergences in a boreal bird supports long-term survival through the Ice Ages // BMC Evol. Biology. 2010. V. 10 (1). № 35. <https://doi.org/10.1186/1471-2148-10-35>
19. Saitoh T., Nishiumi I., Shigeta Y., Ueda K. Re-examination of the taxonomy of the Arctic Warbler *Phylloscopus borealis* (Blasius): Three separate species withing the *Phylloscopus [borealis]* subspecies // Japan. J. Ornithol. 2011. V. 61 (1). P. 46–59. <https://doi.org/10.3838/jjo.61.46>
20. Alström P., Saitoh T. et al. The Arctic Warbler *Phylloscopus borealis* – three anciently separated cryptic species revealed // Ibis. 2011. V. 153. P. 395–410.
21. Dickinson E.C., Christidis L. The Howard & Moore Complete Checklist of the Birds of the World: Passerines. 4th ed. V. 2. Eastbourne: Aves Press, 2014. 752 p.
22. Назаренко А.А. К орнитофауне Хэнтэй-Чикойского нагорья, Южное Забайкалье // Сб. Экология и зоогеография некоторых позвоночных суши Дальнего Востока. Владивосток, 1978. С. 40–56.
23. Лобков Е.Г. Гнездящиеся птицы Камчатки. Владивосток: ДВНЦ АН СССР, 1986. 290 с.
24. Red'kin Ya.A. Geographic variation and reproductive isolation in the eastern populations of arctic warbler // Avian migrants in the Northern Pacific: Breeding and Stopover sites in changing Earth. Scientific conf. Institute of Marine Geology and Geophysics FEB RAS: Abstracts. Yuzhno-Sakhalinsk, September 3–7, 2013. P. 11. /
25. Del Hoyo J., Collar N.J. (Eds). HBW and BirdLife International Illustrated Checklist of the Birds of the World: Passerines. Barcelona, Spain: Lynx Edicions, 2016. V. 2. 1013 p.
26. Bonfield J.K., Smith K.F., Staden R. A New DNA Sequence Assembly Program // Nucl. Acids Res. 1995. V. 23. P. 4992–4999.

27. Kumar S., Stecher G., Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets // *Mol Biol Evol.* 2016. V. 33 №7. P. 1870–1874.
<https://doi.org/10.1093/molbev/msw054>
28. Nei M., Kumar S. *Molecular Evolution and Phylogenetics*. N. Y.: Oxford Univ. Press, 2000. 333 p.
29. Librado P., Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data // *Bioinformatics*. 2009. V. 25. P. 1451–1452.
<https://doi.org/10.1093/bioinformatics/btp187>
30. Goldman N. Statistical tests of models of DNA substitution // *J. Mol. Evol.* 1993. V. 36. P. 182–198.
31. Huelsenbeck J.P., Ronquist F. MrBayes: Bayesian inference of phylogenetic trees // *Bioinformatics*. 2001. V. 17. P. 754–755.
<https://doi.org/10.1093/bioinformatics/17.8.754>
32. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap // *Evolution*. 1985. V. 39. P. 783–791.
33. Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7 // *Systematic Biology*. 2018. V. 67 (5). P. 901–904.
<https://doi.org/10.1093/sysbio/syy032>
34. Bandelt H.J., Forster P., Rohl A. Median-Joining networks for inferring intraspecific phylogenies // *Mol. Biol. Evol.* 1999. V. 16. P. 37–48.
35. Van den Burg M.P., Vieites D.R. Bird genetic databases need improved curation and error reporting to NCBI // *Ibis*. 2022.
<https://doi.org/10.1111/ibi.13143>
36. Стрижикова С.В., Стрижиков В.К., Житенко Н.В. Гистогенез яичников у птиц в пренатальном периоде онтогенеза // *Успехи соврем. естествознания*. 2002. Т. 4. С. 77–78.
37. Triant D.A., DeWoody J.A. Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents // *Genetica*. 2008. V. 132. № 1. P. 21–33.
<https://doi.org/10.1007/s10709-007-9145-6>
38. Hazkani-Covo E., Graur D. A comparative analysis of numt evolution in human and chimpanzee // *Mol. Biol. Evol.* 2007. V. 24. № 1. P. 13–18. <https://doi.org/10.1093/molbev/msl149>
39. Hazkani-Covo E. Nuclear sequences of mitochondrial origin as phylogenetic markers // *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd: Chichester, 2010.
<https://doi.org/10.1002/9780470015902.a0022877>
40. Андрианов Б.В., Романов Д.А., Горелова Т.В. и др. Перенос митохондриальной ДНК в ядерный геном клеток пересееваемой клеточной линии *Drosophila virilis* // *Генетика*. 2013. Т. 49. № 6. С. 788–792.
<https://doi.org/10.7868/S0016675813060027>
41. Martens J., Sun Y.-H., Packert M. Intraspecific differentiation of Sino-Himalayan bish-dwelling *Phylloscopus* leaf warblers, with description of two new taxa (*P. fuscatus*, *P. fuligiventer*, *P. affinis*, *P. armandii*, *P. subaffinis*) // *Vertebrate Zool.* 2008. V. 58. № 2. P. 233–265.
<https://doi.org/10.3897/vz.58.e30935>
42. Alström P., Rheindt F.E., Zhang R. et al. Complete species-level phylogeny of the leaf warbler (Aves: Phylloscopidae) radiation // *Mol. Phylogenet. Evol.* 2018. V. 126. P. 141–152.
<https://doi.org/10.1016/j.ympev.2018.03.0311>

The “Evolution” of the Mitochondrial Genome of the (*Phylloscopus borealis* Sensu Lato) Occurs in its Nuclear Genome

L. N. Spiridonova^{1, *}, O. P. Valchuk¹, Ya. A. Red'kin^{2, 3}

¹Federal Scientific Center for Biodiversity of Terrestrial Biota of Eastern Asia Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690022 Russia

²Zoological Museum, Moscow State University, Moscow, 125009 Russia

³Institute of Biological Problems of the Cryolithozone, Siberian Branch of the Russian Academy of Sciences, Yakutsk, 677000 Russia
e-mail: spiridonova@biosoil.ru

Sequencing of a partial fragment of the *ND5*–*cytb* mitochondrial genes (1553 bp) and its nuclear copies was carried out to *Phylloscopus borealis* sensu lato, belonging to individual taxonomic groups from different parts of the range. It was shown that the majority of taxon-specific and unique mitochondrial substitutions in *examinandus* and *xanthodryas* forms were identical to those in nuclear copies of *borealis* mtDNA. Differences between mitochondrial haplotypes of *examinandus* and nuclear copies of mtDNA *borealis* were low ($p = 0.02$), at the same time the mtDNA genetic divergence in *borealis*–*examinandus*, *borealis*–*xanthodryas* and *examinandus*–*xanthodryas* significantly exceeded this value ($p = 0.035$, 0.044 and 0.046 , respectively). A nuclear copy of the mitochondrial haplotype of the easternmost form of *xanthodryas* was first discovered in the nuclear genome of one *borealis* individual in the western part of the breeding range (Komi Republic). Alongside this, it was shown at the first time, the nuclear copies of *xanthodryas* mtDNA from Toyama Prefecture (Japan) were close to the mitochondrial haplotypes of *borealis* from Kytlym (Sverdlovsk region) ($p = 0.018$). Thus, the mutations emerging in the nuclear copies of mitochondrial genes are the source of most substitutions in the mitochondrial DNA of the studied forms. The origin of the mitochondrial haplotypes *examinandus* and *xanthodryas* from nuclear copies of mtDNA *borealis* and the close similarity of their nuclear genomes gives grounds to consider the mitogenomes of these forms as variants of the haplotype of the single species *Ph. borealis* sensu lato. With a high degree of probability, it can be argued that the divergence time of the haplotypes of the analyzed forms is significantly less than 2.5–3 million years, as previously assumed by a number of authors [Saitoh et al. 2010; Alström et al. 2011], and the “molecular clock” that do not take into account recombination events between the nuclear and mitochondrial genomes cannot be used in this case.

Keywords: mitochondrial genome, nuclear copies of mtDNA, intergenomic recombination, *Phylloscopus borealis*

ГЕНЕТИЧЕСКОЕ РАЗНООБРАЗИЕ МАЛОГО СУСЛИКА *SPERMOPHILUS PYGMAEUS PALLAS, 1779 (SCIURIDAE, RODENTIA)* НА СЕВЕРНОМ КАВКАЗЕ

© 2024 г. Ф. А. Темботова¹, М. С. Гудова¹, А. Х. Амшокова^{1,*}, А. Х. Халидов²

¹Институт экологии горных территорий им. А.К. Темботова Российской академии наук, Нальчик, 360051 Россия

²Дагестанская противочумная станция, Махачкала, 367010 Россия

*e-mail: h.a.amshokova@mail.ru

Поступила в редакцию 20.12.2023 г.

После доработки 13.02.2024 г.

Принята к публикации 20.02.2024 г.

На основании анализа фрагмента гена цитохрома *b* (*cytb*) митохондриальной ДНК (мтДНК) исследовано генетическое разнообразие малого суслика *Spermophilus pygmaeus* Pallas, 1779 Центрального и Восточного Кавказа. С помощью филогенетического анализа выявлено существование двух кластеров А и В в пределах западной клады *S. pygmaeus* 2. Кластер А сформирован гаплотипами сусликов из Восточного Кавказа и правого берега р. Волги, а В — только гаплотипами центральнокавказских зверьков. Дистанция между кластерами А и В достигает 1.3%. Относительно обособленное положение на филогенетическом дереве популяции малого суслика Центрального Кавказа, отсутствие идентичных гаплотипов у центрально- и восточнокавказских животных, а также полученные дистанции указывают на генетическую гетерогенность вида на Северном Кавказе. Результаты молекулярного датирования показали, что эволюционный возраст гаплотипов мтДНК *S. pygmaeus* из исследованных районов Восточного Кавказа составил примерно 260 тыс. лет, а Центрального Кавказа — 163 тыс. лет. Отмечено снижение генетического разнообразия в центральнокавказских популяциях малого суслика по сравнению с популяциями из Восточного Кавказа, что в целом указывает на низкую жизнеспособность *S. pygmaeus*, обитающих в горах Центрального Кавказа.

Ключевые слова: малый суслик, *Spermophilus pygmaeus*, цитохром *b* (*cytb*), митохондриальная ДНК (мтДНК), Центральный и Восточный Кавказ.

DOI: 10.31857/S0016675824070056 **EDN:** BIIBUF

Изучение генетической структуры и генетического разнообразия популяций животных является одним из актуальных направлений современных экологических исследований. Подобные исследования особенно необходимы при изучении видов, резко сокращающих свою численность или находящихся в долговременном депрессивном состоянии. К числу таковых относится малый суслик *Spermophilus pygmaeus* Pallas, 1779, в отношении которого проводились грандиозные по масштабу истребительские работы в аридных ландшафтах бывшего СССР [1]. В частности, на Северном Кавказе в прошлом веке было рекомендовано полное истребление сусликов [2]. В итоге интенсивная сельскохозяйственная деятельность человека, антропогенное преобразование ландшафта на обширной территории, а также снижение пастбищной нагрузки вследствие резкого сокращения поголовья скота и соответственно восстановление высокого травостоя привели к существенному сокращению

численности и площади поселений малого суслика не только на Кавказе, но и во многих регионах России: Самарский край [3], Ростовская область [4, 5], Астраханская область [6], Калмыкия [7] и т. д. К числу факторов, предопределивших современную тотальную депрессию численности малого суслика на европейском юго-востоке России, относятся еще и климатические — повышение среднегодовой температуры в течение XX в., что приводит к нарастанию аридности [8, 9].

По данным службы защиты растений (1958–1984 гг.), практически на все виды сусликов негативно повлияло увеличение площадей распахки. Это привело к постепенному дроблению популяций сусликов на все более мелкие группировки, увеличению их изоляции вплоть до полного прекращения контактов [4]. В результате поселения малого суслика, особенно в пределах степной зоны, приобрели сильно мозаичный характер.

Данное положение еще более усугубляется и тем, что подвижность сусликов невелика, и сезонные перемещения молодняка обычно не превышают 1–2 км (по данным мечения — максимум 5 км) [10]. А как известно, пространственная подразделенность и временная изоляция популяций приводят к ограничению потока генов, снижению уровня гетерозиготности, повышению уровня инбридинга, а также к потере генетического разнообразия. В некоторых случаях возможны как полное вымирание, так и коренная перестройка структуры в связи с прохождением популяциями состояния “бутылочного горлышка” или восстановлением численности при ограниченном числе первопоселенцев (цит. по: Титов с соавт., 2020) [11]. При этом суслики, являясь важным звеном пищевой цепи, служат главным кормовым объектом для множества редких видов змей, птиц и млекопитающих. Колонии этого вида — основа сохранения других редких видов животных, в том числе хищных птиц. Кроме того, их норы, расположенные на целине, часто служат единственным природным укрытием для многих видов позвоночных, не выживающих в условиях трансформированного человеком ландшафта [12–14]. По данным Н.М. Окуловой [15], в норах малого суслика обитают 12 тыс. различных видов животных разных систематических групп.

Крупномасштабное картографирование ареала горного суслика в Центральном-Кавказском очаге чумы показало наличие изоляции отдельных поселений, отсутствие постоянного обмена особями между поселениями и достаточно длительное существование их без притока зверьков извне [16].

Как справедливо отмечают в своей работе В.А. Матросова с соавт. [17], поддержание генетического разнообразия и устойчивости популяций может быть обеспечено реинтродукцией и обменом особями между изолированными популяциями, для чего необходимо знать их генетическую структуру, происхождение, степень внутри- и межпопуляционного полиморфизма. С появлением молекулярно-генетических методов стало возможным не только оценить уровень генетического разнообразия, но и исследовать филогеографическую структуру видов — пространственное распределение генеалогических линий на ареале, выявить генетически обособленные популяции и группировки и т. д. В этом плане наиболее интересен малый суслик, представленный на Кавказе отдельными поселениями, разобщенными физико-географическими барьерами.

Данный вид привлекает внимание исследователей издавна, однако работ, основанных на анализе участка гена *cytb* мтДНК, на обозначенной территории очень мало [18, 19]. Следует отметить, что в базе Генбанк к настоящему времени доступны всего 16 последовательностей *cytb* мтДНК *Spermophilus pygmaeus* Кавказа, из которых восемь происходят из разных точек Восточного Кавказа, в частности

Дагестана, а шесть — из Центрального Кавказа (Эльбрус, Баксан, Джилы-Су, Шаджатмаз). Также еще две последовательности малого суслика из п. Эльбрус депонированы в Генбанк Р. Харрисоном с соавт. в 1993 г. [18]. Работы, основанные на исследовании маркеров митохондриальной (С-регион) и ядерной (интрон 6 гена *p53*) ДНК [20, 21], также выполнены на единичных последовательностях (от 1–3) из Дагестана, Ингушетии, Кабардино-Балкарии [21], что явно недостаточно для столь обширной и географически сложной территории Кавказа. Наибольшее количество экземпляров малого суслика изучено на территории Центрально-Кавказского высокогорного природного очага чумы [22]. В данной работе проанализированы 64 последовательности С-региона горного суслика, собранные на территориях Кабардино-Балкарской и Карачаево-Черкесской республик (29 поселений).

К числу неразрешенных вопросов, как справедливо отмечают О.А. Ермаков с соавт. [21], относятся и филогенетические отношения между “горными” и “равнинными” сусликами Северного Кавказа. Отсутствие подобного рода работ по молекулярной генетике вероятно связано с исчезновением равнинно-предгорных популяций сусликов во многих регионах России, в том числе и на Кавказе [23]. Так, если в 1970–1980 гг. на территории Кабардино-Балкарской Республики равнинно-предгорная форма малого суслика была распространена на левобережье р. Малки, охватывая равнинную степь, предгорные луговые степи и среднегорье в области Мелового (Джинальского) хребта [24], то по последним данным сельскохозяйственное освоение равнинных и предгорных территорий привело к почти полному исчезновению этих популяций малого суслика [23]. С 1990-х гг. предгорно-равнинные популяции малого суслика не регистрируются и занесены в Красную книгу КБР [25] как находящиеся на грани исчезновения. В связи с вышеизложенным сравнение горных и равнинных форм малого суслика в пределах Центрального Кавказа становится невозможным. Отмеченное справедливо и для Восточного Кавказа, где высотные пределы распространения малого суслика не превышают 700 м над ур. моря [26].

Цель исследования — изучить генетическое разнообразие сусликов Центрального и Восточного Кавказа на основании анализа фрагмента гена *cytb* мтДНК и оценить эволюционный возраст восточно- и центральнокавказских группировок *S. pygmaeus*.

МАТЕРИАЛЫ И МЕТОДЫ

Проведен молекулярно-генетический анализ популяций малого суслика в условиях Центрального (Кабардино-Балкарская Республика (КБР): окр. с. Безенги, Актопракский перевал, окр. г. Тырныауз, окр. п. Эльбрус (ущелье Ирикчат, урочище

Джилы-Су)) и Восточного Кавказа (Республика Дагестан): долина Кар-Кар, окр. п. Львовский № 13, окр. с. Зеленоморск, окр. с. Хумтоп. Места сбора материала приводятся на рис. 1.

Изучение генетического разнообразия *Spermophilus pygmaeus* выполнено на основании данных о 66 последовательностях фрагмента гена *cytb* мтДНК длиной 840 пн. Места сбора материала и количество исследованных особей в каждой выборке представлены в табл. 1.

В целях сравнения полученных данных с описанными ранее в литературе в исследование были включены последовательности мтДНК *S. pygmaeus* (OP588846–OP588904) [19], AF157907, AF157910 [18], *S. musicus* Menetries, 1832 (AF157900, AF157904) [18], размещенные в базе GenBank (ncbi.nlm.nih.gov). Также были использованы сиквенсы представителей рода *Spermophilus*: *S. taurensis* Gündüz et al., 2007 (KY938064, KY938069, KY938073) [27], № *S. citellus* Linnaeus, 1766 (AM691632–AM691640), *S. xanthoprymnus* Bennett, 1835 (AM691658–AM691663) [28] и AF157902, AF157909 [18]. Для укоренения филогенетического дерева в качестве

внешней группы использовалась последовательность *Marmota monax* Linnaeus, 1758 – AF157953 [18].

Для молекулярно-генетического анализа использовали участок гена *cytb* ДНК. Тотальную ДНК из образцов мышечной ткани, зафиксированных в 96%-ном этиловом спирте, выделяли вручную с использованием набора реагентов Diatom™ DNA Prep 100 (ООО “Лаборатория Изоген”, Москва) по инструкции производителя. Полученные растворы ДНК хранили при температуре –18 °С.

Амплификацию фрагментов ДНК осуществляли с использованием набора MasterMix X5 (Диалат, Москва). Для полимеразной цепной реакции использовались праймеры: L14725 TGAAAAAYCATCGTTGT [29] и H15915 TCTTCATTTYWGGTTTACAAGAC [18] при параметрах цикла ПЦР, рекомендованных в цитируемой работе. Полученные ПЦР-продукты очищали переосаждением в 0.15 М растворе ацетата натрия в 90%-ном этаноле с последующей промывкой 70%-ном этанолом. Результат визуализировали

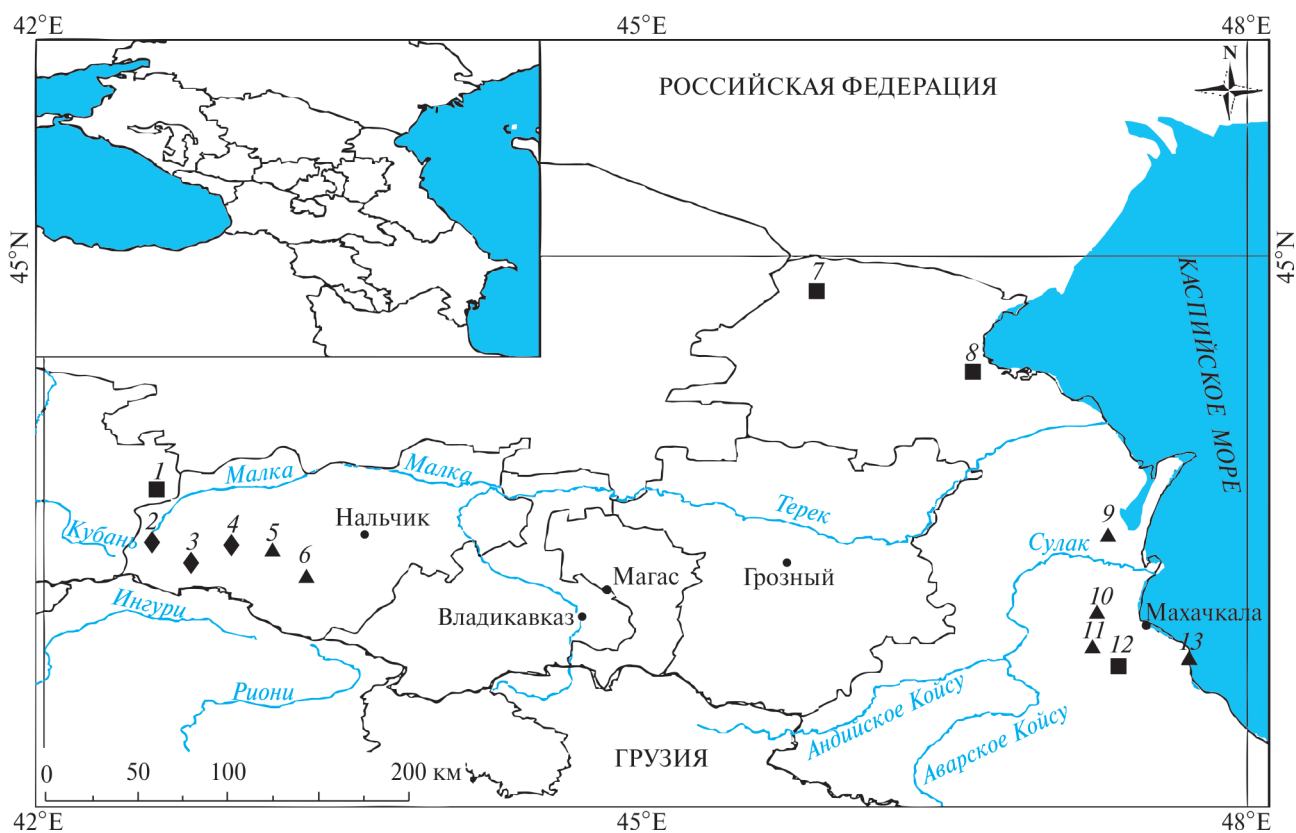


Рис. 1. Карта происхождения материала *Spermophilus pygmaeus* на Северном Кавказе. Центральный Кавказ, Кабардино-Балкарская Республика: 1 – Шаджатмаз, 2 – Джилы-Су, 3 – Эльбрус (Ирик-Чат), 4 – Баксан (Тырныауз), 5 – Актопрак, 6 – Безенги; Восточный Кавказ, Республика Дагестан: 7 – Сухокумск, 8 – Кочубей, 9 – Львовский № 13, 10 – Хумтоп, 11 – Кар-Кар, 12 – Параул, 13 – Зеленоморск. Черными треугольниками обозначены места сбора наших образцов, квадратами – места сбора образцов, использованных в работе [19], ромбами – наши точки сбора, совпавшие с таковыми по [19].

Таблица 1. Характеристика исследованного материала *S. pygmaeus* Центрального и Восточного Кавказа

Гаплотип	Образцы	Место сбора	Координаты	Идентичен гаплотипам из Генбанка
3260Bez	3260Bez–3262Bez, 3735Bez, 3737Bez, 3739Bez–3744Bez	КБР, окр. с. Безенги	N 43.240254° E 43.139148°	AF157904 OP588865 OP588866 - OP588868 OP588903 OP588904
	3719Akto–3727Akto, 3258Akto, 3696Akto, 87Akto–90Akto, 3698Akto, 3697Akto, 3695Akto	КБР, Актопракский перевал	N 43.389237° E 43.139148°	
	132Tyr, 3266Tyr, 3313Tyr, 3701Tyr, 3703Tyr, 3704Tyr	КБР, окр. г. Тырнауз	N 43.412555° E 42.938234°	
	134Elb, 3267Elb	КБР, окр. п. Эльбрус	N 43.265347° E 42.636483°	
	4061Dzhi	КБР, ущелье Джилы-Су	N 43.434402° E 42.540295°	
3744 Bez	3744Bez	КБР, окр. с. Безенги		
3702 Tyr	3702Tyr	КБР, окр. г. Тырнауз		
4057Dzhi	4057Dzhi	КБР, ущелье Джилы-Су		
4058Dzhi	4058Dzhi, 4060Dzhi, 4062Dzhi	»		
4059Dzhi	4059Dzhi	»		
3268Elb	3268Elb	КБР, окр. п. Эльбрус		
3255Akto	3255Akto	КБР, Актопракский перевал		
3694Akto	3694Akto	»		
3693Akto	3693Akto	»		
11Kar	11Kar	РД, долина Кар-Кар	N 42.862909° E 47.225681°	
12Kar,	12Kar, 9Kar	»		OP588899, OP588902
16Zelen	16Zelen	РД, окр. с. Зеленоморск	N 42.791189° E 47.707594°	
19Lvov	19Lvov	РД, окр. п. Львовский № 13	N 43.470495° E 47.305398°	
20Lvov	20Lvov	»		
22Lvov	22Lvov	»		
23Lvov	23Lvov	»		OP588898, OP588900, OP588901
25Lvov	25Lvov	»		
27Lvov	27Lvov	»		
66Hum	66Hum	РД, окр. с. Хумтоп	N 43.045594° E 47.244806°	
	10Kar	РД, долина Кар-Кар		
	13Zelen–15 Zelen	РД, окр. с. Зеленоморск		
68Hum	68Hum	РД, окр. с. Хумтоп		

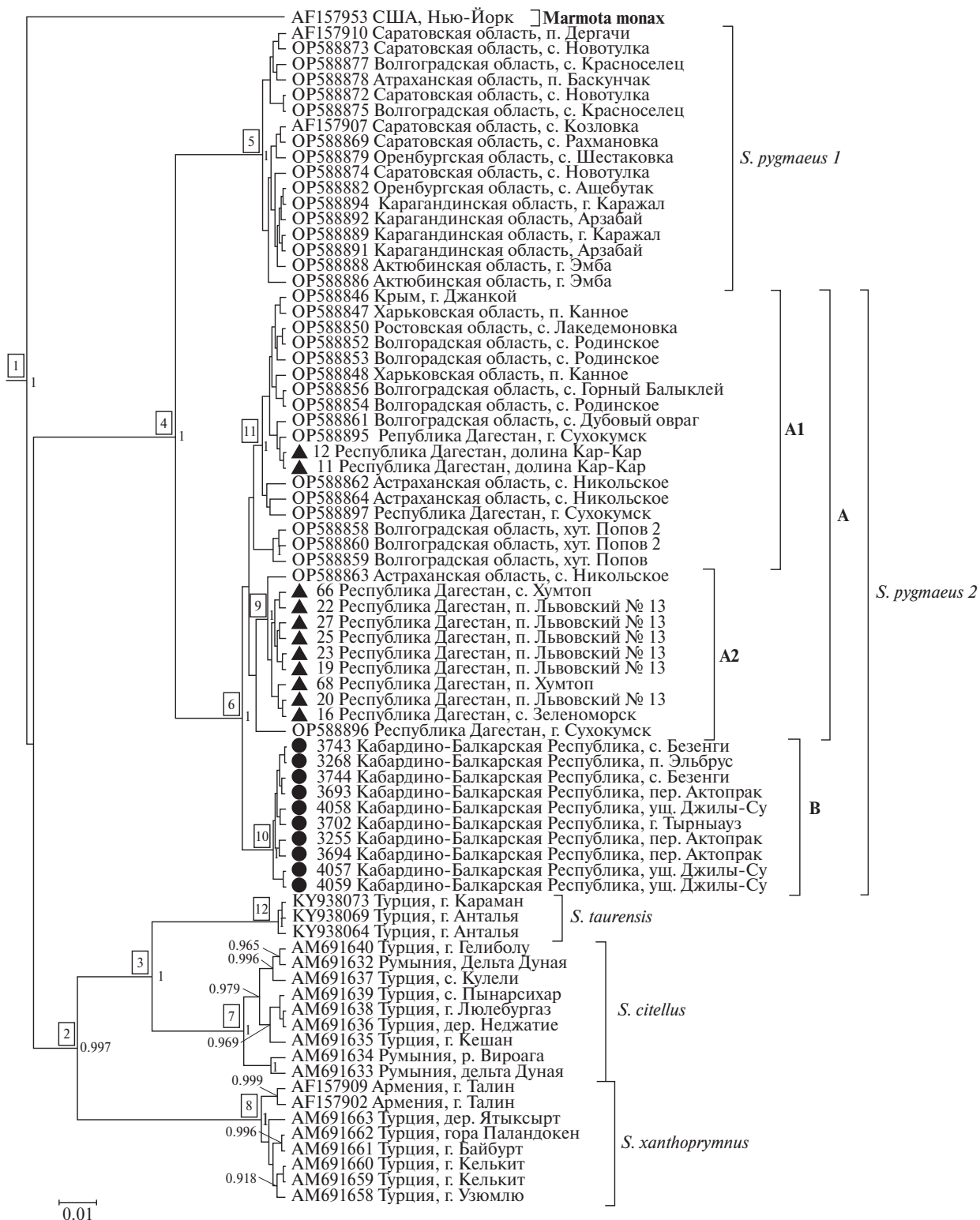


Рис. 2. Байесовское филогенетическое дерево представителей рода *Spermophilus*, основанное на анализе гаплотипов фрагмента гена *cytb* мтДНК (840 пн). Числа в узлах ветвления — значения апостериорных вероятностей (больше 0.900), числа в квадратах — номера узлов. Черными треугольниками обозначены гаплотипы сусликов Восточного Кавказа, а черными кружочками — гаплотипы зверьков Центрального Кавказа. РД — Республика Дагестан, КБР — Кабардино-Балкарская Республика.

электрофорезом в 1.5%-ном агарозном геле в присутствии бромистого этидия.

Секвенирование проводили на базе ООО “Синтол” (Москва).

Редактирование и выравнивание полученных последовательностей проводили с использованием программы BioEdit 7.09.0. [30]. Определение одинаковых последовательностей (гаплотипов) выполняли с помощью онлайн-сервиса FaBox 1.5 [31]. Филогенетический анализ нуклеотидных последовательностей методом Байеса МСМС проводили в MrBayes v3.2.6 [32]. Для построения медианной сети методом Median-Joining использовали программу Network 4.6.1 [33].

Количество гаплотипов, значения гаплотипического (h) и нуклеотидного (π) разнообразия, индексы Таджимы и Фу, а также показатели демографической и пространственной экспансии были рассчитаны в программе Arlequin v3.5 [34].

Время, прошедшее с момента существования последних общих предков (TMRCA) для клад мтДНК, рассчитывали в программе BEAST V1.10.4 [35] методом байесовской цепи Маркова Монте-Карло (МСМС), используя модель замещения НКУ+I, выбранную в Mega 6 [36]. Мы использовали некоррелированные расслабленные часы в качестве модели часов и размер констант в качестве коалесцентной модели. Для молекулярного датирования использовали следующие калибровки: 10.9 млн лет для корневого узла расхождения *Marmota* и других видов *Spermophilus* [37], 5.0 млн лет — время дивергенции между *S. xanthoprymnus* и *S. citellus* + *S. taurensis* и 2.5 млн лет между *S. citellus* и *S. taurensis* [28]. Время расхождения узлов дендрограммы *Spermophilus* рассчитано для шести вариантов замен за миллион лет: 0.5, 0.9, 1.2, 2.4, 3.2 и 6.7%.

РЕЗУЛЬТАТЫ

Среди 66 проанализированных последовательностей фрагмента гена *cytb* мтДНК *S. pygmaeus* Центрального и Восточного Кавказа обнаружен 21 гаплотип (табл. 1). Всего выявлено 36 переменных позиций, из них 21 парсимони-информативные. Усредненный нуклеотидный состав исследованного фрагмента (аденин — 26.7%, тимин — 33.5%, цитозин — 26.5%, гуанин — 13.4%) сходен с таковым у других мелких млекопитающих [38], особенностью которых является низкое содержание гуаниновых оснований [39].

Филогенетическое дерево, полученное по результатам байесовского анализа, демонстрирует существование в пределах вида *S. pygmaeus* двух крупных кластеров с высоким уровнем поддержки (рис. 2).

Данные кластеры соответствуют выделенным О.А. Ермаковым с соавт. [19] группировкам малого

суслика — “восточной” (*S. pygmaeus* 1) с левого берега и “западной” (*S. pygmaeus* 2) с правого берега р. Волги. Как видно из рис. 2, “западный” кластер более крупный и сформирован из двух гаплогрупп (А и В). Первая, более крупная гаплогруппа (А) включает в себя две подгруппы А1 и А2. А1 образована нуклеотидными последовательностями сусликов [19]: Крыма (г. Джанкой), Харьковской (п. Конное), Ростовской (с. Лакедемоновка), Волгоградской (хут. Попов 2, села Родинское, Горный Балыклей, Дубовый овраг), Астраханской (с. Никольское) областей, Дагестана (г. Сухокумск).

Особое внимание обращает на себя положение гаплотипов малого суслика из долины Кар-Кар на филогенетическом дереве (рис. 2). Несмотря на их обособленное положение, следует отметить, что из четырех исследованных гаплотипов малого суслика долины Кар-Кар только три попали в подгруппу А1, тогда как четвертый, идентичный гаплотипу 66Num, вошел в подгруппу А2. Таким образом, в одном из исследованных нами географических пунктов Восточного Кавказа (долина Кар-Кар) отмечены две генетически различающиеся группировки малого суслика.

Большая часть анализируемых нами гаплотипов (9 из 11) из разных географических точек Восточного Кавказа образовали компактную группу (А2) вместе с гаплотипами OP588863 из Астраханской области (с. Никольское) и OP588896 из Дагестана (г. Сухокумск). Один из наших гаплотипов (23) из окр. п. Львовский № 13 был идентичен трем последовательностям OP588898, OP588900 и OP588901 из окр. с. Параул Республики Дагестан [19]. Вторая гаплогруппа (В) сформирована гаплотипами сусликов из разных географических пунктов только Центрального Кавказа (3743Bez, 3268Elb, 3744Bez, 3693Akto, 4058Dzhi, 3702Tyrrn, 3255Akto, 3694Akto, 4057Dzhi и 4059Dzhi). Следует отметить, что все гаплотипы из Кабардино-Балкарской Республики, анализируемые в работе О.А. Ермакова с соавт. [19], оказались полностью идентичны гаплотипу 3743Bez из окр. с. Безенги.

“Восточный” кластер (*S. pygmaeus* 1) образован последовательностями мтДНК сусликов из Волгоградской (с. Красноселец), Астраханской (с. Баскунчак), Оренбургской (с. Шестаковка, с. Ащербутак), Саратовской (с. Рахмановка, п. Дергачи и с. Новотулка), Актыбинской (г. Эмба, с. Алабас) и Карагандинской (г. Каражал и Арзабай) областей. Генетическая дистанция между “западной” и “восточной”кладами составила 4.6%.

Также в анализ были включены последовательности и других представителей рода *Spermophilus*: *S. citellus*, *S. taurensis* и *S. xanthoprymnus*. При парном сравнении отмеченных таксонов дистанции варьировали от 5.9 до 12.1%. Минимальная дистанция 5.9 отмечена между видами *S. citellus*—*S. taurensis*, максимальная — 12.1% между парами

видов *S. pygmaeus* и *S. xanthoprymnus*. Близкая к максимальной дистанция 11.6% выявлена между *S. pygmaeus* и *S. taurensis* (табл. 2). Генетические дистанции, составляющие около 10%, выявлены между парами *S. pygmaeus*—*S. citellus*, *S. citellus*—*S. xanthoprymnus* и *S. xanthoprymnus*—*S. taurensis*.

Анализ медианной сети гаплотипов демонстрирует разделение сусликов Центрального и Восточного Кавказа на две группы. Данные группировки условно названы как восточнокавказская и центральнокавказская соответственно их географическому происхождению (рис. 3). Как видно из рис. 3, для 49 исследованных особей из Центрального Кавказа было отмечено всего 10 гаплотипов. Базовый гаплотип 3743Bez широко представлен на Центральном Кавказе и встречается во всех исследованных географических точках с высокой частотой. Он отмечен у 38 особей сусликов и обнаружен у большинства особей из Актопракского перевала (18), Эльбруса (2), Джилы-Су (1), а также у шести особей из семи окр. г. Тырныауза и 11 особей из 12 окр. с. Безенги.

При сравнении наших последовательностей с таковыми из базы Генбанк выявлено, что гаплотип 3743Bez полностью совпадает со всеми последовательностями из Кабардино-Балкарской Республики: одной последовательностью из Шатджатмаза, Джилы-Су, Баксана и тремя последовательностями из Эльбруса [19]. Гаплотипы девяти особей (3255Akto, 3694Akto, 3693Akto, 3268Elb, 3744Bez, 3702Tyr, 4057Dzhi, 4058Dzhi и 4059Dzhi) являются уникальными и отличаются от широко распространенного на одну-две замены, а гаплотип 4058Dzhi является общим для двух особей из урочища Джилы-Су. В восточнокавказской выборке полученные 17 последовательностей сформировали 11 гаплотипов, из которых преобладающий гаплотип 66Hum отмечен у шести особей из трех географических точек (окр. с. Хумтоп, окр. с. Зеленоморск и окр. п. Львовский № 13). Второй по частоте встречаемости гаплотип 12Каг отмечен у двух особей из долины Кар-Кар. Данный гаплотип отмечен и на других территориях Дагестана (Параул, Кочубей) [19]. В целом в восточнокавказской группировке количество уникальных гаплотипов выше, чем в центральнокавказской, и составляет около 53%. В

выборке из окр. п. Львовский № 13 все гаплотипы были уникальными, т. е. представлены одной особью каждый. Однако, как было отмечено ранее, гаплотип 23 из данной выборки совпал с тремя последовательностями из окр. с. Параул Республики Дагестан. Важно отметить, что полностью идентичных гаплотипов у сусликов Центрального и Восточного Кавказа не выявлено. Между двумякладами выявлено 12 замен и два гипотетических гаплотипа.

Показатели генетического разнообразия приводятся в табл. 3. Как видно, уровень гаплотипической изменчивости отличается в центрально- и восточнокавказских выборках. Среди центральнокавказских выборок наибольшими значениями гаплотипического и нуклеотидного разнообразия характеризуется выборка из урочища Джилы-Су. Большая же часть центральнокавказских выборок характеризовалась очень низкими значениями как гаплотипического (от 0.1667 до 0.2857), так и нуклеотидного разнообразия (от 0.0002 до 0.0007). Несколько выше данные показатели в выборке из окр. п. Эльбрус (ущелье Ирикчат) (табл. 3). В целом суммарная выборка из Центрального Кавказа также характеризуется низкими значениями обоих показателей. Если исходить из того, что популяции с низкими значениями h и π исторически более молодые [40, 41], можно предположить, что по сравнению с восточнокавказской выборкой центральнокавказская более молодая. В отличие от центральнокавказских в большинстве выборок Восточного Кавказа наблюдаются высокие значения гаплотипического и нуклеотидного разнообразия. Исключение составляет выборка из окрестностей с. Зеленоморск, для которой приводимые значения в 2 и более раз ниже, чем в остальных выборках. При объединении всех выборок Восточного Кавказа в одну значения гаплотипического и нуклеотидного разнообразия также были высокими. Согласно данным некоторых авторов [42, 43], наибольшее генетическое разнообразие видов встречается в местах предполагаемых предковых рефугиумов.

Значения критерия Таджимы ($D = -2.137$, $p = 0.003$) и теста Фу ($F_s = -8.676$, $p = 0.000$) для суммарной центральнокавказской выборки малого

Таблица 2. Межвидовые генетические дистанции представителей рода *Spermophilus* для области гена цитохрома b 840 пн

Виды	<i>S. pygmaeus</i>	<i>S. taurensis</i>	<i>S. xanthoprymnus</i>	<i>S. citellus</i>
<i>S. pygmaeus</i>	—	0.0118	0.0121	0.0111
<i>S. taurensis</i>	0.1155	—	0.0114	0.0089
<i>S. xanthoprymnus</i>	0.1212	0.1018	—	0.0120
<i>S. citellus</i>	0.0995	0.0588	0.1006	—

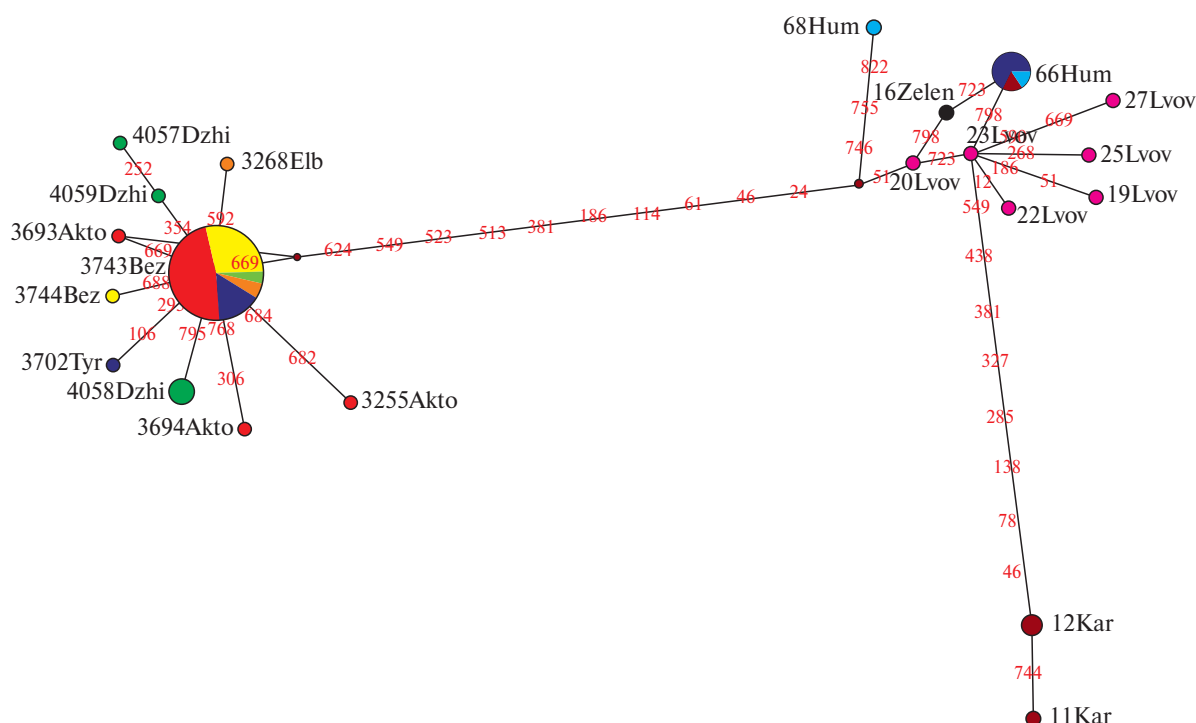


Рис. 3. Медианная сеть гаплотипов мтДНК *S. pygmaeus*, построенная на основе анализа участка *cytb* (алгоритм — Median-Joining, Network 4.6.1.). Величина круга пропорциональна количеству идентичных гаплотипов. Гаплотипы, выявленные на территории Республики Дагестан: Kar (долина Кар-Кар) обозначены коричневым цветом, Zelen (окр. с. Зеленоморск) — черным цветом, Hum (окр. с. Хумтоп) — голубым цветом, Lvov (окр. п. Львовский № 13) — розовым цветом; гаплотипы, выявленные на территории КБР: Tyr (окр. г. Тырнауз) — синим цветом, Bez (окр. с. Безенги) — желтым цветом, Akto (Актопракский перевал) — красным цветом, Elb (окр. п. Эльбрус) — оранжевым цветом, Dzhi (ущелье Джилы-Су) — салатовым цветом. Цифры — позиции нуклеотидных замен, красные ромбики — гипотетические гаплотипы.

суслика оказались отрицательными и статистически значимыми, что свидетельствует о возможном действии отрицательного отбора на данный участок генома в этих популяциях либо о популяционной экспансии.

Результаты молекулярного датирования, основанные на трех калибровочных данных (табл. 4), показали, что эволюционный возраст (для модели, рассчитанной для частоты мутаций 0.5% на миллион лет) исследуемых образцов *S. pygmaeus* в целом как вида составил 2.277 млн. лет (95% HPD: 1.475–3.166 млн лет) (узел 4). Возраст так называемого “восточного” кластера *S. pygmaeus* 1 (узел 5) из левого берега р. Волги [19] составил 349 тыс. лет (95% HPD: 0.186–0.534), “западного” *S. pygmaeus* 2 кластера (узел 6), в который вошли и все анализируемые нами последовательности как из Центрального, так и Восточного Кавказа, составил 739 тыс. лет (95% HPD: 0.451–1.058). Возраст центральнокавказской группировки (узел 10) составил 163 тыс. лет (95% HPD: 0.0698–0.270), а восточнокавказской вместе с одним образцом OP588863 из Астраханской области (узел 9) составил 260 тыс. лет (95% HPD: 0.114–0.431).

ОБСУЖДЕНИЕ

Результаты проведенного анализа последовательностей фрагмента гена *cytb* малого суслика из девяти географических выборок Центрального и Восточного Кавказа свидетельствуют о некоторой степени генетической дифференциации центрально- и восточнокавказских выборок. Видно, что на филогенетическом дереве, как и на медианной сети, гаплотипы малого суслика с Центрального и Восточного Кавказа в основном группируются отдельно друг от друга, т. е. отмечается достаточно четкая связь распределения гаплотипов по регионам. Идентичных гаплотипов у малого суслика из центрально- и восточнокавказских группировок не обнаружено.

Данные о генетической дистанции между анализируемыми группировками малого суслика согласуются с их географическим распространением. Генетические дистанции (Net distance), полученные при попарном сравнении между собой выборок из Центрального и Восточного Кавказа, варьировали от 1.33 до 1.67, что свидетельствует о некоторой степени генетической обособленности

Таблица 3. Показатели гаплотипического (h) и нуклеотидного (π) разнообразия и значения тестов Таджимы и Фу *S. rugtaeus* на Центральном и Восточном Кавказе

Географические группировки (размер выборки)	N	$\pi \pm S.E.$	$h \pm S.E.$	Tajima's D	Fu's F
Центральный Кавказ					
1. КБР, окр. с. Безенги (12)	2	0.0002 ± 0.0003	0.1667 ± 0.134	-1.1405	-0.4757
2. КБР, Актопракский перевал (21)	4	0.0006 ± 0.0006	0.2714 ± 0.1242	-1.7268	-1.6297
3. КБР, окр. г. Тырнауза (7)	2	0.0007 ± 0.0007	0.2857 ± 0.1964	-1.237	0.856
4. КБР, окр. п. Эльбрус (3)	2	0.0008 ± 0.0010	0.6667 ± 0.3143	0.0000	0.20067
5. КБР, ущелье Джилы-Су (6)	4	0.0017 ± 0.0014	0.8000 ± 0.1721	0.6003	-1.072
Для <i>Spermophilus</i> Центрального Кавказа (49) в целом	10	0.0007 ± 0.0006	0.3997 ± 0.0889	-2.1373	-8.676
Восточный Кавказ					
6. РД, долина Кар-Кар (4)	3	0.0060 ± 0.0044	0.8333 ± 0.2224	-0.834	1.506
7. РД, окр. п. Львовский № 13 (6)	6	0.0028 ± 0.0020	1.0000 ± 0.0962	-1.390	-3.927
8. РД, окр. с. Зеленоморск (5)	2	0.0005 ± 0.0006	0.4000 ± 0.2373	-0.817	0.090
Для <i>Spermophilus</i> Восточного Кавказа (15) в целом	10	0.0054 ± 0.0032	0.8952 ± 0.0704	-0.611	-2.443

Примечание. N — число гаплотипов, S. E. — стандартная ошибка, статистически достоверные значения тестов выделены полужирным. РД — Республика Дагестан, КБР — Кабардино-Балкарская Республика.

центрально- и восточнокавказских выборок. На основании выполненного крупномасштабного картографирования поселений сусликов в эльбурской части ареала А.И. Дятлов с соавт. [44] высказали мнение о том, что никаких связей, по крайней мере в последние несколько тысяч лет, нельзя предполагать между сусликами из Приэльбрусья и зверьками, населяющими Алханчуртскую долину (Терско-Сунженское междуречье), а также с поселениями равнинного Дагестана. Географические группировки малого суслика, обитающие на Центральном Кавказе, оказались генетически близки между собой. Следует отметить, что несмотря на генетическую близость восточнокавказских и равнинных правобережных выборок, идентичных гаплотипов выявлено не было.

В пределах вида *S. rugtaeus* между “восточной” и “западной” группировками (*S. rugtaeus* 1 и *S. rugtaeus* 2) (рис. 2) дистанция составила 4.6%. Полученные результаты практически совпадают с данными О.А. Ермакова с соавт. [19], выявившими

между “западной” и “восточной” группами гаплотипов дистанцию 4.9%. Между видами рода *Spermophilus*, анализируемыми в данной работе, отмечается размах дистанции от 5.9—12.1%. По результатам, полученным на основании анализа того же фрагмента мтДНК другими исследователями, средние генетические дистанции между видами сусликов рода *Spermophilus* колебались от 0.0260 (2.6%) до 0.1401 (14%) [45].

Сравнение нуклеотидного и гаплотипического разнообразия в популяциях малого суслика из разных географических точек Центрального и Восточного Кавказа показало, что центральнокавказские выборки характеризуются существенно более низкими показателями генетического разнообразия по сравнению с сусликами Восточного Кавказа. Среди всех анализируемых центральнокавказских выборок самые высокие значения гаплотипического и нуклеотидного разнообразия отмечены в выборке с наибольшей высотой местности — урочище Джилы-Су (около 2400 м над ур. моря).

Таблица 4. Время расхождения (млн лет) таксонов *Spermophilus* с шестью вариантами (%) скоростей эволюции

Номер узла на дереве. Род, вид	0.5	0.9	1.2	2.4	3.1	6.7
1. <i>Marmota/Spermophilus</i>	8.594	7.693	7.426	6.525	6.250	5.506
2. <i>S. xanthoprymnus/S. taurensis</i> + <i>S. citellus</i>	4.642	4.181	4.044	3.568	3.427	3.001
3. <i>S. taurensis/S. citellus</i>	2.606	2.382	2.313	2.077	2.001	1.778
4. <i>S. Pygmaeus</i> 1 (“восточная”)/ <i>S. Pygmaeus</i> 2 (“западная”) группы	2.277	1.984	1.899	1.610	1.517	1.226
5. <i>S. pygmaeus</i> 1 (“восточная” группа)	0.349	0.304	0.294	0.249	0.235	0.193
6. <i>S. pygmaeus</i> 2 (“западная” группа)	0.739	0.644	0.616	0.522	0.494	0.399
7. <i>S. citellus</i>	0.730	0.644	0.620	0.532	0.504	0.414
8. <i>S. xanthoprymnus</i>	0.411	0.360	0.346	0.295	0.280	0.230
9. <i>S. pygmaeus</i> (Восточный Кавказ)	0.260	0.227	0.218	0.186	0.175	0.144
10. <i>S. pygmaeus</i> (Центральный Кавказ) (B)	0.163	0.143	0.138	0.118	0.112	0.0916
11. <i>S. pygmaeus</i> (A1)	0.343	0.299	0.286	0.244	0.231	0.187
12. <i>S. taurensis</i>	0.109	0.097	0.093	0.0806	0.077	0.064

Полученные результаты согласуются с данными Е.С. Котенева с соавт. [21], показавшими наиболее высокое разнообразие “высокогорной” гаплогруппы на основании анализа последовательностей С-региона мтДНК. Как и было отмечено ранее, объединенная выборка из Центрального Кавказа ($n = 49$) также характеризуется низкими значениями обоих показателей. Аналогичные результаты были получены О.А. Ермаковым с соавт. [19], которые отмечали низкое нуклеотидное (0.0003) и гаплотипическое (0.333) разнообразие для центральнокавказской выборки. Низкие значения нуклеотидного и гаплотипического разнообразия, отмечаемые в центральнокавказских группировках, позволяют предположить, что эти популяции проходили через так называемое “бутылочное горлышко”, в результате чего была утеряна значительная часть генетического разнообразия. Что касается численности, согласно данным Н.Н. Мадьянова [46], по сравнению со временем открытия природного очага (в 1971 г.) за последние годы заметно сократилась не только площадь поселений носителя, особенно в рассматриваемой части Центрально-Кавказского высокогорного природного очага чумы, но и снизилась численность, а также существенно изменились количество и конфигурация заселенных сусликами участков. Таким образом, можно предположить, что длительное падение численности и изоляция привели к увеличению доли обычных гаплотипов и в целом к снижению генетического разнообразия. Об обособлении приэльбрусской части ареала исследователи писали еще в 1980 г. и связывали этот процесс с вымиранием поселений на предгорной южной окраине равнин. Уже

в 1980 г. отмечалось, что приэльбрусские поселения изолированы от основной части ареала, расположенного на Прикаспийских равнинах, причем пространственный разрыв между крайними поселениями составлял 30–50 км, тогда как до 1970-х гг. разрыв составлял всего 10–20 км [44]. В сравнении с центральнокавказскими выборками восточнокавказские характеризуются достаточно высокими значениями нуклеотидного и гаплотипического разнообразия. Учитывая, что уровень генетического разнообразия определяется рядом факторов — временем существования вида, численностью, ареалом [47], можно предположить, что наиболее высокий уровень гаплотипического разнообразия в восточнокавказской выборке определяется более длительным временем существования *S. pygmaeus* на отмеченной территории, что подтверждается полученными датировками. Как видно из табл. 4, возраст восточнокавказской группы (узел 9) составил 260 тыс. лет, а центральнокавказской (узел 10) составил 163 тыс. лет (для модели, рассчитанной для частоты мутаций 0.5% на миллион лет), из чего можно предположить, что малый суслик проник в горы с равнины. Точку зрения, что предки современных горных сусликов в разное время проникали в высокогорье из равнинных районов, высказывали и другие исследователи [44, 48], о возможном расселении суслика в горы с равнин Предкавказья говорится и в монографии Н.К. Верещагина [49].

Таким образом, относительно обособленное положение гаплотипов малого суслика Центрального и Восточного Кавказа на филогенетическом дереве, отсутствие идентичных гаплотипов у центрально- и восточнокавказских группировок, а

также генетическая дистанция, полученная между ними (1.34%), указывают на генетическую гетерогенность *S. pygmaeus* на исследуемой территории. Также выявлено, что в долине Кар-Кар встречаются две генетически различающиеся группировки малого суслика, дистанция между которыми составляет 0.0108 (1.1%).

Высокое генетическое разнообразие исследованных восточнокавказских выборок по сравнению с центральнокавказскими свидетельствует, что они, возможно, более благополучны и жизнеспособны, чем центральнокавказские. Снижение популяционного генетического разнообразия может быть связано со снижением приспособленности популяции [50]. Потери аллельного разнообразия могут также негативно повлиять на долгосрочную жизнеспособность популяций, снижая способность популяций адаптироваться к изменяющимся условиям окружающей среды [51].

Работа выполнена по программе Государственного задания Института экологии горных территорий им. А.К. Темботова РАН (FMEU-2023-0001 “Разнообразие (генетическое, морфологическое, таксономическое) позвоночных Северного Кавказа, структура их размещения в регионе как основа долгосрочного мониторинга природных и антропогенных экосистем”).

Исследование одобрено Этическим комитетом ФБУН “Институт экологии горных территорий им. А.К. Темботова РАН”, протокол № 1 от 10.12.2022 г.

Все применимые международные, национальные и/или институциональные принципы ухода и использования животных были соблюдены.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Шилова С.А. Вопросы контроля численности и охраны сусликов России (род *Spermophilus*) // Аридные экосистемы. 2011. Т. 17. № 4 (49). С. 104–112.
2. Калабухов Н.И. Плотность заселения сусликами чумных районов Северо-Кавказского края и возможность сплошной очистки от сусликов // Тр. по защите растений. 1933. Сер. 4. Вып. 2. С. 65–86.
3. Дудников А.А., Чернышова О.В., Симаков М.Д., Титов С.В. История изучения распространения и современное состояние популяций сусликов Самарского края // Russ. J. Ecosystem Ecol. 2022. V. 7. № 1. <https://doi.org/10.21685/2500-0578-2022-1-1>
4. Симонович Е.И., Сидельников В.В. К вопросу о состоянии популяции малого и крапчатого суслика на территории Ростовской области // Междунар. журн. эксперим. образования. 2014. № 5 (2). С. 53–54.
5. Миноранский В.А., Сидельников В.В., Симонович Е.И. Состояние популяций сусликов (*Spermophilus* spp.) на Дону // Фунд. исследования. 2015. № 2. С. 277–280.
6. Магеррамов Ш.В., Марцоха К.С., Манджиева В.С. и др. Влияние повышения температуры зимних месяцев на сроки основных фенологических фаз жизнедеятельности *Spermophilus pygmaeus* Pallas, 1778 на территории Прикаспийской низменности (Астраханская область) // Изв. Саратовского ун-та. Серия: Химия. Биология. Экология. 2021. Т. 21. № 1. С. 72–77. <https://doi.org/10.18500/1816-9775-2021-21-1-72-77>
7. Шилова С.А., Савинецкая Л.Е., Неронов В.В. Динамика численности и биомассы малого суслика (*Spermophilus pygmaeus* Pall., 1778) в пастбищных экосистемах Калмыкии за 28-летний период // Аридные экосистемы. 2009. Т. 15. № 2 (38). С. 28–38.
8. Удовиков А.И. Динамика эпизоотической активности природных очагов чумы Европейского юго-востока России: прогноз на начало XXI столетия: Автореф. дис. докт. биол. наук. Саратов: Росс. научно-исследоват. противочумный ин-т “Микроб”, 2010. 46 с.
9. Евченко Ю.М., Ляпустина Л.В., Грижебовский Г.М., Агапатов Д.С. Чрезвычайные ситуации на природно-очаговых по чуме территориях Северного Кавказа // Вестник Рос. военно-мед. академии. 2014. № 2 (46). С. 120–123.
10. Наумов Н.П. Механизмы взаимодействия популяций (на примере некоторых млекопитающих и птиц) // Рус. орнитол. журнал. 2010. Т. 19. Экспресс-выпуск 576. С. 1003–1024.
11. Титов С.В., Симаков М.Д., Чернышова О.В. и др. Генетические особенности популяционной структуры большого суслика (*Spermophilus major* Pall.) в связи с динамикой населения во времени // Russ. J. Ecosystem Ecol. 2020. V. 5. № 4. <https://doi.org/10.21685/2500-0578-2020-4-3>
12. Абатуров Б.Д. Среодообразующие функции населения малого суслика в полупустыне // Суслики Евразии (роды *Spermophilus*, *Spermophilopsis*): происхождение, систематика, экология, поведение, сохранение видового разнообразия. Мат. Рос. науч. конф. М.: КМК, 2005. С. 8–9.
13. Тимошенков В.А. Перспективы существования колоний малого суслика в отделении Хомутовская степь Украинского степного природного заповедника // Суслики Евразии (роды *Spermophilus*, *Spermophilopsis*). Мат. рос. науч. конф. М.: КМК, 2005. С. 108–110.
14. Бронсков А., Тимошенков В. Распространение и численность малого суслика (*Spermophilus pygmaeus*) в Донецкой области // Мониторинг териофауны: Тр. териологической школы. 2010. Вып. 10. С. 107–114.

15. Окулова Н.М. Нора малого суслика как консорциум // Паразитология. 2003. Т. 37. № 5. С. 361–380.
16. Труфанов Г.В. Экологическое обоснование методов регуляции численности горного суслика и его блох в Центральном-Кавказском очаге чумы: Автореф. дис. канд. биол. наук. Ставрополь: Ставропольский научно-исслед. противочумный ин-т, 1996. 22 с.
17. Матросова В.А., Савинецкая Л.Е., Шикарова О.Н. и др. Внутри- и межпопуляционный полиморфизм контрольного региона митохондриальной ДНК у крапчатого суслика (*Spermophilus suslicus*) // ДАН. Биология. 2014. Т. 455. № 6. С. 721–726. <https://doi.org/10.7868/S0869565214120238>
18. Harrison R.G., Bogdanowicz S.M., Hoffmann R.S. et al. Phylogeny and evolutionary history of the ground squirrels (Rodentia: Marmotinae) // J. Mamm. Evol. 1993. V. 10. № 3. P. 249–276. <https://doi.org/10.1023/B:Jomm.0000015105.96065>
19. Ermakov O.A., Brandler O.V., Ivanov A.Yu. et al. Riverine barriers and geographic variation in little ground squirrel (*Spermophilus pygmaeus*, Sciuridae, Rodentia) based on mitochondrial cytochrome b gene sequences // Russ. J. Theriology. 2023. V. 22. № 1. P. 24–31. <https://doi.org/10.15298/rusjtheriol.22.1.03>
20. Ермаков О.А., Тумов С.В., Савинецкий А.Б. и др. Молекулярно-генетические и палеоэкологические аргументы в пользу конспецифичности малого (*Spermophilus pygmaeus*) и горного (*Spermophilus musicus*) сусликов // Зоол. журнал. 2006. Т. 85. № 12. С. 1474–1483.
21. Ермаков О.А., Симонов Е.П., Сурин В.Л., Тумов С.В. Внутривидовой полиморфизм контрольного региона митохондриальной ДНК и филогеография малого суслика (*Spermophilus pygmaeus*, Sciuridae, Rodentia) // Генетика. 2018. Т. 54. № 11. С. 1316–1326. <https://doi.org/10.1134/S0016675818110048>
22. Котенев Е.С., Котенева Е.А., Кот Л.А. и др. Генетическое разнообразие популяций горного суслика (*Spermophilus musicus* Menetries, 1832; Sciuridae, Rodentia) – основного носителя чумы на территории Центрально-Кавказского высокогорного природного очага // Генетика. 2021. Т. 57. № 6. С. 711–720. <https://doi.org/10.31857/S0016675821060059>
23. Темботова Ф.А., Кононенко Е.П. Млекопитающие степных экосистем под угрозой исчезновения на Северном Кавказе // Изв. Самарского науч. центра РАН. 2017. Т. 19. № 5(2). С. 253–259.
24. Темботов А.К., Шхашимишев Х.Х. Животный мир Кабардино-Балкарии. Нальчик: Эльбрус, 1984. 191 с.
25. Красная книга Кабардино-Балкарской Республики. Нальчик: ООО “Печатный двор”, 2018. 496 с.
26. Израилова Г.Р., Халидов А.Х., Халилов Р.А., Адиева А.А. Экология малого суслика на территории Республики Дагестан // Междунар. журн. приклад. и фундамент. исследований. 2013. № 9. С. 97–99.
27. Gür H., Perktas U., Gür M.K. Do climate-driven altitudinal range shifts explain the intraspecific diversification of a narrow ranging montane mammal, Taurus ground squirrels? // Mammal Res. 2017. V. 63 (2). P. 197–211. <https://doi.org/10.1007/s13364-017-0347-8>
28. Gündüz İ., Jaarola M., Tez C. et al. Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, Sciuridae, Rodentia) in Turkey, with a description of a new species // Mol. Phylogenet. and Evol. 2007. V. 43. № 3. P. 916–935. <https://doi.org/10.1016/j.ympev.2007.02.021>
29. Steppan S.J., Akhverdyan M.R., Lyapunova E.A. et al. Molecular phylogeny of the marmots (Rodentia: Sciuridae): Tests of evolutionary and biogeographic hypotheses // Syst. Biol. 1999. V. 48. № 4. P. 715–734.
30. Hall T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT // Nucl. Ac. Symp. Ser. 1999. V. 41. P. 95–98.
31. Villesen P. FaBox: An online toolbox for FASTA sequences // Mol. Ecol. Notes. 2007. V. 7. № 6. P. 965–968. <https://doi.org/10.1111/j.1471-8286.2007.01821.x>
32. Ronquist F., Huelsenbeck J.P. MRBAYES3: Bayesian phylogenetic inference under mixed models // Bioinformatics. 2003. V. 19. P. 1572–1574.
33. Bandelt H.J., Forster P., Röhl A. Median-Joining Networks for inferring intraspecific phylogenies // Mol. Biol. Evol. 1999. V. 16. № 1. P. 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
34. Excoffier L., Lischer H. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows // Mol. Ecol. Res. 2010. V. 10 (3). P. 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
35. Suchard M.A., Lemey P., Baele G. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10 // Virus Evol. 2018. V. 4(1). <https://doi.org/10.1093/ve/vey016>
36. Tamura K., Stecher G., Peterson D. et al. Mega 6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. and Evol. 2023. V. 30. № 12. P. 2725–2729. <https://doi.org/10.1093/molbel/mst197>
37. Yin Y., Jiang W., Zhang Z. et al. The divergence of small mammals in Xinjiang, China, as revealed by phylogenetic analyses of COI and Cytb // Animal Biology. 2014. V. 64. P. 163–176. <https://doi.org/10.1163/15707563-00002435>
38. Амбрасон Н.И., Родченкова Е.Н., Костыгов А.Ю. Генетическая изменчивость и филогеография рыжей полевки (*Clethrionomys glareolus*, Arvicolinae, Rodentia) на территории России с анализом зоны интрогрессии мтДНК близкородственного вида

- красной полевки (*Cl. rutilus*) // Генетика. 2009. Т. 45. № 5. С. 610–623.
39. Челомина Г.Н., Атопкин Д.М. Молекулярно-генетические свидетельства глубокого филогенетического разрыва между европейской и азиатской расами малой лесной мыши по данным изменчивости гена цитохрома b мтДНК // Мол. биология. 2010. Т. 44. № 5. С. 792–803.
 40. Кимура М. Молекулярная эволюция: теория нейтральности. М.: Мир, 1985. 398 с.
 41. Avise J.C. Molecular markers, natural history and evolution. N.Y.: Chapman Hall, 1994. 511 p.
 42. Nei N., Maruyama T., Chakraborty R. The bottleneck effect and genetic variation in populations // Evolution. 1975. V. 29. № 1. P. 1–10.
 43. Ibragim K.M., Nikols R.A., Hewitt G.M. Spatial patterns of genetic variation generated by different forms of dispersal during range expansions // Heredity. 1996. V. 77. P. 282–291.
 44. Дятлов А.И., Петров П.А., Голубев П.Д., Труфанов Г.В. О структуре ареала малых сусликов (*Citellus pygmaeus* Pall., 1778) в Приэльбрусье // Экология. 1980. № 5. С. 77–83.
 45. Канустина С.Ю., Брандлер О.В., Адиya Ya. Филогения сусликов рода *Spermophilus* и положение алашаньского суслика (*Spermophilus alashanicus*. Büchner, 1888) на филогенетическом древе палеарктических “короткохвостых” сусликов // Мол. биология. 2015. Т. 49. № 3. С. 442–448. <https://doi.org/10.7868/S0026898415030088>
 46. Мадьянов Н.Н. Ландшафтная структура Центрально-Кавказского высокогорного природного очага чумы: на примере Кабардино-Балкарской Республики: Автореф. дис. ... канд. географ. наук. Нальчик: Высокогорный геофиз. ин-т Федер. службы по гидрометеор. и мониторингу окр. среды, 2006. 23 с.
 47. Avise J.C. Phylogeography: The history and formation of Species. Cambridge, London: Harvard Univ. Press, 2000. 447 p.
 48. Цвирка М.В., Кораблев В.П. К вопросу о хромосомном видообразовании на примере малого (Spermophilus pygmaeus (Pallas, 1832) и горного кавказского (*Spermophilus musicus* (Menetries, 1832)) сусликов (Rodentia, Sciuridae) // Зоол. журнал. 2014. Т. 93. № 7. С. 917–925. <https://doi.org/10.7868/S0044513414070150>
 49. Верещагин Н.К. Млекопитающие Кавказа. История формирования фауны. М., Л.: Изд-во АН СССР, 1959. 704 с.
 50. Westemeier R., Brawn J., Simpson S. et al. Tracking the long-term decline and recovery of an isolated population // Science. 1998. V. 282. № 5384. P. 1695–1698. <https://doi.org/10.1126/science.282.5394.1695>
 51. Markert J.A., Champlin D.M., Gutjahr-Gobell R. Population genetic iversity and fitness in multiple environments // BMC Evol. Biol. 2010. V. 10. № 205. P. 1–13.

Genetic Diversity Of The Little Ground Squirrel *Spermophilus pygmaeus* Pallas, 1779 (Sciuridae, Rodentia) In The Northern Caucasus

F. A. Tembotova¹, M. S. Gudova¹, A. Kh. Amshokova^{1, *}, A. Kh. Khalidov²

¹Tembotov Institute of Ecology of Mountain Territories, Russian Academy of Sciences, Nalchik, 360051 Russia

²Dagestan Anti-Plague Station, Makhachkala, 367010 Russia

*e-mail: h.a.amshokova@mail.ru

Based on the analysis of a fragment of the cytochrome b (cytb) gene of mitochondrial DNA (mtDNA), the genetic diversity of the little ground squirrel *Spermophilus pygmaeus* Pallas, 1779 of the Central and Eastern Caucasus was studied. Phylogenetic analysis revealed the existence of two clusters A and B within the western clade of *S. pygmaeus*. Cluster A is formed by haplotypes of ground squirrels from the Eastern Caucasus and the right bank of the river. Volga (Ermakov et al., 2023), and B – only haplotypes of Central Caucasian animals. The distance between clusters A and B reaches 1.3%. The relatively isolated position on the phylogenetic tree of the ground squirrel population of the Central Caucasus, the absence of identical haplotypes in Central and East Caucasian animals, and the distances obtained indicate genetic heterogeneity of the ground squirrel in the North Caucasus. A decrease in haplotypic and nucleotide variability was noted in the Central Caucasian populations of the lesser gopher as compared to those from the Eastern Caucasus, which in general indicates the low viability of *S. pygmaeus* inhabiting the mountains of the Central Caucasus.

Keywords: little ground squirrel, *Spermophilus pygmaeus*, cytochrome b (cytb), mitochondrial DNA, Central and Eastern Caucasus.

ЭКСПРЕССИЯ ГЕНА β 1-АДРЕНОРЕЦЕПТОРА У ПАЦИЕНТОВ С ФИБРИЛЛЯЦИЕЙ ПРЕДСЕРДИЙ ДО И ПОСЛЕ ХИРУРГИЧЕСКОГО ЛЕЧЕНИЯ

© 2024 г. В. О. Попова¹, *, Э. Ф. Муслимова¹, Т. Ю. Реброва¹, Е. А. Арчаков¹, Р. Е. Баталов¹, С. А. Афанасьев¹

¹Научно-исследовательский институт кардиологии, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, 634012 Россия

*e-mail: popovalerie@yandex.ru

Поступила в редакцию 04.12.2023 г.

После доработки 29.01.2024 г.

Принята к публикации 30.01.2024 г.

Фибрилляция предсердий (ФП) является одним из самых частых нарушений ритма сердца и приводит к сердечной недостаточности. Избыточная симпатическая активность усугубляет аритмогенные процессы на клеточном и тканевом уровнях. Гиперсимпатикотония приводит к десенситизации β 1-адренорецепторов (β 1-АР), развивается негативная регуляция вплоть до прекращения синтеза новых рецепторов. Процессы десенситизации могут косвенно отражаться в повышении или снижении экспрессии гена *ADRB1*, кодирующего β 1-АР. Проведен анализ уровня относительной экспрессии гена β 1-АР *ADRB1* в лейкоцитах 52 пациентов с разными формами ФП в динамике: до лечения, через трое суток, 3 и 12 месяцев после аблации. До аблации экспрессия оказалась сниженной у пациентов с персистирующей ФП по сравнению с группами с другими формами ФП ($p = 0.026$). При исследовании динамики выявлено значимое снижение уровня экспрессии у пациентов с пароксизмальной ФП на точке трое суток по сравнению с уровнем до аблации ($p = 0.003$) с дальнейшим повышением значений до исходного уровня на точке 12 месяцев ($p = 0.021$) относительно точки “трое суток”. У пациентов с персистирующей ФП уровень экспрессии значимо повышался через три месяца после аблации ($p = 0.046$) по сравнению с уровнем до операции. Пациенты с длительно персистирующей ФП после аблации отличались тенденцией к снижению экспрессии.

Ключевые слова: фибрилляция предсердий, радиочастотная аблация, β 1-адренорецепторы, ген *ADRB1*, экспрессия.

DOI: 10.31857/S0016675824070067 **EDN:** BIERNQ

Фибрилляция предсердий (ФП) является одним из самых частых нарушений ритма сердца. Распространенность ФП в европейской части РФ составляет 2.04%, однако она увеличивается при наличии у пациентов хронических сердечно-сосудистых заболеваний (ССЗ): 12.3% у больных хронической сердечной недостаточностью (ХСН), 9.4% при наличии стабильной стенокардии, 8.2% у пациентов с сахарным диабетом и 4.2% у пациентов с артериальной гипертензией (АГ). Также частота ФП увеличивается с возрастом: от 0.1% в возрасте 20–29 лет до 9.6% в возрастной группе 80–89 лет [1]. Для пациентов с ФП характерно явное уменьшение толерантности к физическим нагрузкам, а также возникновение и прогрессирование дисфункции левого желудочка и сердечной недостаточности [2].

Показано, что на мембранах кардиомиоцитов преобладают β 1-адренорецепторы (β 1-АР), их активизация сопровождается повышением проводимости, сократимости, возбудимости и автоматизма сердечной мышцы [3, 4]. Стимуляция β 1-АР запускает каскад реакций: аденилатциклаза образует комплекс с G-протеином, повышается содержание циклического аденозин-3,5-монофосфата, активируется протеинкиназа А, которая фосфорилирует фосфоламбан, кальциевые каналы L-типа, тропонин I и калиевые каналы [5, 6]. Фосфорилирование кальциевых каналов L-типа и фосфоламбана, регулирующего работу Ca^{2+} -АТФазы саркоплазматического ретикулума, увеличивает присутствие свободных ионов кальция в миоплазме, что усиливает инотропию. Фосфорилирование тропонина I облегчает распад актин-миозиновых мостиков

и диссоциацию комплекса кальций-тропонина С, что ускоряет расслабление мышц [7].

Гиперсимпатикотония, характерная для многих хронических ССЗ, сопровождается разобщением $\beta 1$ -АР с аденилатциклазой, что может привести к дальнейшему повышению симпатической активности и десенситизации по принципу “обратной связи” [8]. Избыточная симпатическая активность провоцирует усиление тока ионов Na^+ , Ca^{2+} и K^+ , что увеличивает склонность к аритмогенным ранним и отсроченным постдеполяризациям. Высвобождение норадреналина в локализованной области сердца усугубляет аритмогенные процессы на клеточном и тканевом уровнях [9].

Известно, что $\beta 1$ -АР кодируются геном *ADRB1*, локализованным на 10 хромосоме (10q25.3) [10]. Вследствие повышения или снижения экспрессии гена *ADRB1* может изменяться и представленность $\beta 1$ -АР на мембранах клеток, в том числе кардиомиоцитов, что играет значимую роль в формировании сердечно-сосудистой патологии. В свою очередь, развитие ССЗ приводит к изменению экспрессии и гена *ADRB1* [11], и рецептора. При десенситизации развивается негативная регуляция, и синтез новых рецепторов прекращается [12]. Есть данные о том, что уровень экспрессии гена *ADRB1* в мононуклеарах периферической крови коррелирует с уровнем артериального давления у пациентов с гипертонией, а также с экспрессией этого гена в миокарде [13]. Следовательно, изменение экспрессии гена *ADRB1* в клетках периферической крови может отражать особенности течения ССЗ.

Цель нашего исследования — оценка динамики относительной экспрессии гена *ADRB1* в лейкоцитах пациентов с фибрилляцией предсердий до и в разные сроки после аблации и в зависимости от тяжести заболевания.

МАТЕРИАЛЫ И МЕТОДЫ

В исследуемую выборку включено 52 пациента с ФП. Из них 36 (69.2%) мужчин и 16 (30.8%) женщин. Возраст в выборке составил 50.7 (26.0; 77.0) лет. Клиническая характеристика пациентов представлена в табл. 1.

Наличие ФП определяли по результатам суточного мониторирования ЭКГ [14]. Для оценки степени выраженности клинических проявлений ФП была использована классификация European Heart Rhythm Association (EHRA), согласно которой: EHRA I — ФП не вызывает симптомов; EHRA II (IIa/IIb) — легкие/средневыраженные симптомы ФП, не нарушающие нормальную повседневную активность пациента; EHRA III — выраженные симптомы ФП, нарушающие нормальную повседневную активность пациента; EHRA IV — инвалидизирующие симптомы ФП, приводящие

к невозможности нормальной повседневной активности пациента [2]. Для лечения ФП всем пациентам выполняли радиочастотную или криобаллонную аблацию. Объем вмешательства включал в себя антральную изоляцию устьев легочных вен под контролем циркулярного электрода до полного блока электрического проведения от вены в предсердие и наоборот.

На момент поступления в отделение с целью катетерного лечения ФП у всех пациентов был собран фармакологический анамнез. Коррекция терапии согласно современным клиническим рекомендациям производилась во время текущей госпитализации. У 23 пациентов (44.3%) была диагностирована ХСН. В процессе обследования был проведен тест 6-минутной ходьбы для определения функционального класса (ФК) ХСН в соответствии с классификацией Нью-Йоркской кардиологической ассоциации (NYHA) [15].

В качестве материала для исследования использовали цельную венозную кровь. Образцы крови забирали в вакутейнеры с $\text{K}_3\text{ЭДТА}$ в условиях процедурного кабинета до аблации, через трое суток после аблации, через 3 и 12 месяцев после аблации. Образцы крови до аблации получены у 52 (100%) пациентов, в точке “3 суток” после аблации у 41 (78.8%) пациента, в точке “3 месяца” после аблации у 19 (36.5%) пациентов, в точке “12 месяцев” после аблации у 14 (26.9%) пациентов. Уменьшение количества пациентов в последующих точках связано с более ранней выпиской из отделения и отказом пациента прибыть на повторное обследование через 3 и 12 месяцев.

Выделение РНК из лейкоцитов венозной крови проводили с помощью набора реагентов InnuPREP Blood RNA Kit (Analytik Jena, Германия). Качество образцов РНК оценивали спектрофотометрическим методом по отношению $\text{A}_{260}/\text{A}_{280}$ (NanoVue, Healthcare Bio-Science, Швеция), которое варьировало в пределах от 1.93 до 2.3. Для синтеза кДНК использовали реактивы RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, США). Образцы РНК до дальнейшего исследования хранили при -80°C в низкотемпературном морозильнике HAIER DW-86L728 (Haier BioMedical, Китай) Центра коллективного пользования “Медицинская геномика” (<https://www.tnimc.ru/ckp/>).

У всех пациентов оценен относительный уровень экспрессии гена *ADRB1* (ENSG00000043591) методом полимеразной цепной реакции в режиме реального времени. Использовались следующие праймеры и TaqMan зонды: hADRB1-F = CAGGTGAAGCTCGAAGCCC; hADRB1-R = CTCCCATCCCTTCCCAA; hADRB1-probe = FAM-AAAGCCACGGACCGTTGCAC-BHQ1 (ООО “ДНК-Синтез”, Россия). В качестве референсных были использованы ген

глицеральдегид-3-фосфат-дегидрогеназы *GAPDH* (ENSG00000111640) и ген бета-актина *ACTB* (ENSG00000075624) [16]. Смесь праймеров и зондов для одномоментной амплификации двух генов в одной пробирке предоставлены фирмой ООО “ДНК-Синтез” (Россия). Подбор праймеров осуществлялся так, чтобы была исключена возможность амплификации геномной ДНК за счет расположения праймеров/зонда в разных экзонах. Для повышения эффективности реакции подбирались праймеры/зонд, не содержащие димеры и шпильки. Все праймеры/зонды проходили проверку в системе BLAST. Постановку реакции проводили в триплетах, в одном планшете параллельно проводили реакции для целевого и референсных генов. Для мастер-микса использованы 10' DreamTaq Buffer (Thermo Scientific, США), DreamTaq DNA Polymerase (Thermo Scientific, США), Invitrogen dNTP Set 100 mM (Thermo Scientific, США). Объем реакционной смеси в лунке составлял 25 мкл (из них 1 мкл кДНК). Программа для проведения ПЦР: 1 цикл 95 °C 180 с, 45 циклов 95 °C 15 с и 60 °C 60 с (измерение флуоресценции). В качестве калибратора постановок и для стандартных проб (для расчета эффективности реакции) использована кДНК, синтезированная на базе тотальной РНК человека (Human XpressRef Universal Total RNA, Qiagen, Германия). Расчет уровня экспрессии проводили с применением стандартной кривой, с поправкой на эффективность реакции и калибратор постановок по методу Pfaffl [17]. Эксперименты по экспрессии генов проводили в соответствии с Руководством MIQE [18].

Количественные данные были предварительно проверены на соответствие нормальному закону распределения с помощью критерия Шапиро – Уилка. Дальнейший анализ количественных параметров проводили с помощью критерия Манна – Уитни или теста Краскела – Уоллиса. Анализ зависимых данных проводили с помощью критерия Уилкоксона. Результаты представляли в виде медианы и интерквартильного размаха. Уровень значимости различий принимали $p < 0.05$.

РЕЗУЛЬТАТЫ

Уровень относительной экспрессии гена *ADRB1* до абляции ($n = 52$) составил 0.24 (0.09; 0.49), в точке “3 суток” после абляции ($n = 41$) – 0.16 (0.10; 0.26), в точке “3 месяца” после абляции ($n = 19$) – 0.22 (0.13; 0.39), в точке “12 месяцев” после абляции ($n = 14$) – 0.23 (0.20; 0.41). Значимые различия между точками отсутствовали (рис. 1). Уровень экспрессии гена *ADRB1* у пациентов с разными формами ФП представлен в табл. 2. Оказалось, что до абляции пациенты с разным типом ФП имели значимые различия в экспрессии гена. Наиболее низким этот показатель был у пациентов с персистирующей ФП ($p = 0.026$). Последующие наблюдения тоже показали различия между рассматриваемыми группами пациентов.

Так, на точке “3 суток” после абляции медиана экспрессии гена *ADRB1* у пациентов с персистирующей ФП характеризовалась более чем в три раза увеличением своего значения. Напротив, у пациентов групп с пароксизмальной и длительно персистирующей ФП наблюдалось снижение этого

Таблица 1. Клиническая характеристика пациентов с фибрилляцией предсердий

Параметр	Значение, n (%)
Пароксизмальная/персистирующая/длительно персистирующая фибрилляция предсердий	31 (59.6)/12 (23.1)/9 (17.3)
EHRA I/II/III	9 (17.3)/42 (80.8)/1 (1.9)
Радиочастотная абляция/криоабляция	42 (80.8)/10 (19.2)
ХСН ФК I/II/III	13 (25.0)/7 (13.5)/3 (5.8)
Ишемическая болезнь сердца	9 (17.3)
Гипертоническая болезнь	32 (61.5)
Фракция выброса левого желудочка	65.0 (56.3; 69.0)
Ингибиторы АПФ	23 (44.2)
Бета-адреноблокаторы	24 (46.2)
Гиполипидемические препараты	18 (34.6)
Антикоагулянты	52 (100)
Антиаритмические препараты	43 (82.7)
Амиодарон/соталол/пропафенон/лаппаконитина гидробромид	17 (32.7)/13 (25.0)/8 (15.4)/5 (9.6)

Примечание: ХСН – хроническая сердечная недостаточность, ФК – функциональный класс, АПФ – ангиотензинпревращающий фермент.

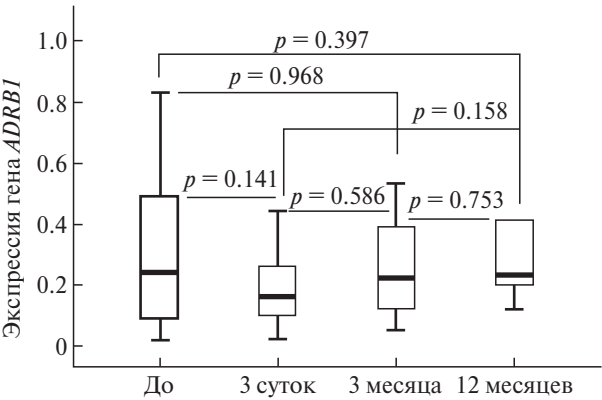


Рис. 1. Относительная экспрессия гена *ADRB1* до и на разных сроках после абляции у пациентов с фибрилляцией предсердий.

показателя более чем в два раза. Изменение экспрессии гена *ADRB1* между точкой до абляции и “3 суток” после у пациентов с пароксизмальной ФП оказалось наиболее выраженным ($p = 0.003$). Однако межгрупповые различия не достигли статистической значимости ($p = 0.067$).

Пациенты с разными формами ФП на сроках в 3 месяца и 12 месяцев также не показали статистически значимых отличий по уровню экспрессии. Но при этом сохранили различие в направленности изменений. Так, через 3 месяца после абляции для группы с персистирующей ФП характерно

дальнейшее увеличение экспрессии гена в 1.6 раз, что превысило значения до абляции в пять раз ($p = 0.046$). Умеренное увеличение экспрессии отмечено и в группе с пароксизмальной ФП. Напротив, в группе с длительно персистирующей ФП продолжалось снижение этого показателя.

На точке “12 месяцев” в группе пациентов, имевших пароксизмальную ФП, было отмечено продолжение увеличения экспрессии. Медиана показателя от падения на третьи сутки после абляции увеличилась уже в три раза ($p = 0.021$), достигнув дооперационного уровня ($p = 0.477$). В группе пациентов, имевших персистирующую ФП, медиана экспрессии снизилась до уровня, наблюдавшегося на третьи сутки после абляции ($p = 0.285$), но оставалась значительно выше своих исходных значений.

ОБСУЖДЕНИЕ

Считается, что около 85% случаев ФП являются вторичными по отношению к другим ССЗ. При этом среди факторов риска возникновения ФП выделяют, прежде всего, ИБС, АГ и сердечную недостаточность. У больных ФП сердечные сокращения становятся неэффективными, прогрессирует дисфункция левого желудочка с развитием сердечной недостаточности [2]. Наджелудочковые аритмии в раннем послеоперационном периоде существенно ухудшают показатели гемодинамики и увеличивают риск тромбоэмболических осложнений.

Таблица 2. Экспрессия гена *ADRB1* (Ме (Q1; Q3)) у пациентов с разными формами фибрилляции предсердий в динамике наблюдений

Форма ФП Срок	Пароксизмальная	Персистирующая	Длительно персистирующая	p
До абляции	0.32 (0.19; 0.73) <i>n</i> = 31	0.09 (0.05; 0.33) <i>n</i> = 12	0.52 (0.21; 0.61) <i>n</i> = 9	0.026
3 сут после	0.13 (0.06; 0.25) <i>n</i> = 22	0.28 (0.12; 1.19) <i>n</i> = 12	0.24 (0.16; 0.53) <i>n</i> = 7	0.067
3 мес. после	0.21 (0.12; 0.40) <i>n</i> = 10	0.47 (0.27; 1.44) <i>n</i> = 6	0.18 (0.05; 1.01) <i>n</i> = 3	0.288
12 мес. после	0.41 (0.23; 1.29) <i>n</i> = 11	0.20 (0.19; 0.81) <i>n</i> = 3	—	0.243
<i>p_x</i> (динамика изменения экспрессии)	<i>p</i> ₁ = 0.003 <i>p</i> ₂ = 0.139 <i>p</i> ₃ = 0.477 <i>p</i> ₄ = 0.327 <i>p</i> ₅ = 0.021 <i>p</i> ₆ = 0.465	<i>p</i> ₁ = 0.099 <i>p</i> ₂ = 0.046 <i>p</i> ₃ = 0.593 <i>p</i> ₄ = 0.753 <i>p</i> ₅ = 0.285 <i>p</i> ₆ = 0.180	<i>p</i> ₁ = 0.237 <i>p</i> ₂ = 1.000 <i>p</i> ₃ = — <i>p</i> ₄ = 1.000 <i>p</i> ₅ = — <i>p</i> ₆ = —	

Примечание: *p*₁ — уровень значимости различий в экспрессии до и 3 сут после операции; *p*₂ — до и 3 мес. после операции; *p*₃ — до и 12 мес. после операции; *p*₄ — через 3 сут и через 3 мес. после операции; *p*₅ — через 3 сут и через 12 мес. после операции; *p*₆ — через 3 мес. и 12 мес. после операции.

Определение уровней нейрогуморальных факторов может стать решающим при определении стратегии лечения пациентов с ФП [19].

Циркулирующие катехоламины стимулируют $\beta 1$ -АР кардиомиоцитов, что приводит к активации аденилатциклазы, увеличению содержания внутриклеточного цАМФ и активации протеинкиназы А. Гиперфосфорилирование рианодиновых рецепторов является одним из механизмов, приводящих к утечке Ca^{2+} из СПР, в результате чего создаются условия для развития систолической и диастолической дисфункций, повышается риск возникновения аритмии [20]. Гиперактивация симпатoadrenalовой системы приводит и к патологическому ремоделированию β -АР. Происходит разобщение рецептора с G-белками, а затем и снижение плотности рецепторов на клеточной мембране вплоть до прекращения синтеза рецепторов [9, 21]. $\beta 1$ -АР, кодируемые геном *ADRB1*, опосредуют широкий спектр физиологических реакций сердечно-сосудистой системы и представляют большой интерес для исследователей в качестве терапевтической мишени, в том числе при нарушениях ритма сердца. Высказывается мнение, что особенности β -адренорецепторного аппарата могут быть фактором, определяющим характер патологии сердца у конкретного пациента, и его ответом на проводимое лечение [22].

Сообщается, что изменение экспрессии гена *ADRB1* может влиять на функциональные возможности миокарда при ССЗ. Так, у пациентов с ХСН одновременно наблюдались повышение экспрессии гена *ADRB1* и увеличение сократительного резерва миокарда после сердечной ресинхронизирующей терапии [23]. Показано, что при ХСН применение метопролола тартрата повышало экспрессию $\beta 1$ -АР [24]. По мнению авторов этой работы, такой результат указывал на восстановление прохождения сигнала с улучшением функции миокарда. Однако есть исследования и с противоположным результатом. Их авторы отмечают, что агонисты β -АР и избыточная экспрессия $\beta 1$ -АР вызывали гипертрофию сердца и фиброз *in vivo* [25]. Именно с этим авторы связывают развитие сердечной недостаточности и потенциально летальных аритмий. Сообщается, что среди больных ХСН ишемического генеза пациенты с гипертрофией левого желудочка отличались сниженным по сравнению с лицами без гипертрофии уровнем экспрессии гена *ADRB1* в миокарде [11].

В настоящем исследовании нами был оценен уровень относительной экспрессии гена $\beta 1$ -АР *ADRB1* в лейкоцитах периферической крови пациентов с диагнозом ФП. В сформированной выборке до операции пациенты с персистирующей ФП характеризовались низким уровнем экспрессии гена *ADRB1* относительно пациентов с другими формами ФП. Это может указывать на значительное

изменение адренореактивности их организма и не противоречит представлениям о том, что в условиях патологии экспрессия *ADRB1* может влиять на состояние миокарда. В нашем исследовании независимо от типа ФП проведение процедуры аблации уже через трое суток ознаменовалось выраженным изменением экспрессии *ADRB1*. Такой результат может быть следствием действия одного и того же фактора. Скорее всего, это обусловлено стрессовым воздействием на организм пациентов при процедуре аблации. С другой стороны, быстрота и выраженность выявленных изменений указывают на то, что независимо от рассматриваемых форм ФП рецепторный аппарат сохраняет достаточные реактивные возможности. Это предположение хорошо согласуется и с последующей динамикой изменений экспрессии *ADRB1*. Для каждого типа ФП нами были зафиксированы выраженные изменения этого показателя.

Можно предположить, что при персистирующей ФП проведенное катетерное лечение способствовало уменьшению патологического влияния на $\beta 1$ -адренорецепторный аппарат клеток и восстановлению экспрессии гена до уровня пароксизмальной ФП. У пациентов с длительно персистирующей ФП наблюдалась тенденция к снижению экспрессии гена *ADRB1*. Однако выявленные различия не были статистически значимыми в связи с малым количеством пациентов с этой формой ФП на поздних этапах наблюдения.

Ранее в той же выборке пациентов мы оценивали уровень β -адренореактивности мембран эритроцитов (β -АРМ) [26]. Этот показатель оказался сопоставим в группах с разными формами ФП как до, так и после хирургического лечения. При этом отсутствовали существенные изменения и в динамике β -АРМ в каждой из групп. Таким образом, несмотря на значимые изменения уровня экспрессии гена *ADRB1* в лейкоцитах, изменения адренореактивности организма у пациентов с ФП выявлено не было.

Ограничением нашего исследования можно считать относительно небольшую выборку и выход пациентов из исследования на длительных сроках наблюдения. Кроме этого, для более полной оценки влияния аблации на $\beta 1$ -адренорецепторный аппарат клеток и адренореактивность организма в дальнейшем необходимо изучить уровень экспрессии самих рецепторов на мембранах лейкоцитов. Совпадение динамик экспрессии гена *ADRB1* и самих рецепторов у пациентов с разными формами ФП позволит выявить фундаментальные механизмы постоперационного течения разных нозологических форм. В свою очередь, полученные результаты могут быть учтены при прогнозе исхода или состояния пациента в отдаленном периоде после аблации.

Таким образом, пациенты с персистирующей ФП, в отличие от пациентов с пароксизмальной или длительно персистирующей ФП, характеризовались сниженным уровнем экспрессии гена $\beta 1$ -адренорецептора *ADRB1*. Уровень экспрессии гена *ADRB1* после процедуры аблации зависел от формы ФП у этих пациентов. При этом для пациентов с пароксизмальной ФП характерно снижение экспрессии на третьи сутки после аблации с дальнейшим возвращением этого показателя к исходному уровню. У пациентов с персистирующей ФП экспрессия, наоборот, характеризовалась выраженным превышением дооперационных значений. Пациенты с длительно персистирующей ФП после аблации отличались прогрессирующей тенденцией к снижению экспрессии гена *ADRB1*.

Работа выполнена в рамках темы ФНИ № 122020300183-4.

Исследование одобрено Этическим комитетом НИИ кардиологии ТНИМЦ, протокол № 208 от 20.01.2021 года. Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников было получено информированное добровольное согласие. Все обследованные — совершеннолетние.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Мареев Ю.В., Поляков Д.С., Виноградова Н.Г. и др. ЭПОХА: Эпидемиология фибрилляции предсердий в репрезентативной выборке европейской части Российской Федерации // Кардиология. 2022. Т. 62. № 4. С. 12–19. <https://doi.org/10.18087/cardio.2022.4.n1997>
2. Hindricks G., Potpara T., Dagres N. et al. 2020 ESC Guidelines for the diagnosis and management of a trial fibrillation developed in collaboration with the EACTS: The Task Force for the diagnosis and management of atrial fibrillation of the ESC Developed with the special contribution of the EHRA of the ESC // Europ. Heart J. 2020. V. 42. № 5. P. 373–498. <https://doi.org/10.1093/eurheartj/ehaa612>
3. Афанасьев С.А., Реброва Т.Ю., Муслимова Э.Ф., Борисова Е.В. Ассоциация полиморфных вариантов гена *ADRB1* с сократительной дисфункцией миокарда и адренореактивностью эритроцитов у пациентов с нарушениями ритма // Росс. Кардиол. журнал. 2019. Т. 24. № 7. С. 47–52. <https://doi.org/10.15829/1560-4071-2019-7-47-52>
4. Афанасьев С.А., Лишманов Ю.Б., Кондратьев Б.Ю. и др. Фармакологические доказательства нарушения адренореактивности миокарда в условиях хронической коронарной ишемии // Физиология человека. 1997. Т. 23. № 5. С. 58–62.
5. Barnes P.J. $\beta 2$ -Agonists, anticholinergics, and other nonsteroid drugs // Clinical Respiratory Medicine. 2008. P. 471–482. <https://doi.org/10.1016/B978-032304825-5.10036-4>
6. Kume H., Takai A., Tokuno H., Tomita T. Regulation of Ca^{2+} dependent K^{+} -channel activity in tracheal myocytes by phosphorylation // Nature. 1989. V. 341. P. 152–154. <https://doi.org/10.1038/341152a0>
7. Bers D.M. Cardiac excitation-contraction coupling // Nature. 2002. V. 415. P. 198–205. <https://doi.org/10.1038/415198a>
8. Стрюк Р.И., Длусская И.Г. Адренореактивность и сердечно-сосудистая система. М.: Медицина, 2003. 160 с.
9. Grandi E., Ripplinger C.M. Antiarrhythmic mechanisms of beta-blocker therapy // Pharmacol. Res. 2019. V. 146. P. 104274. <https://doi.org/10.1016/j.phrs.2019.104274>
10. Yang-Feng T.L., Xue F.Y., Zhong W.W. et al. Chromosomal organization of adrenergic receptor genes // PNAS. 1990. V. 87. № 4. P. 1516–1520. <https://doi.org/10.1073/pnas.87.4.1516>
11. Муслимова Э.Ф., Реброва Т.Ю., Кондратьева Д.С. и др. Экспрессия гена $\beta 1$ -адренорецептора (*ADRB1*) в миокарде больных хронической сердечной недостаточностью // Генетика. 2021. Т. 57. № 11. С. 1297–1305. <https://doi.org/10.1134/S1022795421110089>
12. Леванов А.Н., Игнатъев И.В., Сычев Д.А. и др. Связь генетического полиморфизма бета-адренорецепторов с эффективностью терапии бета-адреноблокаторами у больных с сердечно-сосудистой патологией // Саратовский научно-мед. журнал. 2009. Т. 5. № 1. С. 41–44.
13. Sun X., Zhou M., Wen G. et al. Paroxetine attenuates cardiac hypertrophy via blocking *GRK2* and *ADRB1* interaction in hypertension // J. Am. Heart Association. 2021. V. 10. № 1. <https://doi.org/10.1161/JAHA.120.016364>
14. Аракелян М.Г., Бокерия Л.А., Васильева Е.Ю. и др. Фибрилляция и трепетание предсердий. Клинические рекомендации 2020 // Росс. Кардиол. журнал. 2021. Т. 26. № 7. С. 190–260. <https://doi.org/10.15829/1560-4071-2021-4594>
15. Мареев В.Ю., Фомин И.В., Агеев Ф.Т. и др. Клинические рекомендации ОССН — РКО — РНМОТ. Сердечная недостаточность: хроническая (ХСН) и острая декомпенсированная (ОДСН). Диагностика, профилактика и лечение // Кардиология. 2018. Т. 58. № S6. С. 8–158. <https://doi.org/10.18087/cardio.2475>

16. *Molina C.E., Jacquet E., Ponien P. et al.* Identification of optimal reference genes for transcriptomic analyses in normal and diseased human heart // *Cardiovascular Research*. 2018. V. 114. № 2. P. 247–258. <https://doi.org/10.1093/cvr/cvx182>
17. *Pfaffl M.W.* Quantification strategies in real-time PCR // *A–Z of Quantitative PCR*. CA, USA: IUL, 2004. P. 87–112.
18. *Bustin S., Benes V., Garson J. et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments // *Clinical Chemistry*. 2009. V. 55. № 4. P. 611–622. <https://doi.org/10.1373/clinchem.2008.112797>
19. *Евтушенко В.В., Евтушенко А.В., Павлюкова Е.Н. и др.* Дисфункция синусового узла у пациентов с длительно персистирующей фибрилляцией предсердий: клиничко-лабораторные параллели // *Сиб. журнал клин. и эксперим. медицины*. 2021. Т. 36. № 1. С. 123–128. <https://doi.org/10.29001/2073-8552-2021-36-1-123-128>
20. *Dridi H., Kushnir A., Zalk R. et al.* Intracellular calcium leak in heart failure and atrial fibrillation: A unifying mechanism and therapeutic target // *Nature Reviews. Cardiology*. 2020. V. 17. № 11. P. 732–747. <https://doi.org/10.1038/s41569-020-0394-8>
21. *Вейн А.М., Вознесенская Т.Г., Воробьева О.В. и др.* Вегетативные расстройства: клиника, диагностика, лечение: монография. М: Мед. информ. агентство, 1998. 643 с.
22. *Amare A.T., Schubert K.O., Klingler-Hoffmann M. et al.* The genetic overlap between mood disorders and cardiometabolic diseases: A systematic review of genome wide and candidate gene studies // *Citation: Translational Psychiatry*. 2017. V. 7. № 1. P. 1–12. <https://doi.org/10.1038/tp.2016.261>
23. *Vanderheyden M., Mullens W., Delrue L. et al.* Endomyocardial upregulation of beta1 adrenoreceptor gene expression and myocardial contractile reserve following cardiac resynchronization therapy // *J. of Cardiac Failure*. 2008. V. 14. № 2. P. 172–178. <https://doi.org/10.1016/j.cardfail.2007.10.016>
24. *Белоусов Ю.Б.* Место метопролола в лечении хронической сердечной недостаточности // *Фарматека*. 2002. № 4. С. 33–43.
25. *Engelhardt S., Hein L., Wiesmann F., Lohse M.J.* Progressive hypertrophy and heart failure in beta1-adrenergic receptor transgenic mice // *PNAS of the USA*. 1999. V. 96. № 12. P. 7059–7054. <https://doi.org/10.1073/pnas.96.12.7059>
26. *Муслимова Э.Ф., Попова В.О., Реброва Т.Ю. и др.* В-адренореактивность мембран эритроцитов у пациентов с дилатацией левого или правого предсердий на фоне фибрилляции предсердий // *Бюл. сибирской медицины*. 2023. Т. 22. № 3. С. 61–67. <https://doi.org/10.20538/1682-0363-2023-3-61-6>

Expression of the $\beta 1$ -Adrenoreceptor Gene In Patients With Atrial Fibrillation Before And After Surgical Treatment

V. O. Popova¹*, E. F. Muslimova¹, T. Yu. Rebrova¹,
E. A. Archakov¹, R. E. Batalov¹, S. A. Afanasiev¹

¹Cardiology Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences,
Tomsk, 634045 Russia

*e-mail: popovalerie@yandex.ru

Atrial fibrillation (AF) is one of the most frequent cardiac arrhythmias and leads to heart failure. Excessive sympathetic activity aggravates arrhythmogenic processes at the cellular and tissue levels. Hypersympathicotonia leads to desensitization of $\beta 1$ -adrenoreceptors ($\beta 1$ -AR), negative regulation develops until the synthesis of new receptors stops. Desensitization processes may be indirectly reflected in an increase or decrease in the expression of the *ADRB1* gene encoding $\beta 1$ -AR. The level of relative expression of the $\beta 1$ -AP *ADRB1* gene in leukocytes of 52 patients with different forms of AF was analyzed in dynamics: before treatment, 3 days, 3 and 12 months after ablation. Before ablation, expression was reduced in patients with persistent AF compared to groups with other forms of AF ($p = 0.026$). The study of dynamics revealed a significant decrease in the level of expression in patients with paroxysmal AF at the point of 3 days compared with the level before ablation ($p = 0.003$), with a further increase in values to the baseline at the point of 12 months ($p = 0.021$) relative to the point of 3 days. In patients with persistent AF, the expression level significantly increased 3 months after ablation ($p = 0.046$) compared to the level before surgery. Patients with long-term persistent AF after ablation had a tendency to decrease expression.

Keywords: atrial fibrillation, radiofrequency ablation, $\beta 1$ -adrenergic receptors, *ADRB1* gene, expression.

ПОЛИМОРФИЗМ ГЕНОВ АНТИОКСИДАНТОВ И ИЗБЫТОЧНЫЙ ВЕС У ДЕТЕЙ

© 2024 г. М. А. Шкурят^{1, 2}, Е. В. Машкина^{1*},
Н. П. Милютин¹, Е. Д. Теплякова³, Т. П. Шкурят^{1, 2}

¹Южный федеральный университет, Ростов-на-Дону, 344090 Россия

²Медицинский центр “Наука”, Ростов-на-Дону, 344000 Россия

³Ростовский государственный медицинский университет, Ростов-на-Дону, 344022 Россия

E-mail: lenmash@mail.ru

Поступила в редакцию 30.01.2024 г.

После доработки 21.02.2024 г.

Принята к публикации 29.02.2024 г.

Одним из компонентов патологического процесса, приводящего к развитию ожирения, является окислительный стресс. Уровень образования свободно-радикальных продуктов контролируется антиоксидантной системой. Полиморфизм генов антиоксидантов влияет на уровень и/или активность кодируемых ферментов. Целью работы было исследовать ассоциацию однонуклеотидных замен в генах антиоксидантной системы с риском формирования избыточной массы тела у детей и подростков. Материалом для исследования послужили образцы ДНК 279 детей с избыточным весом и 131 ребенка из контрольной группы. Проводили генотипирование по rs6721961 (–617G>T) *NFE2L2*, rs 4998557 (7958G>A) *SOD1*, rs4880 (47C>T Ala16Val) *SOD2*, rs1001179 (–262C>T) *CAT*, rs713041 (718C>T) *GPX4*, rs662 (Gln192Arg) *PON1*. Показано, что генотип –617GT по rs6721961 *NFE2L2* среди детей с избыточной массой тела регистрируется статистически значимо реже. Повышенный риск формирования избыточной массы тела выявлен для гетерозигот –262CT по rs1001179 *CAT* и аллеля –262T. В результате анализа межгенных взаимодействий выявлен 6-тилокусный генотип, ассоциированный со снижением риска формирования избыточной массы тела.

Ключевые слова: избыточный вес и ожирение у детей, антиоксиданты, полиморфизм генов, *SOD1*, *SOD2*, *CAT*, *GPX4*, *PON1*, *NFE2L2*, межгенные взаимодействия.

DOI: 10.31857/S0016675824070078 **EDN:** BICGJI

Ожирение – одно из самых распространенных метаболических нарушений в мире, которое развивается из-за дисбаланса потребления и расхода энергии. Ожирение распространено среди лиц обоих полов и всех возрастов, социально-экономических слоев и этнических групп. Распространенность детского ожирения во всем мире резко возросла за последние 30 лет. По оценкам Всемирной организации здравоохранения [1] детское ожирение в настоящее время затрагивает более 41 млн детей в возрасте до пяти лет. Среди детей и подростков от 5 до 19 лет в 2016 г. избыточный вес имели 213 млн, а от ожирения страдало до 124 млн [2]. При этом около трети детей с ожирением имеют некоторые признаки метаболического синдрома [3]. По данным мультицентрового исследования средняя частота избыточной массы тела и ожирения среди детей в Российской Федерации составила 19.9 и 5.6% соответственно [4]. Избыточный вес и ожирение у детей и подростков могут повышать

риск развития кардиометаболических синдромов, сахарного диабета 2-го типа, инсулинорезистентности [5].

Патогенез ожирения включает целый комплекс взаимосвязанных факторов риска, в том числе экологические, генетические, метаболические, психосоциальные [6, 7]. Ожирение, в том числе у детей и подростков, может быть ассоциировано с нарушением редокс-статуса, что выражается в снижении уровня эндогенных и экзогенных антиоксидантов и повышении уровня активных форм кислорода [8, 9]. Активные формы кислорода (АФК) играют важную роль в развитии ожирения и его метаболических осложнений. АФК влияют на концентрацию молекул, участвующих в воспалении, что связано с большим количеством адипоцитов, способствующих адипо- и липогенезу, стимулируют дифференцировку адипоцитов, регулируют энергетический баланс в нейронах гипоталамуса, контролирующих аппетит [10].

Контроль за уровнем АФК в тканях осуществляет антиоксидантная система, которая включает эндогенные (например, глутатион, убихинон, тиоредоксин, ураты) и экзогенные (например, аскорбиновая кислота, α -токоферол, каротиноиды) соединения, а также антиоксидантные ферменты. Наиболее важными антиоксидантными ферментами являются супероксиддисмутаза (SOD1, SOD2, SOD3), каталаза (КАТ), глутатионпероксидаза (GPX). Супероксиддисмутаза представляет собой металлоферменты, ответственные за детоксикацию супероксидных радикалов с образованием H_2O_2 . Пероксид водорода, в свою очередь, обезвреживается каталазой и пероксидазами. К антиоксидантам относятся и параоксоназы (PON, арилдиалкилфосфатазы). PON1 представляет собой эстеразу, которая вырабатывается в печени и связана с липопротеинами высокой плотности (ЛПВП). Фермент гидролизует большой спектр субстратов, в том числе и липопероксиды. По данным литературы показано, что генетически обусловленный уровень активности PON1 может быть вовлечен в изменение продукции адипокинов и миокинов у детей при ожирении [11].

Ядерный фактор 2 эритроидного происхождения (Nrf2, Nuclear factor erythroid 2-related factor 2), принадлежащий к семейству транскрипционных факторов Cap'n'Collar (CNC), подсемейству факторов транскрипции лейциновой молнии (bZIP), является ключевым регулятором клеточного ответа на окислительный стресс путем контроля экспрессии антиоксидантных и детоксикационных ферментов для устранения избытка АФК [12, 13]. Nrf2 в ядре связывается с элементом антиоксидантного ответа (ARE; GTGACNNNGC), присутствующим в регуляторной области как гена NFE2L2, так и генов-мишеней Nrf2, кодирующих антиоксидантные и детоксикационные белки, тем самым индуцируя транскрипцию гена. Nrf2 способствует не только системной детоксикации АФК и уменьшению окислительного стресса в различных тканях, но также дифференцировке и метаболизму адипоцитов [14].

Известно, что генетические вариации, такие как однонуклеотидные замены (SNP), могут влиять на функционирование кодируемых белков и изменять риск развития мультифакторных заболеваний человека. В результате исследований ассоциации однонуклеотидных замен с формированием избыточной массы тела и ожирения выявлены многочисленные генетические маркеры полигенной формы ожирения [15–19].

Однако частоты регистрации однонуклеотидных замен, ассоциированных с ожирением, имеют популяционные особенности, что обуславливает необходимость проведения репликативных исследований для конкретных групп населения [20–22].

Цель работы — проанализировать ассоциацию однонуклеотидных замен в генах антиоксидантной системы с риском формирования избыточной массы тела у детей и подростков Ростова-на-Дону.

МАТЕРИАЛЫ И МЕТОДЫ

Материалом для исследования послужили образцы ДНК, выделенные из клеток крови 410 детей и подростков в возрасте от 4 до 16 лет. Дети и подростки были разделены на две группы: контрольная группа (131 человек с нормальным индексом массы тела ($ИМТ = 18.6 \pm 0.1$)) и группа детей с избыточным весом (279 человек, $ИМТ = 25.8 \pm 0.19$). Пороговые значения ИМТ для определения избыточного веса и ожирения определяли с учетом рекомендаций Всемирной организации здравоохранения: дети в возрасте до 5 лет относятся к группе с избыточным весом и ожирением, если ИМТ превышает медианное значение, указанное в Стандартных показателях физического развития детей (ВОЗ), более чем на два стандартных отклонения. У детей старше 5 лет избыточный вес и ожирение диагностируются, если ИМТ превышает медианное значение более чем на одно стандартное отклонение [23]. Распределение детей по полу было следующее: в контроле девочки составили 46%, мальчики — 54%; среди детей с избыточным весом доля девочек составила 45%, мальчиков — 55%.

Формирование исследуемых групп было проведено на базе Детской городской поликлиники № 4 и медицинского центра “Наука” г. Ростова-на-Дону. Все процедуры были выполнены с соблюдением Хельсинкской декларации Всемирной медицинской ассоциации “Этические принципы проведения научных медицинских исследований с участием человека” (World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects) и статей 20, 22, 23 Федерального закона “Об основах охраны здоровья граждан в Российской Федерации” от 21.11.2011 № 323-ФЗ (ред. от 26.05.2021). Согласно нормам биоэтики, родители детей и подростков, участвующих в исследовании, были предварительно подробно проинформированы о цели и содержании исследования и подписали письменную форму информированного согласия. Проведение исследования было утверждено комитетом по биоэтике Академии биологии и биотехнологии им. Д.И. Иванова Южного федерального университета (протокол № 2 от 17.01.2018).

ДНК из клеток крови выделяли термокоагуляционным методом с использованием реагента “ДНК-экспресс—кровь” (Литех, Россия).

В исследование были включены однонуклеотидные замены, влияющие на транскрипционную активность гена или функциональную активность кодируемой белковой молекулы. Кроме того,

согласно данным базы HaploReg (v.4.2), включенные в исследование SNP обладают регуляторным потенциалом, поскольку способны оказывать влияние на структуру хроматина, ДНК-белковые взаимодействия [24].

Анализ однонуклеотидных замен rs4998557 (7958G>A) гена *SOD1* и rs662 (Gln192Arg) гена *PON1* проводили методом аллель-специфичной амплификации с использованием реагентов SNP-экспресс (Литех).

Однонуклеотидные замены rs4880 (47C>T Ala16Val) гена *SOD2*, rs1001179 (–262C>T) гена *CAT*, rs713041 (718C>T) гена *GPX4* анализировали методом ПЦР в реальном времени с использованием коммерческих тест-систем (Синтол, Россия).

Анализ однонуклеотидной замены rs6721961 (–617G>T) гена *NFE2L2* проводили методом PCR-СТРП согласно методике, описанной Shimoyama с коллегами [25].

Анализ распределения частот генотипов и аллелей в исследуемых группах детей и подростков проводили без учета пола ребенка. Тест на соблюдение равновесия по Харди – Вайнбергу был проведен путем сравнения наблюдаемых частот генотипов с ожидаемыми. Частоты аллелей и генотипов генов по исследуемым SNP в группах детей и подростков сравнивали с помощью критерия χ^2 . В качестве поправки на множественность сравнений использовали пермутационный тест. Статистически значимыми различия считали при $p_{\text{perm}} < 0.05$. Потенциальное locus-locus взаимодействие оценивали с использованием непараметрического программного обеспечения MDR (Multifactor Dimensionality Reduction).

РЕЗУЛЬТАТЫ

Распределение частот генотипов по всем исследуемым SNP в контрольной группе соответствует равновесию Харди – Вайнберга.

Для rs6721961 гена *NFE2L2* выявлены статистически значимые различия в частотах генотипов между двумя исследуемыми группами: среди детей с избыточным весом частота гомозигот –617TT составила 1.8%, тогда как в контроле данный генотип составляет всего 0.8% (табл. 1). В то же время в контроле частота гетерозигот –617GT в 1.9 раз выше (табл. 1). Данный генотип ассоциирован с понижением риска формирования избыточной массы тела у детей (OR = 0.48 95%CI 0.26 – 0.88).

Частоты генотипов и аллелей по SNP генов *SOD1* (Cu,ZnSOD) и *SOD2* (MnSOD) не отличаются в двух исследуемых группах детей (табл. 1).

Различия в частотах генотипов и аллелей между группами детей, отличающихся по ИМТ, выявлены для однонуклеотидной замены rs1001179 гена *CAT*. Частота аллеля –262C в контроле составила 0.82, а

генотипа –262CC – 69.5%. Среди детей с избыточной массой тела данные частоты равны 0.75 и 56.1% соответственно (табл. 1). Данный генотип и аллель ассоциированы с понижением риска формирования избыточной массы тела (для генотипа –262CC OR = 0.56 95%CI 0.36 – 0.87, для аллеля –262C OR = 0.66 95%CI 0.46 – 0.96). Повышенный риск формирования избыточной массы тела выявлен для гетерозигот –262CT (OR = 1.83 95%CI 1.15 – 2.91) и аллеля –262T (OR = 1.51 95%CI 1.04 – 2.19).

Для rs662 *PON1*, rs713041 *GPX4* не выявлено ассоциации с изменением риска формирования избыточной массы тела у детей и подростков (табл. 1).

MDR-анализ выявил значимость взаимодействия исследуемых локусов для формирования избыточной массы тела (табл. 2). Модель взаимодействия локусов имеет максимальную воспроизводимость (100%), точность прогнозирования составляет 59% (табл. 2).

На рис. 1 представлен характер взаимодействия между исследуемыми локусами. Выявлен эффект синергизма между локусами *PON1* и *SOD2*, *PON1* и *GPX4*, *GPX4* и *CAT*.

Частота генотипа GG *NFE2L2*/GG *SOD1*/CT *SOD2*/CC *CAT*/CT *GPX4*/GlnGln *PON1* среди детей и подростков с избыточной массой тела составила 1.8%, тогда как в контроле – 7.6% ($\chi^2 = 7.01$, $p = 0.008$; OR = 0.22 95% CI 0.07 – 0.66). Таким образом, выявлен шестилокусный генотип, ассоциированный со снижением риска формирования избыточной массы тела.

ОБСУЖДЕНИЕ

При ожирении окислительный стресс является одним из основных компонентов патологического процесса, что влечет за собой возможное истощение антиоксидантной системы. Степень нарушения работы антиоксидантов может быть обусловлена особенностями генотипа человека, в частности по генам ферментативных антиоксидантов. Мы исследовали частоту генотипов и аллелей для шести генов антиоксидантной системы: *NFE2L2*, *SOD1*, *SOD2*, *CAT*, *GPX4*, *PON1*.

Нами выявлены статистически значимые различия в распределении частот генотипов по rs6721961 гена *NFE2L2* между двумя исследуемыми группами детей. Генотип –617GT ассоциирован со снижением риска формирования избыточной массы тела. Однонуклеотидная замена –617G>T (rs6721961) локализуется в ARE-последовательности промотора гена *NFE2L2* и приводит к уменьшению эффективности связывания транскрипционного фактора и, следовательно, к снижению уровня транскрипции гена *NFE2L2* и Nrf2-контролируемых генов [26]. Эффект Nrf2 определяется его концентрацией,

Таблица 1. Частоты генотипов и аллелей по исследуемым SNP генов антиоксидантной системы

Генотип, аллель	Контроль, абс. (%)	Избыточный вес, абс. (%)	χ^2	p
NFE2L2 (rs6721961 –617G > T)				
GG	107 (81.7)	248 (88.9)	6.24	0.04
GT	23 (17.5)	26 (9.3)		
TT	1 (0.8)	5 (1.8)		
PXB (χ^2)	0.04 (p > 0.05)	14.5 (p < 0.05)		
Аллель –617G	0.905	0.94	2.04	0.15
Аллель –617T	0.095	0.06		
SOD1 (rs4998557 7958G> A)				
GG	94 (71.8)	203 (73.0)	0.22	0.89
GA	33 (25.2)	65 (23.4)		
AA	4 (3.0)	10 (3.6)		
PXB (χ^2)	0.27 (p > 0.05)	2.63 (p > 0.05)		
Аллель G	0.84	0.85	0.001	0.97
Аллель A	0.16	0.15		
SOD2 (rs4880 47C > T Ala16Val)				
CC	37 (28.2)	69 (24.8)	0.95	0.62
CT	58 (44.3)	137 (49.3)		
TT	36 (27.5)	72 (25.9)		
PXB (χ^2)	1.7 (p > 0.05)	0.06 (p > 0.05)		
Аллель C	0.5	0.5	0.03	0.86
Аллель T	0.5	0.5		
CAT (rs1001179 –262C > T)				
CC	91 (69.5)	156 (56.1)	7.04	0.03
CT	33 (25.2)	106 (38.1)		
TT	7 (5.3)	16 (5.8)		
PXB (χ^2)	2.7 (p > 0.05)	0.13 (p > 0.05)		
Аллель C	0.82	0.75	4.43	0.035
Аллель T	0.18	0.25		
GPX4 (rs713041 718C>T)				
CC	51 (38.9)	100 (35.8)	0.77	0.68
CT	56 (42.7)	118 (42.3)		
TT	24 (18.3)	61 (21.9)		
PXB (χ^2)	1.5 (p > 0.05)	5.26 (p < 0.05)		
Аллель C	0.6	0.57	0.67	0.41
Аллель T	0.4	0.43		
PON1 (rs662 Gln192Arg)				
GlnGln	69 (52.7)	131 (46.9)	1.18	0.55
GlnArg	54 (41.2)	128 (45.9)		
ArgArg	8 (6.1)	20 (7.2)		
PXB (χ^2)	0.36 (p > 0.05)	2.26 (p > 0.05)		
Аллель Gln	0.73	0.7	0.84	0.36
Аллель Arg	0.27	0.3		

Примечание: PXB – равновесие по Харди – Вайнбергу.

Таблица 2. Статистические показатели модели межгенных взаимодействий при формировании избыточной массы тела у детей и подростков

Локус	Точность прогнозирования	Воспроизводимость	χ^2 (p)	OR (95%CI)
NFE2L2 (rs6721961) SOD1 (rs4998557) SOD2 (rs4880) CAT (rs1001179) GPX4 (rs713041) PON1 (rs662)	0.59	10/10	112.2 (<0.0001)	15.5 (8.7–27.7)

продолжительностью синтеза и типом клеток. При ожирении активация *NFE2L2* оказывает защитный эффект за счет прежде всего регуляции антиоксидантной защиты. Однако превышение определенного порога в уровне Nrf2 и продолжительности его синтеза может привести к нарушению редокс-гомеостаза, усилить процесс накопления липидов, что, в свою очередь, может привести к перекисному окислению липидов и повреждению тканей [27]. В связи с этим можно предположить, что гетерозиготность по однонуклеотидной замене в промоторе гена *NFE2L2* создает условия для формирования в клетке оптимального гомеостатического уровня Nrf2, достаточного для эффективной работы антиоксидантной системы и не приводящего к активации патологических процессов. Однако эффективность этого “усредненного” уровня может зависеть от особенностей нуклеотидного состава промоторов генов-мишеней.

В настоящем исследовании пониженный риск формирования избыточной массы тела у детей выявлен для гомозигот по аллелю –262C по rs1001179 гена *CAT*. Данная однонуклеотидная замена локализована в промоторе, влияет на уровень транскрипции гена *CAT* и, соответственно, активность каталазы [28–30]. Область локализации данной замены богата сайтами связывания транскрипционных факторов [31]. Анализ *in silico* показал, что rs1001179 в промоторе *CAT* входит в состав консенсусной последовательности для TFII-1 и GATA-1 в присутствии C или для STAT4, ETS1 и GR-β в присутствии T [32]. В ряде работ показан более низкий уровень экспрессии гена *CAT* при наличии C по сравнению с аллелем T [28, 32]. Однако связь между rs1001179 и уровнем и/или активностью каталазы противоречива. В шведских популяциях концентрация эритроцитарной каталазы у лиц с генотипом *TT* была выше по сравнению с таковой у лиц с генотипом *CC* [28]. В российской популяции также было обнаружено, что уровень каталазы в крови у людей с генотипом *CC* низкий, что

приводит к развитию окислительного стресса и способствует развитию диабета 1 типа [33]. С другой стороны, у женщин с раком молочной железы, имеющих генотип *CC*, определяется более высокая активность каталазы в эритроцитах по сравнению с женщинами, имеющими генотипы *TT* или *TC* [34]. Снижение активности каталазы при наличии аллеля *T* выявлено и в других работах [35, 36]. Можно предположить, что уровень каталазы в клетках является результатом сочетанного эффекта генетического (наличие/отсутствие SNP) и биохимического факторов (ткане-специфичный спектр транскрипционных факторов). Необходимо отметить, что в промоторе гена каталазы не обнаружено нуклеотидной последовательности элемента антиоксидантного ответа (ARE), и вследствие этого транскрипционный фактор Nrf2 не может напрямую связываться с промотором *CAT* [37] и модулировать активность транскрипции гена. Возможно, что именно с этим связана независимость локусов *CAT* и *NFE2L2*, выявленная в модели межгенных взаимодействий (рис. 1).

Выявленное нами снижение риска формирования избыточной массы тела у детей с генотипом –262CC может быть связано с наличием базового уровня каталазы в клетках, который обеспечивает совместно с пероксидазами детоксикацию повышенных концентраций пероксида водорода. У гетерозигот по однонуклеотидной замене в промоторе гена *CAT* возможно изменение уровня каталазы в клетках, что влечет за собой изменение уровня АФК и повышение риска ожирения.

Реализация регуляторного потенциала однонуклеотидных замен в генах *CAT* и *NFE2L2* может быть связана с изменением функциональных характеристик затрагиваемого геномного локуса. Анализ *in silico* позволяет оценить возможные функциональные эффекты полиморфных локусов [38]. Проведенный анализ данных по функциональной геномике показал, что rs6721961 *NFE2L2* локализован в области потенциального связывания

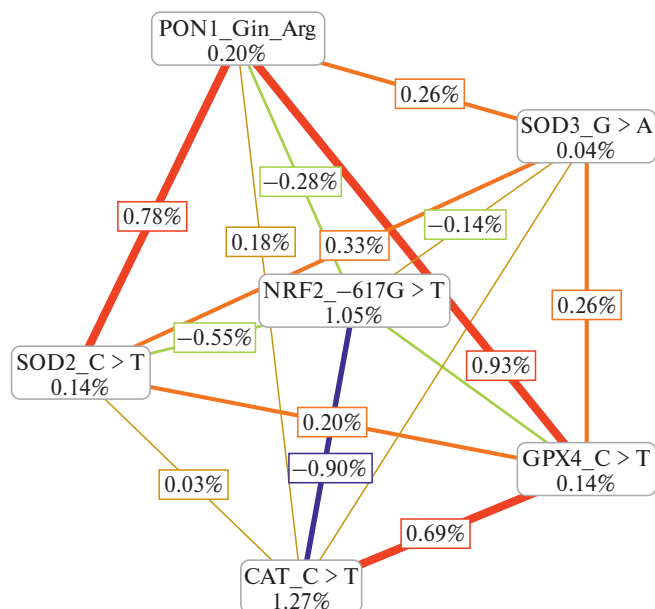


Рис. 1. Характер межгенных взаимодействий при формировании избыточной массы тела у детей (дендрограмма Фрюхтермана – Рейнгольда). Информационная ценность каждого индивидуального локуса представлена на вершинах многогранника; информационная ценность взаимодействия для пары локусов представлена на линиях, их соединяющих (эффект взаимодействия между SNP характеризуется цветом линии: красный – выраженный синергизм, оранжевый – умеренный синергизм, зеленый – умеренный антагонизм, коричневый – аддитивное взаимодействие, синий – избыточность или независимость).

ДНК с 17 различными белковыми молекулами (CTCF, NRSF, POL2, USF1 и др.) [24], а rs1001179 *CAT* локализован в области связывания ДНК для 20 белков (TBP, CMYC, GABP, GATA2 и др.) Регуляторная значимость однонуклеотидной замены может проявляться через модификации гистонов и степень открытия хроматина в области промотора *NFE2L2* или за счет изменения регуляторных мотивов (GATA, STAT, TCF12, Nf1) и тканеспецифичных сайтов гиперчувствительности к ДНКазе для *CAT*.

Совместное функционирование каталазы и глутатионпероксидазы в детоксикации пероксида водорода находит отражение в эффекте синергизма между локусами rs1001179 *CAT* и rs713041 *GPX4*, выявленном нами при анализе межгенных взаимодействий. Глутатионпероксидазы катализируют восстановление H_2O_2 до воды, а органических гидроперекисей до соответствующих спиртов с использованием глутатиона в качестве косубстрата.

Снижение в клетке уровня липопероксидов обеспечивается также функционированием параоксоназы 1. Эффект взаимодействия локусов rs713041

GPX4 и rs662 *PON1* в 3 раза превышает индивидуальный вклад каждого локуса в отдельности (рис. 1).

Уровень образования липопероксидов опосредованно может быть связан с функционированием ферментов первой линии антиоксидантной системы – супероксиддисмутаза. Однонуклеотидная замена rs4880 *SOD2* приводит к аминокислотной замене *Ala16Val* в сигнальном N-концевом пептиде (MTS, Matrix Targeting Signal), который направляет фермент в матрикс митохондрий. Однонуклеотидная замена rs4880 изменяет вторичную структуру белка с образованием β -листа вместо амфифильной α -спирали в сигнальном пептиде [39], что, как следствие, снижает транспорт фермента в митохондрии [40]. В результате у гомозигот по валлину в положении 16 белка *SOD2* уровень фермента в митохондриях снижается, что может повлечь за собой накопление супероксида и активацию свободно-радикальных реакций [41]. При инициации перекисного окисления липидов глутатионпероксидазы становятся основными антиоксидантными ферментами второй линии защиты. Однако некоторые экспериментальные данные позволяют предположить, что генотип *SOD2* дикого типа (гомозиготность по *Ala16*) представляет собой генетический фактор риска повышенной восприимчивости к окислительному стрессу, в то время как мутантный гомозиготный генотип (*ValVal16 SOD2*) проявляет защитный эффект [42]. Гетерозиготный генотип *AlaVal16*, по-видимому, обеспечивает оптимальную ферментативную активность *SOD2* [43–45]. В настоящем исследовании выявлен эффект синергизма между локусами rs4880 *SOD2* и rs662 *PON1*, а генотип *Ala16Val* (47CT) по гену *SOD2* входит в шестилокусный генотип, ассоциированный со снижением риска формирования избыточной массы тела у детей.

Регуляторный потенциал rs4880 *SOD2* достаточно широк и связан с влиянием не только на активность кодируемого фермента *SOD2*. По данным биоинформационного анализа ранее было показано, что rs4880 *SOD2* связан со снижением экспрессии гена *MRPL18*, кодирующего белок L18, участвующий в транспорте из цитозоля в митохондрии 5S рРНК, что может влиять на общий уровень синтеза белка в митохондриях [46]. Митохондриальная дисфункция является одним из основных механизмов развития окислительного стресса при ожирении [47, 48].

Таким образом, полученные нами результаты указывают на потенциальное клиническое влияние двух исследуемых SNP на формирование избыточной массы тела у детей. В то же время выявлен эффект синергизма исследуемых локусов. Статистически значимое снижение риска развития избыточной массы тела выявлено для лиц с наибольшим количеством диких аллелей.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках Государственного задания в сфере научной деятельности № FENW-2023-0018

Проведение исследования одобрено комитетом по биоэтике Академии биологии и биотехнологии им. Д.И. Ивановского Южного федерального университета (протокол № 2 от 17.01.2018). Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников или их законных представителей было получено информированное добровольное согласие.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *World Health Organization*. – Childhood Obesity n.d. <https://www.who.int/dietphysicalactivity/childhood/en/>
2. *NCD Risk Factor Collaboration (NCD-RisC)*. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2416 population-based measurement studies in 128 9 million children, adolescents, and adults // *Lancet*. 2017. V. 390. № 10113. P. 2627–2642. [https://doi.org/10.1016/S0140-6736\(17\)32129-3](https://doi.org/10.1016/S0140-6736(17)32129-3)
3. *Di Cesare M., Sorić M., Bovet P. et al.* The epidemiological burden of obesity in childhood: A worldwide epidemic requiring urgent action // *BMC Med*. 2019. V. 17. P. 212. <https://doi.org/10.1186/s12916-019-1449-8>
4. *Тутельян В.А., Батулин А.К., Конь И.Я. и др.* Распространенность ожирения и избыточной массы тела среди детского населения РФ: мультицентровое исследование // *Педиатрия: журн. им. Г. Н. Сперанского*. 2014. Т. 93. № 5. С. 28–31.
5. *Бочарова О.В., Теплякова Е.Д.* Ожирение у детей и подростков-проблема здравоохранения XXI века // *Казанский мед. журн*. 2020. Т. 101. № 3. С. 381–388. <https://doi.org/10.17816/KMJ2020-381>
6. *Han J.C., Lawlor D.A., Kimm S.Y.* Childhood obesity // *The Lancet*. 2010. V. 375. P. 1737–1748. [https://doi.org/10.1016/S0140-6736\(10\)60171-7](https://doi.org/10.1016/S0140-6736(10)60171-7)
7. *Geng J., Ni Q., Sun W. et al.* The links between gut microbiota and obesity and obesity related diseases // *Biomed Pharmacother*. 2022. V. 147. <https://doi.org/10.1016/j.biopha.2022.112678>
8. *Stenzel A., Carvalho R., Jesus P. et al.* Serum antioxidant associations with metabolic characteristics in metabolically healthy and unhealthy adolescents with severe obesity: an observational study // *Nutrients*. 2018. V. 10. P. 150. <https://doi.org/10.3390/nu10020150>
9. *González-Domínguez A., Visiedo F., Domínguez-Riscart J. et al.* Blunted reducing power generation in erythrocytes contributes to oxidative stress in prepubertal obese children with insulin resistance // *Antioxidants*. 2021. V. 10. P. 244. <https://doi.org/10.3390/antiox10020244>
10. *Savini I., Catani M.V., Evangelista D. et al.* Obesity-associated oxidative stress: strategies finalized to improve redox state // *Int. J. Mol. Sci*. 2013. V. 14. P. 10497–10538. <https://doi.org/10.3390/ijms140510497>
11. *Шестопалов А.В., Давыдов В.В., Туманян Г.Ц. и др.* Содержание адипокинов и миокинов в крови детей и подростков с различным генотипом по полиморфизму rs662 гена параоксоназы-1 // *Ожирение и метаболизм*. 2023. Т. 20. № 3. С. 227–236. <https://doi.org/10.14341/omet13006>
12. *Kaspar J.W., Niture S.K., Jaiswal A.K.* Nrf2: INrf2 (Keap1) signaling in oxidative stress // *Free Radic. Biol. Med*. 2009. V. 47. P. 1304–1309. <https://doi.org/10.1016/j.freeradbiomed.2009.07.035>
13. *Baird L., Yamamoto M.* The molecular mechanisms regulating the KEAP1-NRF2 pathway // *Mol. Cell. Biol*. 2020. V. 40. <https://doi.org/10.1128/MCB.00099-20>
14. *Wang Z., Zuo Z., Li L. et al.* Nrf2 in adipocytes // *Arch. Pharm. Res*. 2020. V. 43. P. 350–360. <https://doi.org/10.1007/s12272-020-01227-0>
15. *Hinney A., Hebebrand J.* Polygenic obesity in humans // *Obes. Facts*. 2008. V. 1. № 1. P. 35–42. <https://doi.org/10.1159/000113935>
16. *Meyre D., Delplanque J., Chèvre J.C. et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations // *Nat. Genetics*. 2009. V. 41. № 2. P. 157–159. <https://doi.org/10.1038/ng.301>
17. *Lu Y., Loos R.J.* Obesity genomics: Assessing the transferability of susceptibility loci across diverse populations // *Genome Med*. 2013. V. 5. № 6. P. 55. <https://doi.org/10.1186/gm459>
18. *Schlauch K.A., Kulick D., Subramanian K. et al.* Single-nucleotide polymorphisms in a cohort of significantly obese women without cardiometabolic diseases // *Int. J. Obesity*. 2019. V. 43. № 2. P. 253–262. <https://doi.org/10.1038/s41366-018-0181-3>
19. *Шкурят М.А., Машкина Е.В., Милютин Н.П., Шкурят Т.П.* Роль полиморфизма

- редокс-чувствительных генов в механизмах окислительного стресса при ожирении и метаболических заболеваниях // Экологическая генетика. 2023. Т. 21. № 3. С. 261–287. <https://doi.org/10.17816/ecogen562714>
20. Bradfield J.P., Voegelzang S., Felix J.F. et al. A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity // Hum. Mol. Genet. 2019. V. 28. № 19. P. 3327–3338. <https://doi.org/10.1093/hmg/ddz161>
21. Трифонова Е.А., Попович А.А., Вагайцева К.В. и др. Метод мультиплексного генотипирования полиморфных вариантов генов, ассоциированных с ожирением и индексом массы тела // Генетика. 2019. Т. 55. № 10. С. 1218–1230. <https://doi.org/10.1134/S001667581910014X>
22. Трифонова Е.А., Попович А.А., Макеева О.А. и др. Репликативный анализ ассоциаций генетических маркеров с ожирением в российской популяции // Генетика. 2021. Т. 57. № 5. С. 604–610. <https://doi.org/10.31857/S0016675821050131>
23. <https://www.who.int/ru/news-room/fact-sheets/detail/obesity-and-overweight>
24. <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>
25. Shimoyama Y., Mitsuda Y., Hamajima N., Niwa T. Polymorphisms of Nrf2, an antioxidative gene, are associated with blood pressure in Japanese // Nagoya J. Med. Sci. 2014. V. 76. № 1–2 P. 113–120.
26. Vasileva L.V., Savova M.S., Amirova K.M. et al. Obesity and NRF2-mediated cytoprotection: Where is the missing link? // Pharmacol Res. 2020. V. 156. № 6. <https://doi.org/10.1016/j.phrs.2020.104760>
27. Xia Y., Zhai X., Qiu Y. et al. The Nrf2 in obesity: A friend or foe? // Antioxidants. 2022. V. 11. № 10. <https://doi.org/10.3390/antiox11102067>
28. Forsberg L., Lyrenäs L., De Faire U., Morgenstern R. A common functional C-T substitution polymorphism in the promoter region of the human catalase gene influences transcription factor binding, reporter gene transcription and is correlated to blood catalase levels // Free Radic Biol Med. 2001. V. 30. P. 500–505. [https://doi.org/10.1016/S0891-5849\(00\)00487-1](https://doi.org/10.1016/S0891-5849(00)00487-1)
29. Quick S.K., Shields P.G., Nie J. et al. Effect modification by catalase genotype suggests a role for oxidative stress in the association of hormone replacement therapy with postmenopausal breast cancer risk // Cancer Epidemiology Biomarkers & Prevention. 2008. V. 17. № 5. P. 1082–1087. <https://doi.org/10.1158/1055-9965.EPI-07-2755>
30. Saify K., Saadat I., Saadat M. Influence of A-21T and C-262T genetic polymorphisms at the promoter region of the catalase (CAT) on gene expression // Environ Health Prev Med. 2016. V. 21. P. 382–386. <https://doi.org/10.1007/s12199-016-0540-4>
31. Galasso M., Gambino S., Romanelli M.G. et al. Browsing the oldest antioxidant enzyme: catalase and its multiple regulation in cancer // Free Radic. Biol. Med. 2021. V. 172. P. 264–272. <https://doi.org/10.1016/j.freeradbiomed.2021.06.010>
32. Galasso M., Pozza E., Chignola R. et al. The rs1001179 SNP and CpG methylation regulate catalase expression in chronic lymphocytic leukemia // Cell Mol. Life Sci. 2022. V. 79. № 10. P. 521. <https://doi.org/10.1007/s00018-022-04540-7>
33. Chistiakov D.A., Savost'yanov K.V., Turakulov R.I. et al. A new type 1 diabetes susceptibility locus containing the catalase gene (chromosome 11p13) in a Russian population // Diabetes/Metabolism Research and Reviews. 2004. V. 20. № 3. P. 219–224. <https://doi.org/10.1002/dmrr.442>
34. Ahn J., Gammon M.D., Santella R.M. et al. Associations between breast cancer risk and the catalase genotype, fruit and vegetable consumption, and supplement use // Am. J. Epidemiol. 2005. V. 162. № 10. P. 943–952. <https://doi.org/10.1093/aje/kwi306>
35. Goth L., Nagy T., Kosa Z. et al. Effects of rs769217 and rs1001179 polymorphisms of catalase gene on blood catalase, carbohydrate and lipid biomarkers in diabetes mellitus // Free Radic Res. 2012. V. 46. P. 1249–1257. <https://doi.org/10.3109/10715762.2012.702899>
36. Goulas A., Agapakis D., Apostolidis A. et al. Association of the common catalase gene polymorphism rs1001179 with glycated hemoglobin and plasma lipids in hyperlipidemic patients // Biochem. Genet. 2017. V. 55. P. 77–86. <https://doi.org/10.1007/s10528-016-9777-2>
37. Glorieux C., Zamocky M., Sandoval J.M. et al. Regulation of catalase expression in healthy and cancerous cells // Free Radic. Biol. Med. 2015. V. 87. P. 84–97. <https://doi.org/10.1016/j.freeradbiomed.2015.06.017>
38. Полоников А.В., Клёсова Е.Ю., Азарова Ю.Э. Биоинформатические инструменты и интернет-ресурсы для оценки регуляторного потенциала полиморфных локусов, установленных полногеномными ассоциативными исследованиями мультифакториальных заболеваний (обзор) // Научные результаты биомедицинских исследований. 2021. Т. 7. № 1. С. 15–31. <https://doi.org/10.18413/2658-6533-2020-7-1-0-2>
39. Shimoda-Matsubayashi S., Matsumine H., Kobayashi T. et al. Structural dimorphism in the mitochondrial targeting sequence in the human manganese superoxide dismutase gene. A predictive evidence for conformational change to influence mitochondrial transport and a study of allelic association in Parkinson's disease // Biochem. Biophys. Res. Commun. 1996. V. 226. № 2. P. 561–565. <https://doi.org/10.1006/bbrc.1996.1394>
40. Sutton A., Khoury H., Prip-Buus C. et al. The Ala16Val genetic dimorphism modulates the import of human manganese superoxide dismutase into rat liver

- mitochondria // *Pharmacogenetics*. 2003. V. 13. № 3. P. 145–157. doi: 10.1097/01.fpc.0000054067.64000.8f
41. Bastaki M., Huen K., Manzanillo P. et al. Genotype-activity relationship for Mn-superoxide dismutase, glutathione peroxidase 1 and catalase in humans // *Pharmacogenet. Genomics*. 2006. V. 16. P. 279–286. [https://doi.org/ 10.1097/01.fpc.0000199498.08725.9c](https://doi.org/10.1097/01.fpc.0000199498.08725.9c)
 42. Cannata A., De Luca C., Andolina G. et al. Influence of the SOD2 A16V gene polymorphism on alterations of redox markers and erythrocyte membrane fatty acid profiles in patients with multiple chemical sensitivity // *Biomed. Rep.* 2021. V. 15. № 6. P. 101. [https://doi.org/ 10.3892/br.2021.1477](https://doi.org/10.3892/br.2021.1477)
 43. Giusti B., Vestrini A., Poggi C. et al. Genetic polymorphisms of antioxidant enzymes as risk factors for oxidative stress-associated complications in preterm infants // *Free Radic. Res.* 2012. V. 46. P. 1130–1139. [https://doi.org/ 10.3109/10715762.2012.692787](https://doi.org/10.3109/10715762.2012.692787)
 44. Palmirotta R., Barbanti P., De Marchis M.L. et al. Is SOD2 Ala16Val polymorphism associated with migraine with aura phenotype? // *Antioxid. Redox Signal.* 2015. V. 22. P. 275–279. [https://doi.org/ 10.1089/ars.2014.6069](https://doi.org/10.1089/ars.2014.6069)
 45. Synowiec E., Wigner P., Cichon N. et al. Single-nucleotide polymorphisms in oxidative stress-related genes and the risk of a stroke in a Polish population – a preliminary study // *Brain Sci.* 2021. V. 11. [https://doi.org/ 10.3390/brainsci11030391](https://doi.org/10.3390/brainsci11030391)
 46. Азарова Ю.Э., Гуреева А.В., Постникова М.И. и др. Связь однонуклеотидного полиморфизма rs4880 гена SOD2 с развитием микрососудистых осложнений сахарного диабета 2-го типа // *Науч. результаты биомед. исследований*. 2023. Т. 9. № 4. С. 461–473. [https://doi.org/ 10.18413/2658-6533-2023-9-4-0-3](https://doi.org/10.18413/2658-6533-2023-9-4-0-3)
 47. De Mello A.H., Costa A.B., Engel J.D.G., Rezin G.T. Mitochondrial dysfunction in obesity // *Life Sci.* 2018. V. 192. P. 26–32. doi: 10.1016/j.lfs.2017.11.019
 48. Prasun P. Mitochondrial dysfunction in metabolic syndrome // *Biochim. Biophys. Acta Mol Basis Dis.* 2020. V. 1866 (10). [https://doi.org/ 10.1016/j.bbadis.2020.165838](https://doi.org/10.1016/j.bbadis.2020.165838)

Polymorphism Of Antioxidant Genes And Overweight In Children

M. A. Shkurat^{1, 2}, E. V. Mashkina^{1*}, N. P. Milyutina¹, E. D. Teplyakova³, T. P. Shkurat^{1, 2}

¹*Southern Federal University, Rostov-on-Don, 344090 Russia*

²*Medical center “Nauka”, Rostov-on-Don, 344000 Russia*

³*Rostov State Medical University, Rostov-on-Don, 344022 Russia*

E-mail: lenmash@mail.ru

Oxidative stress is one of the components of the pathological process leading to the development of obesity. The level of formation of free radical products is controlled by the antioxidant system. Gene polymorphisms influence the level and/or activity of the encoded enzymes. The aim of the work was to investigate the association of SNP in the genes of the antioxidant system with the risk of overweight in children and adolescents. The material for the study were DNA samples from 279 overweight children and 131 children from the control group. Genotyping was performed for rs6721961 (–617G>T) *NFE2L2*, rs4998557 (7958G>A) *SOD1*, rs4880 (47C>T Ala16Val) *SOD2*, rs1001179 (–262C>T) *CAT*, rs713041 (718C>T) *GPX4*, rs66 2 (Gln192Arg) *PON1*. It has been shown that the –617GT genotype (rs6721961) *NFE2L2* is associated with decreased of overweight risk children. An increased risk of developing overweight was detected for heterozygotes –262CT for rs1001179 *CAT* and the –262T allele. As a result of the analysis of intergenic interactions, a 6-locus genotype was identified that is associated with a reduced risk of overweight.

Keywords: overweight and obesity in children, antioxidants, gene polymorphism, *SOD1*, *SOD2*, *CAT*, *GPX4*, *PON1*, *NFE2L2*, intergenic interactions

РЕКОНСТРУКЦИЯ МАТРИЦЫ ГЕНОТИПИЧЕСКИХ КОРРЕЛЯЦИЙ МЕЖДУ ВАРИАНТАМИ ВНУТРИ ГЕНА ДЛЯ СОВМЕСТНОГО АНАЛИЗА ИМПУТИРОВАННЫХ И СЕКВЕНИРОВАННЫХ ДАННЫХ

© 2024 г. Г. Р. Свищёва^{1, 2, *}, А. В. Кириченко¹, Н. М. Белоногова¹,
Е. Е. Елгаева^{1, 3}, Я. А. Цепилов¹, И. В. Зоркольева¹, Т. И. Аксенович¹

¹Федеральный исследовательский центр, Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, 630090 Россия

²Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

³Новосибирский государственный университет, Новосибирск, 630090 Россия

*e-mail: gulsvi@mail.ru

Поступила в редакцию 29.01.2024 г.

После доработки 27.02.2024 г.

Принята к публикации 25.03.2024 г.

При объединении импутированных и секвенированных данных в одном анализе ассоциаций на уровне генов возникает проблема реконструкции матриц генетических корреляций. Она связана с тем, что для гена известны корреляции между всеми импутированными генотипами вариантов и корреляции между всеми секвенированными, но неизвестны корреляции между генотипами вариантов, один из которых импутирован, а другой секвенирован. Для реконструкции этих корреляций мы предлагаем эффективный метод, основанный на максимизации детерминанта матрицы. Этот метод обладает рядом полезных свойств и имеет аналитическое решение для нашей задачи. Апробация предложенного метода была выполнена путем сравнения реконструированных и реальных корреляционных матриц, построенных на индивидуальных генотипах из Биобанка Великобритании. Сравнение результатов анализа ассоциаций на генном уровне, выполненного методами SKAT, VT и PCA на реконструированных и реальных матрицах с использованием смоделированных и вычисленных по реальным фенотипам суммарных статистик, показало высокое качество реконструкции и устойчивость метода к различным структурам гена.

Ключевые слова: импутированные и секвенированные генотипы, анализ ассоциаций на уровне генов, генетические варианты, суммарные статистики.

DOI: 10.31857/S0016675824070089 **EDN:** BHMPLU

Анализ ассоциаций на генном уровне является одним из самых эффективных статистических методов, используемых для идентификации генов, контролируемых различные признаки и болезни человека [1–4]. Для этого анализа используются два типа генотипических данных: импутированные и секвенированные. Первый тип получается путем генотипирования большого числа распространенных вариантов, расположенных равномерно по геному, и последующего восстановления (импутации) генотипов промежуточных вариантов [5]. Второй получается в результате полногеномного секвенирования [6]. Эти два типа данных частично перекрываются, однако по большей части они характеризуют разные варианты в гене. Секвенированные генотипы описывают преимущественно белок-кодирующие последовательности, а импутированные – интронные [7, 8]. Каждый из этих типов широко используется для анализа ассоциаций

на уровне гена. Проблема возникает, когда мы хотим объединить оба типа генотипических данных в одном анализе. Известно, что в отсутствие доступа к индивидуальным генотипам и фенотипам для проведения анализа ассоциаций на генном уровне необходимы суммарные статистики – результаты одноточечного анализа ассоциаций (z-статистики и размеры эффектов) для всех вариантов, идентифицированных в гене, а также матрица корреляций между генотипами всех этих вариантов.

Для генетического варианта, генотипированного двумя способами, результат одноточечного анализа, а именно z-статистика, определяется как взвешенная сумма результатов, полученных для каждого генотипа. Однако восстановление корреляций между вариантами с импутированными и секвенированными генотипами представляет проблему.

Эта проблема, которая определяется как поиск значений, заполняющих недостающую информацию в матрице корреляций, давно изучается в линейной алгебре и в общем случае не имеет единственного допустимого решения. В этой связи были предложены численные методы решения, которые определяют диапазоны неизвестных элементов матрицы [9–11]. Однако эти методы были разработаны для матриц маленькой размерности (< 5). Для больших корреляционных матриц существующие численные решения требуют существенных вычислительных затрат [12, 13]. Учитывая, что для анализа ассоциаций необходимо реконструировать матрицы генотипических корреляций примерно для 20 000 генов, многие из которых содержат большое число генотипированных вариантов, численные методы не подходят, и нам необходимо найти легко реализуемое аналитическое решение.

Особенностью рассматриваемых нами матриц корреляций является то, что все их недостающие элементы можно поместить в один блок. Тогда саму восстанавливаемую матрицу можно представить в блочном виде 3×3 , где неизвестный блок расположен вне главной диагонали, симметрично по обе стороны от нее. Это значительно упрощает задачу.

Цель данного исследования — поиск аналитического решения для восстановления матриц корреляций с учетом их специфической структуры и сравнение свойств реконструированной и реальной матриц с использованием данных из Биобанка Великобритании.

МЕТОД

Алгоритм реконструкции матрицы

Большинство методов реконструкции матрицы основаны на максимизации детерминанта матрицы [12–15]. Основным требованием для использования этого подхода является предположение о корректности реконструированной корреляционной матрицы (в частности, симметричность и положительная полуопределенность ее заданных диагональных подматриц [12]) и многомерной нормальности данных [14]. Метод реконструкции корреляционной матрицы, основанный на максимизации ее детерминанта, имеет ряд полезных теоретических свойств [14]:

существование и уникальность решения: существует ровно одна реконструированная корреляционная матрица, использующая принцип максимума детерминанта;

максимальная энтропия: максимум детерминанта ведет к максимизации энтропии;

максимальное правдоподобие: максимум детерминанта, по сути, это оценка максимального правдоподобия корреляционной матрицы для неизвестной базовой многомерной нормальной модели;

центр области допустимых решений: максимум детерминанта является центром области допустимых решений, ограниченной положительной полуопределенностью, и имеет прямое аналитическое решение для некоторых блочно-структурированных матриц.

На рис. 1 схематически изображены блоки частично заданной матрицы корреляций между тремя неперекрывающимися подмножествами генетических вариантов: M_1 , M_2 и M_3 в рамках одного гена и соответствующие наборы z-статистик, полученные при объединении в один анализ импутированных данных (M_1 и M_2 подмножества вариантов) и экзомных данных (M_2 и M_3 подмножества вариантов).

Для удобства в соответствии с рис. 1 мы представили объединенную корреляционную матрицу U между тремя подмножествами генетических вариантов: M_1 , M_2 и M_3 , внутри гена в виде матричных блоков 3×3 с неизвестным блоком U_{13} (и сопряженным с ним блоком U_{31} , $U_{13} = U_{31}^T$):

$$U = \begin{bmatrix} M_1 & M_2 & M_3 \\ U_{11} & U_{12} & U_{13} - ? \\ U_{12}^T & U_{22} & U_{23} \\ U_{13}^T - ? & U_{23}^T & U_{33} \end{bmatrix} \quad (1)$$

Тогда подматрица U_{13} может быть реконструирована через соседние с ней подматрицы, выделенные пунктирной линией в формуле (1):

$$U_{13} = U_{12} U_{22}^{-1} U_{23}. \quad (2)$$

Этот вывод, полученный в ряде работ [12, 14–16], следует из свойств детерминанта блочной матрицы вида $\begin{pmatrix} A & C \\ C^T & B \end{pmatrix}$, в частности:

$$a) \det \begin{pmatrix} A & C \\ C^T & B \end{pmatrix} = \det(A) * \det(B - CA^{-1}C^T);$$

б) детерминант матрицы будет максимальным, если занулить подматрицу C , т. е. свести исходную матрицу к блочно-диагональному виду. Стоит отметить, что максимизация детерминанта матрицы, которая эквивалентна максимизации произведения собственных значений при сохранении фиксированного среднего собственных значений, равного единице, стремится сделать все собственные

значения равными, улучшая число обусловленности матрицы.

Легко показать, что в терминах принятых обозначений (рис. 1) максимальное значение детерминанта объединенной матрицы U выражается как

$$\begin{aligned} \det(U) &= \det(U_{22}) \\ \det(U_{11} - U_{12}U_{22}^{-1}U_{12}^T) \det(U_{33} - U_{23}^TU_{22}^{-1}U_{23}) &= \\ &= \det(U_{22}) \det(U_{11}|U_{22}) \det(U_{33}|U_{22}) \end{aligned} \quad (3)$$

или в терминах детерминантов исходных матриц корреляций как

$$\det(U) = \frac{\det(U_{imp}) \det(U_{seq})}{\det(U_{shared})},$$

где U_{imp} , как показано на рис. 1, представляет исходную корреляционную матрицу для вариантов гена с импутированными генотипами, U_{seq} — для вариантов гена с секвенированными генотипами и U_{shared} — для общих (перекрывающихся) вариантов гена.

Заметим, что решение (2) относится к аппроксимации, полученной путем удаления столбцов/строк с неизмеренными корреляционными коэффициентами:

$$U = UU^{-1}U \approx U_{*k}U_{kk}^{-1}U_{*k}^T. \quad (4)$$

Здесь U_{kk} заданная обратимая подматрица, где индекс k означает набор вариантов с генотипами, измеренными двумя способами: импутированным и секвенированным, а индекс $*$ означает все варианты гена. Тогда в терминах наших обозначений (рис. 1) из формулы (4) получаем формулу $U_{13} = U_{12}U_{22}^{-1}U_{23}$, идентичную формуле (2).

Важно отметить, что решение (2) совпадает с условно независимым решением (см. формулу (3)), а значит реконструированные корреляции не могут превышать реальные значения [14]. Безусловно, значительное занижение реальных генотипических корреляций может привести к инфляции статистик при анализе ассоциаций на геномном уровне. Однако поскольку секвенированные варианты являются редкими, их корреляции с распространенными импутированными вариантами будут близки к нулю [17], и можно ожидать, что статистическая инфляция будет минимальной.

Сравнение реальных и реконструированных матриц

Свойства предложенного метода, реконструирующего неизвестные коэффициенты корреляции между импутированными и секвенированными вариантами без использования информации о

генотипах, были изучены путем сравнения восстановленных матриц с реальными матрицами, построенными по индивидуальным генотипам.

Оценка эффективности

Для оценки качества реконструкции корреляционных матриц был выбран показатель: оценка функции потерь, традиционно рассчитываемая как среднеквадратичная ошибка между фактическими значениями U_{13} и реконструируемыми $U_{13_{rec}}$:

$$L_1 = \text{mean} \left(\left(U_{13} - U_{13_{rec}} \right)^2 \right).$$

Кроме того, поскольку наиболее популярным методом анализа ассоциаций на геномном уровне является метод SKAT [18], который при использовании корреляционной матрицы ограничивается только собственными значениями ее спектрального разложения, в качестве дополнительных показателей эффективности мы использовали коэффициент регрессии RS_1 и коэффициент детерминации RS_2 регрессионного уравнения, сопоставляющего собственные значения реконструированной матрицы и реальной. В отличие от L_1 коэффициенты RS_1 и RS_2 рассчитываются на полной матрице U . Ожидается, что оба коэффициента будут близки к единице, а оценка функции потерь — к нулю.

Зависимость показателей эффективности была исследована относительно доли неизвестных коэффициентов среди всех (заданных и неизвестных) коэффициентов корреляции между импутированными и секвенированными вариантами. Эту долю можно представить геометрически как отношение площадей, s (рис. 1):

$$d = \frac{s(U_{13})}{s(U_{22}) + s(U_{12}) + s(U_{23}) + s(U_{13})}, d \in [0, 1].$$

МАТЕРИАЛ

Генотипы и матрицы генотипических корреляций

Для построения объединенных матриц корреляций между вариантами внутри генов были использованы секвенированные и импутированные генотипы из Биобанка Великобритании (UK BioBank).

Оригинальные секвенированные генотипы (Data-Field 23156) были представлены в формате vcf в сборке GRCh38/hg38. Контроль качества этих данных осуществлялся с помощью пакета VCFtools версии 1.16.1 и по алгоритму, описанному на сайте <https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=914>. Секвенированные данные были преобразованы в формат rgen с помощью пакета plink версии 2.00a3.3LM.

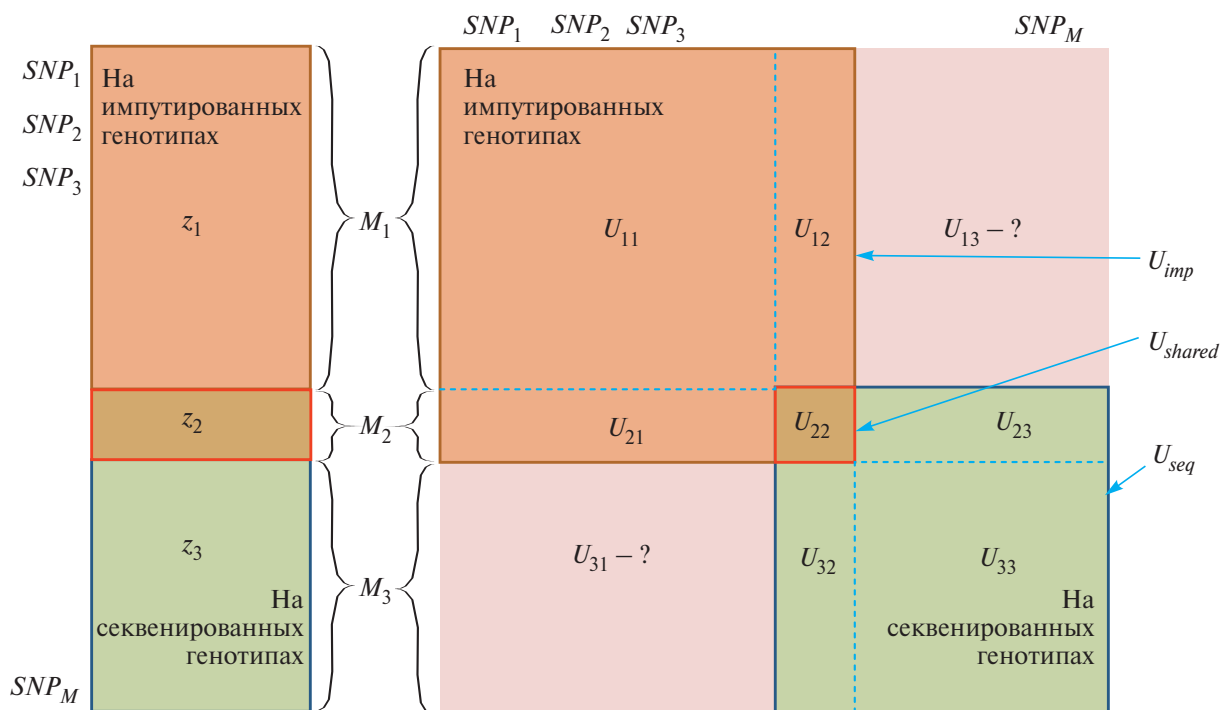


Рис. 1. Схематическое представление объединенной частично заданной матрицы корреляций между генетическими вариантами внутри гена. M_1 и M_3 обозначают неперекрывающиеся подмножества вариантов с импутированными и секвенированными генотипами соответственно, а M_2 — подмножество перекрывающихся вариантов. Суммарные статистики и матрицы генотипических корреляций для $(M_1 + M_2)$ вариантов посчитаны на импутированных данных, а для $(M_2 + M_3)$ вариантов — на секвенированных данных. Матрица корреляций между наборами M_1 и M_3 (блок U_{13}) неизвестна.

Оригинальные импутированные генотипы (Data-Field 22828), представленные в формате bgen в сборке GRCh37/hg19, были конвертированы в сборку GRCh38/hg38 с помощью пакета liftOver и в формат pgen с помощью пакета plink версии 2.00a. Дальнейшие манипуляции с данными проводились также с помощью пакета plink версии 2.00a.

Далее из обоих типов данных были выбраны генотипы только биаллельных маркеров для белых европейцев (Data-Field 21000), не связанных генетическим родством (Data-Field 22021). Из импутированных данных были исключены варианты, позиции которых совпадали с позициями секвенированных вариантов, чтобы отделить предпочтение секвенированным генотипам при объединении их с импутированными. Выборка была ограничена индивидами, у которых были оба типа генотипов, $N \sim 200000$ человек. Данные о генотипах обоих типов были объединены и сохранены в pgen-формате.

Для тестирования предложенного метода были выбраны гены на 21-й хромосоме. Для расчета корреляционных матриц была использована информация о позициях генов из данных Ensembl (<https://www.ensembl.org/>). Матрицы корреляций вычислялись с помощью пакета ldstore версии 2.0 (<http://www.christianbenner.com/>). Для каждого гена были

вычислены три матрицы: по секвенированным, импутированным и объединенным генотипам. Для анализа отбирались гены с общим числом вариантов от 50 до 7000 в объединенной матрице и с обязательным наличием перекрывающихся вариантов. Кроме того, для каждого гена в корреляционную матрицу включали только те варианты, которые прошли установленные пороги по частоте минорного аллеля maf и показателю качества импутации $info$: для импутированных вариантов $info > 0.8$ и $maf > 5 \times 10^{-5}$, а для секвенированных вариантов $maf > 3.8 \times 10^{-5}$, что соответствовало количеству минорных аллелей, $mas > 10$. Всего в анализ были включены корреляционные матрицы для 124 генов. После фильтраций в объединенных матрицах корреляций число импутированных вариантов варьировалось от 2 до 1572, число секвенированных вариантов — от 6 до 352, а число перекрывающихся вариантов — от 1 до 96.

Фенотип и суммарные статистики

Для анализа реальных фенотипических данных были получены суммарные статистики для количественного признака, индекс массы тела (body mass index, BMI) ($N = 187600$). Чтобы привести распределение признака к нормальному, были

исключены из анализа 2247 человек со значениями признака, выходящими за пределы трех среднеквадратических отклонений от среднего значения, и было выполнено преобразование признака с помощью процедуры преобразования рангов (rank-transformation), после чего нормальность распределения признака была подтверждена критерием Колмогорова – Смирнова ($p = 1$).

Для получения суммарных статистик для BMI проводили одноточечный анализ ассоциаций для секвенированных и импутированных данных с помощью пакета fastGWA-GLMM, бета-версия 1.94.0 [19]. Анализ выполнялся с использованием опции – fastGWA-mlm и ограничивался редкими вариантами с частотой минорного аллеля (maf) в диапазоне от 5×10^{-5} до 0.01. Для исключения случайных эффектов, обусловленных родством между индивидами в выборке, была предварительно рассчитана матрица родства с использованием fastGWA-GLMM (опция – make-bK-sparse с параметрами по умолчанию) для всей выборки Биобанка Великобритании ($N = 487000$). В наш анализ в качестве ковариат мы включили пол, возраст и первые 10 главных генетических компонент, предоставленных Биобанком Великобритании.

Симуляционные суммарные статистики

Предложенный метод был также протестирован на симуляционных суммарных статистиках. Для каждого из выбранных генов 21-й хромосомы был смоделирован вектор z-статистик на реальной генотипической матрице корреляций U по определенному сценарию, учитывающему размер эффекта гена на признак, τ (1000 повторов для каждого сценария). Для симуляций суммарных статистик была рассмотрена модель, лежащая в основе метода SKAT, использующего суммарные статистики в качестве входных данных [20]

$$z \sim MN(0, U + \tau U^2).$$

Параметр τ прямо пропорционален наследуемости признака, обусловленной вариантами гена. Мы зафиксировали параметр τ как 0, 0.02, 0.05 и 0.1, где нулевой гипотезе, предполагавшей отсутствие ассоциаций между анализируемым геном и признаком, соответствует $\tau = 0$.

РЕЗУЛЬТАТЫ

Анализ показателей эффективности

В табл. 1 приведена статистика, описывающая структуру распределения выбранных показателей эффективности. Как видно, для большинства генов оценка функции потерь близка к нулю и в целом не превышает 0.11, а коэффициенты RS_1 и RS_2

близки к единице, что свидетельствует об очень хорошей сходимости собственных значений реконструированной корреляционной матрицы к собственным значениям реальной (рис. 2). При сравнении показателей эффективности между собой, как ожидалось, наблюдается сильная корреляция. Парные корреляции Пирсона составили: $\text{cor}(L_1, RS_1) = -0.752$, $\text{cor}(L_1, RS_2) = -0.816$ и $\text{cor}(RS_1, RS_2) = 0.739$.

Следует отметить, что мы не обнаружили явной зависимости показателей эффективности от доли неизвестных коэффициентов среди всех (заданных и неизвестных) коэффициентов корреляции между импутированными и секвенированными вариантами. У генов, для которых наблюдали “выбросы” (сильные отклонения от ожидаемых значений) по всем трем показателям эффективности: L_1 , RS_1 и RS_2), мы обнаружили сильное расхождение между секвенированными и импутированными генотипами для перекрывающихся вариантов, несмотря на фильтрацию импутированных генотипов по показателю info, что свидетельствует скорее о низком качестве импутации генотипов, чем о плохом качестве реконструкции.

Сравнение результатов анализа ассоциаций на геномном уровне

После успешной апробации предложенного метода мы оценили качество реконструкции объединенной матрицы корреляций, выполнив анализ ассоциаций на геномном уровне с использованием реконструированной и реальной матриц корреляций и сравнив результаты анализа – p -значения. Для анализа мы выбрали три популярных теста, SKAT (Sequence kernel association test) [18], BT (Burden test) и PCA (Principle component analysis) [21], использующих различные регрессионные модели и реализованных в R-пакете sumFREGAT [21, 22]. Все тесты выполнялись без взвешивания

Таблица 1. Показатели структуры распределения для оценок качества реконструкции корреляционных матриц

Статистика распределения	L_1	RS_1	RS_2
Минимум	5.00×10^{-8}	0.953	0.927
1-й квартиль	5.59×10^{-5}	0.999	1.000
Медиана	1.02×10^{-4}	1.000	1.000
Среднее	1.48×10^{-3}	0.999	0.998
3-й квартиль	2.32×10^{-4}	1.000	1.000
Максимум	1.06×10^{-1}	1.027	1.000

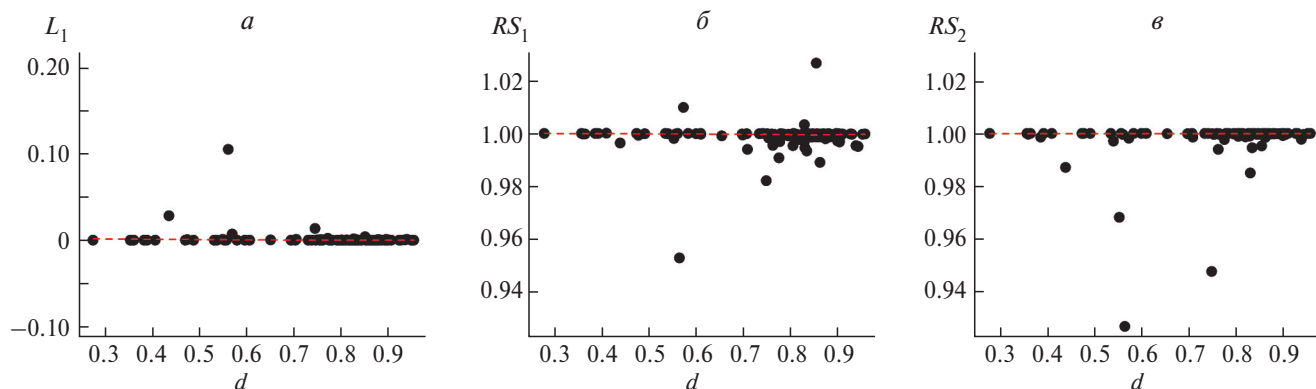


Рис. 2. Зависимости показателей эффективности реконструкции корреляционных матриц от доли неизмеренных коэффициентов среди всех (измеренных и неизмеренных) коэффициентов корреляции между импутированными и экзомными вариантами. По оси X откладывается доля неизмеренных коэффициентов среди всех коэффициентов корреляции между импутированными и экзомными вариантами, а по оси Y откладывается показатель эффективности.

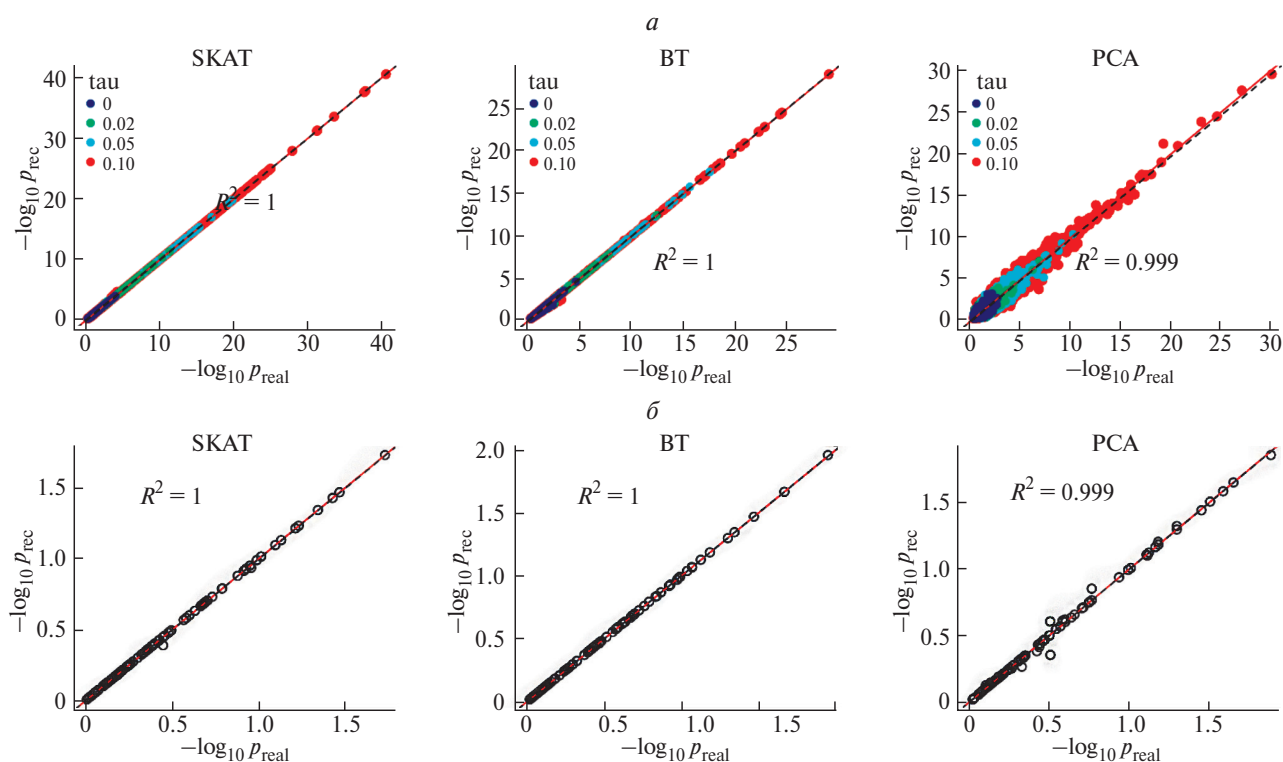


Рис. 3. Сопоставление результатов анализа ассоциаций на генном уровне, выполненного методами SKAT, BT и PCA на реконструированных и исходных матрицах корреляций: *а* – с использованием z -статистик, симулированных для различных сценариев относительно t ; *б* – с использованием z -статистик, вычисленных на реальном признаке BMI.

суммарных статистик. Для PCA долю объясненной дисперсии брали 0.85.

По всем изученным генам 21-й хромосомы было получено полное соответствие p -значений с $R^2 = 1$ для методов SKAT и BT, использующих суммарные статистики как симулированные (рис. 3, а), так и

вычисленные по реальному признаку BMI (рис. 3, б). Для метода PCA хорошее соответствие наблюдалось для всех генов на реальных данных ($R^2 = 0.999$) и для генов с маленькими p -значениями ($p < 1 \times 10^{-7}$) на симулированных данных. Для генов с большими p -значениями наблюдалось небольшое

рассеивание в виде облака без смещения регрессионной линии от ожидаемой ($R^2 = 0.999$).

ОБСУЖДЕНИЕ

Объединение импутированных и секвенированных данных для анализа ассоциаций на уровне генов может помочь обнаружить сигналы в генах, обусловленные совместным влиянием вариантов внутри генов. Если для генов исходные генотипы недоступны, корреляционные матрицы строятся на основе референсной выборки. В этом случае, как правило, корреляции между импутированными и секвенированными вариантами гена неизвестны. Для реконструкции этих корреляций мы предлагаем эффективный математический метод, основанный на максимизации детерминанта матрицы. Этот метод обладает рядом полезных свойств и имеет аналитическое решение для нашей задачи.

Ограничения и возможности предложенного метода были исследованы на реальных корреляционных матрицах, построенных на индивидуальных генотипах из Биобанка Великобритании. Очевидно, что одним из ограничений предложенного метода является невозможность его использования, если импутированные и секвенированные варианты гена не перекрываются. Кроме того, мы обнаружили, что еще одним ограничением метода служит сильное расхождение между импутированными и секвенированными генотипами перекрывающихся вариантов генов. Вероятно, такое расхождение обусловлено использованием в анализе редких вариантов с ошибочно импутированными генотипами. В дополнение следует отметить, что мы не выявили никаких ограничений, касающихся доли неизвестных корреляций в объединенной матрице.

Для апробации предложенного метода был проведен анализ ассоциаций на уровне генов, выполненный как на смоделированных суммарных статистиках, так и на суммарных статистиках, вычисленных на реальном признаке BMI из Биобанка Великобритании. Сравнение результатов анализа ассоциаций на геномном уровне, выполненного популярными методами SKAT, VT и PCA на реконструированных и реальных матрицах, показало высокое качество реконструкции вне зависимости от доли перекрывающихся вариантов в гене и устойчивость метода к различным структурам генов.

Для перекрывающихся вариантов соответствие между корреляционными матрицами, построенными на импутированных данных, и корреляционными матрицами, построенными на секвенированных данных, может быть нарушено по ряду причин, например, использование референсных выборок разного размера или включение в анализ редких вариантов с генотипами низкого уровня импутации. Для решения этой проблемы можно применить процедуру регуляризации, которая

для каждого гена “подгоняет” реконструированную корреляционную матрицу к соответствующим z-статистикам с помощью двух параметров регуляризации, настраиваемых персонально для импутированных и секвенированных данных. Регуляризация по типу Тихонова [23] использует спектральную фильтрацию собственных значений корреляционной матрицы. Подобная процедура нами уже была предложена в работе [21] при использовании референсных генотипических данных для анализа ассоциаций на уровне гена. Общая схема предлагаемой регуляризации для реконструированной матрицы корреляций, объединяющей импутированные и секвенированные генотипы, имеет вид:

$$E\left(zz^T\right)=\begin{bmatrix}\lambda_1 U_{11} & \sqrt{\lambda_1 \lambda_2} U_{12} & \sqrt{\lambda_1 \lambda_2} U_{13} \\ \sqrt{\lambda_2 \lambda_1} U_{12}^T & \lambda_2 U_{22} & \lambda_2 U_{23} \\ \sqrt{\lambda_2 \lambda_1} U_{13}^T & \lambda_2 U_{23}^T & \lambda_2 U_{33}\end{bmatrix} + \begin{pmatrix}(1-\lambda_1) I_{M_1} & 0 & 0 \\ 0 & (1-\lambda_2) I_{M_2} & 0 \\ 0 & 0 & (1-\lambda_2) I_{M_3}\end{pmatrix},$$

здесь $U_{13} = U_{12} U_{22}^{-1} U_{23}$, λ_1 и λ_2 — параметры регуляризации, вычисленные на импутированных и секвенированных данных соответственно, а I_{M_*} — единичные матрицы размерности M_* . Основываясь на предположении о многомерном нормальном распределении z-статистик, регуляризация корреляционных матриц обеспечит их устойчивость.

Таким образом, предложенный метод, который реконструирует неизвестные корреляции между импутированными и секвенированными генотипами вариантов в гене на основе корреляций между импутированными вариантами и корреляций между секвенированными вариантами без учета индивидуальных генотипов, может быть успешно применен для объединения разных типов генотипических данных с целью поиска сигналов, обусловленных совместным влиянием импутированных и секвенированных вариантов в генах.

Данное исследование проводилось с использованием ресурсов Биобанка Великобритании в рамках заявок № 18219 и № 59345.

Работа Г.Р. Свищёвой, И.В. Зоркольева, Н.М. Белоноговой и Е.Е. Елгаевой выполнена при финансовой поддержке гранта Российского научного фонда (РНФ) № 23-25-00209.

Настоящая статья не содержит каких-либо исследований с участием в качестве объектов людей.

Авторы заявляют, что у них нет конфликта интересов.

Вклад авторов. Г.Р. Свищёва: концептуализация, методология и апробация, исследование, визуализация, написание статьи; А.В. Кириченко: подготовка данных, исследование, написание статьи; Н.М. Белоногова: исследование, написание статьи; Е.Е. Елгаева: написание статьи; Я.А. Цепилов: концептуализация, написание статьи; И.В. Зоркольева: подготовка данных, исследование, написание статьи; Т.И. Аксенович: концептуализация: научное редактирование и управление, написание статьи. Все соавторы обсуждали результаты и внесли вклад в подготовку окончательной версии рукописи.

СПИСОК ЛИТЕРАТУРЫ

1. Eichler E.E., Flint J., Gibson G. et al. Missing heritability and strategies for finding the underlying causes of complex disease // *Nat. Rev. Genet.* 2010. V. 11. № 6. P. 446–450. <https://doi.org/10.1038/nrg2809>
2. Li B., Leal S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data // *The Am. J. Hum. Genet.* 2008. V. 83. № 3. P. 311–321. <https://doi.org/10.1016/j.ajhg.2008.06.024>
3. Cirulli E.T. The increasing importance of gene-based analyses // *PloS Genetics.* 2016. V. 12. № 4. <https://doi.org/10.1371/journal.pgen.1005852>
4. Kang G., Jiang B., Cui Y. Gene-based genomewide association analysis: A comparison study // *Curr. Genomics.* 2013. V. 14. № 4. P. 250–255. <https://doi.org/10.2174/13892029113149990001>
5. Li Y., Willer C., Sanna S., Abecasis G. Genotype imputation // *Ann. Rev. Genomics and Hum. Genet.* 2009. V. 10. P. 387–406. <https://doi.org/10.1146/annurev.genom.9.081307.164242>
6. Uffelmann E., Huang Q.Q., Munung N.S. et al. Genome-wide association studies // *Nat. Rev. Methods Primers.* 2021. V. 1. № 59. P. 1–21. <https://doi.org/10.1038/s43586-021-00056-9>
7. Guo Y., Long J., He J. et al. Exome sequencing generates high quality data in non-target regions // *BMC Genomics.* 2012. V. 13. № 1. P. 1–10. <https://doi.org/10.1186/1471-2164-13-194>
8. Clark M.J., Chen R., Lam H.Y. et al. Performance comparison of exome DNA sequencing technologies // *Nat. Biotechnol.* 2011. V. 29. № 10. P. 908–914. <https://doi.org/10.1038/nbt.1975>
9. Stanley J.C., Wang M.D. Restrictions on the possible values of r_{12} , given r_{13} and r_{23} // *Educational and Psychol. Measurement.* 1969. V. 29. № 3. P. 579–581.
10. Glass G.V., Collins J.R. Geometric proof of the restriction on the possible values of r_{xy} when r_{xz} and r_{yz} are fixed // *Educational and Psychol. Measurement.* 1970. V. 30. № 1. P. 37–39.
11. Budden M., Hadavas P., Hoffman L., Pretz C. Generating valid 4×4 correlation matrices // *Applied Mathemat. E-Notes.* 2007. V. 7. P. 53–59.
12. Glunt W., Hayden T., Johnson C.R., Tarazaga P. Positive definite completions and determinant maximization // *Linear Algebra and its Applications.* 1999. V. 288. P. 1–10. [https://doi.org/10.1016/S0024-3795\(98\)10211-2](https://doi.org/10.1016/S0024-3795(98)10211-2)
13. Vandenberghe L., Boyd S., Wu S.-P. Determinant maximization with linear matrix inequality constraints // *SIAM J. Matrix Analysis and Applications.* 1998. V. 19. № 2. P. 499–533. <https://doi.org/10.1137/S0895479896303430>
14. Georgescu D.I., Higham N.J., Peters G.W. Explicit solutions to correlation matrix completion problems, with an application to risk management and insurance // *Royal Soc. Open Sci.* 2018. V. 5. № 3. P. 172348.
15. Grone R., Johnson C.R., Sá E.M., Wolkowicz H. Positive definite completions of partial Hermitian matrices // *Linear Algebra and its Applications.* 1984. V. 58. P. 109–124.
16. Popescu O., Rose C., Popescu D.C. Maximizing the determinant for a special class of block-partitioned matrices // *Mathem. Problems in Engineering.* 2004. V. 2004. P. 49–61. <https://doi.org/10.1155/S1024123X04307027>
17. Li B., Liu D.J., Leal S.M. Identifying rare variants associated with complex traits via sequencing // *Curr. Protocols in Hum. Genet.* 2013. V. 78. № 1. P. 1–26. <https://doi.org/10.1002/0471142905.hg0126s78>
18. Wu M.C., Lee S., Cai T. et al. Rare-variant association testing for sequencing data with the sequence kernel association test // *The Am. J. Hum. Genet.* 2011. V. 89. № 1. P. 82–93. <https://doi.org/10.1016/j.ajhg.2011.05.029>
19. Jiang L., Zheng Z., Fang H., Yang J. A generalized linear mixed model association tool for biobank-scale data // *Nat. Genet.* 2021. V. 53. № 11. P. 1616–1621. <https://doi.org/10.1038/s41588-021-00954-4>
20. Svishcheva G.R. A generalized model for combining dependent SNP-level summary statistics and its extensions to statistics of other levels // *Scientific Reports.* 2019. V. 9. № 1. P. 1–8. <https://doi.org/10.1038/s41598-019-41827-5>
21. Svishcheva G.R., Belonogova N.M., Zorkoltseva I.V. et al. Gene-based association tests using GWAS summary statistics // *Bioinformatics.* 2019. V. 35. № 19. P. 3701–3708. <https://doi.org/10.1093/bioinformatics/btz172>
22. Belonogova N.M., Svishcheva G.R., Kirichenko A.V. et al. sumSTAAR: A flexible framework for gene-based

association studies using GWAS summary statistics // PloS Comput. Biology. 2022. Т. 18. № 6.
<https://doi.org/10.1371/journal.pcbi.1010172>

23. Тихонов А.Н. О решении некорректно поставленных задач и методе регуляризации // ДАН. 1963. Т. 151. № 3. С. 501–504.

Reconstruction Of a Matrix Of Genotypic Correlations Between Variants Within A Gene For Joint Analysis Of Imputed And Sequenced Data

G. R. Svishcheva^{1,2,*}, A. V. Kirichenko¹, N. M. Belonogova¹, E. E. Elgaeva^{1,3},
Ya. A. Tsepilov¹, I. V. Zorkoltseva¹, T. I. Axenovich¹

¹ Federal Research Centre, Institute of Cytology and Genetics, Siberian Branch
of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

² Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia

³ Novosibirsk State University, 630090 Novosibirsk, Russia

*e-mail: gulsvi@mail.ru

When combining imputed and sequenced data in a single gene-based association analysis, the problem of reconstructing genetic correlation matrices arises. It is related to the fact that for a gene, we know the correlations between genotypes of all imputed variants and the correlations between genotypes of all sequenced variants, but we do not know the correlations between genotypes of variants, one of which is imputed and the other is sequenced. To recover these correlations, we propose an efficient method based on maximising the determinant of the matrix. This method has a number of useful properties and has an analytical solution for our task. Approbation of the proposed method was performed by comparing reconstructed and real correlation matrices constructed on individual genotypes from the UK biobank. Comparison of the results of gene-based association analysis performed by the SKAT, BT and PCA methods on reconstructed and real matrices, using modelled summary statistics and calculated summary statistics on real phenotypes, showed high quality of reconstruction and robustness of the method to different gene structures.

Keywords: imputed and sequenced genotypes, gene-based association analysis, genetic variants, summary statistics

УДК 577.21:616.132-007.64

РАЗНОНАПРАВЛЕННОЕ ИЗМЕНЕНИЕ УРОВНЯ МЕТИЛИРОВАНИЯ CpG-САЙТОВ В 5' РЕГИОНЕ ГЕНА *TBX20* В ВОСХОДЯЩЕЙ АОРТЕ ПРИ АТЕРОСКЛЕРОЗЕ И АНЕВРИЗМЕ

© 2024 г. Ю. А. Королёва¹*, И. А. Гончарова¹, А. А. Зарубин¹, С. А. Шипулина¹,
А. А. Слепцов¹, Д. С. Панфилов², Б. Н. Козлов², М. С. Назаренко¹

¹Научно-исследовательский институт медицинской генетики, Томский национальный
исследовательский медицинский центр Российской академии наук, Томск, 634050 Россия

²Научно-исследовательский институт кардиологии, Томский национальный
исследовательский медицинский центр Российской академии наук,
Томск, 634012 Россия

*e-mail: yuliya.koroleva@medgenetics.ru

Поступила в редакцию 30.11.2023 г.

После доработки 28.12.2023 г.

Принята к публикации 10.01.2024 г.

Выявлено гипометилирование пяти CpG-сайтов в 5'-регионе гена *TBX20* (локус 7p14.2) в тканях атеросклеротической бляшки аорты по сравнению с ее дилатированным участком у больных с аневризмой восходящей аорты. С привлечением внешних данных мы установили, что при диссекции аорты и атеросклерозе аорты уровень метилирования ДНК региона chr7:35253926-35262250 изменяется разнонаправленно. Полученные результаты свидетельствуют об изменении эпигенетической регуляции как при атеросклеротическом поражении аорты, так и при ее аневризме.

Ключевые слова: *TBX20*, аневризма аорты, атеросклероз аорты, метилирование.

DOI: 10.31857/S0016675824070091 **EDN:** BHIOTN

Метилирование ДНК тесно связано с развитием и прогрессией сердечно-сосудистых заболеваний, в том числе обусловленных атеросклеротическим поражением артерий. Опубликованные клинические и экспериментальные наблюдения показывают, что метилирование ДНК может быть вовлечено в патогенез такой жизнеугрожающей патологии как аневризма аорты (АА) [1, 2].

Между АА и атеросклеротическим поражением сердечно-сосудистой системы наблюдаются сложные взаимоотношения. С одной стороны, оба состояния имеют общие факторы риска, такие как гипертензия и курение, а наличие атеросклеротических бляшек аорты (АБА) в ее восходящем отделе считается фактором риска развития аневризмы [3, 4]. Однако, с другой стороны, в некоторых исследованиях высказано предположение о протективном эффекте аневризмы восходящего отдела грудной аорты в отношении риска развития и осложненного течения атеросклероза артерий различной локализации [5, 6]. Этот эффект может быть обусловлен не только генетическими особенностями, но и эпигенетическими модификациями, в том числе и

разнонаправленным изменением метилирования каких-либо генов или их регуляторных элементов. В связи с этим интересным представлялся анализ уровня метилирования ДНК в тканях восходящей аорты и выявление его паттернов при аневризме и атеросклеротическом поражении сосуда.

Уровень метилирования ДНК был проанализирован у шести мужчин с несиндромальными спорадическими формами аневризмы восходящей аорты без сочетанного поражения аортального клапана. Возраст обследованных составил 48–64 г. У всех пациентов наблюдалась артериальная гипертензия, у четырех — ожирение, у трех — стенокардия и у одного пациента — сахарный диабет второго типа. Атеросклероз коронарных или сонных артерий (степень стеноза больше 60%) по результатам ультразвукового исследования выявлен у трех пациентов, в доклинической стадии (30%) — у двух пациентов. У трех пациентов отмечены фиброзные атеросклеротические бляшки восходящей аорты, еще у одного пациента обнаружено атеросклеротическое поражение восходящей аорты на стадии липидных полос.

Всем пациентам проведено открытое хирургическое лечение в НИИ кардиологии г. Томска в период 2020–2023 гг. У четырех пациентов взяты три фрагмента ткани аорты: атеросклеротическая бляшка дилатированной части аорты (АБА); дилатированная часть, не пораженная атеросклерозом (АА) и нерасширенная проксимальная часть дуги аорты (относительно интактная – НА). У двух пациентов не наблюдалось атеросклеротического поражения аорты, поэтому взяты только два фрагмента сосуда – АА и НА.

От всех участников было получено информированное письменное согласие на обследование. Исследование одобрено этическим комитетом Научно-исследовательского института медицинской генетики Томского национального исследовательского медицинского центра Российской академии наук (регистрационный номер 191, протокол № 13 от 15.11.2021 г.).

Метилирование ДНК оценивали методом бисульфитного секвенирования ограниченных наборов геномных локусов (RRBS). Для выполнения анализа ДНК в количестве 100 мкг на образец фрагментировали расщеплением рестриктазой Msp I (Сибэнзим). Репарацию концов и А-хвостов фрагментов ДНК осуществляли с использованием набора для подготовки библиотеки ДНК NEBNext Ultra для Illumina с последующим лигированием метилированных адаптеров (IDT) и очисткой продукта. Далее проводили отбор фрагментов размером 150–500 п.н. при помощи магнитных шариков (Agencourt AM Pure XP, Beckman Coulter). Лигированную с адаптером и очищенную ДНК обрабатывали бисульфитом натрия с помощью набора EZ DNA Methylation Kit (Zymo Research) и амплифицировали в течение 14 циклов ПЦР (NEBNext Q5U Master Mix, New England Biolabs, Inc) с использованием универсальных и индексных праймеров (NEBNext Multiplex Oligos для Illumina). Полученный ПЦР-продукт очищали на магнитных шариках (Agencourt AMPure XP, Beckman Coulter). После контроля качества методом капиллярного электрофореза (Bioanalyzer 2100, Agilent) библиотеку RRBS секвенировали на приборе Illumina HiSeq1500 (2 × 150 п.н.).

Результаты RRBS секвенирования были обработаны с помощью платформы DRAGEN Bio-IT v.3.9.5 (Illumina) и сопоставлены с геномом человека (сборка GRCh38). Для оценки качества использовали программу MultiQC v.1.11. На один образец приходилось 51.9 [47.4; 63.6] млн прочтений. Анализ дифференциального метилирования ДНК проводился в программной среде R с использованием пакетов methylKit и limma. Минимальное покрытие составило 10х. Дифференциально метилированными сайтами (ДМС) считались CpG-сайты с разницей среднего уровня метилирования между группами образцов $|\Delta\beta| \geq 20\%$ и уровнем

значимости с поправкой на множественные сравнения $p_{FDR} < 0.05$.

В результате исследования было установлено гипометилирование 5 CpG-сайтов локуса 7p14.2 в тканях атеросклеротической бляшки по сравнению с участком дилатированной аорты (табл. 1). Уровень метилирования в АБА по сравнению с АА снижен на 44.1–58.4% ($p_{FDR} < 0.05$). Данные CpG-сайты расположены в 2558 п.н. друг от друга и на расстоянии 5179–7757 п.н. от 5'UTR гена *TBX20* (рис. 1).

Согласно геномному браузеру UCSC [7], четыре из выявленных нами дифференциально метилированных CpG-сайтов приходятся на сайты связывания транскрипционных факторов (ТФ) (табл. 1). При этом число ТФ, способных связываться с областью, включающей конкретный CpG-сайт, варьирует от 2 (для chr7:35259293) до 24 (для chr7:35261291). Кроме этого, два CpG-сайта локализованы непосредственно в последовательности энхансера EN38E2547742 и три в области гена нкРНК *ENSG00000226063* (табл. 1, рис. 1). Один из ДМС (chr7:35259405) представляет собой однонуклеотидный полиморфизм с высокой частотой мутационного аллеля среди представителей европейской популяции (rs78661208, MAF C = 0.3876).

Учитывая то, что между группами АА и НА, а также АБА и НА различия в уровне метилирования отдельных CpG-сайтов превышали 20% (однако группы сравнения были малы по размеру), мы рассмотрели внешние данные по анализу метилирования ДНК с помощью метилочипов Illumina в тканях восходящей аорты при ее диссекции ($n = 12$) по сравнению с нормальной тканью аорты ($n = 6$) (датасет GSE84274, загружен из базы данных GEO [8]), а также атеросклерозе аорты ($n = 15$) по сравнению с непораженной тканью ($n = 15$) (датасет GSE46394), загружен из базы данных GEO [8]).

При диссекции аорты в отдельных CpG-сайтах региона chr7:35253926–35262250 выявлено статистически значимое гиперметилирование (разница уровня метилирования по сравнению с непораженной тканью составляла от 1.39 до 17.70%), при этом гиперметилированные сайты группировались в три региона (рис. 1) (GSE84274 [8]). При атеросклерозе аорты, напротив, выявлены три региона, в которых CpG-сайты были гипометилированы (с разницей уровня метилирования от –3.16 до –15.50%) по сравнению с непораженной тканью аорты (GSE46394 [8]) (рис. 1). Таким образом, при диссекции аорты и атеросклерозе аорты уровень метилирования ДНК региона chr7:35253926–35262250 (в области генов нкРНК *ENSG00000289335* и *ENSG00000226063*) изменяется разнонаправленно.

Ген *TBX20* является членом подсемейства Tbx1 Т-бокс-содержащих генов, контролирующих различные факторы транскрипции, необходимые для

Таблица 1. Дифференциально метилированные CpG-сайты, локализованные в 5' регионе гена *TVX20* в тканях из различных отделов восходящей аорты

CpG-сайты	Локализация CpG-сайтов	Предсказанные сайты связывания ТФ (UCSC/Haploregv4.2)	SNP/MAF (GnomAD v. 4.0.0)	Уровень метилирования				Разница уровня метилирования		
				АБА (n = 6)	АА (n = 4)	НА (n = 4)	АБА/АА (p _{FDR})	АБА/НА (p _{FDR})	АА/НА (p _{FDR})	
chr7:35258799	Интрон гена нкРНК AC009531.2 (ENSG00000226063.1)	EBF1, Ebf2, EBF3	rs78696287/ T = 0.0001	39.4 ± 17.9	83.5 ± 4.6	46.1 ± 13.9	-44.1 (0.025)	-6.7 (-)	37.4 (0.384)	
chr7:35259293	Интрон гена нкРНК AC009531.2 (ENSG00000226063.1)	ZNF528, ZNF343	Нет	20.7 ± 10.2	78.5 ± 7.4	54.5 ± 17.1	-57.8 (0.001)	-33.8 (0.241)	24.0 (0.936)	
chr7:35259405	Экзон 1-го гена нкРНК AC009531.2 (ENSG00000226063.1)	IRF7, IRF8/AP-1, SETDB1, Znf143	rs78661208/ C = 0.3876	22.4 ± 11.4	70.7 ± 12.0	54.7 ± 14.9	-48.3 (0.024)	-32.3 (0.415)	16.0 (0.936)	
chr7:35261291	энхансер EN38E2547742	KLF1, KLF2, KLF4, KLF5, KLF7, KLF10, KLF12, KLF14, KLF15, MAZ, PATZ1, PRDM9, SPI, SP2, SP3, SP4, SP9, FA-P2A, TFAP2B, TFAP2C, Wt1, ZNF148, ZNF263, ZNF281	нет	4.8 ± 6.7	63.2 ± 5.9	32.7 ± 23.3	-58.4 (0.017)	-27.9 (0.688)	30.5 (0.936)	
chr7:35261357	энхансер EN38E2547742	нет	нет	14.3 ± 5.8	69.3 ± 10.3	33.7 ± 20.0	-55.0 (0.010)	-19.4 (-)	35.6 (0.299)	

Примечание. ТФ – транскрипционные факторы, SNP – однонуклеотидные полиморфизмы, MAF – частота минорного аллеля (для европеоидных популяций), GnomAD – база данных генетических вариантов (<https://gnomad.broadinstitute.org/>), UCSC – геномный браузер University of California Santa Cruz (<https://genome.ucsc.edu/>), Haploreg – программа по функциональному анализу генетических вариантов и их расположению в галоглилах (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>), p_{FDR} – уровень значимости р с поправкой на множественные сравнения (FalseDiscoveryRate), АБА – атеросклеротическая бляшка дилатированной части аорты, АА – дилатированная часть аорты, не пораженная атеросклерозом, НА – нерасширенная проксимальная (относительно интактная) часть дуги аорты, нкРНК – некодирующая РНК.

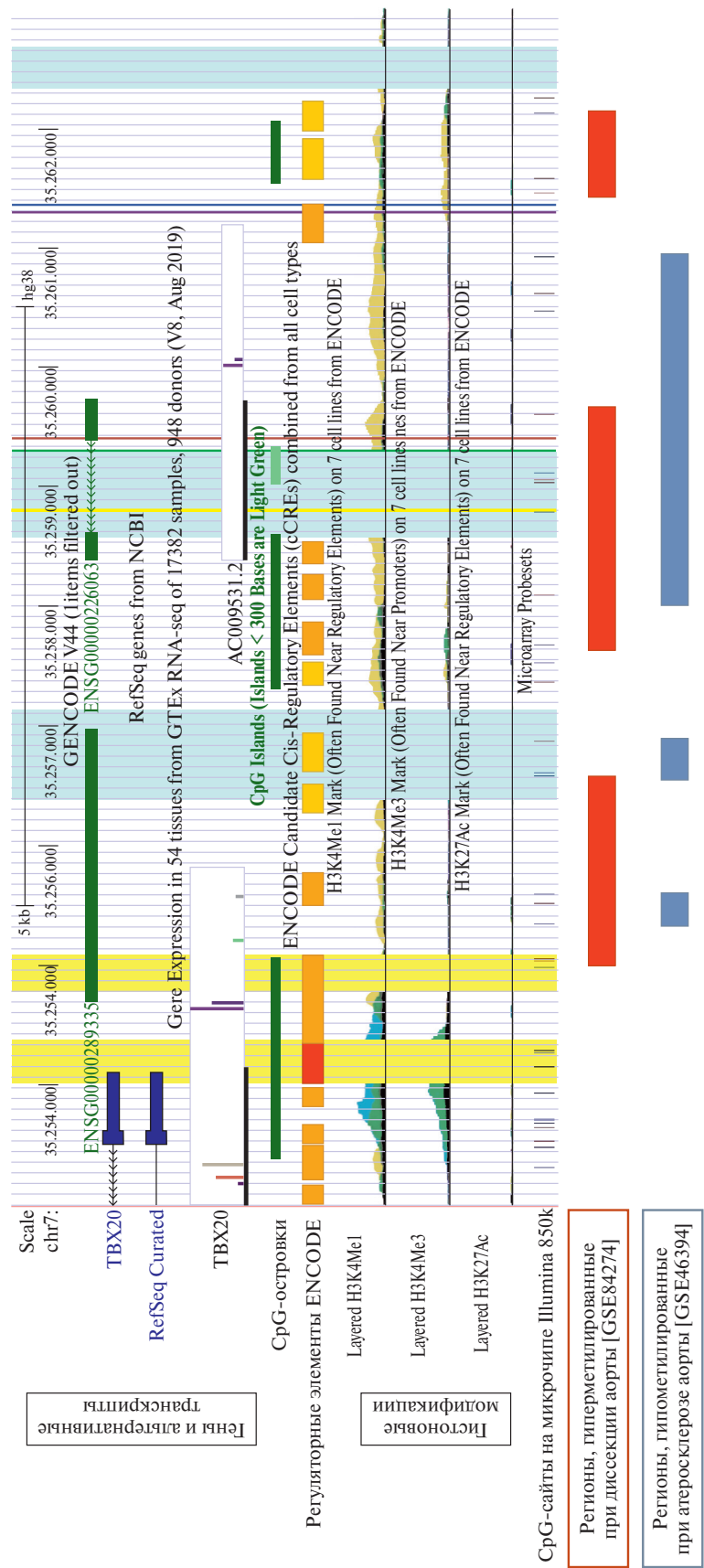


Рис. 1. Локализация и эпигенетический контекст исследуемой 5'-области гена *TBX20*. Тонкими цветными вертикальными линиями отмечены пять ДМС между атеросклеротической близкой дилатированной части аорты, не пораженной атеросклерозом, дилатированной частью аорты и нерасширенной проксимальной частью аорты (данное исследование). Широкие вертикальные полосы обозначены: желтым – регионы, гипометилированные в тканях миокарда при тетраде Фалло [18, 19], голубым – регионы, гипометилированные в тканях аорты при атеросклерозе [17]. Различия по уровню метилирования между непораженными тканями аорты и атеросклеротически измененной аортой для трех регионов (слева направо) составили 41, 29 и 30%. Красные горизонтальные линии обозначают регионы гиперметилированных CpG-сайтов при диссекции аорты (GSE84274 [8]), синие горизонтальные линии – регионы гипометилированных CpG-сайтов при атеросклерозе аорты по сравнению с непораженной тканью аорты (GSE46394 [8]). Все треки получены и/или выровнены на UCSC Genome Browser (сборка генома GRCh38).

эмбрионального развития и органогенеза. *Tbx20* важен для развития сердечно-сосудистой системы и участвует в ремоделировании сердца в ответ на патофизиологические стрессы [9]. Редкие патогенные варианты *TBX20* у человека ассоциированы со сложным спектром врожденных пороков сердца, включающим дефекты формирования перегородок и клапанов, коарктацию аорты, а также с аневризмой грудной аорты в сочетании с двустворчатым аортальным клапаном [10–12]. В исследованиях GWAS в области гена *TBX20* выявлены 16 частых SNP, ассоциированных с диаметром восходящей аорты, аневризмой аорты и атеросклерозом сонных и коронарных артерий [13].

Как генетические варианты, так и изменение эпигенетической регуляции, включая метилирование генов, связанных с развитием сосудистой системы и сердца (в том числе таких, как Т-бокс-содержащие гены), могут способствовать нарушению целостности аорты и патогенезу аневризмы [2]. Высокая функциональная активность гена *TBX20* выявлена в фибробластах и гладкомышечных клетках аорты у пациентов с аневризмой восходящей аорты и контрольной группы [14, 15]. В то же время данные по изменению метилирования в области гена *TBX20* в клетках и тканях восходящей аорты при ее аневризме отсутствуют.

Также мало изучена роль *TBX20* в развитии атеросклероза артерий различной локализации. Однако недавно с помощью технологии секвенирования транскриптома отдельных клеток была выявлена преимущественная экспрессия *TBX20* в гладкомышечных клетках и фибробластах пораженных атеросклерозом коронарных артерий человека. Более того, показано, что в гладкомышечных клетках и фибробластах область данного гена “насыщена” участками открытого хроматина [16]. Ранее в тканях грудной аорты, пораженных атеросклерозом, относительно непораженной ткани были выявлены несколько протяженных гипометилированных областей, затрагивающих 5'-UTR гена *TBX20* [17].

Результаты настоящего исследования согласуются с представленными выше, поскольку гипометилированные в атеросклеротической бляшке аорты CpG-сайты локализируются в том же регионе или непосредственной близости от него (рис. 1). Показано, что и промоторная область гена *TBX20*, находящаяся в 5000–8000 п.н. от изученного в данном исследовании региона, гипометилирована в тканях миокарда при тетраде Фалло [18, 19]. Выявленные в настоящем исследовании гипометилированные в атеросклеротической бляшке аорты CpG-сайты локализованы в регионе концентрации регуляторных элементов, в частности энхансеров, и являются сайтами связывания различных ТФ. В частности, среди ТФ, предположительно способных связываться с областью CpG-сайта chr7:35261291, отмечен SP1, который способствует активации гена

TBX20. Можно предположить, что снижение метилирования данного сайта при атеросклерозе аорты приведет к повышению активности гена *TBX20* в клетках атеросклеротической бляшки (преимущественно в гладкомышечных клетках и фибробластах). Это предположение косвенно подтверждается данными о повышении экспрессии гена *TBX20* при снижении уровня метилирования другого сайта связывания ТФ SP1 в области промотора *TBX20* в тканях миокарда при тетраде Фалло [19].

В настоящем исследовании показано, что ДМС chr7:35259405 является полиморфным сайтом с достаточно высокой частотой (36%) в европеоидных популяциях (табл. 1). Поскольку метилированию в ДНК подвергается цитозин, в случае однонуклеотидных полиморфизмов замена нуклеотида может приводить к утрате существующего или формированию нового сайта метилирования. Для установления роли однонуклеотидных полиморфизмов, а также участков связывания с транскрипционными факторами в данном локусе генома требуется проведение дополнительных исследований.

Таким образом, в данном исследовании выявлено гипометилирование пяти CpG-сайтов, расположенных на расстоянии примерно 5000 п.н. от гена *TBX20* (в области гена нкРНК *ENSG00000226063*), в тканях атеросклеротической бляшки аорты по сравнению с ее дилатированным участком у больных с аневризмой восходящей аорты. В то же время показано, что при диссекции аорты и атеросклерозе аорты уровень метилирования ДНК региона chr7:35253926–35262250 (в области генов нкРНК *ENSG00000289335* и *ENSG00000226063*) изменяется разнонаправленно. Полученные результаты свидетельствуют об изменении эпигенетической регуляции как при атеросклеротическом поражении аорты, так и при ее аневризме и выступают подтверждением возможной роли гена *TBX20* и генов нкРНК в процессах атерогенеза у лиц с аневризмой восходящей аорты и аортальным клапаном нормального строения.

Исследование выполнено за счет гранта РНФ № 22-25-00701.

Исследование одобрено Этическим комитетом Научно-исследовательского института медицинской генетики Томского национального исследовательского медицинского центра Российской академии наук — протокол № 13 от 15 ноября 2021 г., а также Этическим комитетом Научно-исследовательского института кардиологии Томского национального исследовательского медицинского центра Российской академии наук — протокол № 213 от 12 мая 2021 г.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской

декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики. От каждого из включенных в исследование участников было получено информированное добровольное согласие. Все обследованные — совершеннолетние.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Portelli S.S., Robertson E.N., Malecki C. et al.* Epigenetic influences on genetically triggered thoracic aortic aneurysm // *Biophys. Rev.* 2018. V. 10. № 5. P. 1241–1256.
<https://doi.org/10.1007/s12551-018-0460-1>
2. *Liu P., Zhang J., Du D. et al.* Altered DNA methylation pattern reveals epigenetic regulation of Hox genes in thoracic aortic dissection and serves as a biomarker in disease diagnosis // *Clin. Epigenetics.* 2021. V. 13. № 1. P. 124.
<https://doi.org/10.1186/s13148-021-01110-9>
3. *Leone O., Corsini A., Pacini D. et al.* The complex interplay among atherosclerosis, inflammation, and degeneration in ascending thoracic aortic aneurysms // *J. Thorac. Cardiovasc. Surg.* 2020. V. 160. № 62. P. 1434–1443.
<https://doi.org/10.1016/j.jtcvs.2019.08.108>
4. *Isselbacher E.M., Preventza O., H. Black J. 3rd. et al.* ACC/AHA guideline for the diagnosis and management of aortic disease: A report of the American Heart Association // *Circulation.* 2022. V. 146. № 24. P. e334–e482.
<https://doi.org/10.1161/CIR.0000000000001106>
5. *Chau K.H., Bender J.R., Eleftheriades J.A.* Silver lining in the dark cloud of aneurysm disease // *Cardiology.* 2014. V. 128. № 4. P. 327–332.
<https://doi.org/10.1159/000358123>
6. *Weininger G., Ostberg N., Shang M. et al.* Lipid profiles help to explain protection from systemic atherosclerosis in patients with ascending aortic aneurysm // *J. Thorac. Cardiovasc. Surg.* 2022. V. 163. № 2. P. e129–e132.
<https://doi.org/10.1016/j.jtcvs.2021.09.031>
7. *Nassar L.R., Barber G.P., Benet-Pagès A. et al.* The UCSC Genome Browser database: 2023 update // *Nucl. Ac. Res.* 2023. V. 51. № D1. P. D1188–D1195.
<https://doi.org/10.1093/nar/gkac1072>
8. *Barrett T., Wilhite S.E., Ledoux P. et al.* NCBI GEO: archive for functional genomics data sets—update // *Nucl. Ac. Res.* 2013. V. 41. P. D991–D995.
<https://doi.org/10.1093/nar/gkac1072>
9. *Chen Y., Xiao D., Zhang L. et al.* The role of Tbx20 in cardiovascular development and function // *Front. Cell Dev. Biol.* 2021. V. 9.
<https://doi.org/10.3389/fcell.2021.638542>
10. *Kirk E.P., Sunde M., Costa M.W. et al.* Mutations in cardiac T-box factor gene *TBX20* are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy // *Am. J. Hum. Genet.* 2007. V. 81. № 2. P. 280–291.
<https://doi.org/10.1086/519530>
11. *Luyckx I., Kumar A.A., Reyniers E. et al.* Copy number variation analysis in bicuspid aortic valve-related aortopathy identifies *TBX20* as a contributing gene // *Eur. J. Hum. Genet.* 2019. V. 27. № 7. P. 1033–1043.
<https://doi.org/10.1038/s41431-019-0364-y>
12. *Tcheandjieu C., Xiao K., Tejeda H. et al.* High heritability of ascending aortic diameter and trans-ancestry prediction of thoracic aortic disease // *Nat. Genet.* 2022. V. 54. № 6. P. 772–782.
<https://doi.org/10.1038/s41588-022-01070-7>
13. *Sollis E., Mosaku A., Abid A. et al.* The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource // *Nucl. Ac. Res.* 2023. V. 51. № D1. P. D977–D985.
<https://doi.org/10.1093/nar/gkac1010>
14. *Li Y., Ren P., Dawson A. et al.* Single-cell transcriptome analysis reveals dynamic cell populations and differential gene expression patterns in control and aneurysmal human aortic tissue // *Circulation.* 2020. V. 142. № 14. P. 1374–1388.
<https://doi.org/10.1161/CIRCULATIONAHA.120.046528>
15. *Ma W.F., Hodonsky C.J., Turner A.W. et al.* Enhanced single-cell RNA-seq work-flow reveals coronary artery disease cellular cross-talk and candidate drug targets // *Atherosclerosis.* 2022. V. 340. P. 12–22.
<https://doi.org/10.1016/j.atherosclerosis.2021.11.025>
16. *Hodonsky C.J., Turner A.W., Khan M.D. et al.* Integrative multi-ancestry genetic analysis of gene regulation in coronary arteries prioritizes disease risk loci // *medRxiv.* 2023. V. 2.
<https://doi.org/10.1101/2023.02.09.23285622>
17. *Lacey M., Baribault C., Ehrlich K.C., Ehrlich M.* Atherosclerosis-associated differentially methylated regions can reflect the disease phenotype and are often at enhancers // *Atherosclerosis.* 2019. V. 280. P. 183–191.
<https://doi.org/10.1016/j.atherosclerosis.2018.11.031>
18. *Yang X., Kong Q., Li Z. et al.* Association between the promoter methylation of the *TBX20* gene and tetralogy of fallot // *Scand. Cardiovascular J.* 2018. V. 52. № 5. P. 287–291.
<https://doi.org/10.1080/14017431.2018.1499955>
19. *Gong J., Sheng W., Ma D. et al.* DNA methylation status of *TBX20* in patients with tetralogy of Fallot // *BMC Med. Genomics.* 2019. V. 12. № 1. P. 75.
<https://doi.org/10.1186/s12920-019-0534-3>

Methylation Levels In The 5' Region Of The *TBX20* Gene In The Ascending Aorta Change In Opposite Direction In Atherosclerosis And Aneurysm

**Yu. A. Koroleva^{a, *}, I. A. Goncharova^a, A. A. Zarubin^a, S. A. Shipulina^a,
A. A. Sleptsov^a, D. S. Panfilov^b, B. N. Kozlov^b, M. S. Nazarenko^a**

*^aResearch Institute of Medical Genetics, Tomsk National Research Medical Center,
Russian Academy of Sciences, Tomsk, 634050 Russia*

*^bResearch Institute of Cardiology, Tomsk National Research Medical Center,
Russian Academy of Sciences, Tomsk, 634012 Russia*

**e-mail: yuliya.koroleva@medgenetics.ru*

We found hypomethylation of 5 CpG sites in the 5' region of *TBX20* gene (7p14.2) in the tissues of atherosclerotic aortic plaque compared to dilated part of aorta in patients with ascending aortic aneurysm. Using GEO database, we found that the DNA methylation level in the chr7:35253926-35262250 region changes in opposite direction in aortic dissection and aortic atherosclerosis. The results suggest an alteration in epigenetic regulation both in aortic atherosclerosis and aortic aneurysm.

Keywords: *TBX20*, aortic aneurysm, atherosclerosis, DNA methylation.

Funding: The research was supported by the grant of the Russian Science Foundation (project No. 22-25-00701).

КАРИОТИП И МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКАЯ ДИФФЕРЕНЦИАЦИЯ 24-ХРОМОСОМНОЙ ФОРМЫ СЕРОГО ХОМЯЧКА *NOTHOCRICETULUS MIGRATORIUS* ИЗ ТЯНЬ-ШАНЯ

© 2024 г. О. В. Брандлер¹*, А. В. Блехман¹

¹Институт биологии развития им. Н.К. Кольцова Российской академии наук, Москва, 119334 Россия

*e-mail: rasmarmot@yandex.ru

Поступила в редакцию 05.12.2023 г.

После доработки 12.01.2024 г.

Принята к публикации 16.01.2024 г.

Серый хомячок *Nothocricetulus migratorius*, являясь широкоареальным палеарктическим видом грызунов, обладает кариотипом со стабильным числом хромосом $2n = 22$ на всей территории обитания. Нами найдены серые хомячки с диплоидным числом хромосом $2n = 24$, локально распространенные на Кураминском хребте Тянь-Шаня. Впервые дано описание нового кариотипа и анализ G и NORs — дифференциально окрашенных хромосомных наборов. От 22-хромосомного кариотипа серых хомячков описываемый кариотип отличается морфологией Y-хромосомы и наличием дополнительной пары гетероморфных мелких хромосом. Молекулярно-генетический анализ выявил генетическую дивергенцию 24- и 22-хромосомных форм *N. migratorius*, различия между которыми по митохондриальным маркерам сопоставимы, а по ядерным превышают различия между *Cricetulus barabensis* ($2n = 20$) и *C. pseudogriseus* ($2n = 24$). Полученные нами данные позволяют обсуждать таксономический статус 24-хромосомной формы серых хомячков с Кураминского хребта и рассматривать дифференциацию кариоморф *N. migratorius* как этап хромосомного видообразования.

Ключевые слова: серый хомячок, кариотип, молекулярно-генетическая изменчивость, хромосомное видообразование.

DOI: 10.31857/S0016675824070102 **EDN:** BGWVSN

Серый хомячок *Nothocricetulus migratorius* Pallas, 1773 имеет один из самых больших ареалов среди грызунов Старого Света, занимающий территорию от Восточных Балкан на западе до Центрального Китая и Монголии на востоке и от Волжско-Камского междуречья на севере до Ирана, Афганистана и Пакистана на юге [1–3]. Обитает в равнинной и горной степи. Обладает широкой морфологической изменчивостью, что нашло отражение в описании более полутора десятков подвидов [1], которые образуют две группы — западную *phaeus* и восточную *migratorius* [2], которые также поддерживаются молекулярно-генетическими данными [4]. В отличие от морфологических и молекулярно-генетических признаков кариотип *N. migratorius*, изученный в разных частях ареала, стабилен и имеет $2n = 22$, $NF = 44$ (Nombre Fundamentale) [5–7]. В связи с этим было неожиданным обнаружение нами серых хомячков в Тянь-Шане с отличающимся кариотипом [8]. В настоящей работе мы впервые описываем характеристики нового кариотипа с использованием дифференциальных окрасок хромосом и оцениваем молекулярно-генетические

отличия и филогенетические связи данной хромосомной формы в пределах вида *N. migratorius* и с рядом других видов хомячков рода *Cricetulus*.

Были исследованы хромосомные наборы трех самцов серых хомячков с Ангренского плато, расположенного на Кураминском хребте Тянь-Шаня (41.20° с. ш., 70.62° в. д.), отловленных нами в 1995 г. Для получения хромосомных препаратов использована стандартная методика, обычно применяющаяся при работе с мелкими млекопитающими [9]. Для G-окрашивания хромосом применялся метод, предложенный М. Сибрайтом [10]. Дифференциальное окрашивание ядрышкообразующих районов (ЯОР) хромосом выполнено по методу окраски Ag-NOR У. Хоуэла и Д. Блэка [11].

Для оценки филогенетических связей тянь-шаньских хомячков с другими представителями *Cricetulus* был проведен сравнительный молекулярно-генетический анализ этих животных с тремя особями *N. migratorius* из разных частей ареала (Семипалатинской обл. Казахстана, Монголии и Украины), а также с представителями двух

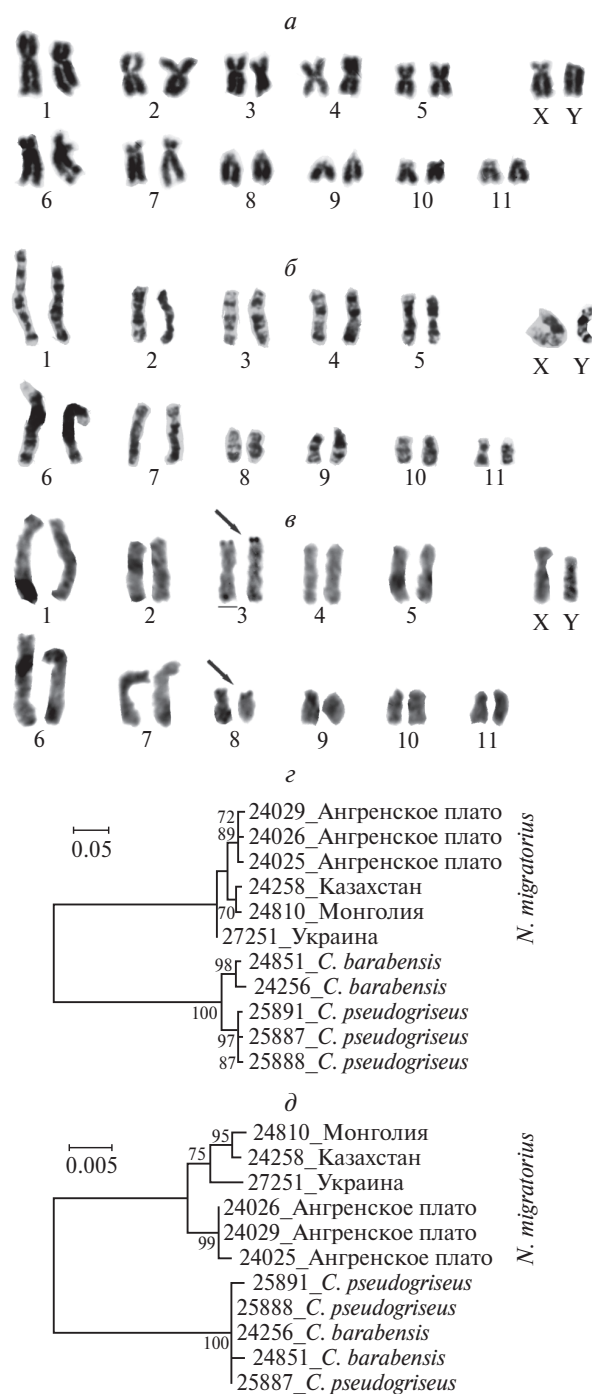


Рис. 1. Кариотип самца серого хомячка *N. migratorius* $2n = 24$. *a* — рутинная окраска, *б* — G-окраска, *в* — Ag-NOR-окраска. Стрелками указано положение ЯОР, подчеркнута хромосома с ЯОР на обоих плечах. Кладогаммы объединенных нуклеотидных последовательностей хомячков по результатам анализа ML: *г* — мтДНК (*cytb* и *COI*), *д* — яДНК (*BGN* и *C-myc*). Цифры в наименовании ветвей соответствуют регистрационным номерам образцов в КТЖ ИБР. Цифрами у узлов обозначены индексы бутстрэп-поддержки.

других видов хомячков — барабинского *Cricetulus barabensis* Pallas, 1770 (две особи из аймака Ховсгел Монголии и Павлодарской области Казахстана) и забайкальского *C. pseudogriseus* Orlov and Iskhakova, 1975 (три особи из аймака Дорнод Монголии). Материалом для данного исследования послужили образцы тканей хомячков, хранящихся в УНУ “Коллекция тканей диких животных для генетических исследований” при Федеральном государственном бюджетном научном учреждении Институт биологии развития им. Н.К. Кольцова РАН (КТЖ ИБР, рег. номер 3579666).

Геномную ДНК выделяли солевым методом [12]. Для анализа использовали последовательности двух митохондриальных молекулярно-генетических маркеров: ген цитохрома *b* (*cytb*) — 1140 пн и фрагмент первой единицы гена цитохромоксидазы (*COI*) — 662 пн, а также два ядерных маркера: фрагмент гена *BGN* — 750 пн и фрагмент гена *C-myc* — 685 пн. Реакции амплификации проводили с опубликованными ранее парами праймеров в соответствии с предложенными для них протоколами: *cytb* — L14727-SP и H15915-SP [13]; *COI* — L5310 и R6036R [14]; *BGN* — BGN-f и BGN-r [15]; *C-myc* — S92-F и S91-R [16]. ПЦР проводили в конечном объеме 25 мкл, содержащем 30–50 нг тотальной ДНК, 0.2 mM dNTP, 4 pmol каждого праймера, ПЦР буфер в однократной финальной концентрации и 1 единицу HS-Taq ДНК полимеразы (Евроген, Россия) на термоциклере AB Veriti 96 well TC (Applied Biosystems). Результаты ПЦР оценивались электрофорезом 2 мкл реакционной смеси в 1.5% агарозном геле, окрашенном бромистым этидием. Для очистки продукта ПЦР использовались наборы для выделения ДНК из агарозного геля и реакционных смесей Cleanup Standart (Евроген, Россия) в соответствии с инструкцией производителя. Реакцию секвенирования проводили в 10 мкл реакционной смеси с теми же праймерами, что и в реакциях амплификации, с использованием набора BigDye v.3.1 (“Applied Biosystems”, США) по протоколам производителя. Полученные фрагменты анализировали в ЦКП ИБР РАН на автоматическом генетическом анализаторе ABI 3500 (“Applied Biosystems”).

Корректировку хроматограмм проводили с помощью программы SeqMan пакета Lasergene 11 (DNASTAR, США). Неоднозначно расшифрованные позиции редактировали вручную. Последовательности по всем четырем маркерам выравнивались независимо с помощью алгоритма MUSCLE [17] в пакете программ MEGA X [18]. Все вновь полученные последовательности депонированы в GenBank NCBI с номерами доступа: PP457730–PP457740, PP481634–PP481644, PP477416–PP477426, PP477427–PP477437. Выравненные фрагменты генов каждого образца были объединены в суммарные последовательности мтДНК (1802

пн) и яДНК (1435 пн). Выбор наилучших моделей эволюции объединенных нуклеотидных последовательностей для анализа по методу максимального правдоподобия (ML) выполнен в программе MEGA X на основе байесовского информационного критерия (*BIC*). Филогенетический анализ (ML) проводили в программе MEGA X с использованием HKY+G модели нуклеотидной эволюции для объединенных последовательностей мтДНК и Kimura 2-параметрической модели для объединенных последовательностей яДНК. Генетические различия оценивались по попарным дистанциям (*p*-distance, d_p). Устойчивость узлов филогенетических деревьев оценивали, применяя бутстрэп-анализ по 1000 репликам.

Исследованные кариотипы трех самцов серого хомячка с Ангренского плато не отличаются друг от друга. Все три особи имеют $2n = 24$, $NF = 46$. Аутосомы представлены пятью парами крупных и средних мета- и субметацентриков, составляющих постепенно убывающий по величине ряд; двумя парами крупных субтелоцентриков; тремя парами мелких субмета- субтелоцентриков и одной самой мелкой гетероморфной парой набора, состоящей из субметацентрика и акроцентрика. X-хромосома — средней величины субметацентрик; Y-хромосома — акроцентрик, практически равный по длине X-хромосоме (рис. 1, а–в). Данный кариотип отличается от ранее описанного для *N. migratorius* [5 и др.] наличием дополнительной самой мелкой гетероморфной пары аутосом и морфологией Y-хромосомы, которая у всех других серых хомячков субметацентрическая. При сравнении G-окрашенных хромосом 24- и 22-хромосомных кариотипов *N. migratorius* [19] прослеживается парная гомология аутосом, за исключением 11-й пары. Степень детализации G-бэндов полученных препаратов также не позволила идентифицировать возможные гомологичные 11-й паре участки в 22-хромосомном кариотипе с более детальной G-окраской [20]. В то же время элементы 11-й пары проявляют сходство с коротким плечом субметацентрической Y-хромосомы *N. migratorius*.

Предположение о возможности происхождения крупных Y-хромосом у хомячков в результате слияния с аутосомами дискутируется длительное время [21, 22], однако прямых доказательств таких хромосомных перестроек до настоящего времени не получено. Образование неополовой хромосомы путем слияния предковой Y-хромосомы и аутосомы обнаружено у рыб с множественной системой половых хромосом [23]. Вопрос, имели ли место подобные перестройки в эволюции половых хромосом у хомячков, требует дальнейшего изучения.

В 24-хромосомном наборе гетероморфная 11-я пара состоит из двуплечей и одноплечей хромосом. Ранее описанный для хомячков из Армении [5] и Таджикистана [7] гетероморфизм по величине

двуплечих хромосом одной из мелких мета-, субметацентрических пар (№ 8) связывался с изменчивостью гетерохроматинового материала. Гомологичная пара (№ 8) в 24-хромосомном кариотипе является субметацентрической и изоморфной (рис. 1, а).

Ag-NOR-окраска стабильно выявляла ЯОР на 3-й и 8-й парах аутосом (рис. 1, в). На гомологах этих хромосом в 22-хромосомном кариотипе обнаружена аналогичная локализация ЯОР [22]. В той же работе ЯОР обнаружены еще на двух парах хромосом, на гомологах которых в описываемом нами кариотипе ЯОР проявлялись нестабильно, что может быть связано с недостаточно высоким качеством препаратов.

Результаты филогенетического анализа, представленные на кладограммах (рис. 1, г, д), показывают сходные результаты по митохондриальным (г) и ядерным (д) маркерам. Все последовательности образуют две клады, одна из которых включает всех *N. migratorius*, а вторая объединяет *C. barabensis* и *C. pseudogriseus*. При этом все три хомячка с описанным нами кариотипом ($2n = 24$) с Ангренского плато образуют компактные кластеры с высокой бутстрэп-поддержкой внутри клад *N. migratorius*. Митохондриальные последовательности *C. barabensis* и *C. pseudogriseus*, имеющих разные кариотипы $2n = 20$ и $2n = 24$ соответственно [24], также образуют отдельные видовые кластеры с высокой бутстрэп-поддержкой внутри общей клад, но последовательности яДНК этих видов объединены в единый слабо дифференцированный кластер. Генетические различия по мтДНК между хромосомными формами *N. migratorius* в два раза ниже ($d_p = 0.02$ по *cytb*), чем между *C. barabensis* и *C. pseudogriseus* ($d_p = 0.04$). При этом различия по ядерным маркерам *N. migratorius* с разными кариотипами на порядок выше ($d_p = 0.01$), чем между *C. barabensis* и *C. pseudogriseus* ($d_p = 0.001$).

В проведенных ранее молекулярно-генетических исследованиях *N. migratorius*, в которых использовались образцы двух из трех хомячков с Ангренского плато, эта форма образовывала отдельную филетическую линию с базальным расположением на дереве [4]. Среди других видов Cricetidae только *C. longicaudatus* и *C. pseudogriseus* имеют $2n = 24$ [25], и лишь у последнего Y-хромосома является относительно мелким акроцентриком, что рассматривается как архаичный признак для хомячков [26] и для Muroidea в целом [25]. Можно предположить, что серые хомячки из западного Тянь-Шаня являются предковой формой для всех остальных *N. migratorius*, 22-хромосомный кариотип которых произошел от описываемого 24-хромосомного набора.

Место находки 24-хромосомных хомячков расположено в пределах распространения группы

подвидов *coerulescens* [4]. Однако отсутствие морфологического материала от кариотипированных особей не позволяет отнести хомячков Кураминского хребта к какой-либо из центральноазиатских морфологических групп. Тем не менее полученные данные позволяют обсуждать таксономический статус 24-хромосомной формы серых хомячков с Кураминского хребта и рассматривать дифференциацию *N. migratorius* как пример хромосомного видообразования [27].

Работа финансировалась из средств Договора НИР № 261 от 08.11.2021, выполняемого в рамках проекта № 075-15-2021-1069 от 28.09.2021 г. ФНТП развития генетических технологий на 2019 – 2027 гг. (II очередь. Биоресурсные коллекции) Минобрнауки России.

Исследование одобрено Этическим комитетом ИБР РАН им. Н.К. Кольцова, протокол № 70 от 25.05.2023 г.

Все применимые международные, национальные и/или институциональные принципы ухода и использования животных были соблюдены.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Громов И.М., Ербаева М.А. Млекопитающие фауны России и сопредельных территорий. Зайцеобразные и грызуны. СПб.: Наука. 1995. 522 с.
2. Лебедев В.С. *Cricetinae* – Млекопитающие России. Систематико-географический справочник // Сб. тр. Зоол. музея МГУ. Т. 52 / под ред. Павлинов И.Я., Лисовский А.А. М.: Т-во научн. изданий КМК, 2012. С. 211–220.
3. Lebedev V.S., Bannikova A.A., Neumann K. et al. Molecular phylogenetics and taxonomy of dwarf hamsters *Cricetulus* Milne-Edwards, 1867 (*Cricetidae*, Rodentia): Description of a new genus and reinstatement of another // *Zootaxa*. 2018. V. 4387. P. 331–349. <https://doi.org/10.11646/zootaxa.4387.2.5>
4. Lebedev V., Poplavskaya N., Bannikova A. et al. Genetic differentiation in *Cricetulus migratorius* Pallas, 1773 (Rodentia, *Cricetidae*) // *Mammalian Biology*. 2018. V. 92. P. 115–119. <https://doi.org/10.1016/j.mambio.2018.05.001>
5. Yerganian G., Papoyan S. Isomorphic sex chromosomes, autosomal heteromorphism, and telomeric associations in the grey hamster of Armenia, *Cricetulus migratorius*, Pall. // *Hereditas*. 1965. V. 52. № 3. P. 307–319. <https://doi.org/10.1111/j.1601-5223.1965.tb01963.x>
6. Гайченко В.А. Хромосомный набор серого хомячка с территории Украины // *Вестник зоологии*. 1974. № 2. С. 79–80.
7. Картавецца И.В. Внутривидовой хромосомный полиморфизм широкоареальных видов: краснохвостой песчанки и серого хомячка // *Фенетика популяций*. Саратов: 1985. С. 84–85.
8. Брандлер О.В. Новая хромосомная форма серого хомячка *Cricetulus migratorius* Pallas, 1773 из Тянь-Шаня // VI съезд Териологического об-ва. Москва, 13–16 апреля 1999. Тез. докл. М.: 1999. С. 37.
9. Ford C.E., Hamerton J.L. A colchicine, hypotonic citrate, squash sequence for mammalian chromosomes // *Stain Technol.* 1956. V. 31. № 6. P. 247–251.
10. Seabright M. A rapid banding technique for human chromosomes // *Lancet*. 1971. V. 11. P. 971–972.
11. Howell W.M., Black D.A. Controlled silver staining of nucleolus: A 1-step method // *Experientia*. 1980. V. 36. P. 1014–1015.
12. Aljanabi S.M., Martinez I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques // *Nucl. Ac. Res.* 1997. V. 25 (22). <https://doi.org/10.1093/nar/25.22.4692>
13. Jaarola M., Searle J.B. Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences // *Mol. Ecol.* 2002. V. 11. P. 2613–2621. <https://doi.org/10.1046/j.1365-294X.2002.01639.x>
14. Nakamura I., Ohnuma A., Ichihashi T. Mus musculus mitochondrial COI gene for cytochrome c oxidase subunit I, partial cds. 2008. <http://www.ncbi.nlm.nih.gov/nuccore/193248428>, дата обращения 30.11.2023 г.
15. Lyons L.A., Laughlin T.F., Copeland N.G. et al. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes // *Nature Genetics*. 1997. V. 15. № 1. P. 47–56.
16. Steppan S.J., Storz B.L., Hoffmann R.S. Nuclear DNA phylogeny of the squirrels (Mammalia: Rodentia) and the evolution of arboreality from c-myc and RAG1 // *Mol. Phylog. and Evol.* 2004. V. 30. № 3. P. 703–719. [https://doi.org/10.1016/S1055-7903\(03\)00204-5](https://doi.org/10.1016/S1055-7903(03)00204-5)
17. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // *Nucl. Ac. Res.* 2004. V. 32 № 5. P. 1792–1797. <https://doi.org/10.1093/nar/gkh340>
18. Kumar S., Stecher G., Li M. et al. *MEGA X: Molecular evolutionary enetics nalysis* across computing platforms // *Mol. Biol. and Evol.* 2018. V. 35. P. 1547–1549. <https://doi.org/10.1093/molbev/msy096>
19. Lavappa K.S. Chromosome banding patterns and idiogram of the Armenian hamster, *Cricetulus migratorius* // *Cytologia*. 1977. V. 42. № 1. P. 65–72.
20. Radjabli S.I., Sablina O.V., Graphodatsky A.S. Selected karyotypes // *Atlas of Mammalian Karyotypes* / Eds O'Brien S.J., Nash W.G., Menninger J.C. Chichester: John Wiley, 2006. P. 209–221.

21. Matthey R. Cytologie comparee des Cricetinae palearctiques et americains // Rev. Suisse Zool. 1961. V. 68. P. 41–61.
22. Ахвердян М.Р. Особенности поведения половых хромосом в мейозе у серого хомячка (*Cricetulus migratorius* Pallas, 1770) // Генетика. 1993. Т. 29. № 6. С. 950–959.
23. Kitano J., Peichel C.L. Turnover of sex chromosomes and speciation in fishes // Environ. Biol. of Fishes. 2012. V. 94. P. 549–558.
<https://doi.org/10.1007/s10641-011-9853-8>
24. Poplavskaya N., Bannikova A., Neumann K. et al. Phylogeographic structure in the chromosomally polymorphic rodent *Cricetulus barabensis* sensu lato (Mammalia, Cricetidae) // J. Zool. Syst. and Evol. Res. 2019. V. 57. № 3. P. 679–694.
<https://doi.org/10.1111/jzs.12251>
25. Romanenko S.A., Volobouev V.T., Perelman P.L. et al. Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison // Chromosome Research. 2007. V. 15. P. 283–298.
<https://doi.org/10.1007/s10577-007-1124-3>
26. Орлов В.Н., Исакова Э.Н. Таксономия надвида *Cricetulus barabensis* (Rodentia, Cricetidae) // Зоол. журн. 1975. Т. 54. В. 4. С. 597–604.
27. Воронцов Н.Н. Развитие эволюционных идей в биологии. М.: И-во Отдел УНЦ ДО МГУ. Прогресс-Традиция, АБФ. 1999. 640 с.

Karyotype and Molecular Genetic Differentiation of a 24-Chromosomal form of the Gray Hamster *Nothocricetulus migratorius* from the Tien Shan

O. V. Brandler^{1,*}, A. V. Blekhman¹

¹Koltzov Institute of Developmental Biology of Russian Academy of Sciences, Moscow, 119334 Russia

*e-mail: rusmarmot@yandex.ru

The widespread Palaearctic rodent species gray hamster *Nothocricetulus migratorius* has a karyotype with a stable number of chromosomes $2n = 22$ throughout the entire range of its habitat. We found gray hamsters with diploid number of chromosomes $2n = 24$ locally distributed in the Qurama Ridge of the Tien Shan. A new karyotype and analysis of G- and NORs-bands of differentially stained chromosome sets were described for the first time. The described karyotype differs from the 22-chromosomal karyotype of gray hamsters by the Y-chromosome morphology and the presence of an additional pair of heteromorphic small chromosomes. Molecular genetic analysis revealed genetic divergence of 24- and 22-chromosomal forms of *N. migratorius*, and the differences between them in mitochondrial markers are comparable, and in nuclear markers exceed the differences between *C. barabensis* ($2n = 20$) and *C. pseudogriseus* ($2n = 24$). The data obtained give grounds to discuss the taxonomic status of the 24-chromosomal form of gray hamsters from the Qurama Ridge and consider the differentiation of *N. migratorius* karyomorphs as a stage of chromosomal speciation.

Keywords: gray hamster, karyotype, molecular genetic variability, chromosomal speciation.

УДК 577.21

ПРОФИЛИ ЭКСПРЕССИИ ГЕНОВ, ВОВЛЕЧЕННЫХ В СИНТЕЗ ЛИГНАНОВ, В РАЗВИВАЮЩИХСЯ СЕМЕНАХ ЛЬНА

© 2024 г. Е. Н. Пушкова¹, Е. М. Дворянинова¹, Л. В. Повхова¹, Т. А. Рожмина^{1, 2},
Р. О. Новаковский¹, Е. А. Сигова¹, А. А. Дмитриев¹, Н. В. Мельникова^{1, *}

¹Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991 Россия

²Федеральный научный центр лубяных культур, Торжок, 172002 Россия

*e-mail: mnv-4529264@yandex.ru

Поступила в редакцию 11.12.2023 г.

После доработки 13.03.2024 г.

Принята к публикации 20.03.2024 г.

Семена льна являются богатейшим растительным источником лигнанов, препятствующих развитию многих заболеваний. Среди лигнанов в семени культивируемого вида *Linum usitatissimum* преобладает диглюкозид секоизоларицирезинола (SDG). Нами выполнено секвенирование транскриптомов семян льна на пяти стадиях развития для восьми сортов/линий, различающихся по содержанию лигнанов, для трех вариантов условий выращивания и проведена оценка экспрессии генов *PLR1* и *UGT74S1*, играющих ключевую роль в синтезе SDG. Выявлены коэкспрессия генов *PLR1* и *UGT74S1* и изменение уровня экспрессии этих генов в десятки и сотни раз в процессе развития семян, что подтверждает их роль в синтезе SDG льняного семени. Пониженная температура (16 °C) и избыточный полив приводили к сдвигу максимального уровня экспрессии обоих генов на более поздние сроки (14-й день после раскрытия цветка) по сравнению с условиями недостаточного полива и повышенной температуры (24 °C) и оптимальными условиями (20 °C) (7-й день после раскрытия цветка). При этом при повышенной температуре и недостаточном поливе уровень экспрессии генов *PLR1* и *UGT74S1* был ниже, чем при оптимальных условиях. Не выявлено ассоциации между содержанием лигнанов в семенах исследованных сортов/линий льна и уровнем экспрессии генов *PLR1* и *UGT74S1*. Наши результаты дают важную информацию о вкладе генотипа и среды в экспрессию ключевых генов синтеза SDG, что в том числе необходимо для разработки оптимальных подходов для получения семян льна с высоким содержанием лигнанов.

Ключевые слова: лен, *Linum usitatissimum*, экспрессия генов, *PLR1*, *UGT74S1*, секоизоларицирезинол.

DOI: 10.31857/S0016675824070113 **EDN:** BGNLFZ

Семена льна содержат биологически активные вещества и все шире используются для производства полезных для здоровья продуктов питания и биологически активных добавок [1–5]. Льняное семя — один из богатейших растительных источников лигнанов, препятствующих развитию рака, сердечно-сосудистых заболеваний и сахарного диабета [1, 6–14]. Среди лигнанов в семенах *Linum usitatissimum* L. преобладает диглюкозид секоизоларицирезинола (secoisolariciresinol diglucoside, SDG) [11]. У разных сортов льна различия в содержании SDG достигают нескольких раз, и данная характеристика может определять потенциал использования сорта для лечебного питания или производства лекарственных средств [11, 15–18]. Известно, что пинорезинол-ларицирезинол редуктазы (pinoresinol-lariciresinol reductases, PLRs) играют ключевую роль в синтезе лигнанов растений

[19]. В семенах льна *PLR1* сначала катализирует превращение (–)-пинорезинола в (–)-ларицирезинол, а затем в (+)-секоизоларицирезинол [20–23]. Уридин-гликозилтрансферазы (uridine glycosyl transferases, UGTs) катализируют образование гликозидных связей и играют важную роль в синтезе SDG льна, причем наибольший вклад вносит *UGT74S1* [24, 25].

Целью работы являлось выявление закономерностей в экспрессии генов *UGT74S1* и *PLR1* при развитии семян льна в разных условиях для сортов, различающихся по содержанию лигнанов.

В анализе использовались растения восьми сортов/линий льна с различным содержанием лигнанов в семенах: AGT 427, Atalante, AGT 981, Entre-Rios, Raciol, AGT 422, Lola, AGT 1535 (табл. 1). Данные о содержании секоизоларицирезинола в

Таблица 1. Содержание секоизоларицирезинола в семенах восьми сортов/линий льна

Сорт/линия	Секоизоларицирезинол, мг/кг
AGT 427	5125
Atalante	4650
AGT 981	4300
Entre-Rios	3900
Raciol	3900
AGT 422	3625
Lola	2900
AGT 1535	2125

семенах исследованных сортов/линий льна представлены Институтом льна (г. Торжок, Россия) и получены совместно с чешской компанией Agritek (не опубликованы). Растения льна выращивали в 15-литровых горшках с почвой в течение месяца в оптимальных условиях (20 °C и полив через день), а затем переносили в три климатические камеры. В первой камере растения выращивали при 16 °C и ежедневном поливе (далее – 16 °C), во второй камере – при 20 °C и поливе через день (далее – 20 °C), в третьей камере – при 24 °C и поливе раз в три дня (далее – 24 °C). Режим освещения: 16 часов – день, 8 часов – ночь. Сбор семян проводили на 3, 7, 14, 21 и 28 день после цветения (ДПЦ, день после раскрытия цветка). Выделение РНК выполняли по методике, описанной в работе L. Wang и соавт. [26], с рядом модификаций. Оценку качества и концентрации РНК выполняли методом гель-электрофореза, а также на приборах 2100 Bioanalyzer (Agilent Technologies, США) и Qubit (Thermo Fischer Scientific, США).

Для подготовки кДНК-библиотек для высокопроизводительного секвенирования применяли набор QIAseq Stranded mRNA Select Kit (Qiagen, США). Использовали пулы РНК, полученные от пяти одинаковых образцов (один и тот же сорт/линия, условия выращивания, стадия развития). Контроль качества кДНК-библиотек выполняли на 2100 Bioanalyzer (Agilent Technologies) и Qubit (Thermo Fischer Scientific). В результате получены 212 кДНК-библиотек высокого качества (для линии AGT 422 на 21 и 28 ДПЦ при 24 °C и сорта Lola на 28 ДПЦ при 24 °C пригодные для дальнейшего анализа библиотеки получить не удалось). Для 3, 7, 14 и 21 ДПЦ кДНК-библиотеки подготовлены в двухкратной биологической повторности, а для 28 ДПЦ – в однократной повторности. Секвенирование полученных транскриптомных библиотек выполняли на приборе NextSeq 2000 (Illumina, США) с использованием набора NextSeq 2000 P3 Reagents (100 Cycles) (Illumina), прочтения по 51 нуклеотиду с двух сторон. В среднем для каждой кДНК-библиотеки получено 2 млн парноконцевых прочтений.

Данные секвенирования депонированы в базе NCBI Sequence Read Archive (SRA), номер биопрокта PRJNA1039849.

Полученные прочтения Illumina обрезают по качеству и фильтровали по длине с использованием Trimmomatic [27]. Для анализа экспрессии генов использовали приложение PPliner [28]. Прочтения картировали на геном льна сорта Атлант (GCA_014858635.1 в базе NCBI Genome) [29], после чего определяли число прочтений для каждого гена в расчете на 1 млн прочтений (counts per million, CPM). Для дальнейшего анализа использовали данные CPM для транскриптов H1233_034242 + H1233_034241 (соответствуют *UGT74S1* из работы K. Ghose и соавт. [24]) и H1233_076413 (соответствует *PLR1* из работы D. Dalisay и соавт. [30]).

В результате проведенного анализа получили данные об уровне экспрессии генов *UGT74S1* и *PLR1* в семенах восьми сортов/линий льна, выращенных в трех вариантах температуры и полива, для 3, 7, 14, 21 и 28 ДПЦ (рис. 1). Профили экспрессии *UGT74S1* и *PLR1* были весьма похожи между собой при одних и тех же условиях выращивания для каждого из генотипов, что может свидетельствовать о коэкспрессии этих генов. О коэкспрессии *PLR1* и *UGT74S1* уже сообщалось ранее [24, 30], и наши данные согласуются с теми результатами.

Экспрессия изучаемых генов в целом изменялась сходным образом для разных сортов/линий при развитии семян в одних и тех же условиях. На 3 ДПЦ уровень экспрессии *UGT74S1* и *PLR1* был низким для всех генотипов в трех условиях выращивания. Для всех исследованных сортов/линий наиболее высокого уровня экспрессия изучаемых генов достигала на 7 ДПЦ в условиях 20 и 24 °C. Затем на 14 ДПЦ в условиях 20 и 24 °C происходило снижение уровня экспрессии, которое продолжалось при дальнейшем развитии семян. В условиях 16 °C уровень экспрессии генов начинал повышаться на 7 ДПЦ и достигал максимального значения на 14 ДПЦ в большинстве сортов/линий (исключение – Atalante для гена *UGT74S1*, для которого уровень экспрессии был близким на 7 и 14 ДПЦ). На 21 и 28 ДПЦ в условиях 16 °C наблюдалось снижение экспрессии обоих генов, однако имелись различия между генотипами – у одних сортов/линий уровень экспрессии достигал минимума на 28 ДПЦ, а у других все еще оставался достаточно высоким на этой стадии развития.

Таким образом, нами выявлены сходные закономерности в динамике уровня экспрессии *UGT74S1* и *PLR1* в большинстве исследованных генотипов льна. При развитии семян наблюдалось изменение экспрессии этих генов в десятки и сотни раз, что подтверждает их важность в синтезе SDG в семенах льна. Различные условия

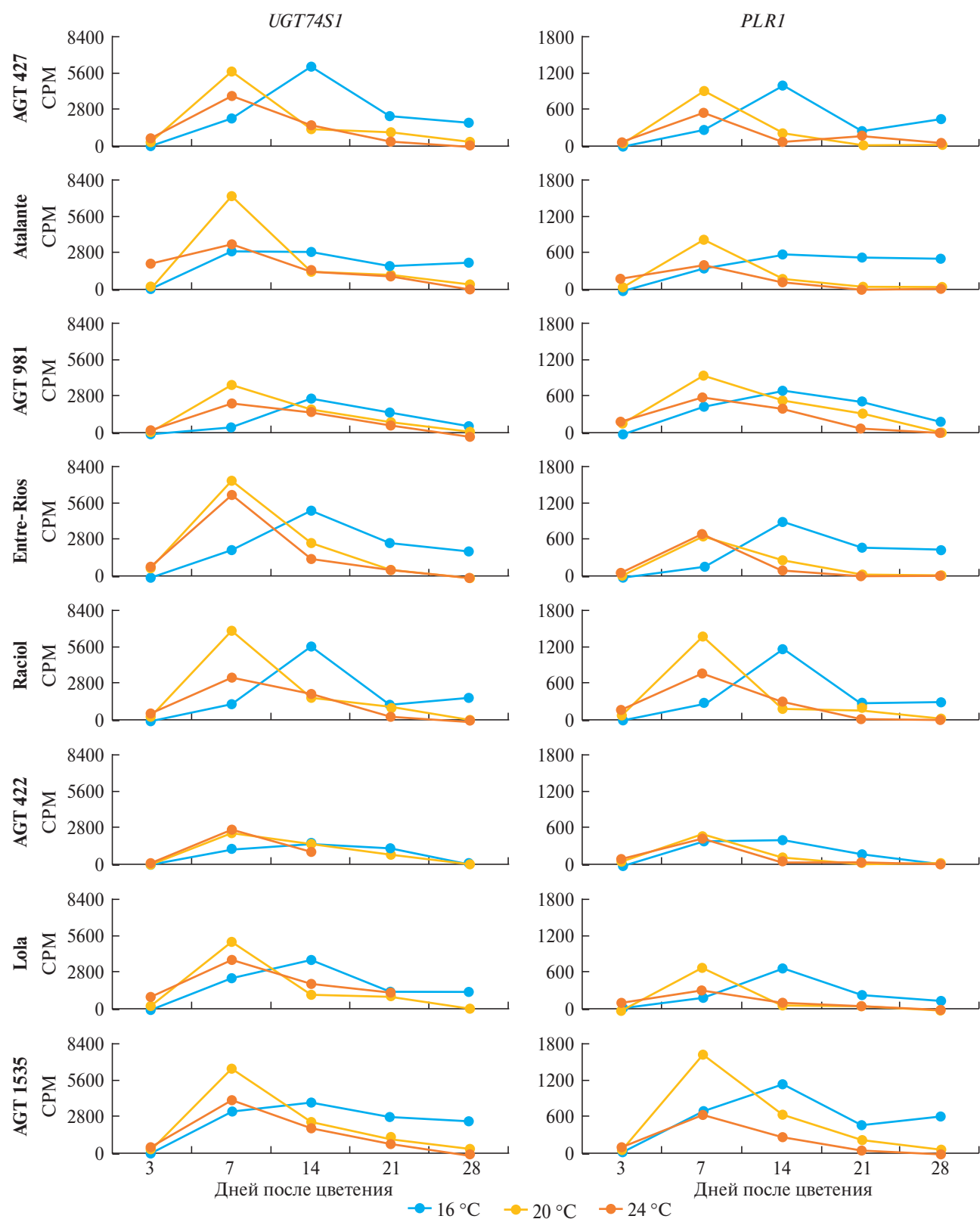


Рис. 1. Профили экспрессии генов *UGT74S1* и *PLR1* при развитии семян льна (3, 7, 14, 21 и 28 ДПЦ) для сортов/линий AGT 427, Atalante, AGT 981, Entre-Rios, Raciol, AGT 422, Lola, AGT 1535, выращенных при 16 °C и избыточном поливе (16 °C), 20 °C и оптимальном поливе (20 °C), 24 °C и недостаточном поливе (24 °C). Отсутствуют данные для AGT 422 для 21 и 28 ДПЦ при 24 °C и для Lola для 28 ДПЦ при 24 °C.

температуры и полива влияли на профили экспрессии анализируемых генов — пониженная температура и избыточный полив приводили к сдвигу максимального уровня экспрессии на более поздние сроки (14 ДПЦ) по сравнению с условиями недостаточного полива и повышенной температуры и оптимальными условиями (7 ДПЦ). Кроме того, в условиях 16 °С в большей степени проявлялись межсортовые различия в профилях экспрессии изучаемых генов по сравнению с условиями 20 и 24 °С. Наши результаты согласуются с данными, полученными в работе D. Dalisay и соавт. [30], где также анализировалась динамика изменения экспрессии генов *PLR*, и для *PLR1* наиболее высокий уровень экспрессии в исследуемых условиях был отмечен на 6 ДПЦ. Однако в настоящей работе впервые показано, как контролируемые условия внешней среды отражаются на профилях экспрессии генов *UGT74S1* и *PLR1*, играющих ключевую роль в синтезе SDG льняного семени.

При сравнении уровня экспрессии генов *UGT74S1* и *PLR1* между условиями 20 и 24 °С на 7 ДПЦ (в этот срок в этих условиях достигался максимальный уровень экспрессии) для большинства генотипов экспрессия была выше в условиях 20 °С. Это может свидетельствовать о том, что повышенная температура и недостаточный полив способны отрицательно влиять на экспрессию генов *UGT74S1* и *PLR1*. Сравнение уровня экспрессии этих генов между условиями 16 и 20 °С провести сложнее, так как профили экспрессии для этих условий существенно различаются. Однако можно отметить значительные межсортовые различия — для одних генотипов максимальный уровень экспрессии в условиях 20 °С превышал таковой в условиях 16 °С, а для других генотипов наблюдалось обратное.

Мы также провели сравнение уровня экспрессии генов *UGT74S1* и *PLR1* между генотипами с высоким и низким содержанием лигнанов в семени. Так, высокий уровень лигнанов имеют AGT 427 и Atalante, а низкий — Lola и AGT 1535 (табл. 1). Мы не обнаружили существенных различий в максимальном уровне экспрессии изучаемых генов для этих сортов/линий в условиях 20 и 24 °С. В условиях 16 °С сравнение проводить несколько сложнее из-за более выраженных межсортовых различий, однако проследить ассоциации между максимальным уровнем экспрессии и содержанием секоизолярицирезинола в семенах разных генотипов нам также не удалось. В то же время в работе L. Garros и соавт. [17] выявлена положительная корреляция между содержанием SDG и уровнем экспрессии *UGT74S1* и *PLR1*. Однако в том исследовании данные по экспрессии представлены только для одной временной точки, а не в процессе развития семян.

Таким образом, на представленной выборке сортов/линий льна нами показано влияние условий

выращивания на экспрессию генов *UGT74S1* и *PLR1*, вовлеченных в синтез лигнанов льняного семени. В то же время нами не обнаружена связь между различным содержанием лигнанов в семенах разных сортов/линий льна и уровнем экспрессии генов *UGT74S1* и *PLR1*. Наша работа расширяет знания о вкладе генотипа и среды в экспрессию ключевых генов синтеза SDG, что в том числе необходимо для разработки оптимальных подходов для выращивания льна с целью получения богатых лигнанами семян.

Работа выполнена при финансовой поддержке Российского научного фонда, грант № 21-16-00111.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Goyal A., Sharma V., Upadhyay N. et al. Flax and flaxseed oil: An ancient medicine & modern functional food // J. Food Sci. and Technology. 2014. V. 51. P. 1633–1653.
<https://doi.org/10.1007/s13197-013-1247-9>
2. Fombuena V., Petrucci R., Dominici F. et al. Maleinized linseed oil as epoxy resin hardener for composites with high bio content obtained from linen byproducts // Polymers. 2019. V. 11. P.
<https://doi.org/10.3390/polym11020301>
3. Corino C., Rossi R., Cannata S. et al. Effect of dietary linseed on the nutritional value and quality of pork and pork products: Systematic review and meta-analysis // Meat Science. 2014. V. 98. P. 679–688.
<https://doi.org/10.1016/j.meatsci.2014.06.041>
4. Singh K.K., Mridula D., Rehal J. et al. Flaxseed: A potential source of food, feed and fiber // Crit. Rev. in Food Sci. and Nutrition. 2011. V. 51.
<https://doi.org/10.1080/10408390903537241>
5. Akter Y., Junaid M., Afrose S.S. et al. A comprehensive review on *Linum usitatissimum* medicinal plant: Its phytochemistry, pharmacology, and ethnomedicinal uses // Mini Rev. in Med. Chemistry. 2021. V. 21. P. 2801–2834.
<https://doi.org/10.2174/1389557521666210203153436>
6. Imran M., Ahmad N., Anjum F.M. et al. Potential protective properties of flax lignan secoisolariciresinol diglucoside // Nutrition J. 2015. V. 14. P. 71.
<https://doi.org/10.1186/s12937-015-0059-3>
7. Parikh M., Netticadan T., Pierce G.N. Flaxseed: Its bioactive components and their cardiovascular benefits // Am. J. of Physiology. Heart and Circulatory Physiology. 2018. V. 314. P. H146–H159.
<https://doi.org/10.1152/ajpheart.00400.2017>

8. *Kezimana P., Dmitriev A.A., Kudryavtseva A.V. et al.* Secoisolariciresinol diglucoside of flaxseed and its metabolites: Biosynthesis and potential for nutraceuticals // *Front. in Genetics*. 2018. V. 9. <https://doi.org/10.3389/fgene.2018.00641>
9. *Mali A.V., Padhye S.B., Anant S. et al.* Anticancer and antimetastatic potential of enterolactone: Clinical, preclinical and mechanistic perspectives // *Europ. J. Pharmacology*. 2019. V. 852. P. 107–124. <https://doi.org/10.1016/j.ejphar.2019.02.022>
10. *Cullis C.A.* Genetics and Genomics of *Linum*. Cham, Switzerland: Springer Int. Publ., 2019. 270 p.
11. *Muir A.D., Westcott N.D.* Flax: The genus *Linum*. Boca Raton, FL, USA: CRC Press, 2003. 320 p.
12. *Locke A., Schneiderhan J., Zick S.M.* Diets for health: goals and guidelines // *Am. Family Physician*. 2018. V. 97. P. 721–728.
13. *Tse T.J., Guo Y., Shim Y.Y. et al.* Availability of bioactive flax lignan from foods and supplements // *Crit. Rev. Food Sci. Nutr.* 2023. V. 63. P. 9843–9858. <https://doi.org/10.1080/10408398.2022.2072807>
14. *Chhillar H., Chopra P., Ashfaq M.A.* Lignans from linseed (*Linum usitatissimum* L.) and its allied species: Retrospect, introspect and prospect // *Crit. Rev. Food Sci. Nutr.* 2021. V. 61. P. 2719–2741. <https://doi.org/10.1080/10408398.2020.1784840>
15. *Johnsson P., Kamal-Eldin A., Lundgren L.N. et al.* HPLC method for analysis of secoisolariciresinol diglucoside in flaxseeds // *J. Agricultural and Food Chemistry*. 2000. V. 48. P. 5216–5219. <https://doi.org/10.1021/jf0005871>
16. *Ezzat S.M., Shouman S.A., Elkhoely A. et al.* Anticancer potentiality of lignan rich fraction of six flaxseed cultivars // *Sci. Reports*. 2018. V. 8. P. 544. <https://doi.org/10.1038/s41598-017-18944-0>
17. *Garros L., Drouet S., Corbin C. et al.* Insight into the influence of cultivar type, cultivation year, and site on the lignans and related phenolic profiles, and the health-promoting antioxidant potential of flax (*Linum usitatissimum* L.) seeds // *Molecules*. 2018. V. 23. <https://doi.org/10.3390/molecules23102636>
18. *Diederichsen A., Fu Y.-B.* Flax genetic diversity as the raw material for future success // *Genus*. 2008. V. 32. P. 33.
19. *Markulin L., Corbin C., Renouard S. et al.* Pinorensin-lariciresinol reductases, key to the lignan synthesis in plants // *Planta*. 2019. V. 249. P. 1695–1714. <https://doi.org/10.1007/s00425-019-03137-y>
20. *Hemmati S., von Heimendahl C.B., Klaes M. et al.* Pinorensin-lariciresinol reductases with opposite enantiospecificity determine the enantiomeric composition of lignans in the different organs of *Linum usitatissimum* L. // *Planta Medica*. 2010. V. 76. P. 928–934. <https://doi.org/10.1055/s-0030-1250036>
21. *Hano C., Martin I., Fliniaux O. et al.* Pinorensin-lariciresinol reductase gene expression and secoisolariciresinol diglucoside accumulation in developing flax (*Linum usitatissimum*) seeds // *Planta*. 2006. V. 224. P. 1291–1301. <https://doi.org/10.1007/s00425-006-0308-y>
22. *Von Heimendahl C.B., Schafer K.M., Eklund P. et al.* Pinorensin-lariciresinol reductases with different stereospecificity from *Linum album* and *Linum usitatissimum* // *Phytochemistry*. 2005. V. 66. P. 1254–1263. <https://doi.org/10.1016/j.phytochem.2005.04.026>
23. *Hemmati S., Schmidt T.J., Fuss E.* (+)-Pinorensinol/(–)-lariciresinol reductase from *Linum perenne* Himmelszelt involved in the biosynthesis of justicidin B // *FEBS Letters*. 2007. V. 581. P. 603–610. <https://doi.org/10.1016/j.febslet.2007.01.018>
24. *Ghose K., Selvaraj K., McCallum J. et al.* Identification and functional characterization of a flax UDP-glycosyltransferase glucosylating secoisolariciresinol (SECO) into secoisolariciresinol monoglucoside (SMG) and diglucoside (SDG) // *BMC Plant Biology*. 2014. V. 14. <https://doi.org/10.1186/1471-2229-14-82>
25. *Fofana B., Ghose K., McCallum J. et al.* UGT74S1 is the key player in controlling secoisolariciresinol diglucoside (SDG) formation in flax // *BMC Plant Biology*. 2017. V. 17. P. 35. <https://doi.org/10.1186/s12870-017-0982-x>
26. *Wang L., Stegemann J.P.* Extraction of high quality RNA from polysaccharide matrices using cetyltrimethylammonium bromide // *Biomaterials*. 2010. V. 31. P. 1612–1618. <https://doi.org/10.1016/j.biomaterials.2009.11.024>
27. *Bolger A.M., Lohse M., Usadel B.* Trimmomatic: A flexible trimmer for Illumina sequence data // *Bioinformatics*. 2014. V. 30. P. 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
28. *Krasnov G.S., Dmitriev A.A., Kudryavtseva A.V. et al.* PPLine: An automated pipeline for SNP, SAP, and splice variant detection in the context of proteogenomics // *J. of Proteome Res.* 2015. V. 14. P. 3729–3737. <https://doi.org/10.1021/acs.jproteome.5b00490>
29. *Dmitriev A.A., Pushkova E.N., Novakovskiy R.O. et al.* Genome sequencing of fiber flax cultivar Atlant using Oxford Nanopore and Illumina platforms // *Front. Genetics*. 2020. V. 11. <https://doi.org/10.3389/fgene.2020.590282>
30. *Dalisay D.S., Kim K.W., Lee C. et al.* Dirigent protein-mediated lignan and cyanogenic glucoside formation in flax seed: Integrated omics and MALDI mass spectrometry imaging // *J. Nat. Products*. 2015. V. 78. P. 1231–1242. <https://doi.org/10.1021/acs.jnatprod.5b00023>

Expression Profiles Of Genes Involved In Lignan Synthesis In Developing Flax Seeds

E. N. Pushkova¹, E. M. Dvorianinova¹, L. V. Povkhova¹, T. A. Rozhmina^{1, 2}, R. O. Novakovskiy¹,
E. A. Sigova¹, A. A. Dmitriev¹, N. V. Melnikova^{1, *}

¹Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

²Federal Research Center for Bast Fiber Crops, Torzhok, 172002 Russia

*e-mail: mnv-4529264@yandex.ru

Flax seeds are the richest plant source of lignans, which prevent the development of many diseases. Secoisolariciresinol diglucoside (SDG) is the predominant lignan in seeds of the cultivated species *Linum usitatissimum*. We sequenced transcriptomes of flax seeds at five developmental stages for 8 varieties differing in lignan content grown under three different conditions and evaluated the expression of *PLR1* and *UGT74S1* genes, which play a key role in SDG synthesis. The co-expression of *PLR1* and *UGT74S1* genes was detected, and the expression level of these genes was observed to change tens and hundreds of times during seed development, confirming their role in SDG synthesis in flax seeds. Low temperature (16 °C) and abundant watering resulted in a shift of the maximum expression level of both genes to later dates (14th day after flowering) compared to poor watering and high temperature (24 °C) and optimal conditions (20 °C) (7th day after flowering). Meanwhile, the expression level of *PLR1* and *UGT74S1* genes was lower under high temperature and poor watering than under optimal conditions. No association was found between lignan content in seeds of the studied flax varieties and the expression level of *PLR1* and *UGT74S1* genes. Our results provide important information on the contribution of genotype and environment to the expression of key genes of SDG synthesis, which is also necessary for the development of optimal approaches to obtain lignan-rich flax seeds.

Keywords: flax, *Linum usitatissimum*, gene expression, *PLR1*, *UGT74S1*, secoisolariciresinol.

Acknowledgements: This work was financially supported by the Russian Science Foundation, grant 21-16-00111.

СЕКВЕНИРОВАНИЕ И АННОТАЦИЯ ХЛОРОПЛАСТНОГО ГЕНОМА *TRITICUM TIMONOVUM* HESLOT ET FERRARY

© 2024 г. А. Р. Кулуев¹, Р. Т. Матниязов¹, Б. Р. Кулуев¹*, Л. Ю. Привалов¹, А. В. Чемерис¹

¹Институт биохимии и генетики — обособленное структурное подразделение Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук, Уфа, 450054 Россия

*e-mail: kuluev@bk.ru

Поступила в редакцию 14.12.2023 г.

После доработки 10.01.2024 г.

Принята к публикации 16.01.2024 г.

Впервые был секвенирован хлоропластный геном синтетической октаплоидной пшеницы *Triticum timonovum* Heslot et Ferrary (линия к-43065, Франция). Секвенирование пластома *T. timonovum* проводилось на секвенаторе Genolab M (GeneMind, Китай). Сборка кольца хлоропластного генома осуществлялась с помощью программы NOVOwarp. Размер хлоропластного генома *T. timonovum* составил 136158 пн. Длина области инвертированных повторов составила 21552 пн, области SSC — 12795 пн и области LSC — 80257 пн. Были сравнены хлоропластные геномы *T. timonovum* и различных линий *T. timopheevii* из GenBank. Хлоропластный геном *T. timonovum* линии к-43065 оказался наиболее близок к пластоми *T. timopheevii* с номером доступа AB976560.1 и отличался от него лишь наличием одной вставки А в позиции 47891.

Ключевые слова: пшеница, октаплоид, синтетические виды, *Triticum timopheevii*, *T. timonovum*, хлоропластный геном.

DOI: 10.31857/S0016675824070124 **EDN:** BGHYZA

Triticum timonovum Heslot et Ferrary (пшеница тимоновум) — октаплоидная пшеница, искусственно созданная путем удвоения числа хромосом *T. timopheevii* Zhuk. с использованием колхичина. Авторами [1] это растение было обозначено как абберантный тетраплоид. Известно, что пшеница тимоновум — это синтетический октаплоид ($2n = 56$), который характеризуется такими полезными признаками, как высокое содержание белка в зерне и хорошие хлебопекарные качества, а также способностью вызывать у межвидовых гибридов цитоплазматическую мужскую стерильность. Между тем эта пшеница характеризуется медленным темпом развития и меньшей урожайностью чем *T. timopheevii*, что, по-видимому, связано с высокой плоидностью *T. timonovum* [2].

Принимая во внимание генотип исходной пшеницы, геном *T. timonovum* должен быть обозначен как GGA¹A¹GGA¹A¹, но по данным Е. Бадаевой и др. [3, 4] следует обозначать как GGA¹A¹BBA¹A¹, так как несколько хромосом генома А и одна хромосома генома G были заменены гомеологами геномов А и В от некоей неидентифицированной полбы. В дополнение к этому, в исследовании кариотипа *T. fungicidum* Zhuk. геномы А, В, А¹ и G этой пшеницы были идентичны родительским видам *T.*

timopheevii и *T. carthlicum* Nevsk., но в то же время геномы А и В отличались от соответствующих геномов *T. timonovum*. Авторы исследования предположили, что геномы А и В в геномы *T. fungicidum* и *T. timonovum* были переданы разными видами пшениц. Синтетический октаплоид *T. timonovum* — малоизученное растение, и его геном пока не исследован, к тому же не были изучены возможные изменения хлоропластного генома этой пшеницы под действием такого сильного мутагена как колхичин при создании *T. timonovum* путем полиплоидизации *T. timopheevii*.

Цель настоящей работы — секвенирование и аннотация хлоропластного генома *T. timonovum*.

Семена *T. timonovum* были предоставлены Федеральным исследовательским центром Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР, Санкт-Петербург). Растения проращивались в теплице при 21 °С в течение трех недель, затем было собрано около 20 г зеленых листьев и помещено на 48 ч в холод (+4 °С) и темноту для уменьшения содержания крахмала. Гомогенизацию листьев проводили при помощи фарфоровой ступки и пестика в жидком азоте. Хлоропластная ДНК была выделена из

зеленых листьев с помощью метода сахарозного градиента [5] с некоторыми модификациями.

Подготовка библиотек ДНК для секвенирования проводилась методом shotgun при помощи набора SG GM Plus (Raissol, Россия). Сначала проводилась ферментативная фрагментация хлоропластной ДНК до размера фрагментов 350 пн с одновременным восстановлением концов. Далее проводили лигирование адаптеров и очистку продукта реакции на магнитных частицах Smart beads (Raissol, Россия). На заключительном этапе проводилась индексная ПЦР и очистка продукта реакции на магнитных частицах Smart beads (Raissol, Россия). Библиотеки ДНК пулировали и проводили секвенирование при помощи набора Genolab M Kit V1.0 100 M reads/flow cell на секвенаторе Genolab M (GeneMind, Китай). Режим секвенирования — 2' 150 пн, файлы .base были демультиплексированы на приборе Genolab M с получением файлов с расширением fastq. В общей сложности было получено 13 млн ридов. Удаление адаптеров приготовленной библиотеки из файлов формата fastq проводили при помощи программы Trimmomatic v. 0.22 [6]. Обрезанные риды были собраны в контиги, которые объединялись и записывались в файл формата fasta, содержащий информацию о полной кольцевой молекуле хлоропластной ДНК, которая была сгенерирована при помощи программы NOVOwrap [7].

В качестве референсного генома был использован хлоропластный геном *T. timopheevii* с номером доступа KJ614408.1. Полный хлоропластный геном *T. timonovum* был аннотирован при помощи ресурса Chloroplast Genome Annotation, Visualization, Analysis, and GenBank Submission 2 (CPGAVAS2) (<http://47.96.249.172:16019/analyzer/home>) [8]. Кольцевую карту хлоропластного генома визуализировали при помощи ресурса Chloroplast (<https://irscope.shinyapps.io/Chloroplast>) [9]. Выравнивание нуклеотидных последовательностей хлоропластных геномов проводили при помощи MAFFT v.7 [10]. Филогенетическое дерево строили и визуализировали при помощи Archaeopteryx (bootstrap = 1000) [11]. В качестве внешнего вида для филогенетического анализа был взят *Secale cereale* subsp. segetale (MZ507427.1) из GenBank.

Нуклеотидная последовательность хлоропластного генома *T. timonovum* была депонирована в GenBank (номер доступа OR936056). Аннотацию хлоропластного генома проводили при помощи ресурса CPGAVAS 2, при этом размер хлоропластного генома *T. timonovum* составил 136158 пн. Длина области инвертированных повторов 21552 пн, области SSC — 12795 пн, а области LSC — 80257 пн. Содержание GC-пар во всем пластидном геноме — 38.3%. В области SSC содержание GC составляет 32.17%, а в области LSC — 36.25%. В области IR *T. timonovum* GC-содержание доходит до 43.92%.

У пшеницы тимоновум аннотировали 132 структурных гена, из которых 85 — белок-кодирующие гены, 31 ген тРНК и 4 гена рРНК. В том числе 7 генов, кодирующих белки (*rps19*, *rpl2*, *rpl23*, *ndhB*, *rps7*, *rps12*, *rps15*), 8 генов тРНК (*trnH-GUG*, *trnM-CAU*, *trnL-CAA*, *trnV-GAC*, *trnT-CGU*, *trnA-UGC*, *trnR-ACG*, *trnN-GUU*), 4 гена рРНК (*rRNA4.5*, *rRNA23*, *rRNA16* и *rRNA5*) были дублированы из-за того, что находятся в области повтора IR. При этом в SSC-области обнаружено 10 белок-кодирующих генов и 1 ген тРНК, в LSC-области — 68 белок-кодирующих генов и 22 гена тРНК. Кроме того, из 132 генов 11 имеют по одному интрону (*atpF*, *ndhB*, *petB*, *petD*, *rpl2*, *trnI-GAU*, *ndhA*, *rpl16*, *rps12A*, *rps16*, *trnG-UCC*) и 9 генов — по 2 интрона (*trnK-UUU*, *trnS-CGA*, *trnL-UAA*, *trnV-UAC*, *trnT-CGU*, *trnA-UGC*, *ycf3*). Самый крупный интрон (2559 пн) находится в гене *trnK-UUU*, внутри которого располагается другой ген — *matK*. Все полученные результаты представлены на рис. 1, а в виде кольцевой структуры, полученной при помощи онлайн-ресурса Chloroplast. Различные цветовые блоки на внешнем круге кольца отображают принадлежность генов к тем или иным функциональным группам.

На основе анализа нуклеотидных последовательностей полных хлоропластных геномов различных образцов *T. timopheevii* из GenBank, *T. turgidum* MG958546.1 и *T. timonovum* k-43065 (секвенированный нами) было построено филогенетическое дерево, которое показало, что *T. timonovum* k-43065 наиболее близок к *T. timopheevii* (AB976560.1) (рис. 1, б), при этом другие образцы *T. timopheevii* расположились на древе немного дальше. *T. turgidum* MG958546.1 и *Secale cereale* subsp. segetale (MZ507427.1), как и ожидалось, оказались внешними видами по отношению к исследованным пшеницам. Пластомы *T. timonovum* и *T. timopheevii* (AB976560.1) отличались только одной вставкой A в положении 47891, при этом другие образцы *T. timopheevii* отличались от этих близких по хлоропластному геному образцов пшениц различными делециями и вставками.

Ранее японскими авторами [12] на основе микросателлитного анализа хлоропластной ДНК большого числа образцов *T. timopheevii* из дикой флоры были выявлены три подгруппы, которые можно видеть также на построенном нами древе (рис. 1, б). Лишь образец с номером KJ614407.1 расположился отдельно от этих трех групп.

К сожалению, французские авторы [1] в своей статье не указали, какой конкретно образец тетраплоидной пшеницы *T. timopheevii* с геномом GA они использовали при создании октаплоидного вида *T. timonovum* с геномом GGAA. Однако по нуклеотидной последовательности пластидного генома этот октаплоид оказался наиболее близок к образцу пшеницы Тимофеева (AB976560.1), секвенированного грузинскими авторами [13]; однако

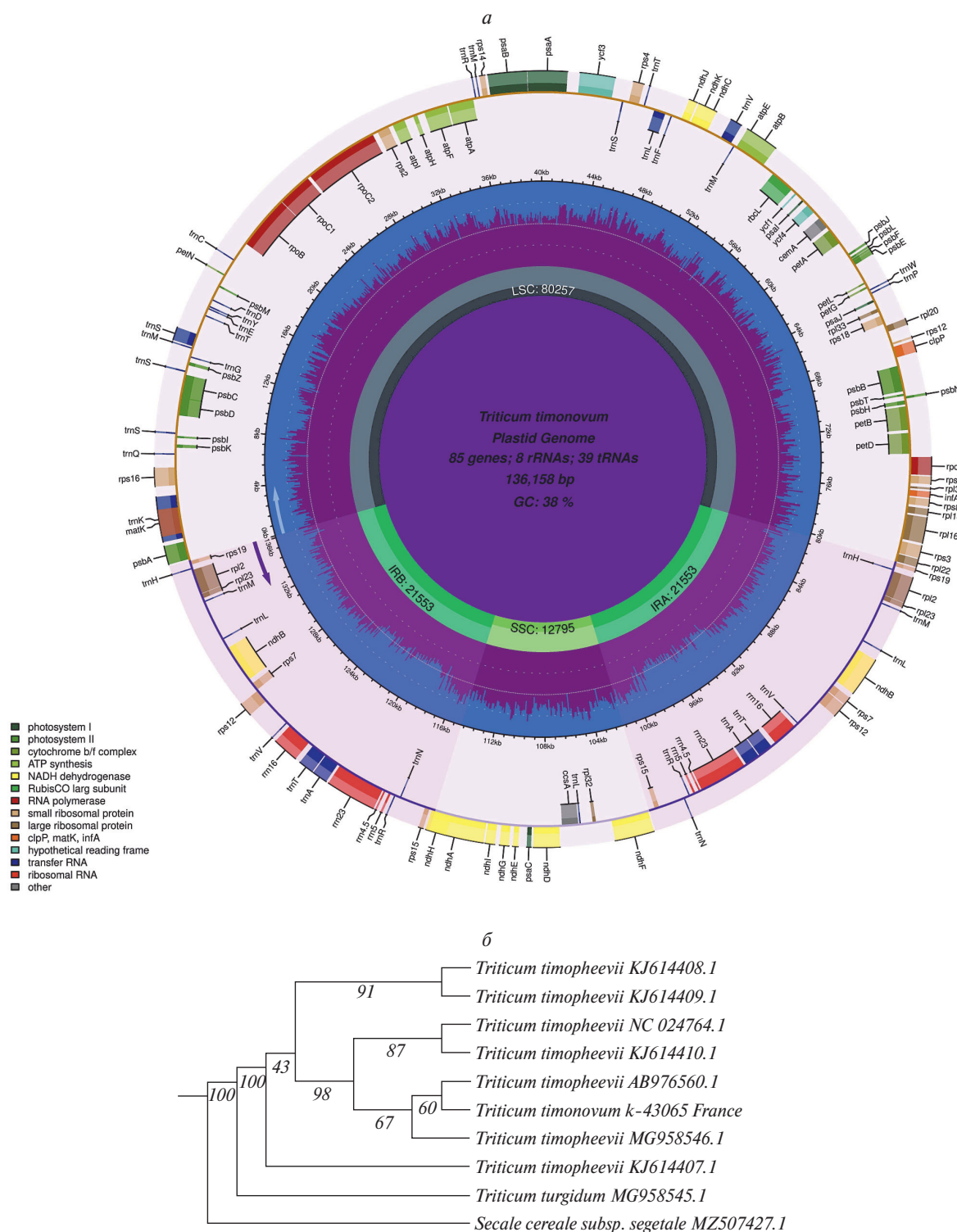


Рис. 1. *а* – визуальное представление в виде кольца секвенированного хлоропластного генома *T. timonovum* к-43065. Разными цветами отображены гены, синий круг посередине отображает уровень GC. IRA – инвертированная область повтора А, IRB – инвертированная область повтора В. Гены, расположенные за пределами внешнего круга, транскрибируются по часовой стрелке, а расположенные внутри гены – против часовой стрелки. *б* – филогенетическое древо, построенное на основе выравнивания нуклеотидных последовательностей хлоропластных геномов различных образцов *T. timopheevii* из GenBank, *T. turgidum* MG958546.1 и *T. timonovum* к-43065. В качестве внешнего вида представлен *Secale cereale* subsp. *segetale* (MZ507427.1).

в их статье также не сообщается ни каталожный номер, ни точное географическое происхождение секвенированного образца данной пшеницы.

Тем не менее, исходя из того, что от других представителей *T. timopheevii* хлоропластный геном пшеницы тимоновум отличается наличием ряда инделов, можно предположить, что именно образец *T. timopheevii* (AB976560.1) или близкий к нему был исходным при получении *T. timonovum*, при этом хлоропластный геном полученного октаплоида под действием колхицина сильным изменениям не подвергся.

Авторы выражают искреннюю признательность коллективу Отдела генетических ресурсов пшениц Всероссийского института генетических ресурсов растений им. Н.И. Вавилова (ВИР, Санкт-Петербург) за предоставление семенного материала для исследований.

Работа выполнена в рамках Государственного контракта по проекту РНФ № 23-24-00275 “Фило-генетические взаимоотношения отдельных видов пшенично-эгилопсного комплекса разных уровней плоидности через призму их полных хлоропластных геномов с прицелом на происхождение В- и G-субгеномов полиплоидных форм пшениц линий *turgidum-aestivum* и *timopheevii-zhukovskyi*”.

Настоящая статья не содержит каких-либо исследований с участием животных и людей в качестве объектов исследований.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Heslot H., Raymond R. Obtention experimentale d'un autotetraploide aberrant (*Triticum timonovum*) a partir de *Triticum timopheevi* Zhuk. // Compt. Rend. Hebd. Séances Acad. Sci. 1959. V. 248. P. 452–455.
2. Мурашёв В.В., Морозова З.А. Сравнительный морфогенез *Triticum timopheevii* (Zhuk.) и синтезированного октоплоидного вида *T. timonovum* Heslot et Ferrary // Вест. Моск. у-та. Серия 16. 2008. Т. 63. № 3. С. 127–133.
3. Badaeva E.D., Badaev N.S., Filatenko A.A. et al. Cytological investigation of cereal, hexa- and octoploid species containing G genome // Genetika (Mos.). 1990. V. 26. № 4. P. 708–716.
4. Badaeva E.D., Filatenko A.A., Badaev N.S. Cytogenetic investigation of *Triticum timopheevii* (Zhuk.) Zhuk. and related species using the C-banding technique // Theor. Appl. Genet. 1994. V. 89. P. 622–628.
5. Shi C., Hu N., Huang H. et al. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing // PLoS One. 2012. V. 7. № 2. <https://doi.org/10.1371/journal.pone.0031468>
6. Bolger A.M., Lohse M., Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data // Bioinformatics. 2014. V. 30. P. 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
7. Wu P., Xu C., Chen H. et al. NOVOWrap: An automated solution for plastid genome assembly and structure standardization // Mol. Ecol. Resour. 2021. V. 21. № 6. P. 2177–2186. <https://doi.org/10.1111/1755-0998.13410>
8. Shi L., Chen H., Jiang M. et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer // Nucl. Ac. Res. 2019. V. 47. P. W65–W73. <https://doi.org/10.1093/nar/gkz345>
9. Zheng S., Poczar P., Hyvönen J. et al. Chloroplast: An online program for the versatile plotting of organelle genomes // Front Genet. 2020. V. 11. <https://doi.org/10.3389/fgene.2020.576124>
10. Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability // Mol. Biol. Evol. 2013. V. 30. № 4. P. 772–780. <https://doi.org/10.1093/molbev/mst010>
11. Han M.V., Zmasek C.M. phyloXML: XML for evolutionary biology and comparative genomics // BMC Bioinformatics. 2009. V. 10. P. 1–6. <https://doi.org/10.1186/1471-2105-10-356>
12. Mori N., Kondo Y., Ishii T. et al. Genetic diversity and origin of timopheevi wheat inferred by chloroplast DNA fingerprinting // Bred. Sci. 2009. V. 59. P. 571–578. <https://doi.org/10.1270/jsbbs.59.571>
13. Gogniashvili M., Naskidashvili P., Bedoshvili D. et al. Complete chloroplast DNA sequences of Zanduri wheat (*Triticum* spp.) // Genet. Resour. Crop. Evol. 2015. V. 62. P. 1269–1277. <https://doi.org/10.1007/s10722-015-0230-x>

Sequencing And Annotation Of The Chloroplast Genome Of *Triticum timonovum* Heslot et Ferrary

A. R. Kuluev¹, R. T. Matniyazov¹, B. R. Kuluev^{1, *}, L. Yu. Privalov¹, A. V. Chemeris¹

¹*Institute of Biochemistry and Genetics - Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, 450054 Russia*

**e-mail: kuluev@bk.ru*

The chloroplast genome of the synthetic octaploid *Triticum timonovum* Heslot et Ferrary k-43065 (France) was sequenced for the first time. Plastome sequencing was carried out on a Genolab M sequencer (GeneMind, China). The genome assembly was carried out using the NOVOwrap program. The size of the chloroplast genome of *T. timonovum* was 136158 bp. Meanwhile, the length of the inverted repeat region was 21552 bp, the SSC region was 12795 bp. and LSC – 80257 bp. The chloroplast genomes of *T. timonovum* and different *T. timopheevii* accessions from the GenBank database were compared. As for the chloroplast genome, *T. timonovum* was closer to *T. timopheevii* (AB976560.1), but differed from it by the presence of one insert A at position 47891.

Keywords: wheat, octaploid, synthetic species, *Triticum timopheevii*, chloroplast genome.