BABLIOBBCKIII ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ VAVILOV JOURNAL OF GENETICS AND BREEDING

2024 • 28 • 6



Молекулярная и клеточная биология / Генетика растений / Генетика человека / Медицинская генетика

vavilov.elpub.ru vavilovj-icg.ru vavilov_journal@bionet.nsc.ru Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ VAVILOV IOURNAL OF GENETICS AND BREEDING

Основан в 1997 г. Периодичность 8 выпусков в год DOI 10.18699/vjgb-24-64

Учредители

Сибирское отделение Российской академии наук

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

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

Главный редактор

А.В. Кочетов – академик РАН, д-р биол. наук (Россия)

Заместители главного редактора

Н.А. Колчанов – академик РАН, д-р биол. наук, профессор (Россия) И.Н. Леонова – д-р биол. наук (Россия) Н.Б. Рубцов – д-р биол. наук, профессор (Россия) В.К. Шумный – академик РАН, д-р биол. наук, профессор (Россия)

Ответственный секретарь

Г.В. Орлова – канд. биол. наук (Россия)

Редакционная коллегия

Е.Е. Андронов – канд. биол. наук (Россия) Ю.С. Аульченко – д-р биол. наук (Россия) О.С. Афанасенко – академик РАН, д-р биол. наук (Россия) Д.А. Афонников – канд. биол. наук, доцент (Россия) Л.И. Афтанас – академик РАН, д-р мед. наук (Россия) Л.А. Беспалова – академик РАН, д-р с.-х. наук (Россия) А. Бёрнер – д-р наук (Германия) Н.П. Бондарь – канд. биол. наук (Россия) С.А. Боринская – д-р биол. наук (Россия) П.М. Бородин – д-р биол. наук, проф. (Россия) А.В. Васильев – чл.-кор. РАН, д-р биол. наук (Россия) М.И. Воевода – академик РАН, д-р мед. наук (Россия) Т.А. Гавриленко – д-р биол. наук (Россия) И. Гроссе – д-р наук, проф. (Германия) Н.Е. Грунтенко – д-р биол. наук (Россия) С.А. Демаков – д-р биол. наук (Россия) И.К. Захаров – д-р биол. наук, проф. (Россия) И.А. Захаров-Гезехус – чл.-кор. РАН, д-р биол. наук (Россия) С.Г. Инге-Вечтомов – академик РАН, д-р биол. наук (Россия) А.В. Кильчевский – чл.-кор. НАНБ, д-р биол. наук (Беларусь) С.В. Костров – чл.-кор. РАН, д-р хим. наук (Россия) А.М. Кудрявцев – чл.-кор. РАН, д-р биол. наук (Россия) И.Н. Лаврик – д-р биол. наук (Германия) Д.М. Ларкин – канд. биол. наук (Великобритания) Ж. Ле Гуи – д-р наук (Франция)

И.Н. Лебедев – д-р биол. наук, проф. (Россия) Л.А. Лутова – д-р биол. наук, проф. (Россия) Б. Люгтенберг – д-р наук, проф. (Нидерланды) В.Ю. Макеев – чл.-кор. РАН, д-р физ.-мат. наук (Россия) В.И. Молодин – академик РАН, д-р ист. наук (Россия) М.П. Мошкин – д-р биол. наук, проф. (Россия) С.Р. Мурсалимов – канд. биол. наук (Россия) Л.Ю. Новикова – д-р с.-х. наук (Россия) Е.К. Потокина – д-р биол. наук (Россия) В.П. Пузырев – академик РАН, д-р мед. наук (Россия) Д.В. Пышный – чл.-кор. РАН, д-р хим. наук (Россия) И.Б. Рогозин – канд. биол. наук (США) А.О. Рувинский – д-р биол. наук, проф. (Австралия) Е.Ю. Рыкова – д-р биол. наук (Россия) Е.А. Салина – д-р биол. наук, проф. (Россия) В.А. Степанов – академик РАН, д-р биол. наук (Россия) И.А. Тихонович – академик РАН, д-р биол. наук (Россия) Е.К. Хлесткина – д-р биол. наук, проф. РАН (Россия) Э.К. Хуснутдинова – д-р биол. наук, проф. (Россия) М. Чен – д-р биол. наук (Китайская Народная Республика) Ю.Н. Шавруков – д-р биол. наук (Австралия) Р.И. Шейко – чл.-кор. НАНБ, д-р с.-х. наук (Беларусь) С.В. Шестаков – академик РАН, д-р биол. наук (Россия) Н.К. Янковский – академик РАН, д-р биол. наук (Россия)

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

Founded in 1997 Published 8 times annually DOI 10.18699/vjqb-24-64

Founders

Siberian Branch of the Russian Academy of Sciences

Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences The Vavilov Society of Geneticists and Breeders

Editor-in-Chief

A.V. Kochetov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

Deputy Editor-in-Chief

N.A. Kolchanov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia *I.N. Leonova*, Dr. Sci. (Biology), Russia *N.B. Rubtsov*, Professor, Dr. Sci. (Biology), Russia *V.K. Shumny*, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

Executive Secretary

G.V. Orlova, Cand. Sci. (Biology), Russia

Editorial board

- O.S. Afanasenko, Full Member of the RAS, Dr. Sci. (Biology), Russia D.A. Afonnikov, Associate Professor, Cand. Sci. (Biology), Russia L.I. Aftanas, Full Member of the RAS, Dr. Sci. (Medicine), Russia E.E. Andronov, Cand. Sci. (Biology), Russia Yu.S. Aulchenko, Dr. Sci. (Biology), Russia L.A. Bespalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia N.P. Bondar, Cand. Sci. (Biology), Russia S.A. Borinskaya, Dr. Sci. (Biology), Russia P.M. Borodin, Professor, Dr. Sci. (Biology), Russia A. Börner, Dr. Sci., Germany M. Chen, Dr. Sci. (Biology), People's Republic of China S.A. Demakov, Dr. Sci. (Biology), Russia T.A. Gavrilenko, Dr. Sci. (Biology), Russia I. Grosse, Professor, Dr. Sci., Germany N.E. Gruntenko, Dr. Sci. (Biology), Russia S.G. Inge-Vechtomov, Full Member of the RAS, Dr. Sci. (Biology), Russia E.K. Khlestkina, Professor of the RAS, Dr. Sci. (Biology), Russia E.K. Khusnutdinova, Professor, Dr. Sci. (Biology), Russia A.V. Kilchevsky, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology), **Belarus** S.V. Kostrov, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia A.M. Kudryavtsev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia D.M. Larkin, Cand. Sci. (Biology), Great Britain I.N. Lavrik, Dr. Sci. (Biology), Germany J. Le Gouis, Dr. Sci., France I.N. Lebedev, Professor, Dr. Sci. (Biology), Russia B. Lugtenberg, Professor, Dr. Sci., Netherlands L.A. Lutova, Professor, Dr. Sci. (Biology), Russia V.Yu. Makeev, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.), Russia
- V.I. Molodin, Full Member of the RAS, Dr. Sci. (History), Russia M.P. Moshkin, Professor, Dr. Sci. (Biology), Russia S.R. Mursalimov, Cand. Sci. (Biology), Russia L.Yu. Novikova, Dr. Sci. (Agricul.), Russia E.K. Potokina, Dr. Sci. (Biology), Russia V.P. Puzyrev, Full Member of the RAS, Dr. Sci. (Medicine), Russia D.V. Pyshnyi, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia I.B. Rogozin, Cand. Sci. (Biology), United States A.O. Ruvinsky, Professor, Dr. Sci. (Biology), Australia E.Y. Rykova, Dr. Sci. (Biology), Russia E.A. Salina, Professor, Dr. Sci. (Biology), Russia Y.N. Shavrukov, Dr. Sci. (Biology), Australia R.I. Sheiko, Corr. Member of the NAS of Belarus, Dr. Sci. (Agricul.), Belarus S.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology), Russia V.A. Stepanov, Full Member of the RAS, Dr. Sci. (Biology), Russia I.A. Tikhonovich, Full Member of the RAS, Dr. Sci. (Biology), Russia A.V. Vasiliev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia M.I. Voevoda, Full Member of the RAS, Dr. Sci. (Medicine), Russia N.K. Yankovsky, Full Member of the RAS, Dr. Sci. (Biology), Russia I.K. Zakharov, Professor, Dr. Sci. (Biology), Russia I.A. Zakharov-Gezekhus, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

Молекулярная и клеточная биология

- 583 оригинальное исследование Влияние ауксин-зависимой деградации когезина и конденсинов на репарацию двуцепочечных разрывов ДНК в эмбриональных стволовых клетках мыши. А.В. Смирнов, А.С. Рыжкова, А.М. Юнусова (на англ. языке)
- 592 обзор Структура и эволюция метаполицентромер. Е.О. Гришко, П.М. Бородин (на англ. языке)

Генетика растений

- 602 оригинальное исследование Новый ген опушения листа *Hl1th*, интрогрессированный в мягкую пшеницу от *Thinopyrum ponticum*, и его фенотипическое проявление при гомеологичных хромосомных замещениях. *А.В. Симонов, Е.И. Гордеева, М.А. Генаев, В. Ли, И.О. Булатов, Т.А. Пшеничникова*
- 610 оригинальное исследование Влияние хромосом 1А и 1D *T. aestivum* на фертильность аллоплазматических рекомбинантных линий (*H. vulgare*)-*T. aestivum* в зависимости от цитоядерной совместимости. *Л.А. Першина, Н.В. Трубачеева, В.К. Шумный*

619 оригинальное исследование Содержание метаболитов и профиль экспрессии генов соответствующих метаболических путей в контрастных по окраске плодах баклажана (Solanum melongena L.). М.А. Филюшин, Е.А. Джос, А.В. Щенникова, Е.З. Кочиева

Генетика человека

628	обзор Импутация генотипов в геномных исследованиях человека. А.А. Бердникова, И.В. Зоркольцева, Я.А. Цепилов, Е.Е. Елгаева
640	оригинальное исследование Поиск сигналов положительного отбора генов циркадных ритмов PER1, PER2, PER3 в различных популяциях людей. А.И. Мишина, С.Ю. Бакоев, А.Ю. Ооржак, А.А. Кескинов, Ш.Ш. Кабиева, А.В. Коробейникова, В.С. Юдин, М.М. Боброва, Д.А. Шестаков, В.В. Макаров, Л.В. Гетманцева
650	обзор Генетические аспекты лактазной недостаточности у коренного населения Сибири. Б.А. Малярчук
659	оригинальное исследование Следы палеолитической экспансии в генофонде нивхов по данным

в генофонде нивхов по данным о полиморфизме аутосомных SNP и Y-хромосомы. В.Н. Харьков, Н.А. Колесников, Л.В. Валихова, А.А. Зарубин, А.Л. Сухомясова, И.Ю. Хитринская, В.А. Степанов

Медицинская генетика

667 оригинальное исследование Полиморфные варианты гена рецептора дофамина *DRD2* (rs6277, rs1800497) у подростков с проблемным использованием компьютерных видеоигр. С.Ю. Терещенко, К.В. Афоничева, И.В. Марченко, М.В. Шубина, М.В. Смольникова

© Сибирское отделение Российской академии наук, 2024

© Институт цитологии и генетики СО РАН, 2024

Вавиловский журнал генетики и селекции, 2024

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII • VAVILOV JOURNAL OF GENETICS AND BREEDING CONTENTS • $2024 \cdot 28 \cdot 6$

Molecular and cell biology

583	ORIGINAL ARTICLE Effects of the auxin-dependent degradation of the cohesin and condensin complexes on the repair of distant DNA double-strand breaks in mouse embryonic stem cells. <i>A.V. Smirnov, A.S. Ryzhkova, A.M. Yunusova</i>
500	

592 REVIEW Structure and evolution of metapolycentromeres. E.O. Grishko, P.M. Borodin

Plant genetics

- 602 ORIGINAL ARTICLE A new leaf pubescence gene, *Hl1th*, introgressed into bread wheat from *Thinopyrum ponticum* and its phenotypic manifestation under homoeologous chromosomal substitutions. *A.V. Simonov, E.I. Gordeeva, M.A. Genaev, W. Li, I.O. Bulatov, T.A. Pshenichnikova*
- 610 ORIGINAL ARTICLE

The effect of *T. aestivum* chromosomes 1A and 1D on fertility of alloplasmic recombinant (*H. vulgare*)-*T. aestivum* lines depending on cytonuclear compatibility. *L.A. Pershina, N.V. Trubacheeva, V.K. Shumny*

619 ORIGINAL ARTICLE

Metabolite concentrations and the expression profiles of the corresponding metabolic pathway genes in eggplant (*Solanum melongena* L.) fruits of contrasting colors. *M.A. Filyushin, E.A. Dzhos, A.V. Shchennikova, E.Z. Kochieva*

Human genetics

- 628 REVIEW Genotype imputation in human genomic studies. A.A. Berdnikova, I.V. Zorkoltseva, Y.A. Tsepilov, E.E. Elgaeva
 - 640 ORIGINAL ARTICLE

Search for signals of positive selection of circadian rhythm genes *PER1*, *PER2*, *PER3* in different human populations. *A.I. Mishina, S.Y. Bakoev, A.Y. Oorzhak, A.A. Keskinov, Sh.Sh. Kabieva, A.V. Korobeinikova, V.S. Yudin, M.M. Bobrova, D.A. Shestakov, V.V. Makarov, L.V. Getmantseva*

650 REVIEW

Genetic aspects of lactase deficiency in indigenous populations of Siberia. *B.A. Malyarchuk*

659 ORIGINAL ARTICLE

Traces of Paleolithic expansion in the Nivkh gene pool based on data on autosomal SNP and Y chromosome polymorphism.V.N. Kharkov, N.A. Kolesnikov, L.V. Valikhova, A.A. Zarubin, A.L. Sukhomyasova, I.Yu. Khitrinskaya, V.A. Stepanov

Medical genetics

667 ORIGINAL ARTICLE Polymorphic variants of the dopamine receptor gene *DRD2* (rs6277, rs1800497) in adolescents with problematic video game use. *S.Yu. Tereshchenko, K.V. Afonicheva, I.V. Marchenko, M.V. Shubina, M.V. Smolnikova*

© Siberian Branch RAS, 2024

© Institute of Cytology and Genetics, SB RAS, 2024 Vavilov Journal of Genetics and Breeding, 2024 DOI 10.18699/vjgb-24-65

Effects of the auxin-dependent degradation of the cohesin and condensin complexes on the repair of distant DNA double-strand breaks in mouse embryonic stem cells

A.V. Smirnov 🔟 🖾, A.S. Ryzhkova 🕕, A.M. Yunusova 🕕

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia Shldn89@gmail.com

Abstract. The SMC protein family, including cohesin and condensin I/II, plays a pivotal role in maintaining the topological structure of chromosomes and influences many cellular processes, notably the repair of double-stranded DNA breaks (DSBs). The cohesin complex impacts DSB repair by spreading yH2AX signal and containing DNA ends in close proximity by loop extrusion. Cohesin supports DNA stability by sister chromatid cohesion during the S/G2 phase, which limits DNA end mobility. Cohesin knockdown was recently shown to stimulate frequencies of genomic deletions produced by distant paired DSBs, but does not affect DNA repair of a single or close DSBs. We examined how auxin-inducible protein degradation of Rad21 (cohesin) or Smc2 (condensins I+II) changes the frequencies of rearrangements between paired distant DSBs in mouse embryonic stem cells (mESCs). We used Cas9 RNP nucleofection to generate deletions and inversions with high efficiency without additional selection. We determined optimal Neon settings and deletion appearance timings. Two strategies for auxin addition were tested (4 independent experiments in total). We examined deletion/inversion frequencies for two regions spanning 3.5 and 3.9 kbp in size. Contrary to expectations, in our setting, Rad21 depletion did not increase deletion/inversion frequencies, not even for the region with an active Ctcf boundary. We actually observed a 12 % decrease in deletions (but not inversions). At the same time, double condensin depletion (Smc2 degron line) demonstrated high biological variability between experiments, complicating the analysis, and requires additional examination in the future. TIDE analysis revealed that editing frequency was consistent (30–50 %) for most experiments with a minor decrease after auxin addition. In the end, we discuss the Neon/ddPCR method for deletion generation and detection in mESCs.

Key words: CRISPR/Cas9; mouse embryonic stem cells; auxin; cohesin; condensin; DNA repair.

For citation: Smirnov A.V., Ryzhkova A.S., Yunusova A.M. Effects of the auxin-dependent degradation of the cohesin and condensin complexes on the repair of distant DNA double-strand breaks in mouse embryonic stem cells. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(6):583-591. DOI 10.18699/vjgb-24-65

Funding. This work was supported by Russian Science Foundation grant No. 22-74-00084. Cell culture was performed at the Collective Center of ICG SB RAS "Collection of Pluripotent Human and Mammalian Cell Cultures for Biological and Biomedical Research", project number FWNR-2022-0019 (https://ckp.icgen.ru/cells/; http://www.biores.cytogen. ru/brc_cells/collections/ICG_SB_RAS_CELL). Droplet digital PCR was performed using the QX100 equipment (project number FWNR-2022-0015). Sanger DNA sequencing was performed at the Genomics Core Facility (ICBFM SB RAS, Novosibirsk).

Влияние ауксин-зависимой деградации когезина и конденсинов на репарацию двуцепочечных разрывов ДНК в эмбриональных стволовых клетках мыши

А.В. Смирнов 问 🖾, А.С. Рыжкова 🕕, А.М. Юнусова 问

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия 🖾 hldn89@gmail.com

Аннотация. Семейство SMC-белков, включающее когезин и конденсины I/II, играет ключевую роль в формировании топологической структуры хромосом и косвенно влияет на широкий спектр клеточных процессов, в том числе и на репарацию двуцепочечных разрывов ДНК (DSB). Комплекс когезина регулирует репарацию DSB на нескольких уровнях, например, распространяя сигнал γH2AX и удерживая концы ДНК в непосредственной близости за счет экструзии петель возле разрыва. Когезин также скрепляет сестринские хроматиды во время фазы S/G2, что ограничивает потенциальную подвижность концов ДНК. По имеющимся данным, в фибробластах человека нокдаун когезина стимулирует образование геномных делеций между удаленными DSB (3.2 тыс. п. о.), но не влияет на репарацию одиночных или близких DSB (34 п. о.). Мы решили проверить это наблюдение на эмбриональных стволовых клетках мыши, несущих ауксин-индуцибельный дегрон Rad21 (субъединица когезина) или Smc2 (субъединица конденсинов I+II). Для этого мы использовали нуклеофекцию RNP Cas9 и пары гайдовых РНК для генерации делеций и инверсий с высокой эффективностью без дополнительной селекции. Мы определили оптимальные условия для эффективной электропорации, включая настройки Neon, а также тайминги появления делеций. Были протестированы две стратегии добавления ауксина (суммарно четыре независимых эксперимента). Были исследованы частоты перестроек в двух сайтах размером около 3.5 и 3.9 тыс. п. о. Вопреки ожиданиям, деплеция Rad21 не увеличивала частоту делеций/инверсий, даже для региона с активной границей Ctcf. Фактически наблюдалось снижение частоты делеций (но не инверсий) на 12 %. Деплеция Smc2 не приводила к заметному увеличению частот делеций/инверсий, возможно, из-за высокой биологической изменчивости между экспериментами. Анализ TIDE показал, что частота редактирования была постоянной для большинства экспериментов (30–50 %), с незначительным снижением после добавления ауксина. В статье также обсуждается применимость метода Neon/ddPCR для создания и детекции делеций в эмбриональных стволовых клетках мыши.

Ключевые слова: CRISPR/Cas9; эмбриональные стволовые клетки мыши; ауксин; когезин; конденсин; репарация ДНК.

Introduction

Properly joining the two ends of a double-strand break (DSB) is crucial for preserving genome integrity. Unprocessed DNA ends can degrade, leading to loss of genetic information. Moreover, because DNA repair occurs in the vast space of the nucleus, incorrect ligation of multiple DSBs can result in chromosomal rearrangements, such as translocations, inversions, deletions, mitotic bridges and even chromothripsis. One-sided breaks that arise during replication are also highly dangerous and must be restrained and connected to the appropriate DNA molecule.

SMC complexes (cohesin, condensin-I, condensin-II) consist of several proteins organized into a ring-shaped structure (Kabirova et al., 2023). They utilize ATP-driven motor activity to shape and organize the genome into topological domains (TADs). Cohesin is an integral part of cellular homeostasis, regulating DNA conformation and topology, thus governing most vital processes from replication and cell division to gene expression and programmed DNA breaks in meiosis or V(D)J recombination. Although many reports have linked cohesin to DNA repair, its exact role therein remains unclear. Cohesin is attracted to DSB foci (Ström et al., 2004; Ünal et al., 2004), but is probably not essential for DNA end-joining per se (Gelot et al., 2016). Early cytogenetic and microscopic evidence suggests that cohesin limits DNA end mobility and prevents genomic rearrangements (Wu N., Yu, 2012). A study using a genetic reporter showed that cohesin knockdown leads to an increased frequency of deletions when two DSBs are introduced at a distance of 3.2 kbp but does not affect the ligation of closely located breaks (34 bp) (Gelot et al., 2016).

Importantly, these observations were only relevant to the S phase, where cohesin is required for sister chromatid cohesion. Knockdown of cohesin in G1-synchronized cells did not have an effect on deletion frequencies, probably because cohesin molecules physically limit the mobility of the DSB ends to preserve genome integrity only during S phase (Supplementary Material 1)¹. Generally speaking, cohesin removal does not affect deletion frequencies in G1, because it does not hold fragments together; but in the S phase, the excised fragments and the DSB ends are "stapled" to a sister chromatid (Supplementary Material 1). Cohesin acts on multiple levels to organize DSB repair, including retaining sister chromatids for homologous recombination (HR) and replication fork re-

start (Wu, Yu, 2012); limiting the mobility of the DSB for a better homology search within confined "repair domains" (Piazza et al., 2021); and amplifying the γ H2AX signal by asymmetrically extruding flanking chromatin in the vicinity of Ataxia-telangiectasia mutated (ATM) kinase at the DSB (Arnould et al., 2021).

At the same time, cohesin promotes replication stress by interfering with replication during loop extrusion (Minchell et al., 2020), which complicates the picture even further. Another insight comes from the DIvA (DSB Inducible via AsiSI) U2OS cells. This cell line expresses AsiSI restrictase with attached estrogen receptor ligand-binding domain (Aymard et al., 2014). After induction by 4-hydroxytamoxifen (4OHT), AsiSI translocates into the nucleus and introduces around 100-200 DSBs in annotated genomic loci (Dobbs et al., 2022). Multiple breaks induced with AsiSI tend to cluster together and form special kinds of D-compartments, but cohesin is not required for this process or any other kind of chromatin compartmentalization (Schwarzer et al., 2017; Arnould et al., 2023). Trans-interactions of multiple AsiSI-induced DSBs were also not affected by the Rad21 knockdown, but cohesin was required to reinforce affected TADs locally (Arnould et al., 2023). Thus, the connections between cohesin, chromatin compartmentalization, sister chromatid cohesion, DSB restraining and end joining are highly complex.

The role of condensins, another SMC family of DNA organizers, in DSB repair is still unclear. Their primary function is genome compaction before mitosis and they are mostly not active during interphase, although the complex resides in the nucleus throughout the cell cycle. Recent photobleaching experiments indicated that during interphase condensin II is very efficiently blocked from chromatin by the primary binding partner - the microcephalin protein (Mcph1) (Houlard et al., 2021). Mcph1 plays an important but poorly understood role in DSB repair, such as facilitating HR repair through Rad51 filament stabilization (Wu X. et al., 2009; Chang et al., 2020). Defects in condensin assembly lead to chromosomal aberrations and sister chromatid interlinks in mitotic chromatin (Wu N., Yu, 2012; Baergen et al., 2019). Evidence suggests that yeast condensin cooperates with topoisomerase-II to dissolve DNA knots (Dyson et al., 2021) and condensin II could be directly or indirectly involved in homology-directed repair (Wood et al., 2008).

Does cohesin directly impact the joining of close and distant DSBs? How do TAD features (size, borders, chromatin) influence deletion frequencies? Do condensin complexes play any

¹ Supplementary Materials 1–7 are available at:

https://vavilovj-icg.ru/download/pict-2024-28/appx20.pdf

role in DSB repair? In a series of pilot experiments presented here, we begin to explore some of these glaring questions.

Previously, we obtained and extensively characterized mouse embryonic stem cells (mESCs) with auxin-inducible degron (AID) knock-ins for Rad21 (cohesin) and Smc2 (both condensins I+II). These cells exhibit rapid depletion of the target protein within 1-2 hours of auxin addition (Yunusova et al., 2021). Cas9 activity generates blunt ends at the target sites. Using a pair of gRNA frequently leads to an excision of an intermediate DNA segment, which could lead to deletion or inversion after non-homologous or microhomology-mediated end joining (NHEJ/MMEJ) (Canver et al., 2014; Watry et al., 2020; Li et al., 2021). The cell lines were nucleofected with Cas9 and paired gRNAs, and studied using droplet digital PCR (ddPCR) to detect deletions and inversions in the mESCs population. This approach allowed us to assess the influence of the spatial organization of DNA and chromatin on the joining of two distant DNA ends. Overall, the method demonstrated high efficiency and sensitivity for detecting deletions and inversions. At the same time, the results were somewhat inconsistent, and the method we used might be more challenging than we had anticipated. We discuss its potential and limitations in the following chapters.

Materials and methods

gRNA design and cloning. We selected two genomic regions to induce paired DSBs: the *Ace2* gene locus (ChrX: 162.922.328–162.971.416) and a distinct TAD border that shows strong Ctcf signals in ChIP-seq data for mESCs (Chr5:

49.487.342–49.557.342) (GRCm39). High scoring gRNA sites were chosen using Benchling and Aidit algorithms. The sequences of the optimal gRNAs are listed in the Table. All oligonucleotides used in the study were purchased from DNA-Synthesis (Russia). 100 nt gRNAs were synthesized by the T7 *in vitro* transcription system from a PCR product amplified from a gRNA vector with the T7-primer (overhang 5'-GTTAATACGACTCACTATAG-20nt(gRNA)-3') and the reverse primer (see the Table) (HiScribe[®] T7 High Yield RNA Synthesis Kit, E2040S, protocol for short products). After 4 hours at 37 °C, the reaction volume (20 µl) was diluted to 100 µl and treated with 2 µl (4U) of DNaseI (NEB #M0303) in the corresponding buffer. RNA was purified with Monarch[®] RNA Cleanup Kit (50 µg) (T2040L) and diluted in 30 µl water to achieve concentrations of 2 µg/µl or higher.

mESCs nucleofections. Both mESCs auxin degron cell lines were characterized in our laboratory earlier (Rad21miniIAA7-eGFP, Smc2-miniIAA7-eGFP) (Yunusova et al., 2021). Cells were cultured on plates coated with a 1 % gelatin solution under 2i conditions (1 μ M PD, 3 μ M CHIR) in DMEM (Thermo Fisher, USA), supplemented with 7.5 % ES FBS (Gibco, USA), 7.5 % KSR (Gibco), 1 mM L-glutamine (Sigma, USA), NEAA (Gibco), 0.1 mM β -mercaptoethanol, LIF (1000 U/ml, Polygen), and penicillin/streptomycin (100 U/ml each). Upon reaching appropriate confluence (70–80 %), the cells were passaged every two days.

Single nucleofection sample consisted of 5 μ l Buffer R with 300000 cells which were mixed with 5 μ l of RNP complex diluted in Buffer R in a 10 μ l tip. Nucleofections were carried

Oligonucleotides used in the study					
Oligonucleotide	Sequence 5'–3'	Application			
gRNA Ace2 F	cacctgataaagtcagctgt	gRNA sequence			
gRNA Ace2 R2	ataagggcaacgaattgaca				
gRNA Ctcf F	ccttgacaagggcaccatgg				
gRNA Ctcf R2	aagaggctcatcagggactc				
T7 Ace2-F F	gttaatacgactcactatagcacctgataaagtcagctgt	T7 in vitro transcription			
T7 Ace2-R2 F	gttaatacgactcactatagataagggcaacgaattgaca				
T7 Ctcf-F F	gttaatacgactcactatagccttgacaagggcaccatgg				
T7 Ctcf-R2 F	gttaatacgactcactatagaaaagcaccgactcggtgcc				
gRNA31 Rev	aaaagcaccgactcggtgcc				
Ace2 F	gcagagtcattattacttccttg	ddPCR for deletions (149 bp)			
Ace2 R	caacctgggttcagaccctc	and inversions (154 bp) at the Ace2 locus			
Ace2 Inv R	ggcacaaagagttcatattacttac				
Ace2 Probe	HEX-tacctgcttacaactcagctgagaac-BHQ2				
Ctcf F	ggaggcataataacaactgctc	ddPCR for deletions (205 bp)			
Ctcf R	cagaggttagaacctatgaatcgg	and inversions (227 bp) at the Ctcf locus			
Ctcf Inv R	ggcacaaagagttcatattacttac				
Ctcf Probe	HEX-agacagagctgatcaagacagcatggt-BHQ2				
Emid1 F	gccaggactgggtagcac	ddPCR for the reference region (79 bp)			
Emdi1 R	aggaggctcctgaatttgtgacaag				
Emid1 Probe	FAM-cctgggtcatctgagctgagtcc-BHQ1				
Ace2 TIDE F	gtcatggatgcgctttggat	TIDE PCR (412 bp)			
Ace2 TIDE R	aatggagagaatggggcagg				

out at Neon preset condition #10 (Pulse voltage 1000, Pulse width 100, Pulse No. 1). Other tested conditions included #2 (1400, 20, 1), #6 (1100, 30, 1), #7 (1200, 30, 1), #13 (1100, 20, 2), #17 (850, 30, 2). The RNP mix consisted of 0.2 µl of concentrated Cas9-NLS protein (30 pmoles) (Biolabmix, Russia) and 2000 ng of each gRNA (1:2 ratio each). We aimed to set two replicates for the technical experiments (see Fig. 2) and three replicates for deletion/inversion frequencies (DIF) measurements (see Fig. 3). Auxin (500 µM of indole-3-acetic acid (IAA)) was added either 2 hours before nucleofection or right after cell plating after nucleofection, and was kept in the culture medium for the whole period. Target protein degradation was confirmed by microscopic analysis of GFP fluorescence loss (Supplementary Material 2). Cells were collected 24 hours after nucleofection. Genomic DNA was isolated from cells using phenol-chloroform extraction.

ddPCR assays. Droplet digital PCR (ddPCR) was performed using a QX100 system (Bio-Rad, USA) with primers and probes specific for the Ace2 and Ctcf regions, as well as the reference gene Emid1 (see the Table). ddPCR reactions were set in 20 µl volumes containing 1× ddPCR Supermix for Probes (no dUTP), 900 nM primers and 250 nM probes, and 50 ng genomic DNA. ddPCR reactions for each sample were performed in duplicates. PCR was conducted according to the following program: 95 °C for 10 min, then 45 cycles of 94 °C for 30 s and 58 °C for 1 min, with a ramp rate of 2 °C per second, and a final step at 98 °C for 10 min. The results were analyzed using QuantaSoft 1.7.4 (Bio-Rad). The resulting number was presented as mean +/- combined SEM. For Ace2 DIF calculations, the initial ddPCR results were multiplied by two, because the gene is located at the X chromosome (the mESCs DGES-1 line used in the study has male XY origin). Statistical analysis for relative differences between DIF across the experiments was performed with the Student test (control sample frequencies were set as 1).

TIDE sequencing. We PCR-amplified genomic site corresponding to gRNA sites F for *Ace2* (412 bp) from 50 ng of mESCs genomic DNA (samples from Fig. 3) (see the Table). PCR products were purified at 2 % agarose gel and Sanger sequenced using forward primer (reverse primer produced similar estimates in small-scale experiment). Sanger files were compared with wild-type control locus in the TIDE application with mostly default parameters (the start of the alignment window was switched to 91 instead of 100 bp) (http://shinyapps.datacurators.nl/tide/) (Brinkman et al., 2014). Average mutation percentage was calculated for three replicates for each degron.

Results

Implementing ddPCR assay at the mouse Ace2 locus

First, we set out to optimize the Neon nucleofection parameters for mouse embryonic stem cells (mESCs). The outline of a typical experiment is shown in Fig. 1. Following pilot tests, we estimated the average deletion frequency at multiple sites (based on two replicates) across two genomic regions (Fig. 2*a*). We selected one site (Ace2 F/R2) for Neon optimization. Initially, we tested the nucleofection parameters for wildtype mESCs on a Neon device across 24 basic settings with an EGFP plasmid (data not shown). From this experiment, we identified six conditions demonstrating higher survival rates and GFP fluorescence (conditions #2, 6, 7, 10, 13, 17, 18). Control mESCs were then nucleofected with Cas9 RNP, and the frequencies of deletions were analyzed by droplet digital PCR (ddPCR) (Fig. 2b). We observed a general inverse correlation between cell survival and the efficiency of deletion generation (Fig. 2b; Supplementary Material 3). Consequently, we selected condition #10 for further experiments, since high cell mortality is undesirable in our approach. Overall, the detection of deletion alleles with ddPCR proved to be specific, enabling reliable analysis of genomic DNA from the total mESCs population (Supplementary Materials 4, 5).

To optimize auxin addition time points, we conducted a small experiment to evaluate the timing of the appearance of deletions after nucleofection. It is known that in RNP nucleofection experiments mutations accumulate gradually. In our observations with mESCs, a small percentage of deletions (~2 % of the 48-hour level) appeared already in the first 3 hours after nucleofection (Fig. 2c). After 24 hours, approximately 63 % of deletions from the 48-hour level were observed. Considering the limited survival of mESCs beyond 24 hours without Rad21 or Smc2, this time frame was selected for subsequent experiments with all auxin degron lines. We also performed additional tests of the RNP stability during pre-incubation at 25 °C, revealing that DPBS buffer could effectively substitute the original Buffer R (Neon) without diminishing efficiency (Fig. 2d).

Deletion/inversion frequencies in mESCs degron lines

Using the established protocol, we measured deletion/inversion frequencies (DIF) at two genomic sites in various chromatin contexts. The Ace2 region was considered a "neutral" region located in the middle of a large TAD and showing no expression in mESCs. Here we focused on the 3495 bp deletion (F-R2) (Fig. 2a). We also analyzed DIF at another genomic site - a strong Ctcf boundary (Chr5:49.487.342-49.557.342) (Fig. 2a; Supplementary Material 6). To account for biological variability, we analyzed two independent experiments that were set with different cell batches and gRNA preparations (Day A, B). We validated these observations with two alternative auxin treatment strategies (Fig. 1b, c). In the first strategy, we first nucleofected the cells, then plated them in six wells and added auxin to half of them. This way, degradation starts simultaneously with Cas9 cutting. In the second strategy, auxin was added 2 hours prior to nucleofection to reliably remove all protein complexes (Fig. 1c). However, this necessitates separate nucleofections for control and treated samples, introducing additional handling variability. Furthermore, protein depletion prior to nucleofection could potentially increase cellular sensitivity to the procedure, possibly affecting ddPCR outcomes.

Surprisingly, we did not observe a Rad21-dependent DIF increase (Fig. 3*a*). More specifically, we documented a small but reproducible decrease in *Ace2* deletion frequencies in all experiments (mean relative decrease across four experiments: -12.1 %, p = 0.0423). Inversion frequencies remained unaffected. It is noteworthy that despite the distinct topological characteristics of the examined regions, there was no visible difference for Rad21-related effects, as the Ctcf region also showed minor and not statistically significant alterations in



Fig. 1. Experimental approach to study deletion/inversion frequencies in mESCs.

a – mESCs degrons lines were nucleofected with Cas9 and paired gRNAs; After 24 hours, genomic DNA was extracted and analyzed with ddPCR: we measured relative concentrations of the deletion and inversion alleles against the reference gene (*Emid1*). Two different nucleofection strategies with respect to auxin addition were tested; b – in the first approach, all cells were mixed together after nucleofection and then split in two sample groups (3+3 × 24 w). Auxin was added immediately after plating to half of the wells; c – in the second strategy, cells were preincubated with auxin for 2 hours and then nucleofected independently of control cells. In both cases, auxin was kept in culture medium for the duration of the experiment (24 hours). Degradation of the Rad21 and Smc2 proteins could be tracked by the loss of eGFP fluorescence (Supplementary Material 2); d – scheme of the droplet digital PCR modification designed to detect genomic rearrangements (ddXR method) (Watry et al., 2020). Induction of paired DNA breaks could lead to excision of the intermediate fragment, resulting in deletion or inversion. The loss or inversion of the fragment allows to efficiently amplify PCR product, activating probe fluorescence.

deletion frequencies (mean relative decrease across four experiments: -10 %, p = 0.109) (Fig. 3*a*).

Conversely, Smc2 depletion showed a trend towards DIF increase in one of the days (Day A, auxin added 2 hours before nucleofection (Fig. 3)), where it reached +33 % (*Ace2* deletions), +75 % (*Ace2* inversions), +61 % (Ctcf deletions), +63 % (Ctcf inversion). This effect, however, was not replicated in the subsequent trial (Day B, auxin added 2 hours before nucleofection) (Fig. 3b) upon switching to a different Cas9 batch. Depletion effect on *Ace2* deletions was not statistically significant, nor were changes in *Ace2* inversion frequencies at a significance level of 0.05 (mean relative

increase across four experiments: +39 %, p = 0.088). These discrepancies could be caused by some unaccounted biological factors, such as varying Cas9 batch efficiencies or differences in cell survival post-nucleofection between experiments (see Discussion). The role of the condensin complexes in distant end joining needs additional examinations in the future.

DSB repair efficiency in degron lines

Our objective was to investigate the impact of SMC protein depletion on DSB repair efficiency, particularly at a single DSB site or at closely positioned pairs of DSBs. Previous research indicated that 34 bp deletions are repaired differently



Fig. 2. Optimization of Neon conditions for deletion generation.

a – deletions examined in the study. Primers and the probe for ddPCR are shown for the Ace2 locus; b–d – optimizing mESCs Neon nucleofection conditions with the F-R2 (Ace2) gRNA pair.



Fig. 3. Deletion/inversion frequencies (DIF) for different genomic regions before and after addition of auxin.

a – DIF for Ace2 F-R2 and Ctcf F-R2 regions in Rad21 degron line; b – DIF for Ace2 F-R2 and Ctcf F-R2 regions in Smc2 degron line. Data presented as average between three nucleofection replicates and combined SEM. Statistical analysis for mean relative values across four biological experiments is provided in the main text.



Fig. 4. DSB repair efficiency at a single site Ace2 F, measured by TIDE.

a – demonstration of Sanger data for the control unedited locus and the mutated locus in the cells from Fig. 2, *d*. Cut site is marked with a dotted line. Editing efficiency measured with TIDE is shown as %; *b* – frequencies of site modifications in various degron lines from Fig. 3. Data shown as average and SEM.

from larger 3200 bp deletions in Rad21-deficient cells (Gelot et al., 2016). One of the drawbacks of the ddPCR method is its inability to detect small deletions, due to interference with the wild-type locus amplification. The authors of the ddXR method recommend digesting genomic DNA to selectively eliminate wild-type genomic loci from ddPCR amplification. With this trick, they were able to amplify deletions as small as 91 bp (Watry et al., 2020).

We managed to apply restriction to a region of 192 bp at the Ace2 locus (Supplementary Material 7), although attempts to apply it to other short deletions were less successful (data not shown). Given these limitations, the ddPCR method was deemed unsuitable for analyzing 34 bp deletions. Instead, to screen how SMC-protein depletion affects DSB repair at a single end we utilized the Tracking of Indels by DEcomposition (TIDE) method (Brinkman et al., 2014), a straightforward approach based on Sanger sequencing of the break site. This method facilitates the demultiplexing and calculation of Cas9 cut signatures at the break, thereby estimating DSB repair efficiency as a percentage of mutant alleles. Estimating indel mutation signatures at the break site serves not only as an indicator of Cas9 activity but also as a measure of nucleofection efficiency (Fig. 4). We PCR amplified and sequenced regions at the Cas9 target site for the Ace2 F gRNA (Fig. 2a) (the same samples analyzed with ddPCR (Fig. 3)).

For the Smc2 experiments, we did not detect any significant differences in editing efficiency. The slight decrease in efficiency post-auxin treatment was counterbalanced by an increase in DIF (since deletion/inversion events eliminate Ace2 F sites from PCR amplification in TIDE analysis) (Fig. 4b). For Rad21 depletions, a decrease in editing efficiencies was noted, potentially reflecting increased cell vulnerability under high RNP loads in the absence of Rad21. Notably, one experimental condition (Rad21/Smc2, Day A, auxin added 2 h before nucleofection) exhibited a 2-fold reduction in editing efficiencies. In this scenario, auxin addition paradoxically enhanced Cas9 editing for both degron lines (Fig. 4b), yet DIF were impacted differently in Rad21 and Smc2 lines (Fig. 3). This suggests that at lower editing efficiencies (RNP load), cells might respond differently to protein depletion. For example, Smc2 depletion could permeabilize cells for nucleofection, possibly due to a cell cycle shift or chromosome decondensation. We plan to perform nucleofections with various RNP concentrations in the future to verify this effect.

Discussion

We have performed a series of experiments with auxin degradation and CRISPR/Cas9-induced DSBs using a collection of mESCs with the SMC degrons. mESCs represent an interesting object for studying DNA repair. For instance, SMC complexes depletion and the repair of distant DNA double-strand breaks

mESCs mostly rely on HR to preserve genome stability (Choi et al., 2017) and have different end-joining mechanisms based on specialized polymerases (Schimmel et al., 2017). Since mESCs are difficult to edit with lipofection, we adapted a protocol to generate deletions with Neon nucleofections. This method, in conjunction with ddPCR, demonstrated high efficiency and sensitivity in detecting deletions and inversions, with an average modification rate of 60 % for the *Ace2* locus after Neon nucleofection (TIDE at the F site + deletions + insertions) (Fig. 3, 4). This level of editing is notable compared to plasmid transfection outcomes without selection. However, we encountered significant variability in deletion/ inversion frequencies (DIF) across experiments, highlighting the influence of numerous biological factors on experimental outcomes.

Cas9, a crucial component in our experiments, can significantly impact DSB repair dynamics. Variations in the Cas9:gRNA ratio can dramatically alter editing outcomes (Chenouard et al., 2023), with repair processes potentially delayed up to 20 hours due to persistent Cas9-DNA binding (Kim et al., 2014; Brinkman et al., 2018). Furthermore, Cas9 retention at break sites can modify blunt ends into 3'-overhang trimmed ends (Stephenson et al., 2018; Jones et al., 2021), necessitating different polymerases for non-homologous endjoining. Variability was also observed between different lots of Cas9-NLS (Biolabmix) even at identical molar concentrations. To account for all these issues, we performed experiments with two strategies of auxin addition and set three nucleofection replicates. We also performed two biological replicates with different mESCs batches, gRNA preps and Neon tips. From our experience, such experiments require very careful examination of the optimal experimental conditions, especially when the gene of interest has strong pleiotropic effects on cell homeostasis.

Our timing analysis indicated that cells accumulate 70 % of deletions within 24 hours, and only 2 % in the first 3 hours, suggesting that auxin could be added within a 0–3 hour window after nucleofection without significantly compromising deletion generation. Furthermore, we confirmed that DPBS incubation does not compromise RNP activity, providing a viable alternative to Buffer R. Notably, immediate post-nucleofection auxin addition exhibited lesser variability compared to a 2-hour pre-incubation strategy (Fig. 3), demonstrating the feasibility of its use in future setups due to its uniform experimental conditions.

Analysis of data on the frequencies of deletions and inversions for various mESCs clones with degrons allowed us to draw the following conclusions. We expected that Rad21 depletion will cause elevated rates of deletions and inversions due to unconstrained movement of the DSB ends, as it was reported by another group. In their report, there was a 30 % increase in the amount of cells with a 3 kb deletion (Gelot et al., 2016) after Rad21 siRNA knockdown. Some other reports using cytogenetic and microscopic analysis also suggested that Rad21 knockdown provokes DNA rearrangements (Wu N., Yu, 2012). So far, we have not found any significant stimulatory effects of Rad21 depletion on DIF (Fig. 3). Given that the authors of the initial report (Gelot et al., 2016) worked with a different experimental setting (plasmid transfection with inducible I-SceI, siRNA Rad21 knockdown, SV40-transformed GM639 human WT fibroblasts) and had an alternative detection strategy, our results may reflect differences between the experimental systems. In our setting, the protein was removed almost completely after 2 hours (Yunusova et al., 2021) and Cas9 RNP was active from the beginning (see timings, Fig. 2c). Also, mESCs are more sensitive to DNA damage and may react to DSB differently than immortalized fibroblasts (Choi et al., 2017).

It is possible that the absence of Rad21 sensibilizes cells to DNA damage resulting in a decreased opportunity for distant end-joining events to happen, hiding the stimulatory effect. This would lead to a lower amount of TIDE signal, as we see in our data (Fig. 4b). However, this does not explain why inversion frequencies are not negatively affected (Fig. 3). In theory, the effect of Rad21 degradation may be more noticeable for extremely distant DSBs, such as a 26 kbp deletion that we plan to analyze in the future (Supplementary Material 6). Correlation between topology and DSB is another long-standing question. In our case, deletion over the Ctcf site at the TAD border was not noticeably affected by cohesin depletion.

Unexpectedly, our findings hint at a significant role of Smc2 depletion in promoting genomic rearrangements, although data variability necessitates further investigation. Condensins are not directly involved in DNA repair, but could affect it via side effects (defects in chromosome segregation, chromatin decondensation in G2/M). Cell cycle is an important determinant of a DSB repair outcome. It is well known that G1 DSBs are repaired with slower kinetics (Arnould et al., 2023). Synchronization of human fibroblasts in the G1 phase showed no end-joining stimulation from Rad21 knockdown (Gelot et al., 2016). We and others analyzed cell cycles in mESCs with Rad21 and Smc2 depletion and found that after 6 hours they accumulate in the G2/M phase (manuscript under preparation). Judging from these data, Rad21 and Smc2 clones have the same cell cycle profile. Thus, cell cycle shift alone would not explain the difference between Rad21 and Smc2 depletion effects. In this study, we could only work with an unsynchronized mESCs population. Synchronization of mESCs is very challenging and imposes additional cell lethality making this approach unsuitable for our goal.

We plan to expand our investigations with the repertoire of deletions at other genomic regions with interesting topological organization. We will also try other improvements, such as NGS sequencing with Unique Molecular Identifiers (UMIs) for Cas9 target sites to account for editing efficiency. In the future, we will also extend our findings to simpler, synchronizable human cell lines such as HAP1 and HCT116, which also harbor Rad21/Smc2 degrons, to further dissect these complex dynamics.

Conclusion

Cohesin facilitates genome stability by limiting DNA movements during replication. By this logic, supported by experimental data, the frequencies of deletion between paired distant breaks will increase after cohesin removal. We could not reproduce these findings in the Rads21 auxin-degron cell line as we did not see an increase in deletion or inversion frequencies. This may reflect differences between experimental systems. Both Rad21 and Smc2 degron studies will require more iterations to account for biological variability.

References

- Arnould C., Rocher V., Finoux A.-L., Clouaire T., Li K., Zhou F., Caron P., Mangeot P.E., Ricci E.P., Mourad R., Haber J.E., Noordermeer D., Legube G. Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature*. 2021;590(7847):660-665. DOI 10.1038/s41586-021-03193-z
- Arnould C., Rocher V., Saur F., Bader A.S., Muzzopappa F., Collins S., Lesage E., Le Bozec B., Puget N., Clouaire T., Mangeat T., Mourad R., Ahituv N., Noordermeer D., Erdel F., Bushell M., Marnef A., Legube G. Chromatin compartmentalization regulates the response to DNA damage. *Nature*. 2023;623(7985):183-192. DOI 10.1038/ s41586-023-06635-y
- Aymard F., Bugler B., Schmidt C.K., Guillou E., Caron P., Briois S., Iacovoni J.S., Daburon V., Miller K.M., Jackson S.P., Legube G. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat. Struct. Mol. Biol.* 2014; 21(4):366-374. DOI 10.1038/nsmb.2796
- Baergen A.K., Jeusset L.M., Lichtensztejn Z., McManus K.J. Diminished condensin gene expression drives chromosome instability that may contribute to colorectal cancer pathogenesis. *Cancers (Basel)*. 2019;11(8):1066. DOI 10.3390/cancers11081066
- Brinkman E.K., Chen T., Amendola M., van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 2014;42(22):e168. DOI 10.1093/nar/gku936
- Brinkman E.K., Chen T., de Haas M., Holland H.A., Akhtar W., van Steensel B. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Mol. Cell.* 2018;70(5):801-813.e6. DOI 10.1016/j.molcel.2018.04.016
- Canver M.C., Bauer D.E., Dass A., Yien Y.Y., Chung J., Masuda T., Maeda T., Paw B.H., Orkin S.H. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. 2014;289(31):21312-21324. DOI 10.1074/jbc. M114.564625
- Chang H.-Y., Lee C.-Y., Lu C.-H., Lee W., Yang H.-L., Yeh H.-Y., Li H.-W., Chi P. Microcephaly family protein MCPH1 stabilizes RAD51 filaments. *Nucleic Acids Res.* 2020;48(16):9135-9146. DOI 10.1093/nar/gkaa636
- Chenouard V., Leray I., Tesson L., Remy S., Allan A., Archer D., Caulder A., Fortun A., Bernardeau K., Cherifi Y., Teboul L., David L., Anegon I. Excess of guide RNA reduces knockin efficiency and drastically increases on-target large deletions. *iScience*. 2023;26(4): 106399. DOI 10.1016/j.isci.2023.106399
- Choi E.-H., Yoon S., Park K.-S., Kim K.P. The homologous recombination machinery orchestrates post-replication DNA repair during self-renewal of mouse embryonic stem cells. *Sci. Rep.* 2017;7(1): 11610. DOI 10.1038/s41598-017-11951-1
- Dobbs F.M., van Eijk P., Fellows M.D., Loiacono L., Nitsch R., Reed S.H. Precision digital mapping of endogenous and induced genomic DNA breaks by INDUCE-seq. *Nat. Commun.* 2022;13(1): 3989. DOI 10.1038/s41467-022-31702-9
- Dyson S., Segura J., Martínez-García B., Valdés A., Roca J. Condensin minimizes topoisomerase II-mediated entanglements of DNA *in vivo*. *EMBO J.* 2021;40(1):e105393. DOI 10.15252/embj.2020105393
- Gelot C., Guirouilh-Barbat J., Le Guen T., Dardillac E., Chailleux C., Canitrot Y., Lopez B.S. The cohesin complex prevents the end joining of distant DNA double-strand ends. *Mol. Cell.* 2016;61(1): 15-26. DOI 10.1016/j.molcel.2015.11.002
- Houlard M., Cutts E.E., Shamim M.S., Godwin J., Weisz D., Presser Aiden A., Lieberman Aiden E., Schermelleh L., Vannini A., Nasmyth K. MCPH1 inhibits condensin II during interphase by regulating its SMC2-Kleisin interface. *eLife*. 2021;10:e73348. DOI 10.7554/eLife.73348

- Kabirova E., Nurislamov A., Shadskiy A., Smirnov A., Popov A., Salnikov P., Battulin N., Fishman V. Function and evolution of the loop extrusion machinery in animals. *Int. J. Mol. Sci.* 2023;24(5):5017. DOI 10.3390/ijms24055017
- Kim S., Kim D., Cho S.W., Kim J., Kim J.-S. Highly efficient RNAguided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 2014;24(6):1012-1019. DOI 10.1101/gr.171322.113
- Li D., Sun X., Yu F., Perle M.A., Araten D., Boeke J.D. Application of counter-selectable marker PIGA in engineering designer deletion cell lines and characterization of CRISPR deletion efficiency. *Nucleic Acids Res.* 2021;49(5):2642-2654. DOI 10.1093/nar/gkab035
- Minchell N.E., Keszthelyi A., Baxter J. Cohesin causes replicative DNA damage by trapping DNA topological stress. *Mol. Cell.* 2020; 78(4):739-751.e8. DOI 10.1016/j.molcel.2020.03.013
- Piazza A., Bordelet H., Dumont A., Thierry A., Savocco J., Girard F., Koszul R. Cohesin regulates homology search during recombinational DNA repair. *Nat. Cell Biol.* 2021;23(11):1176-1186. DOI 10.1038/s41556-021-00783-x
- Schimmel J., Kool H., van Schendel R., Tijsterman M. Mutational signatures of non-homologous and polymerase theta-mediated endjoining in embryonic stem cells. *EMBO J.* 2017;36(24):3634-3649. DOI 10.15252/embj.201796948
- Schwarzer W., Abdennur N., Goloborodko A., Pekowska A., Fudenberg G., Loe-Mie Y., Fonseca N.A., Huber W., Haering C.H., Mirny L., Spitz F. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. 2017;551(7678):51-56. DOI 10.1038/nature24281
- Stephenson A.A., Raper A.T., Suo Z. Bidirectional degradation of DNA cleavage products catalyzed by CRISPR/Cas9. J. Am. Chem. Soc. 2018;140(10):3743-3750. DOI 10.1021/jacs.7b13050
- Ström L., Lindroos H.B., Shirahige K., Sjögren C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell.* 2004;16(6):1003-1015. DOI 10.1016/j.molcel. 2004.11.026
- Ünal E., Arbel-Eden A., Sattler U., Shroff R., Lichten M., Haber J.E., Koshland D. DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. *Mol. Cell.* 2004;16(6):991-1002. DOI 10.1016/j.molcel.2004.11.027
- Watry H.L., Feliciano C.M., Gjoni K., Takahashi G., Miyaoka Y., Conklin B.R., Judge L.M. Rapid, precise quantification of large DNA excisions and inversions by ddPCR. *Sci. Rep.* 2020;10(1):14896. DOI 10.1038/s41598-020-71742-z
- Wood J.L., Liang Y., Li K., Chen J. Microcephalin/MCPH1 associates with the Condensin II complex to function in homologous recombination repair. J. Biol. Chem. 2008;283(43):29586-29592. DOI 10.1074/jbc.M804080200
- Wu N., Yu H. The Smc complexes in DNA damage response. Cell Biosci. 2012;2(1):5. DOI 10.1186/2045-3701-2-5
- Wu X., Mondal G., Wang X., Wu J., Yang L., Pankratz V.S., Rowley M., Couch F.J. Microcephalin regulates BRCA2 and Rad51-associated DNA double-strand break repair. *Cancer Res.* 2009;69(13):5531-5536. DOI 10.1158/0008-5472.CAN-08-4834
- Yunusova A., Smirnov A., Shnaider T., Lukyanchikova V., Afonnikova S., Battulin N. Evaluation of the OsTIR1 and AtAFB2 AID systems for genome architectural protein degradation in mammalian cells. *Front. Mol. Biosci.* 2021;8:757394. DOI 10.3389/fmolb.2021. 757394

Jones S.K., Hawkins J.A., Johnson N.V., Jung C., Hu K., Rybarski J.R., Chen J.S., Doudna J.A., Press W.H., Finkelstein I.J. Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. *Nat. Biotechnol.* 2021;39(1):84-93. DOI 10.1038/s41587-020-0646-5

Conflict of interest. The authors declare no conflict of interest. Received March 29, 2024. Revised April 23, 2024. Accepted July 22, 2024.

DOI 10.18699/vjgb-24-66

Structure and evolution of metapolycentromeres

E.O. Grishko (D, P.M. Borodin (D 🖾

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

Abstract. Metapolycentromeres consist of multiple sequential domains of centromeric chromatin associated with a centromere-specific variant of histone H3 (CENP-A), functioning collectively as a single centromere. To date, they have been revealed in nine flowering plant, five insect and six vertebrate species. In this paper, we focus on their structure and possible mechanisms of emergence and evolution. The metapolycentromeres may vary in the number of centromeric domains and in their genetic content and epigenetic modifications. However, these variations do not seem to affect their function. The emergence of metapolycentromeres has been attributed to multiple Robertsonian translocations and segmental duplications. Conditions of genomic instability, such as interspecific hybridization and malignant neoplasms, are suggested as triggers for the de novo emergence of metapolycentromeres. Addressing the "centromere paradox" – the rapid evolution of centromeric DNA and proteins despite their conserved cellular function – we explore the centromere drive hypothesis as a plausible explanation for the dynamic evolution of centromeres in general, and in particular the emergence of metapolycentromeres and holocentromeres. Apparently, metapolycentromeres are more common across different species than it was believed until recently. Indeed, a systematic review of the available cytogenetic publications allowed us to identify 27 candidate species with metapolycentromeres. The list of the already established and newly revealed candidate species thus spans 27 species of flowering plants and eight species of gymnosperm plants, five species of insects, and seven species of vertebrates. This indicates an erratic phylogenetic distribution of the species with metapolycentromeres and may suggest an independent emergence of the metapolycentromeres in the course of evolution. However, the current catalog of species with identified and likely metapolycentromeres remains too short to draw reliable conclusions about their evolution, particularly in the absence of knowledge about related species without metapolycentromeres for comparative analysis. More studies are necessary to shed light on the mechanisms of metapolycentromere formation and evolution. Key words: centromere; centromere size; centromere type; metapolycentromeres.

For citation: Grishko E.O., Borodin P.M. Structure and evolution of metapolycentromeres. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2024;28(6):592-601. DOI 10.18699/vjgb-24-66

Funding. This research was funded by the Russian Science Foundation, grant number 23-24-00304.

Acknowledgements. The authors express sincere gratitude to A. Torgasheva and L. Malinovskaya for their assistance in preparing the article and to A. Nurislamov, G. Koksharova and M. Deryuzhenko for the helpful comments.

Структура и эволюция метаполицентромер

Е.О. Гришко 🛈, П.М. Бородин 🛈 🖾

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

Аннотация. Метаполицентромеры состоят из нескольких последовательных доменов центромерного хроматина, связанных со специфичным для центромеры вариантом гистона H3 – CENP-A, которые вместе функционируют как одна центромера. Они были открыты недавно и обнаружены у девяти видов цветковых растений, пяти видов насекомых и шести видов позвоночных животных. В данном обзоре рассматриваются структура метаполицентромер и возможные механизмы их возникновения и эволюции. Метаполицентромеры могут различаться по количеству центромерных доменов, последовательностям ДНК и эпигенетическим модификациям. Однако эти различия, вероятно, не влияют на их функцию. Появление метаполицентромер объясняют множественными робертсоновскими транслокациями и сегментными дупликациями. В условиях геномной нестабильности (при межвидовой гибридизации и в ходе канцерогенеза) метаполицентромеры могут возникать *de novo*. Гипотеза центромерного драйва представляется убедительным объяснением эволюции центромер в целом и образования метаполицентромер и голоцентромер в частности. По-видимому, метаполицентромер ры встречаются чаще, чем принято считать. Систематический обзор доступных цитогенетических публикаций позволил нам дополнительно идентифицировать 27 видов-кандидатов с метаполицентромерами. Таким образом, список уже установленных и вновь найденных видов-кандидатов охватывает 27 видов цветковых и восемь видов голосеменных растений, пять видов насекомых и семь видов позвоночных животных. Виды, включенные в этот список, спорадически распределены по филогенетическому древу. Это может указывать на независимое эволюционное возникновение метаполицентромер. Однако существующий список видов с идентифицированными и предполагаемыми метаполицентромерами слишком короткий, чтобы сделать надежные выводы об их эволюции, особенно в отсутствие знаний о родственных видах без метаполицентромер для сравнительного анализа. Необходимы дополнительные исследования для того, чтобы пролить свет на механизмы образования и эволюции метаполицентромер.

Ключевые слова: центромера; размер центромеры; тип центромеры; метаполицентромеры.

Four main types of centromeres

The centromere is the region of the chromosome to which spindle filaments attach during mitosis and meiosis. It consists of centromeric DNA and a kinetochore protein complex through which the spindle microtubules attach to the chromosome. Centromeres play a critical role in maintaining chromosome integrity and controlling chromosome segregation during cell division. Disruption of the structure and function of centromeres in mitosis can lead to cell death, and in meiosis, to the formation of unbalanced gametes and sterility. Despite this conserved function, common to all eukaryotes, the centromeres of different organisms can vary significantly in both structure and size (Talbert, Henikoff, 2020). The only epigenetic mark of the centromere, characteristic of the vast majority of species, is the centromere variant of histone H3, the CENP-A protein (Mendiburo et al., 2011).

There are four main types of centromeres: regional centromeres, point centromeres, metapolycentromeres and holocentromeres (Talbert, Henikoff, 2020) (Fig. 1).

Regional centromeres are the most common type of centromere. Cytologically, the regional centromere can be detected as a primary constriction (Flemming, 1882). It is built on centromeric chromatin, marked by CENP-A. Based on centromeric chromatin, the kinetochore is assembled (Cleveland et al., 2003) (Fig. 2). The length of centromeric chromatin varies significantly among different species and can range from several thousand to millions of base pairs (bp) (Haupt et al., 2001; Kanesaki et al., 2015). Usually, centromeric and pericentromeric chromatin consists of highly repeated DNA sequences: satellite DNA or mobile genetic elements. However, centromeres based on non-repeated sequences have also been found (Glöckner, Heidel, 2009; Kanesaki et al., 2015; Talbert et al., 2018). The centromeric sequences of most species consist predominantly of satellite DNA.

Centromeric tandem repeats vary in the number, length, and nucleotide composition of repeating fragments (monomers), but usually have a length of 100–400 bp (Melters et al., 2013). This size ensures DNA coiling around 1–2 nucleosomes. The monomers of the satellite DNA are often A/T rich (Melters et al., 2013), which presumably reduces DNA bending energy and promotes nucleosome folding. The sequences of centromeric repeats can vary even between closely related species (Lee et al., 2005; Talbert et al., 2018). Moreover, even within the same species, centromeres of different chromosomes can consist of either tandem repeats belonging to the same family or completely different repeats (Ahmad et al., 2020; Balzano, Giunta, 2020).

It is known that centromeric repeats are actively transcribed, and the resulting transcripts play an important role in maintaining centromere structure (Talbert, Henikoff, 2018). **Point centromeres** are found only in the chromosomes of the budding yeast *Saccharomyces cerevisiae* (Nagpal, Fierz, 2021). They contain only one centromeric nucleosome, the so-called hemisome (heminucleosome), consisting of histones H4, H2A, H2B, and Cse4 (CENP-A homolog) in a single copy (Furuyama, Biggins, 2007; Henikoff et al., 2014). Only one spindle microtubule is attached to the point centromeres (Winey et al., 1995).



Fig. 1. Four main types of centromeres: regional centromeres (*a*); point centromeres (*b*); metapolycentromeres (*c*), and holocentromeres (*d*).



Fig. 2. Regional centromere structure, according to H. Nagpal and B. Fierz (2021), modified.

Holocentromeres do not form a primary constriction since the spindle microtubules have attachment points along the entire length of the chromosome. Some holocentromeres have no centromeric chromatin at all and CENP-A is distributed evenly along the entire length of the chromosome. In other holocentromeres, the centromeric chromatin forms small, equally spaced, repeated clusters along the entire length of the chromosome (Senaratne et al., 2022). Holocentromeres were detected in 700 species of plants and animals with holocentromeres (Melters et al., 2012). More information about this topic can be found in the reviews (Senaratne et al., 2022; Wang et al., 2022; Castellani et al., 2024; Kuo et al., 2024).

Metapolycentromeres consist of several sequential domains of centromeric chromatin associated with CENP-A and functioning as a single centromere. They are considered a transitional type between regional centromeres and holocentromeres (Neumann et al., 2012).

Here we review the structural features and evolution of metapolycentromeres, the most recently discovered and extremely rare type of centromere.

CENP-A as a centromere identifier

The position of the centromere is determined epigenetically, not by a specific DNA sequence, and the centromeric variant of histone H3 is considered the universal epigenetic mark of a functional centromere (Mendiburo et al., 2011). Centromeric histone H3 has several taxon-specific synonyms: CENP-A in animals, CENH3 in plants, CID (centromere identifier) in drosophila, HCP-3 in nematodes, Cnp1 in fission yeast, and Cse4 in budding yeast. In this article, for convenience, we will use the term CENP-A, as it is the most commonly used. CENP-A or its homologues are found in the centromeres of all eukaryotic species studied, with very rare exceptions including some species of lepidopterans and hemipterans, trypanosomes, and fungi (Drinnenberg et al., 2014; van Hooff et al., 2017; Navarro-Mendoza et al., 2019; Senaratne et al., 2021). The presence of CENP-A on a chromosomal site is necessary and sufficient for the formation of a functional centromere and for ensuring its inheritance (Mendiburo et al., 2011).

CENP-A, like canonical histone H3, includes two domains: an N-terminal domain and a C-terminal domain. The latter is integrated into the nucleosomal octamer and forms the nucleosome body (Sullivan K.F. et al., 1994). This domain contains the following regions (from the N end to the C end): α N-helix, α 1-helix, Loop1, α 2-helix, Loop2, α 3-helix, and C-terminal disordered tail (Black et al., 2004; Tachiwana et al., 2012). Human CENP-A shows 48 % homology with the canonical histone H3, making it the most distinct histone H3 variant. The N-terminal domain of human CENP-A is much shorter than that of canonical H3, and the amino acid sequence has the least homology to the canonical H3 sequence of all regions of the protein. The C-terminal domain is 68 % identical to the canonical one (Sullivan K.F. et al., 1994).

Typically, histones are highly conserved, but the amino acid composition of CENP-A varies significantly between different species (Maheshwari et al., 2015). The N-terminal domain and loop 1 of the C-terminal domain interact with centromeric DNA and show signs of positive selection in some organisms, for example, in *Drosophila melanogaster* and *Arabidopsis thaliana*. The main part of the C-terminal domain (except loop 1) is typically conserved (Malik, Henikoff, 2001; Talbert et al., 2002; Maheshwari et al., 2015).

Thus, the centromere's position is epigenetically marked by CENP-A, which is essential for centromere function across eukaryotes. It shows significant interspecies variation and adaptive evolution, highlighting its critical role in centromere functionality and inheritance.

Structure and characteristics of metapolycentromeres

Metapolycentromeres are found in a few species (see the Table). The number of chromosomes containing metapolycentromeres differs between species. In some species, all chromosomes contain metapolycentromeres. In others, metapolycentromeres are present on a few chromosomes or on just one, while the remaining chromosomes contain regional centromeres (Huang Y.-C. et al., 2016; Malinovskaya et al., 2022). Moreover, between populations of the ant species *Trachymyrmex holmgreni*, variation in the number of chromosome pairs containing metapolycentromeres was observed, from 1 to all 20 pairs (Cardoso et al., 2018). Metapolycentromeres also vary in size. They may occupy from 5 to 40 % of the chromosome length (Malinovskaya et al., 2022).

On routinely stained preparations of metaphase chromosomes, metapolycentromeres appear as elongated primary constrictions (Fig. 3*a*) (Drpic et al., 2018; Malinovskaya et al., 2022). Immunolocalization of CENP-A provides a more accurate identification of metapolycentromeres. This method of identification has been applied to the metaphase chromosomes of *Pisum sativum*, *P. fulvum*, *Lathyrus* spp., *Tribolium castaneum*, and *Muntiacus muntjak* (Neumann et al., 2012, 2015; Drpic et al., 2018; Gržan et al., 2020). Recently, L.P. Malinovskaya et al. (2022) and E. Grishko et al. (2023) detected metapolycentromeres in five species of songbirds: Gouldian finch, European pied flycatcher, Eurasian bullfinch, domestic canary, and common linnet, using non-specific antibodies to the human centromere (ACA) on preparations of surfacespread synaptonemal complexes (Fig. 3*b*).

In all cases, the signals from centromeric chromatin domains were distributed in a paired bead-like pattern, with anticentromere antibodies always binding to the outer side of the primary constriction. In some cases, in legumes and songbirds, centromeric chromatin domains were fused, forming a linear structure (Neumann et al., 2012, 2015; Malinovskaya et al., 2022; Grishko et al., 2023). In the songbirds, unequal spacing between domains and unequal numbers of domains on homologous chromosomes of the same karyotype were observed (Grishko et al., 2023).

The use of ChIP-seq with antibodies to CENP-A showed that the centromeric chromatin of peas metapolycentromeres consists predominantly of AT-rich satellite DNA 150–400 bp long. A combination of ChIP-seq with long-read sequencing demonstrated that the centromeric chromatin of metapolycentromeres also contains various retrotransposons. At the moment, the sequence of only one metapolycentromere has been established. The metapolycentromere of *P. sativum* chromosome 6 is 81.6 Mb long and includes nine families of satellite DNA. Satellites from three of these families form up to 1 Mb clusters of centromeric chromatin marked by CENP-A. Except for the enrichment with satellite DNA, the

Species	Reference	Species	Reference
Flowering plants		Flowering plants	
Allium cepa*	Fiskesjö et al., 1981	Strophanthus divaricates*	Beentje, 1982
Allium erdelii*	Kollmann, 1970	Strophanthus sarmentosus*	Beentje, 1982
Allium neapolitanum*	Badr, Elkington, 1977	Gymnosperm plants	
Allium qasyunense*	Kollmann, 1970	Cryptomeria japonica*	Schlarbaum, Tsuchiya, 1984b
Allium sativum*	Panda et al., 1979	Cunninghamia lanceolata*	Schlarbaum, Tsuchiya, 1984a
Allium subhirsutum*	Badr, Elkington, 1977	Metasequoia glyptostroboides*	Schlarbaum, Tsuchiya, 1984b
Allium trifoliatus*	Miceli et al., 1984	Phyllocladus trichomanoides*	Davies et al., 1997
Allium trioliatum*	Badr, Elkington, 1977	Sequoiadendron giganteum*	Schlarbaum, Tsuchiya, 1984b
Arachis viliosa*	Stalker, Dalmacio, 1981	Taiwania cryptomerioides*	Schlarbaum, Tsuchiya, 1984a
Chamaelirium luteum*	Tanaka, 2020	Taxodium distiehum*	Schlarbaum, Tsuchiya, 1984b
Colchicum ritchii [*]	Feinbrun, 1958	Tsuga longibracteata*	Li, 1991
Colchicum schimperi*	Feinbrun, 1958	Insects	
Dioscorea deltoidea*	Bhat, Bindroo, 1980	Mycetomoellerius urichii	Teixeira et al., 2022
Filipendula ulmaria*	Baker H.G., Baker I., 1967	Solenopsis geminate	Huang YC. et al., 2016
Filipendula vulgaris*	Baker H.G., Baker I., 1967	Solenopsis invicta	Huang YC. et al., 2016
Lathyrus clymenum	Neumann et al., 2015	Trachymyrmex holmgreni	Cardoso et al., 2018
Lathyrus latifolius	Neumann et al., 2015	Tribolium castaneum	Gržan et al., 2020
Lathyrus niger	Neumann et al., 2015	Vertebrates	
Lathyrus ochrus	Neumann et al., 2015	Chloebia gouldiae	Malinovskaya et al., 2022
Lathyrus sativus	Neumann et al., 2015	Ficedula hypoleuca	Malinovskaya et al., 2022
Lathyrus sylvestris	Neumann et al., 2015	Linaria cannabina	Grishko et al., 2023
Lathyrus vernus	Neumann et al., 2015	Mesoplodon carlhubbsi*	Kurihara et al., 2017
Pisum fulvum	Neumann et al., 2015	Muntiacus muntjak	Comings, Okada, 1971
Pisum sativum	Neumann et al., 2012	Pyrrhula pyrrhula	Grishko et al., 2023
Rutidosis leiolepis*	Young et al., 2002	Serinus canaria	Malinovskaya et al., 2022

Note. The newly mined species with potential metapolycentromeres are indicated by asterisks.



Fig. 3. Mitotic metaphase (*a*) and synaptonemal complexes (*b*, *c*) of the male domestic canary after DAPI staining (*a*) and immunostaining using antibodies against SYCP3, the main protein of the lateral elements of the synaptonemal complex (red), human centromere proteins (green) (*b*) and SYCP3 (red), and H3K9me2/3, histone H3, di- and trimethylated at lysine 9 (green) (*c*).

Numbers indicate macrochromosomes with metapolycentromeres. Arrows indicate extended primary constrictions (*a*) and metapolycentromeres (*b*, *c*). GRC indicates germline restricted chromosome. Bar 5 µm. After L.P. Malinovskaya et al. (2022), modified with permission.

metapolycentromere does not differ from the adjacent regions of the chromosome in DNA methylation patterns, the location of transcriptionally active genes, and retrotransposons (Macas et al., 2023).

E. Grishko et al. (2023) and L.P. Malinovskaya et al. (2022) demonstrated that the metapolycentromeres of the songbirds do not differ from their regional centromeres in the H3K9 methylation patterns (Fig. 3c). For example, all macrochromosomes of the domestic canary contain metapolycentromeres, and all of them except the Z chromosome are hypermethylated at H3K9, as well as the regional centromeres of all macrochromosomes except the Z chromosome in several other songbird species studied.

Meanwhile, P. Neumann et al. (2016) revealed a striking similarity between metapolycentromeres and holocentromeres in the patterns of histone modifications H3S10ph, H3S28ph, and H3T3ph distributions in *L. sativus* and *P. sativum* chromosomes. The metapolycentromeres showed a unique pattern of H2AT120ph distribution, significantly different from that of both regional and holocentromeres. The genomes of *Pisum* and *Lathyrus* contain two variants of the CENP-A gene, named CenH3-1 and CenH3-2 (Neumann et al., 2012), the sequences of which show 55 % homology, while corresponding proteins differ in length and amino acid sequence and show 72 % homology (Neumann et al., 2012). Both forms of CENP-A are localized on functional chromatin clusters of metapolycentromeres in these species (Neumann et al., 2015).

Simultaneous immunodetection of CENP-A and tubulin in *P. sativum* revealed colocalization of these proteins in the centromeric region, indicating that each cluster of centromeric chromatin within the metapolycentromere forms a functional kinetochore (Neumann et al., 2012). Regional and metapolycentromeres do not differ in the strength of their suppressive effect on meiotic recombination in the pericentromeric chromosome regions (Grishko et al., 2023).

Thus, metapolycentromeres may vary in the number of centromeric domains and in their genetic content and epigenetic modifications. However, these variations do not seem to affect their function.

Origin of metapolycentromeres

At the moment, several mechanisms for the formation of metapolycentromeres were suggested: multiple Robertsonian translocations in the Indian muntjac (Huang L. et al., 2006), segmental duplications in legumes (Macas et al., 2023), epigenetic charges in the interspecies marsupial hybrids (O'Neill et al., 1998) and expansion of centromeric chromatin and over-expression of the CENP-A protein in the malignant neoplasms (Sullivan L.L. et al., 2011, 2016; Perpelescu et al., 2015).

In the Indian muntjac (*M. muntjak vaginalis*), the metapolycentromere is located on the X chromosome (Drpic et al., 2018). This species has the smallest number of chromosomes among mammals: 2n = 6 in females and 2n = 7 in males (Wurster, Benirschke, 1970). The reduction of the chromosome was determined by a fusion of chromosomes in an ancestor with a karyotype of 2n = 70 (Yang et al., 1997; Chi et al., 2005). The elongated centromere of the X chromosome was suggested to result from several successive Robertsonian translocations (Chang et al., 2001; Huang L. et al., 2006). However, it remains unclear why all autosomes of this species, which also resulted from multiple Robertsonian translocations, have the standard regional centromeres.

In legumes, metapolycentromeres may have arisen through a mechanism associated with the duplication of the centromeric histone H3 gene (Neumann et al., 2015). However, the presence of two CENP-A variants is not a determinant of the presence of metapolycentromeres in *Pisum* and *Lathyrus*. Several plant species have two CENP-A variants but no metapolycentromeres, for example, *A. lyrata* and *Mimulus* spp. (Kawabe et al., 2006; Finseth et al., 2015). Thus, in peas, sequencing of long reads combined with ChIP-seq with antibodies to CENP-A showed the emergence of the newest domain of centromeric chromatin through segmental duplication and subsequent inversion of an existing domain 5.2 Mb long. However, the origin of the remaining domains of centromeric chromatin is unclear (Macas et al., 2023).

Multiple tandem duplications play a major role in the homogenization of centromeric repeat monomers in rice (Ma J., Jackson, 2006). They might result from unequal crossing over, gene conversion, duplicate transposition, satellite transposition, and illegitimate recombination (Copenhaver et al., 1999; Ma J., Jackson, 2006).

Typically, dicentric and polycentric chromosomes cannot ensure the attachment of unipolar spindle microtubules to their chromatids, which causes chromosome breakage and nondisjunction. Thus, there are mechanisms that select against such a chromosome structure (for example, the elimination of one of the centromeres) (Zhang et al., 2010). However, this does not occur in the case of metapolycentromeres due to the close proximity of the centromeric domains (Neumann et al., 2012). It is known that the distance between two functional centromeres should not exceed 20 Mbp for them to function as one centromere during cell division (Higgins et al., 2005). Apparently, this condition is also satisfied for metapolycentromere domains.

Metapolycentromeres can arise *de novo* from regional centromeres under conditions of genomic instability. Such destabilizing conditions may include interspecific hybridization and malignant neoplasms (Metcalfe et al., 2007; Sullivan L.L. et al., 2011).

The elongated centromeres have been observed in some chromosomes of interspecific hybrids of several marsupial species (kangaroos and wallabies), while the chromosomes of the parental species contained regional centromeres. Interestingly, the elongated centromeres were only present on the maternally derived chromosomes (O'Neill et al., 1998, 2001; Metcalfe et al., 2007; Schroeder-Reiter, Wanner, 2009). This phenomenon was observed in hybrids between the closely related species Macropus rufogriseus and M. agilis, as well as in those between the phylogenetically distant species M. eugenii and Wallabia bicolor (O'Neill et al., 1998; Metcalfe et al., 2007). In all these hybrids, the expansion of centromeric chromatin occurred due to an uncontrolled increase in the number of copies of centromeric retrotransposons, and for different hybrids, the families of retrotransposons that facilitated the expansion differed (O'Neill et al., 1998; Metcalfe et al., 2007). Apparently, the changes in the epigenetic context due to hybridization disrupt DNA methylation patterns that normally restrain the activity of centromeric retrotransposons. This, in turn, leads to their repeated copying and the expansion of the centromeric region (O'Neill et al., 1998). However, it is still not clear why this phenomenon is limited to maternally derived chromosomes.

Expansion of centromeric chromatin also occurs in some human cancer cells (Sullivan L.L. et al., 2011, 2016; Perpelescu et al., 2015). Thus, in cell line GM08148, a rearrangement on chromosome 17 resulted in the centromere entering the euchromatic environment; as a result, CENP-A spread into the short arm and formed an elongated functional centromere on a non-centromeric DNA sequence (Sullivan L.L. et al., 2016). Additionally, overexpression of the CENP-A protein and its chaperone HJURP, along with the disruption of the interaction of the tumor suppressor protein Rb with chromatin in cancer cells, can lead to centromere elongation (Sullivan L.L. et al., 2011; Perpelescu et al., 2015). Altered epigenetic landscapes and uncontrolled proliferation of centromeric sequences may trigger dysregulated expansion of centromeric chromatin.

Metapolycentromere evolution and the centromere drive hypothesis

The conservative centromere function – the attachment of spindle microtubules and subsequent chromosomal segregation – implies strict purifying selection on the components of the centromere: centromere DNA and centromere proteins. However, in reality, we observe a completely opposite picture – both centromeric DNA and centromeric proteins evolve rapidly and often differ significantly even between closely related species. This contradiction is called the "centromere paradox" (Henikoff et al., 2001).

To resolve the centromere paradox, S. Henikoff et al. (2001) suggested the centromere drive hypothesis. This hypothesis suggests that in asymmetric female meiosis, the centromeres segregating in the egg rather than in the polar body ("the strong centromeres") would be favored. However, male meiosis is symmetric. In this case, inequality in centromere strength might lead to chromosome nondisjunction and spermatogenic arrest (Malik, Henikoff, 2001). The resulting conflict might be resolved by a selection for centromeric proteins, which are able to equalize the centromeres and compensate for the fitness costs (Fig. 4). This perpetual tug-of-war between male

and female meiosis should result in the rapid evolution of centromeric sequences and proteins (Dawe, Henikoff, 2006).

Selection for "stronger centromeres" in female meiosis might favor variants of the centromeric DNA sequences with enhanced potential to recruit centromeric proteins (in particular CENP-A) and form kinetochores to which more microtubules are attached. This effect may also be enhanced by a selection for an increase in the copy number of such sequences. These processes could cause the occurrence of metapolycentromeres. The suppression of centromeric drive in male meiosis may limit centromere size. Probably, this is why metapolycentromeres are so rare. They have been found in several ant species (Huang Y.-C. et al., 2016; Cardoso et al., 2018) with haploid males. For this reason, there should be no selection for suppression of centromeric drive in male meiosis. This makes Hymenoptera a promising group for the search for new metapolycentromeres.

Thus, the centromere drive hypothesis provides a plausible explanation for the dynamic evolution of centromeres in general and the emergence of metapolycentromeres in particular.

Do metapolycentromeres represent an intermediate stage of evolution between regional centromeres and holocentromeres?

P. Neumann et al. (2012) suggested that metapolycentromeres might represent an intermediate stage of evolution between regional centromeres and holocentromeres. According to this hypothesis, the satellite DNA sequences of the regional centromere, under the influence of centromere drive, might expand so much that they capture the entire chromosome, rendering it holocentric.

During evolution, holocentromeres arose from regional centromeres at least 13 times: four times in plants and nine times in animals (Melters et al., 2012). Despite the common morphological feature (i. e. the absence of a primary constriction for the attachment of spindle filaments), holocentric chromosomes differ from each other in their origin and structure (Melters et al., 2012; Senaratne et al., 2022). Holocentromere centromeric units (chromosomal regions marked with CENP-A) can be based on either satellite or non-repeated



Fig. 4. The model of centromere drive according to S. Henikoff et al. (2001), modified.

DNA sequences (Gassmann et al., 2012; Marques et al., 2015). In turn, satellite holocentromeres are divided into holocentromeres with a large number of small centromeric units and holocentromeres with a small number of large centromeric units (Kuo et al., 2024). Large centromeric units comparable in size to regional centromeres have been discovered in the plants *Chionographis japonica* and *Morus notabilis* (Kuo et al., 2023; Ma B. et al., 2023). It was suggested that holocentromeres in *C. japonica* formed through multiple misrepaired DNA double-strand breaks associated with the insertion of extra-chromosomal circular DNA (Kuo et al., 2024). These insertions of regions of centromeric chromatin might not occur simultaneously throughout the genome, but evolve from metapolycentromeres.

The genera Juncus, Drosera and Cuscuta include both species with holocentromeres and species with regional centromeres (Pazy, Plitmann, 1994; Shirakawa et al., 2011a, b; Guerra et al., 2019; Neumann et al., 2021; Mata-Sucre et al., 2023). Recently, using ChIP-seq with anti-CENP-A antibodies, it was found that the chromosomes of J. effusus bear both regional centromeres and polycentromeres with multiple CENP-A domains (Dias et al., 2024). Such centromeres are similar in structure to metapolycentromeres, but they do not form elongated primary constrictions due to the small number of centromeric domains and their close proximity to each other. The presence of holocentromere and regional centromere species in the genus Juncus led to the suggestion that this species represents a transitional form from regional centromeres to holocentromeres. However, not a single "transitive karyotype" containing both metapolycentric and holocentric chromosomes has been discovered.

Even if this hypothesis holds true, it would only explain the origin of holocentricity in a small number of species with holocentric chromosomes, because most holocentric chromosomes do not possess centromere-specific DNA sequences (Talbert, Henikoff, 2020; Senaratne et al., 2021, 2022).

Backward and forward search for metapolycentromeres

We suspect metapolycentromeres are more common than believed. However, finding them is problematic. They can be reliably revealed by immunostaining chromosomes with antibodies to CENP-A or by ChIP-seq with anti-CENP-A antibodies. Metapolycentromeres may also be indirectly detected by the analysis of the copy number of centromeric repeats, by immunostaining for kinetochore proteins, and, in the case of particularly large metapolycentromeres, by routine chromosome staining, which reveals them as elongated primary constrictions. However, indirect methods do not reveal the actual number of functional domains of centromeric chromatin.

The term metapolycentromere was suggested by P. Neumann et al. (2012), and before that date, elongated primary constrictions were not termed metapolycentromeres and often were not mentioned at all. In the backward search for potential metapolycentromeres, we carried out data mining for the cytogenetic articles in the scholar.google.com database (last access: 7th of July 2023) using 18 keywords (Supplementary Material)¹. We selected all articles written in English that mentioned long primary constrictions in the text or showed them in the micrographs. Table shows the list of already known and newly mined candidate species with metapolycentromeres.

It spans 27 species of flowering and eight species of gymnosperm plants, five species of insects and seven species of vertebrates. It indicates an erratic phylogenetic distribution of the species with metapolycentromeres. This, in turn, may suggest independent evolutionary occurrences of metapolycentromeres. However, the current catalog of species with identified and suspected metapolycentromeres remains too short to draw reliable conclusions about their evolution, particularly in the absence of knowledge about related species without metapolycentromeres for comparative analysis. More studies are necessary to shed light on the mechanisms of metapolycentromere formation and evolution.

Conclusion

The systematic study of new species with and without metapolycentromeres is important for understanding their evolution. Species with karyotypes containing both regional centromeres and metapolycentromeres are especially interesting. A comparison between the centromeric DNA of metapolycentromeres and regional centromeres may shed light on the mechanisms of metapolycentromere formation.

References

- Ahmad S.F., Singchat W., Jehangir M., Suntronpong A., Panthum T., Malaivijitnond S., Srikulnath K. Dark matter of primate genomes: Satellite DNA repeats and their evolutionary dynamics. *Cells*. 2020; 9(12):2714. DOI 10.3390/cells9122714
- Badr A., Elkington T.T. Variation of Giemsa C-band and fluorochrome banded karyotypes, and relationships in *Allium* subgen. *Molium. Pl. Syst. Evol.* 1977;128(1-2):23-35. DOI 10.1007/BF00985168
- Baker H.G., Baker I. The cytotaxonomy of *Filipendula* (Rosaceae) and its implications. *Am. J. Bot.* 1967;54(8):1027-1034. DOI 10.1002/ j.1537-2197.1967.tb10729.x
- Balzano E., Giunta S. Centromeres under pressure: Evolutionary innovation in conflict with conserved function. *Genes (Basel)*. 2020; 11(8):912. DOI 10.3390/genes11080912
- Beentje H.J. A Monograph on Strophanthus DC. (Apocynaceae). Wageningen, 1982
- Bhat B.K., Bindroo B.B. Sex chromosomes in *Dioscorea deltoidea* Wall. *Cytologia* (*Tokyo*). 1980;45(4):739-742. DOI 10.1508/cytologia. 45.739
- Black B.E., Foltz D.R., Chakravarthy S., Luger K., Woods V.L., Cleveland D.W. Structural determinants for generating centromeric chromatin. *Nature*. 2004;430(6999):578-582. DOI 10.1038/nature02766
- Cardoso D.C., Heinze J., Moura M.N., Cristiano M.P. Chromosomal variation among populations of a fungus-farming ant: implications for karyotype evolution and potential restriction to gene flow. *BMC Evol. Biol.* 2018;18(1):146. DOI 10.1186/s12862-018-1247-5
- Castellani M., Zhang M., Thangavel G., Mata-Sucre Y., Lux T., Campoy J.A., Marek M., Huettel B., Sun H., Mayer K.F.X., Schneeberger K., Marques A. Meiotic recombination dynamics in plants with repeat-based holocentromeres shed light on the primary drivers of crossover patterning. *Nat. Plants.* 2024;10:423-438. DOI 10.1038/ s41477-024-01625-y
- Chang S.D., Chao A.S., Lai Y.M., Liu H.Y., Soong Y.K. Interphase FISH-assisted second-trimester termination of a trisomy 21 fetus in an IVF-ET twin pregnancy. A case report. *J. Reprod. Med.* 2001; 46(12):1063-1066
- Chi J.X., Huang L., Nie W., Wang J., Su B., Yang F. Defining the orientation of the tandem fusions that occurred during the evolution of Indian muntjac chromosomes by BAC mapping. *Chromosoma*. 2005;114(3):167-172. DOI 10.1007/s00412-005-0004-x

¹ Supplementary Material is available at:

https://vavilovj-icg.ru/download/pict-2024-28/appx21.pdf

- Cleveland D.W., Mao Y., Sullivan K.F. Centromeres and kinetochores. Cell. 2003;112(4):407-421. DOI 10.1016/S0092-8674(03)00115-6
- Comings D.E., Okada T.A. Fine structure of kinetochore in Indian muntjac. *Exp. Cell Res.* 1971;67(1):97-110. DOI 10.1016/0014-4827(71)90625-2
- Copenhaver G.P., Nickel K., Kuromori T., Benito M.I., Kaul S., Lin X., Bevan M., Murphy G., Harris B., Parnell L.D., McCombie W.R., Martienssen R.A., Marra M., Preuss D. Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science*. 1999; 286(5449):2468-2474. DOI 10.1126/science.286.5449.2468
- Davies B.J., O'Brien I.E.W., Murray B.G. Karyotypes, chromosome bands and genome size variation in New Zealand endemic gymnosperms. *Plant Syst. Evol.* 1997;208(3-4):169-185. DOI 10.1007/ BF00985440
- Dawe R.K., Henikoff S. Centromeres put epigenetics in the driver's seat. *Trends Biochem. Sci.* 2006;31(12):662-669. DOI 10.1016/j.tibs. 2006.10.004
- Dias Y., Mata-Sucre Y., Thangavel G., Costa L., Baez M., Houben A., Marques A., Pedrosa-Harand A. How diverse a monocentric chromosome can be? Repeatome and centromeric organization of *Juncus effusus* (Juncaceae). *Plant J.* 2024;118(6):1832-1847. DOI 10.1111/ tpj.16712
- Drinnenberg I.A., deYoung D., Henikoff S., Malik H.S. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *eLife*. 2014;3:e03676. DOI 10.7554/eLife.03676
- Drpic D., Almeida A.C., Aguiar P., Renda F., Damas J., Lewin H.A., Larkin D.M., Khodjakov A., Maiato H. Chromosome segregation is biased by kinetochore size. *Curr. Biol.* 2018;28(9):1344-1356. DOI 10.1016/j.cub.2018.03.023
- Feinbrun N. Chromosome numbers and evolution in the genus Colchicum. Evolution (N.Y.). 1958;12(2):173. DOI 10.2307/2406028
- Finseth F.R., Dong Y., Saunders A., Fishman L. Duplication and adaptive evolution of a key centromeric protein in *Mimulus*, a genus with female meiotic drive. *Mol. Biol. Evol.* 2015;32(10):2694-2706. DOI 10.1093/molbev/msv145
- Fiskesjö G., Lassen C., Renberg L. Chlorinated phenoxyacetic acids and chlorophenols in the modified *Allium* test. *Chem. Biol. Interact.* 1981;34(3):333-344. DOI 10.1016/0009-2797(81)90105-8
- Flemming W. Zellsubstanz, Kern und Zelltheilung. Leipzig: F.C.W. Vogel, 1882. DOI 10.5962/bhl.title.168645
- Furuyama S., Biggins S. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proc. Natl. Acad. Sci.* USA. 2007;104(37):14706-14711. DOI 10.1073/pnas.0706985104
- Gassmann R., Rechtsteiner A., Yuen K., Muroyama A., Egelhofer T., Gaydos L., Barron F., Maddox P., Essex A., Monen J., Ercan S., Lieb J.D., Oegema K., Strome S., Desai A. An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans. Nature.* 2012;484:534-537. DOI 10.1038/nature10973
- Glöckner G., Heidel A.J. Centromere sequence and dynamics in *Dic*tyostelium discoideum. Nucleic Acids Res. 2009;37(6):1809-1816. DOI 10.1093/nar/gkp017
- Grishko E., Malinovskaya L., Slobodchikova A., Kotelnikov A., Torgasheva A., Borodin P. Cytological analysis of crossover frequency and distribution in male meiosis of cardueline finches (Fringillidae, Aves). *Animals*. 2023;13(23):3624. DOI 10.3390/ani13233624
- Gržan T., Despot-Slade E., Meštrović N., Plohl M., Mravinac B. CenH3 distribution reveals extended centromeres in the model beetle *Tribolium castaneum*. *PLoS Genet*. 2020;16(10):e1009115. DOI 10.1371/ journal.pgen.1009115
- Guerra M., Ribeiro T., Felix L.P. Monocentric chromosomes in Juncus (Juncaceae) and implications for the chromosome evolution of the family. Bot. J. Linn. Soc. 2019;191(4):475-483. DOI 10.1093/ botlinnean/boz065
- Haupt W., Fischer T.C., Winderl S., Fransz P., Torres-Ruiz R.A. The CENTROMERE1 (CEN1) region of *Arabidopsis thaliana*: architecture and functional impact of chromatin. *Plant J.* 2001;27(4):285-296. DOI 10.1046/j.1365-313x.2001.01087.x

- Henikoff S., Ahmad K., Malik H.S. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science*. 2001;293(5532): 1098-1102. DOI 10.1126/science.1062939
- Henikoff S., Ramachandran S., Krassovsky K., Bryson T.D., Codomo C.A., Brogaard K., Widom J., Wang J.-P., Henikoff J.G. The budding yeast centromere DNA element II wraps a stable Cse4 hemisome in either orientation *in vivo. eLife.* 2014;3:e01861. DOI 10.7554/eLife.01861
- Higgins A.W., Gustashaw K.M., Willard H.F. Engineered human dicentric chromosomes show centromere plasticity. *Chromosom. Res.* 2005;13(8):745-762. DOI 10.1007/s10577-005-1009-2
- Huang L., Chi J., Nie W., Wang J., Yang F. Phylogenomics of several deer species revealed by comparative chromosome painting with Chinese muntjac paints. *Genetica*. 2006;127(1-3):25-33. DOI 10.1007/s10709-005-2449-5
- Huang Y.-C., Lee C.-C., Kao C.-Y., Chang N.-C., Lin C.-C., Shoemaker D., Wang J. Evolution of long centromeres in fire ants. *BMC Evol. Biol.* 2016;16(1):189. DOI 10.1186/s12862-016-0760-7
- Kanesaki Y., Imamura S., Matsuzaki M., Tanaka K. Identification of centromere regions in chromosomes of a unicellular red alga, *Cyanidioschyzon merolae*. *FEBS Lett.* 2015;589(11):1219-1224. DOI 10.1016/j.febslet.2015.04.009
- Kawabe A., Nasuda S., Charlesworth D. Duplication of centromeric histone H3 (*HTR12*) gene in *Arabidopsis halleri* and *A. lyrata*, plant species with multiple centromeric satellite sequences. *Genetics*. 2006;174(4):2021-2032. DOI 10.1534/genetics.106.063628
- Kollmann F. Karyotypes of three *Allium* species of the *erdelii* group. *Caryologia*. 1970;23(4):647-655. DOI 10.1080/00087114.1970. 10796400
- Kuo Y.T., Câmara A.S., Schubert V., Neumann P., Macas J., Melzer M., Chen J., Fuchs J., Abel S., Klocke E., Huettel B., Himmelbach A., Demidov D., Dunemann F., Mascher M., Ishii T., Marques A., Houben A. Holocentromeres can consist of merely a few megabasesized satellite arrays. *Nat. Commun.* 2023;14:3502. DOI 10.1038/ s41467-023-38922-7
- Kuo Y.T., Schubert V., Marques A., Schubert I., Houben A. Centromere diversity: How different repeat-based holocentromeres may have evolved. *BioEssays*. 2024;46(6):2400013. DOI 10.1002/bies. 202400013
- Kurihara N., Tajima Y., Yamada T.K., Matsuda A., Matsuishi T. Description of the karyotypes of stejneger's beaked whale (*Mesoplodon stejnegeri*) and hubbs' beaked whale (*M. carlhubbsi*). *Genet. Mol. Biol.* 2017;40(4):803-807. DOI 10.1590/1678-4685-gmb-2016-0284
- Lee H.R., Zhang W., Langdon T., Jin W., Yan H., Cheng Z., Jiang J. Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. *Proc. Natl. Acad. Sci. USA.* 2005;102(33):11793-11798. DOI 10.1073/pnas. 0503863102
- Li L.C. The karyotype analysis of *Tsuga longibracteata* and its taxonomic significance. *Acta Bot. Yunnan.* 1991;13(3):309-313
- Ma B., Wang H., Liu J., Chen L., Xia X., Wei W., Yang Z., Yuan J., Luo Y., He N. The gap-free genome of mulberry elucidates the architecture and evolution of polycentric chromosomes. *Hortic. Res.* 2023;10(7):uhad111. DOI 10.1093/hr/uhad111
- Ma J., Jackson S.A. Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. *Genome Res.* 2006;16(2):251-259. DOI 10.1101/gr.4583106
- Macas J., Ávila Robledillo L., Kreplak J., Novák P., Koblížková A., Vrbová I., Burstin J., Neumann P. Assembly of the 81.6 Mb centromere of pea chromosome 6 elucidates the structure and evolution of metapolycentric chromosomes. *PLoS Genet.* 2023;19(2):e1010633. DOI 10.1371/journal.pgen.1010633
- Maheshwari S., Tan E.H., West A., Franklin F.C.H., Comai L., Chan S.W.L. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet*. 2015;11(1):e1004970. DOI 10.1371/journal.pgen.1004970

- Malik H.S., Henikoff S. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. *Genetics*. 2001;157(3):1293-1298. DOI 10.1093/genetics/157.3.1293
- Malinovskaya L.P., Slobodchikova A.Y., Grishko E.O., Pristyazhnyuk I.E., Torgasheva A.A., Borodin P.M. Germline-restricted chromosomes and autosomal variants revealed by pachytene karyotyping of 17 avian species. *Cytogenet. Genome Res.* 2022;162(3):148-160. DOI 10.1159/000524681
- Marques A., Ribeiro T., Neumann P., Macas J., Novák P., Schubert V., Pellino M., Fuchs J., Ma W., Kuhlmann M., Brandt R., Vanzela A.L.L., Beseda T., Šimková H., Pedrosa-Harand A., Houben A. Holocentromeres in Rhynchospora are associated with genomewide centromere-specific repeat arrays interspersed among euchromatin. *Proc. Natl. Acad. Sci. USA*. 2015;112(44):13633-13638. DOI 10.1073/pnas.1512255112
- Mata-Sucre Y., Matzenauer W., Castro N., Huettel B., Pedrosa-Harand A., Marques A., Souza G. Repeat-based phylogenomics shed light on unclear relationships in the monocentric genus *Juncus* L. (Juncaceae). *Mol. Phylogenet. Evol.* 2023;189:107930. DOI 10.1016/ j.ympev.2023.107930
- Melters D.P., Paliulis L.V., Korf I.F., Chan S.W.L. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosom. Res.* 2012;20(5):579-593. DOI 10.1007/ s10577-012-9292-1
- Melters D.P., Bradnam K.R., Young H.A., Telis N., May M.R., Ruby J., Sebra R., Peluso P., Eid J., Rank D., Garcia J., DeRisi J.L., Smith T., Tobias C., Ross-Ibarra J., Korf I., Chan S.W. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* 2013;14(1):R10. DOI 10.1186/gb-2013-14-1-r10
- Mendiburo M.J., Padeken J., Fülöp S., Schepers A., Heun P. Drosophila CENH3 is sufficient for centromere formation. *Science*. 2011; 334(6056):686-690. DOI 10.1126/science.1206880
- Metcalfe C.J., Bulazel K.V., Ferreri G.C., Schroeder-Reiter E., Wanner G., Rens W., Obergfell C., Eldridge M.D.B., O'Neill R.J. Genomic instability within centromeres of interspecific marsupial hybrids. *Genetics*. 2007;177(4):2507-2517. DOI 10.1534/genetics. 107.082313
- Miceli P., Ficini G., Garbari F. The genus «Allium» L. in Italy. XIII. Morphological, caryological and leaf anatomical study in some C-W Mediterranean triploid populations of «Allium trifoliatum» Cyr. *Webbia*. 1984;38(1):793-803. DOI 10.1080/00837792.1984. 10670350
- Nagpal H., Fierz B. The elusive structure of centro-chromatin: Molecular order or dynamic heterogenetity. J. Mol. Biol. 2021;433(6): 166676. DOI 10.1016/j.jmb.2020.10.010
- Navarro-Mendoza M.I., Pérez-Arques C., Panchal S., Nicolás F.E., Mondo S.J., Ganguly P., Pangilinan J., Grigoriev I.V., Heitman J., Sanyal K., Garre V. Early diverging fungus *Mucor circinelloides* lacks centromeric histone CENP-A and displays a mosaic of point and regional centromeres. *Curr. Biol.* 2019;29(22):3791-3802. DOI 10.1016/j.cub.2019.09.024
- Neumann P., Navrátilová A., Schroeder-Reiter E., Koblížková A., Steinbauerová V., Chocholová E., Novák P., Wanner G., Macas J. Stretching the rules: monocentric chromosomes with multiple centromere domains. *PLoS Genet.* 2012;8(6):e1002777. DOI 10.1371/ journal.pgen.1002777
- Neumann P., Pavlíková Z., Koblížková A., Fuková I., Jedličková V., Novák P., Macas J. Centromeres off the hook: Massive changes in centromere size and structure following duplication of *Cenh3* gene in *Fabeae* species. *Mol. Biol. Evol.* 2015;32(7):1862-1879. DOI 10.1093/molbev/msv070
- Neumann P., Schubert V., Fuková I., Manning J.E., Houben A., Macas J. Epigenetic histone marks of extended meta-polycentric centromeres of *Lathyrus* and *Pisum* chromosomes. *Front. Plant Sci.* 2016;7(MAR2016):234. DOI 10.3389/fpls.2016.00234
- Neumann P., Oliveira L., Čížková J., Jang T.S., Klemme S., Novák P., Stelmach K., Koblížková A., Doležel J., Macas J. Impact of parasitic

lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus *Cuscuta*. *New Phytologist*. 2021;229(4):2365-2377. DOI 10.1111/nph.17003

- O'Neill R.J.W., O'Neill M.J., Marshall Graves J.A. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*. 1998;393(6680): 68-72. DOI 10.1038/29985
- O'Neill R.J.W., Eldridge M.D.B., Graves J.A.M. Chromosome heterozygosity and *de novo* chromosome rearrangements in mammalian interspecies hybrids. *Mamm. Genome.* 2001;12(3):256-259. DOI 10.1007/s003350010270
- Panda B.B., Sahu R.K., Sharma C.B.S.R. Cytogenetic hazards from agricultural chemicals. 2. Selective clastogenesis and spindle inhibition in some plant mitotic systems by the β-exotoxin and the general ineffectiveness of the δ-endotoxin protein of *Bacillus thuringiensis. Mutat. Res. Toxicol.* 1979;67(2):161-166. DOI 10.1016/0165-1218(79)90127-7
- Pazy B., Plitmann U. Holocentric chromosome behaviour in *Cuscuta* (*Cuscutaceae*). *Plant Syst. Evol.* 1994;191:105-109. DOI 10.1007/ BF00985345
- Perpelescu M., Hori T., Toyoda A., Misu S., Monma N., Ikeo K., Obuse C., Fujiyama A., Fukagawa T. HJURP is involved in the expansion of centromeric chromatin. *Mol. Biol. Cell.* 2015;26(15): 2742-2754. DOI 10.1091/mbc.E15-02-0094
- Schlarbaum S.E., Tsuchiya T. The chromosomes of *Cunninghamia* konishii, C. lanceolata, and *Taiwania cryptomerioides* (*Taxodiaceae*). *Plant Syst. Evol.* 1984a;145(3-4):169-181. DOI 10.1007/ BF00983946
- Schlarbaum S.E., Tsuchiya T. Cytotaxonomy and phylogeny in certain species of *Taxodiaceae*. *Plant Syst. Evol.* 1984b;147(1-2):29-54. DOI 10.1007/BF00984578
- Schroeder-Reiter E., Wanner G. Chromosome centromeres: Structural and analytical investigations with high resolution scanning electron microscopy in combination with focused ion beam milling. *Cytogenet. Genome Res.* 2009;24(3-4):239-250. DOI 10.1159/ 000218129
- Senaratne A.P., Muller H., Fryer K.A., Kawamoto M., Katsuma S., Drinnenberg I.A. Formation of the CenH3-deficient holocentromere in Lepidoptera avoids active chromatin. *Curr. Biol.* 2021;31(1):173-181.e7. DOI 10.1016/j.cub.2020.09.078
- Senaratne A.P., Cortes-Silva N., Drinnenberg I.A. Evolution of holocentric chromosomes: Drivers, diversity, and deterrents. *Semin. Cell Dev. Biol.* 2022;127:90-99. DOI 10.1016/j.semcdb.2022.01.003
- Shirakawa J., Hoshi Y., Kondo K. Chromosome differentiation and genome organization in carnivorous plant family Droseraceae. *Chromosome Bot.* 2011a;6(4);111-119. DOI 10.3199/iscb.6.111
- Shirakawa J., Katsuya N., Yoshikazu H. A chromosome study of two centromere differentiating *Drosera* species, *D. arcturi* and *D. regia*. *Caryologia*. 2011b;64(4):453-463. DOI 10.1080/00087114.2011. 10589813
- Stalker H.T., Dalmacio R.D. Chromosomes of Arachis species, section Arachis. J. Hered. 1981;72(6):403-408. DOI 10.1093/oxford journals.jhered.a109541
- Sullivan K.F., Hechenberger M., Masri K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J. Cell Biol. 1994;127(3):581-592. DOI 10.1083/ jcb.127.3.581
- Sullivan L.L., Boivin C.D., Mravinac B., Song I.Y., Sullivan B.A. Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. *Chromosom. Res.* 2011. DOI 10.1007/s10577-011-9208-5
- Sullivan L.L., Maloney K.A., Towers A.J., Gregory S.G., Sullivan B.A. Human centromere repositioning within euchromatin after partial chromosome deletion. *Chromosom. Res.* 2016. DOI 10.1007/ s10577-016-9536-6
- Tachiwana H., Kagawa W., Kurumizaka H. Comparison between the CENP-A and histone H3 structures in nucleosomes. *Nucleus*. 2012; 3(1):6-11. DOI 10.4161/nucl.18372

- Talbert P.B., Henikoff S. Transcribing centromeres: Noncoding RNAs and kinetochore assembly. *Trends Genet.* 2018;34(8):587-599. DOI 10.1016/j.tig.2018.05.001
- Talbert P.B., Henikoff S. What makes a centromere? *Exp. Cell Res.* 2020;389(2):111895. DOI 10.1016/j.yexcr.2020.111895
- Talbert P.B., Masuelli R., Tyagi A.P., Comai L., Henikoff S. Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell*. 2002;14(5):1053-1066. DOI 10.1105/ tpc.010425
- Talbert P.B., Kasinathan S., Henikoff S. Simple and complex centromeric satellites in *Drosophila* sibling species. *Genetics*. 2018; 208(3):977-990. DOI 10.1534/genetics.117.300620
- Tanaka N. Chromosomal traits of *Chamaelirium luteum* (Melanthiaceae) with particular focus on the large heterochromatic centromeres. *Taiwania*. 2020;65(3):286-294. DOI 10.6165/tai.2020.65.286
- Teixeira G.A., Barros L.A.C., de Aguiar H.J.A.C., Lopes D.M. Multiple heterochromatin diversification events in the genome of fungus-farming ants: insights from repetitive sequences. *Chromosoma*. 2022;131(1-2):59-75. DOI 10.1007/s00412-022-00770-7
- van Hooff J.J., Tromer E., van Wijk L.M., Snel B., Kops G.J. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep.* 2017;18(9):1559-1571. DOI 10.15252/embr.201744102

Conflict of interest. The authors declare no conflict of interest. Received April 16, 2024. Revised July 15, 2024. Accepted July 18, 2024.

- Wang Y., Wu L., Yuen K.W.Y. The roles of transcription, chromatin organisation and chromosomal processes in holocentromere establishment and maintenance. *Semin.Cell Dev. Biol.* 2022;127:79-89. DOI 10.1016/j.semcdb.2022.01.004
- Winey M., Mamay C.L., O'Toole E.T., Mastronarde D.N., Giddings T.H., McDonald K.L., McIntosh J.R. Three-dimensional ultrastructural analysis of the *Saccharomyces cerevisiae* mitotic spindle. *J. Cell Biol.* 1995;129(6):1601-1615. DOI 10.1083/jcb.129.6.1601
- Wurster D.H., Benirschke K. Indian muntjac, *Muntiacus muntjak*: a deer with a low diploid chromosome number. *Science*. 1970; 168(3937):1364-1366. DOI 10.1126/science.168.3937.1364
- Yang F., O'Brien P.C.M., Wienberg J., Neitzel H., Lin C.C., Ferguson-Smith M.A. Chromosomal evolution of the Chinese muntjac (*Muntiacus reevesi*). Chromosoma. 1997;106(1):37-43. DOI 10.1007/ s004120050222
- Young A., Hill J., Murray B., Peakall R. Breeding system, genetic diversity and clonal structure in the sub-alpine forb *Rutidosis leiolepis* F. Muell. (Asteraceae). *Biol. Conserv.* 2002;106(1):71-78. DOI 10.1016/S0006-3207(01)00230-0
- Zhang W., Friebe B., Gill B.S., Jiang J. Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres. *Chromosoma*. 2010;119(5):553-563. DOI 10.1007/s00412-010-0278-5

DOI 10.18699/vjgb-24-67

Новый ген опушения листа *Hl1th*, интрогрессированный в мягкую пшеницу от *Thinopyrum ponticum*, и его фенотипическое проявление при гомеологичных хромосомных замещениях

А.В. Симонов (D¹ 🖾, Е.И. Гордеева (D¹, М.А. Генаев (D^{1, 2}, В. Ли (D^{1, 2}, И.О. Булатов^{1, 3}, Т.А. Пшеничникова (D¹

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

³ Новосибирский государственный аграрный университет, Новосибирск, Россия

⊠ sialexander@bionet.nsc.ru

Аннотация. На основе сорта яровой мягкой пшеницы Саратовская 29 (С29) были созданы голубозерные линии C29_4Th(4B) и C29_4Th(4D) с соответствующим замещением хромосом 4B и 4D хромосомой 4Th от пырея вида Thinopyrum ponticum. У этих линий опушение листа отличается от реципиента и различается между собой, в связи с чем нами проведено исследование эффекта замещений на проявление данного признака. Для количественной оценки опушения была применена программа LHDetect2, определяющая длину и число трихом на микрофотографиях. Опушение листа у сорта C29 определяется главным геном *H*/1 в хромосоме 4В и еще одним геном со слабым эффектом с неизвестной хромосомной локализацией. Их взаимодействие приводит к формированию трихом длиной до 300 мкм. Замещение пары хромосом 4В на пару хромосом 4Th пырея модифицирует опушение листа у линии C29 4Th(4B). Характерное для сорта C29 опушение листа у линии C29 4Th(4B) становится реже, при этом образуются трихомы длиной до 600–700 мкм. Замещение гена *HI1 на HI1th у* линии C29_4Th(4B) также подтверждается аллельным состоянием сцепленного с геном НІ1 микросателлитного маркера Хдит 538. Нами была описана модификация опушения у замещенной линии C29_4Th(4D), где произошло замещение пары хромосом 4D, не содержащей гена опушения. Экспрессирующиеся совместно гены *HI1* и *HI1th* у линии C29_4Th(4D) в хромосомах 4B и 4Th соответственно формируют трихомы длиной более 400 мкм. Однако в таком генотипе снижается средняя длина трихом в сравнении с реципиентом. Таким образом, в результате проведенных исследований идентифицирован новый ген опушения листа, интрогрессированный из вида Th. ponticum в мягкую пшеницу, который мы обозначили как H11th. Для ведения отбора мы предлагаем использовать находящиеся в открытом доступе информативные микросателлитные маркеры Xgwm538 и Xgwm165, позволяющие различать хромосомы 4A, 4B, 4D и 4Th.

Ключевые слова: трихомы; цифровые характеристики опушения; фенотипические маркеры; микросателлитные маркеры; взаимодействие генов.

Для цитирования: Симонов А.В., Гордеева Е.И., Генаев М.А., Ли В., Булатов И.О., Пшеничникова Т.А. Новый ген опушения листа *HI1th*, интрогрессированный в мягкую пшеницу от *Thinopyrum ponticum*, и его фенотипическое проявление при гомеологичных хромосомных замещениях. *Вавиловский журнал генетики и селекции*. 2024;28(6):602-609. DOI 10.18699/vjgb-24-67

Финансирование. Работа проведена в рамках бюджетного проекта № FWNR-2022-0017. При анализе данных использованы вычислительные ресурсы ЦКП «Биоинформатика» при поддержке бюджетного проекта № FWNR-2022-0020.

Благодарности. Выражаем благодарность ЦКП репродукции растений и ЦКП микроскопического анализа ИЦиГ СО РАН. Работа поддержана грантом РНФ № 23-24-10029 и соглашением с администрацией НСО № Р-63.

A new leaf pubescence gene, $Hl1^{th}$, introgressed into bread wheat from *Thinopyrum ponticum* and its phenotypic manifestation under homoeologous chromosomal substitutions

A.V. Simonov (D¹ 🖾, E.I. Gordeeva (D¹, M.A. Genaev (D^{1, 2}, W. Li (D^{1, 2}, I.O. Bulatov^{1, 3}, T.A. Pshenichnikova (D¹

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

³ Novosibirsk State Agrarian University, Novosibirsk, Russia

⊠ sialexander@bionet.nsc.ru

Abstract. Blue-grain lines were created on the basis of the spring bread wheat variety Saratovskaya 29 (S29) with chromosome 4B or 4D replaced with chromosome 4Th from *Thinopyrum ponticum*. The leaf pubescence of the two lines differs from S29 and from each other. In this work, we studied the effect of these substitutions on the manifestation of this trait. To quantify pubescence, the LHDetect2 program was used to determine trichome length and number on the leaf fold

² Novosibirsk State University, Novosibirsk, Russia

microphotographs. The key gene *H1* on chromosome 4B and another unidentified gene with a weak effect determine the leaf pubescence of the recipient S29. Their interaction leads to the formation of trichomes of up to 300 microns in length. Replacement of both copies of chromosome 4B with two copies of wheatgrass chromosome 4Th modifies leaf pubescence in line S29_4Th(4B) so that the leaf pubescence characteristic of S29 becomes more sparse, and trichomes of up to 600–700 µm in length are formed. Additionally, we described modification of pubescence in the substitution line S29_4Th(4D) where chromosome 4D that does not carry any pubescence gene was replaced. Under this substitution, trichomes of up to 400 µm in length were formed and the average length of trichomes on the underside of the leaf was reduced. The replacement of the *H11* gene in the lines was also confirmed by the allelic state of the linked microsatellite marker *Xgwm538*. Thus, as a result of the studies, a new leaf pubescence gene introgressed from *Th. ponticum* into bread wheat was identified. We designated it as *H11th*. For the purpose of selection, we propose to use the unlicensed informative microsatellite markers *Xgwm538* and *Xgwm165*, allowing chromosomes 4A, 4B, 4D and 4Th to be distinguished. **Key words:** trichome; digital characteristics of pubescence; phenotypic markers; microsatellite markers; interactions of genes.

For citation: Simonov A.V., Gordeeva E.I., Genaev M.A., Li W., Bulatov I.O., Pshenichnikova T.A. A new leaf pubescence gene, *Hl1th*, introgressed into bread wheat from *Thinopyrum ponticum* and its phenotypic manifestation under homoeologous chromosomal substitutions. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024; 28(6):602-609. DOI 10.18699/vjgb-24-67

Введение

Отдаленная гибридизация широко применяется в селекционных программах для передачи в мягкую пшеницу (Triticum aestivum, AABBDD, 2n = 6x = 42) новых полезных качеств. Для этого используются как близкородственные виды из рода Triticum L. со схожими геномами, такие как Aegilops, так и виды из других родов семейства Poaceae. Декаплоидный вид пырея Thinopyrum ponticum (Podp.) Barkworth & D.R. Dewey (2n = 10x = 70,StStStStEeEeEbEbExEx син. Agropyron elongatum Host., Elvtrigia pontica (Podp.) Holub) принадлежит к третичному генофонду родственников пшеницы и с середины XX в. служит источником полезных генов в селекции пшеницы (Крупин и др., 2019). Имея высокую устойчивость к биотическим и абиотическим факторам стресса, Th. ponticum стал донором эффективных генов устойчивости к различным заболеваниям пшеницы – корневой гнили, листовой и стеблевой ржавчинам, полосатой ржавчине, мучнистой poce (Li H. et al., 2004; Li H., Wang, 2009; Niu et al., 2014; Wang et al., 2019; Li M. et al., 2021; Yang et al., 2023).

В Институте цитологии и генетики СО РАН на протяжении 30 лет пополняется коллекция замещенных, изогенных и аллоплазматических линий мягкой пшеницы на основе ярового сорта Саратовская 29 (С29) и других сортов. Они несут либо отдельные хромосомы, либо определенные перестройки в пшеничных хромосомах, либо цитоплазму родственных видов, приобретенных посредством отдаленной гибридизации. Многие из таких интрогрессий были идентифицированы с использованием цитологических или молекулярных методов (Leonova et al., 2008; Adonina et al., 2021; Shchukina et al., 2022; Першина и др., 2023). Пара хромосом 4Th вида Th. ponticum была перенесена в геном сорта С29 от озимой пшеницы сорта Meropa, созданной в Болгарии (Gordeeva et al., 2019). В результате была получена замещенная линия с голубой антоциановой окраской зерна. Установлено, что ген голубой окраски алейронового слоя Ba (Blue aleurone) расположен в хромосоме 4Th пырея *Th. ponticum* (Zeven, 1991).

С помощью анализа GISH (genome in situ hybridization) было показано, что центромерные и перицентромерные области хромосомы 4Th происходят из хромосомы Е-генома, дистальные области двух его плеч – из хромосомы

St-генома (Zheng et al., 2006). После отбора гибридных растений в поколениях BC_7F_{2-3} по результатам цитологического и молекулярного анализа не было найдено рекомбинаций между пшеничными хромосомами и хромосомой пырея и, следовательно, произошло полное замещение пары хромосом 4B или 4D парой хромосом 4Th (Gordeeva et al., 2022). Помимо голубой окраски зерна, у замещенных линий C29_4Th(4B) и C29_4Th(4D) визуально и тактильно были обнаружены контрастные изменения опушения листовых пластинок в сравнении с реципиентом (Gordeeva et al., 2022).

Известно, что опушение листа является адаптивным признаком (Kaur, Kariyat, 2020). Опушение у риса влияет на транспирацию и засухоустойчивость, тем самым повышая урожайность (Hamaoka et al., 2017). Например, показано положительное влияние данного признака на урожайность растений пшеницы в условиях засухи (Pshenichnikova et al., 2019; Осипова и др., 2020). Опушение листьев злаков характеризуется выростами эпидермальных клеток – трихомами несекретирующего типа, длина и плотность распределения которых сильно варьирует у носителей разных геномов (Pshenichnikova et al., 2017). Так, например, для сортов озимой пшеницы опушение на листьях не характерно (неопубликованные данные), но среди яровых фенотипическое разнообразие по этому признаку может зависеть от региона создания сортов (Генаев и др., 2011).

Встречаемость этого признака у злаков соответствует закону гомологических рядов, сформулированному в 1920 г. Н.И. Вавиловым (1935). Среди видов и родов злаковых растений – ржи, ячменя, риса и других, более отдаленных, видов можно найти образцы с опушением листьев, аналогичным тому, которое имеется у пшеницы (Швачко и др., 2020). У пшеницы сорта С29 в хромосоме 4BL находится основной доминантный ген *H*11, отвечающий за формирование частых трихом средней длины (Майстренко, 1976; Dobrovolskaya et al., 2007). Также известен нелокализованный минорный ген *H*13, формирующий мелкие трихомы и немного усиливающий эффект гена *H*11 (Майстренко, 1976). В диплоидном геноме ячменя (*Hordeum vulgare* L.) ген-регулятор опушения листовой пластинки картирован на длинном плече хромосомы 3H, ген опушения листового влагалища – на длинном плече хромосомы 4H (Saade et al., 2017; Швачко и др., 2020). У синтетических гексаплоидных пшениц опушение листового влагалища и края листа было связано с геномом *A. tauschii* Coss., обнаруженным в длинном плече хромосомы 4D (Dobrovolskaya et al., 2007; Wan et al., 2015).

Настоящая работа направлена на изучение фенотипического проявления нового аллеля гена *Hll* опушения листа, перенесенного с хромосомой 4Th от вида *Th. ponticum* в геном яровой мягкой пшеницы сорта C29. Одновременно была проведена работа по идентификации замещения хромосомной пары 4B или 4D пырейной парой хромосом 4Th с помощью молекулярных маркеров. Целью работы было изучение взаимодействия гена опушения *Hll* с введенным геном пырея, обозначенным нами *Hll*th, при различных замещениях хромосом 4B и 4D путем количественной оценки длины и числа трихом.

Материалы и методы

Растительный материал был представлен яровым сортом-реципиентом мягкой пшеницы C29, а также двумя замещенными линиями этого сорта, C29_4Th(4B) и C29_4Th(4D) (другие ранее использованные синонимы: s:S29_4Th(4B) и s:S29_4Th(4D) соответственно) (Gordeeva et al., 2019, 2022). Согласно цитологическим и молекулярным данным (Gordeeva et al., 2019, 2022), замещенные линии стабильны.

Анализ изображений сгиба листа. Микрофотографии поперечных сгибов с верхней и нижней стороны предфлагового листа были использованы для определения количества и длины трихом по разработанному в ИЦиГ СО РАН протоколу (Дорошков и др., 2009). Изображения получены в ЦКП микроскопического анализа биологических объектов СО РАН ИЦиГ СО РАН на микроскопе Carl Zeiss Axioscop 2 plus через объектив 5х/0.12. Микроскоп оснащен цифровой камерой AxoCam HRc с адаптером TV2/3C 0.63х. Физический размер поля зрения при съемке составлял 2730×2163 мкм, разрешение цифровой фотографии 1300 × 1030 пикселей. Физический размер пикселя составил 2.1 мкм. Изображения для получения цифровых характеристик опушения листа анализировали на вычислительных ресурсах ЦКП «Биоинформатика» при помощи разработанной в лаборатории эволюционной биоинформатики и теоретической генетики ИЦиГ СО РАН программы LHDetect 2 (Genaev et al., 2012). Программа опознает трихомы, определяет их длину и выдает результат в виде последовательности чисел в текстовом файле.

Проанализировано по 12 микрофотографий с шести сгибов с верхней и нижней стороны предфлаговых листьев. Трихомы в пределах каждого класса, формируемые под воздействием разных генов, различаются многократно по длине. Поэтому значения длин были прологарифмированы. Вычисление средней длины представлено как в абсолютных значениях (микроны), так и в виде десятичного логарифма. Также было построено распределение трихом по длине и количеству.

Статистическая обработка. Оценка достоверности различий между генотипами по длине и числу трихом проведена с помощью *t*-критерия Стьюдента, для чего использован MS Excel со статистической надстройкой

АgCStat (Гончар-Зайкин, Чертов, 2003). Критерий значимости различий p < 0.05, 0.01 и 0.001 обозначали одним, двумя и тремя символами соответственно: ^а разница между замещенной линией и реципиентом; ^b разница между двумя замещенными линиями; * разница с верхней и нижней стороны листа в пределах генотипа. Диаграммы распределения трихом по длинам построены в статистическом пакете PAST v. 3.0, для чего использованы данные распознанных трихом суммарно с шести кадров для каждого генотипа.

Генотипирование. ДНК выделяли из молодых листьев по (Plaschke et al., 1995). ДНК образцов диагностировали с помощью ПЦР, применяя микросателлитные маркеры (SSR, simple sequence repeats) для хромосом четвертой гомеологической группы по рекомендованным программам амплификации (Röder et al., 1998). Для этого были выбраны маркеры *Xgwm538* и *Xgwm165*.

Маркер Xgwm538 расположен в длинном плече на хромосоме 4В примерно в 2.1 сМ проксимально от гена Hll у пшеницы (Dobrovolskaya et al., 2007). Он показывал продукты амплификации размером 157 п.н. для генома сорта С29 и 155 п.н. для сорта пшеницы Purple Feed (Dobrovolskaya et al., 2007). Данный маркер у Chinese Spring демонстрировал три фрагмента, 137, 147 и 152 п.н., причем последний продукт соответствует хромосоме 4В, а другие амплифицируются с 4D, что показано на нуллитетрасомных линиях (Brooks et al., 2006). Этот маркер часто применяется, например, для картирования генов устойчивости к инфекциям (Sukhwinder-Singh et al., 2003; Brooks et al., 2006; Singh et al., 2012). В нашей работе использованы классические праймеры маркера Xgwm538 (Röder et al., 1998), показывающие продукты с хромосом 4В и 4D. Этот маркер к тому же проявлял множественный полиморфизм у двух видов пырея и пшенично-пырейных гибридов (Крупин и др., 2011). Маркер Xgwm165 расположен на длинных плечах в хромосомах 4B, 4D и коротком плече хромосомы 4А с перицентрической инверсией (Röder et al., 1998), он часто применяется для картирования разнообразных генов и QTL (Pshenichnikova et al., 2012; Salem, Mattar, 2014; Shchukina et al., 2018).

Для проведения ПЦР задействована готовая смесь реагентов БиоМастер HS-Taq от ООО «Биолабмикс». Продукты ПЦР разделяли электрофорезом в 3.5 % агарозном геле с добавлением бромистого этидия. Для электрофореза использовали буфер ТВЕ (Трис-борат-ЭДТА) и ДНК-маркер длины фрагментов Step50+ («Биолабмикс», Новосибирск, Россия).

Условия выращивания. Растительный материал выращен в гидропонной теплице ЦКП репродукции растений Института цитологии и генетики СО РАН. Условия выращивания: освещение лампами ДНаТ 600 Вт с регулируемой высотой подвески (до 45–50 000 люкс на уровне верхних листьев) длительностью 12–14 ч при температуре 18–20 °С ночью и 24–26 °С днем. Почвенный субстрат – керамзит, увлажняемый питательным раствором Кнопа.

Результаты

Опушение линии с замещением хромосомы 4В на 4Th тактильно отличало ее от реципиента. По результатам микроскопических наблюдений и детального исследова-

ния морфологии опушения листьев (метод описан выше), обе замещенные линии, C29_4Th(4B) и C29_4Th(4D), отличались от сорта-реципиента C29 и друг от друга. На рис. 1 представлены микрофотографии сгибов листьев трех генотипов, на которых визуально можно выделить трихомы, принадлежащие разным ярусам. Цифровая обработка микрофотографий листьев C29 показала, что средняя длина трихом составляла 64.5 мкм на нижней стороне листа и 67.1 мкм – на верхней (см. таблицу). Их максимальная длина не превышала 306 мкм. Несмотря на то что наиболее длинные трихомы формировались на верхней стороне листа, плотность опушения и сумма длин всех трихом с нижней стороны листа были в полтора раза выше, чем с верхней (см. таблицу).

Количество трихом у линии C29_4Th(4B) в сравнении с сортом-реципиентом C29 было снижено в пять раз (см.

таблицу). На верхней стороне наблюдались единичные трихомы длиной до 705.3 мкм, а на нижней – до 539.8 мкм, что в два раза превосходило максимальную длину трихом у сорта С29. С нижней стороны листа у данной линии трихомы встречались чаще, чем с верхней стороны, и сумма всех длин вдвое превышала сумму длин всех трихом сверху. Средние длины трихом на верхней стороне листа у линии и у сорта С29 достоверно не различались, однако с нижней стороны листа разница средних длин трихом была достоверной. Сумма длин всех трихом листа у линии С29_4Th(4B) была значительно снижена по сравнению с С29.

На сгибе листа линии C29_4Th(4B) (см. рис. 1, δ) видно, что мелкие и крупные трихомы различаются по длине многократно. На рис. 2 представлены распределения трихом различной длины в соответствии с их количеством.



Рис. 1. Влияние на фенотип опушения листа замещения хромосом 4В и 4D у сорта C29 на хромосому 4Th. На фотографиях представлены трихомы со сгибов верхней части листа в проходящем свете.

Генотип	C29	C29_4Th(4B)	C29_4Th(4D)
	Верхняя сторона	а листа	
Количество трихом	32.8±2.7	6.8 ± 1.1^{aaa}	39.8±3.13 ^{bbb}
Длина трихом, мкм	67.1±2.7	71.5±16.5	46.2 ± 1.97^{aaa}
Логарифмированная длина трихом	1.72±0.02	1.41 ± 0.06^{aaa}	1.56 ± 0.01^{aaabb}
Пределы длин трихом, мкм	8.1–306.6	8.4–705.3	9.5–426.8
Сумма длин трихом, мкм	2204 ± 240	482 ± 104^{aaa}	1840 ± 140^{bbb}
	Нижняя сторона	листа	
Количество трихом	49.8±2.2	10.0 ± 1.07^{aaa}	52.6±4.5 ^{bbb}
Длина трихом, мкм	64.5±1.56	100.8 ± 13.91^{aa}	54.4 ± 2.0^{aaabbb}
Логарифмированная длина трихом	1.74±0.01	1.59 ± 0.05^{aa}	1.63 ± 0.01^{aaa}
Пределы длин трихом, мкм	11.2–265.3	9.5–539.8	10.4–421.3
Сумма длин трихом, мкм	3215±173	1008 ± 133^{aaa}	2861 ± 239^{bbb}
Достоверность раз	зличий между верхне	ей и нижней сторонами ли	ста
Количество трихом	***	×	*
Длина трихом, мкм	_	_	**
Сумма длин трихом	**	**	**

Морфометрические характеристики опушения листа сорта С29 и замещенных линий с интрогрессией от *Th. ponticum*

Примечание. Значения с надстрочными буквами "а" существенно различаются между замещенными линиями и реципиентом C29; значения с надстрочными буквами "b" существенно различаются у замещенных линий; значения с надстрочными символами "*" существенно различаются между сторонами листа у одного генотипа; число символов соответствует уровням значимости $p < 0.05^{ab*}, p < 0.01^{aabb**}, p < 0.001^{aaabbb***}$.





Рис. 2. Распределение плотности трихом по длинам у разных генотипов.

Шкала по оси *X* логарифмическая. Шкала по оси *Y* отражает количество трихом по классам с шести снимков суммарной шириной примерно с 13 мм (ширина одного снимка составляет 2.163 мм). Данные представлены для нижней поверхности листа.

У линии C29_4Th(4B) (см. рис. 2, красные столбцы) количество трихом с длиной от 30 до 300 мкм многократно снижено, но появился класс трихом с длиной более 300 мкм, отсутствующий у C29. Таким образом, распределение трихом по длине у этой замещенной линии выглядит сглаженным. Разница средних логарифмированных длин у линии C29_4Th(4B) и реципиента оказывается существенной с обеих сторон листа (см. таблицу). Сумма длин трихом у данной замещенной линии также с обеих сторон меньше в 3–4 раза по сравнению с сортом C29.

Иная морфология трихом наблюдалась на микрофотографиях листьев у линии C29_4Th(4D). Плотный полог опушения сорта C29 у нее сохраняется, но дополнительно образовались более длинные трихомы. Тактильно они были малозаметны на общем фоне, но на микрофотографиях выдавались над основной массой трихом, типичных для C29 (см. рис. 1, *в*). Максимальная длина трихом у линии C29_4Th(4D) превышала 400 мкм на обеих сторонах листа, тогда как у C29 она была чуть больше 300 мкм на верхней стороне и больше 200 мкм – на нижней. В сравнении с сортом C29 общее количество трихом увеличилось недостоверно. В то же время разница по данному показателю между линиями C29_4Th(4D) и C29_4Th(4B) была достоверной с обеих сторон листа в пользу первой линии. Сумма длин трихом у линии C29_4Th(4D) в сравнении с C29 уменьшилась незначительно, однако она в 3–4 раза была выше, чем у линии C29_4Th(4B). Тем не менее у линии C29_4Th(4D) наблюдалась достоверно меньшая средняя длина трихом, чем у сорта-реципиента C29. По показателю средней логарифмированной длины трихомы линии C29_4Th(4D) отличались и от C29, и от линии C29_4Th(4B). На рис. 2 распределение трихом у линии C29_4Th(4D) наглядно демонстрирует кратное снижение числа трихом средней длины и наличие класса трихом длиннее, чем у C29.

С геном *Hll* на хромосоме 4В у мягкой пшеницы сорта С29 тесно сцеплен микросателлитный маркер *Хдwm538*, амплифицирующий продукт размером 157 п.н. С использованием нулли-тетрасомных линий сорта Chinese Spring показано, что данный маркер амплифицирует фрагмент размером 174 п.н., специфичный для хромосомы 4В, и два фрагмента, 147 и 137 п.н., – для хромосомы 4D. У С29 детектируется только один продукт меньше 150 п. н., соответствующий хромосоме 4D. Маркер Хдwm538 в нашей работе подтвердил наличие хромосомного замещения в геноме сорта С29 у обеих замещенных линий (рис. 3, *a*). У линии C29 4Th(4B) отсутствовал фрагмент размером >150 п.н., что соответствует диагностическому фрагменту для хромосомы 4В. У линии C29 4Th(4D), напротив, не было фрагмента меньше 150 п.н., который свидетельствует о наличии хромосомы 4D. Таким образом, данный полиморфный маркер Xgwm538 детектирует пшеничные хромосомы 4В и 4D сорта C29 и не амплифицируется на хромосоме 4Th пырея.

Микросателлитный маркер Xgwm165, используемый в нашей работе, амплифицирует фрагменты хромосом четвертой гомеологичной группы 4А, 4В и 4D. На агарозном геле (см. рис. 3, б) после электрофореза в геноме сорта C29 мы детектировали яркие сигналы продуктов амплификации данного маркера размерами ~200, 260 п. н., а также менее выраженный сигнал размером ~ 350 п. н.

Фрагмент ПЦР размером около 200 п.н. наблюдался и в геноме обеих замещенных линий, что соответствует наличию хромосомы 4А. В геноме линии C29_4Th(4B) отсутствовал продукт ПЦР размером 260 п.н., однако



Рис. 3. Электрофореграмма продуктов ПЦР, полученная при амплификации ДНК сорта Саратовская 29 и замещенных линий C29_4Th(4B) и C29_4Th(4D) с помощью микросателлитных маркеров Xgwm538 (a) и Xgwm165 (6).

был обнаружен фрагмент размером 180 п. н. Фрагмент такой же длины (180 п. н.) выявляется в геноме линии C29_4Th(4D) в сочетании с продуктами ПЦР размером 200 и 260 п. н., как у сорта C29, но сигнала размером 350 п. н. не было (см. рис. 3, δ). Результаты ПЦР, полученные с использованием маркера *Xgwm165*, позволяют предположить, что фрагмент 180 п. н. синтезируется с хромосомы 4Th и потому может быть использован в определении данных хромосомных замещений.

Обсуждение

Первым идентифицированным геном опушения листа пшеницы с установленной хромосомной локализацией был ген *Hl1* в хромосоме 4В сорта C29 (Майстренко, 1976). Замещение хромосомы 4В этого сорта на хромосому неопушениого сорта Янецкис Пробат изменяет морфологию опушения. У такого генотипа заметно снижены число трихом и их размер (Doroshkov et al., 2016). Этот фенотип обусловлен наличием гена со слабым эффектом *Hl3*. В отсутствие генов *Hl1* и *Hl3* трихомы на листе изогенной линии C29 практически не формируются, как было показано созданием изогенной линии без опушения (Doroshkov et al., 2016).

Длинные редкие трихомы не характерны для листьев сорта С29. Их появление у двух замещенных линий, очевидно, определяется новым вариантом гена опушения, перенесенного из генома Th. ponticum. Проведенные нами исследования указывают на то, что хромосома 4Th пырея несет этот новый аллельный вариант гена, ортологичный гену пшеницы *Hl1*, но с иным фенотипическим проявлением. Ген пырея, заместивший пшеничный Hll у линии C29 4Th(4B) или дополнивший его у C29 4Th(4D), не только формирует длинные трихомы, но и снижает их общее количество. В соответствии с правилами каталога генных символов (McIntosh et al., 2003) новый аллель был обозначен нами символом Hllth. Ранее у ячменя (H. vulgare L.) в хромосоме 4H в сопоставимой области с хромосомами 4В и 4D был найден ген опушения края листа Hsh (иначе Hs) (Korzun et al., 1999). Локус количественного признака, ассоциированного с опушением края листа, обнаружен в хромосоме 4D с помощью картирующей популяции ITMI (Dobrovolskaya et al., 2007). В настоящей работе мы дополнили гомологический ряд генов опушения для четвертой группы хромосом злаковых растений.

Ранее мы предположили, что ген Hl1 ответственен за число трихом на поверхности листа, т. е. за плотность опушения (Дорошков и др., 2014). В замещенной линии C29_4Th(4B) ген пырея $Hl1^{th}$ в отсутствие пшеничного гена Hl1 стимулировал закладку единичных длинных трихом. При объединении в одном генотипе линии C29_4Th(4D) ген $Hl1^{th}$ оказывает, по-видимому, супрессивное действие по отношению к Hl1, снижая среднюю длину трихом и сумму их длин на листе. Гены, влияющие на формирование длинных редких трихом на поверхности листа у разных сортов, были локализованы и в других хромосомах. Ген Hl2, расположенный в хромосоме 7В пшеницы, был найден у китайского сорта Hong-mangmai (Taketa et al., 2002). Ген $Hl2^{aesp}$ был интрогрессирован из хромосомы 7S A. speltoides Taush. в сорт Родина (Pshenichnikova et al., 2007). У вида *Triticum timopheevii* ген Hl^{tt} с аналогичным фенотипическим проявлением обнаружен в хромосоме 5A (Simonov et al., 2021).

Полученные замещенные линии сорта C29 были созданы для изучения генов регуляции биосинтеза антоцианов. Хромосома 4Th пырея несет ген голубой окраски алейронового слоя зерновки, *Ba* (Gordeeva et al., 2019), что также является фенотипическим маркером присутствия этой хромосомы в геноме. Однако окраска зерна одинаково проявляется при замещении хромосомы как 4B, так и 4D. Фенотипический эффект интрогрессированного опушения может служить маркером для отбора растений при создании голубозерных форм с определенной хромосомной заменой у сортов с опушением, подобным C29.

В нашей работе изучен полиморфизм по микросателлитным маркерам, которые ранее были ассоциированы с хромосомами четвертой группы, и с геном *Hll* в хромосоме 4B в частности (Dobrovolskaya et al., 2007). На рис. 3, *а* представлен маркер *Xgwm538*, рядом с которым был картирован ген *Hll*, у C29 он демонстрировал фрагмент 157 п. н. (Dobrovolskaya et al., 2007). Этот маркер может однозначно показывать, какая из хромосом пшеницы, 4B или 4D, замещена на 4Th от *Th. ponticum*. Хотя было отмечено, что маркер *Xgwm538* демонстрирует специфические для геномов пырея среднего и пырея удлиненного фрагменты (Крупин и др., 2011), они не амплифицировались с перестроенной хромосомы 4Th, состоящей из фрагментов хромосом геномов St и E пырея, перенесенной в геном пшеницы (Zheng et al., 2006).

Поскольку хромосома пырея 4Th не рекомбинирует с пшеничными хромосомами, в работе использован сцепленный с *Hl1* маркер *Xgwm165*, амплифицирующий продукты, специфичные для хромосом 4A, 4B и 4D (Röder et al., 1998). Согласно различным молекулярным картам базы данных GrainGenes (https://graingenes.org/cgi-bin/ GG3/browse.cgi), этот маркер на хромосоме 4В расположен проксимальнее Xgwm538, на расстоянии порядка 20-30 сМ. В геноме ярового сорта пшеницы С29 данный маркер синтезировал разные продукты ПЦР (см. рис. 3, δ) для хромосом четвертой гомеологической группы: 4А около 200 п.н., 4В – около 260 п.н. и слабый сигнал порядка 350 п.н., предположительно, для хромосомы 4D. Также Xgwm165 демонстрировал фрагмент 180 п.н. для хромосомы 4Th. Это позволяет дифференцировать растения пшеницы с разными хромосомными заменами в пределах четвертой гомеологической группы в том случае, если родительским сортам не свойственно опушение С29.

Трихомы формируют особый микроклимат у листьев, способны влиять на стабильность поверхностного воздушного слоя, изменяя ламинарные течения на турбулентные (Schreuder et al., 2001). Турбулентные течения, в свою очередь, способствуют более динамичному газообмену. Соответственно, изменение параметров опушения поверхности должно сказываться на параметрах устьичной проводимости, усвоении углекислого газа и интенсивности испарения влаги. В будущем планируется изучить данные линии по динамике фотосинтетических показателей в различных условиях выращивания, в частности при адаптации к засухе.

Заключение

В нашей работе впервые обнаружен и с помощью цифрового фенотипирования описан новый аллельный вариант гена опушения листа $Hl1^{th}$, перенесенный из декаплоидного вида *Th. ponticum* в мягкую пшеницу. Наблюдаемое фенотипическое проявление нового аллеля гена опушения на фоне пшеничного генома значительно отличалось от действия гена пшеницы Hl1 и по степени выраженности данного признака, скорее, напоминало проявление других генов, Hl^{tt} и $Hl2^{aesp}$, локализованных в хромосомах 5A и 7S родственных злаков *T. timopheevii* и *Ae. speltoides*. Созданные линии позволяют провести сравнение адаптивной ценности схожих признаков опушения листа, контролируемых различными генами в пределах одного модельного генотипа.

Список литературы / References

- Вавилов Н.И. Научные основы селекции пшеницы. М.; Л.: Сельхозгиз, 1935;70-87
- [Vavilov N.I. Scientific Foundations of Wheat Breeding. Moscow– Leningrad: Selkhozgiz Publ., 1935;70-87 (in Russian)]
- Генаев М.А., Дорошков А.В., Морозова Е.В., Пшеничникова Т.А., Афонников Д.А. Компьютерная система WheatPGE для анализа взаимосвязи фенотип–генотип–окружающая среда у пшеницы. *Вавиловский журнал генетики и селекции*. 2011;15(4):784-793 [Genaev M.A., Doroshkov A.V., Morozova E.V., Pshenichnikova T.A., Afonnikov D.A. WheatPGE: A system for analysis of relationships among the phenotype, genotype, and environment in wheat. *Russ. J. Genet. Appl. Res.* 2012;2(3):262-269. DOI 10.1134/ S2079059712030045]
- Гончар-Зайкин П.П., Чертов В.Г. Надстройка к Excel для статистической оценки и анализа результатов полевых и лабораторных опытов. В: Рациональное природопользование и сельскохозяйственное производство в южных регионах Российской Федерации. М.: Современные тетради, 2003;559-565

[Gonchar-Zaykin P.P., Chertov V.G. Nadstroyka k Excel dlya statisticheskoy ocenki i analiza rezul'tatov polevyh i laboratornyh opytov [Elektronnyy resurs]. Available at: URL: http://vniioh.ru/nadstrojkak-excel-dlya-statisticheskoj-ocenki-i-analiza-rezultatov-polevyx-ilaboratornyx-opytov (Accessed 26.09.2021) (in Russian)]

Дорошков А.В., Арсенина С.И., Пшеничникова Т.А., Афонников Д.А. Применение компьютерного анализа микроизображений листа для оценки характеристик опушения пшеницы *Triticum aestivum* L. *Информационный вестник ВОГиС*. 2009;13(1): 218-226

[Doroshkov A.V., Arsenina S.I., Pshenichnikova T.A., Afonnikov D.A. The use of computer-based image processing to leaf hairiness analysis in wheat *Triticum aestivum* L. *Informatsionniy Vestnik VOGiS* = *The Herald of Vavilov Society for Geneticists and Breeding Scientists.* 2009;13(1):218-226 (in Russian)]

Дорошков А.В., Афонников Д.А., Пшеничникова Т.А. Генетический анализ опушения листа у изогенных линий мягкой пшеницы сорта Новосибирская 67. *Генетика*. 2014;50(2):172-180. DOI 10.7868/S0016675813120023

[Doroshkov A.V., Afonnikov D.A., Pshenichnikova T.A. Genetic analysis of leaf pubescence in isogenic lines of bread wheat Novosibirskaya 67. *Russ. J. Genet.* 2014;50:153-160. DOI 10.1134/S1022 795413120028]

Крупин П.Ю., Дивашук М.Г., Фесенко И.А., Карлов Г.И. Адаптация микросателлитных SSR-маркеров пшеницы для анализа геномов пырея среднего, пырея удлиненного и пшенично-пырейных гибридов. *Изв. TCXA*. 2011;3:49-57

[Kroupin P.Yu., Divashuk M.G., Fesenko I.A., Karlov G.I. Adaptation of microsatellite SSR-markers of wheat for the genome analysis of wheatgrass, intermediate wheatgrass, and wheat-wheatgrass hybrids. Izvestiya Timiryazevskoy Sel'skohozjajstvennoy Akademii = Izvestiya of Timiryazev Agricultural Academy. 2011;3:49-57 (in Russian)]

Крупин П.Ю., Дивашук М.Г., Карлов Г.И. Использование генетического потенциала многолетних дикорастущих злаков в селекционном улучшении пшеницы. *С.-х. биология*. 2019;54(3):409-425 [Kroupin P.Yu., Divashuk M.G., Karlov G.I. Gene resources of pe-

rennial wild cereals involved in breeding to improve wheat crop. Sel'skokhozyaistvennaya Biologiya = Agricultural Biology. 2019; 54(3):409-425. DOI 10.15389/agrobiology.2019.3.409eng]

Майстренко О.И. Идентификация и локализация генов опушения листа у молодых растений мягкой пшеницы. *Генетика*. 1976; 12(1):5-15

[Maistrenko O.I. Identification and localization of genes controlling the pubescence of the leaf of young soft wheat plants. *Genetika* (*Moscow*). 1976;12(1):5-15 (in Russian)]

Осипова С.В., Рудиковский А.В., Пермяков А.В., Рудиковская Е.Г., Пермякова М.Д., Верхотуров В.В., Пшеничникова Т.А. Физиологические реакции линий пшеницы (*Triticum aestivum* L.) с генетически различным опушением листа на водный дефицит. *Вавиловский журнал генетики и селекции*. 2020;24(8):813-820. DOI 10.18699/VJ20.678

[Osipova S.V., Rudikovskii A.V., Permyakov A.V., Rudikovskaya E.G., Permyakova M.D., Verkhoturov V.V., Pshenichnikova T.A. Physiological responses of wheat (*Triticum aestivum* L.) lines with genetically different leaf pubescence. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):813-820. DOI 10.18699/VJ20.678]

Першина Л.А., Трубачеева Н.В., Шумный В.К., Бадаева Е.Д. Получение и изучение линии с замещением хромосомы 4В пшеницы *Triticum aestivum* L. на хромосому 4Н^{mar} дикого ячменя *Hordeum marinum* ssp. gussoneanum (4x). Вавиловский журнал генетики и селекции. 2023;27(6):545-552. DOI 10.18699/VJGB-23-66

[Pershina L.A., Trubacheeva N.V., Shumny V.K., Badaeva E.D. Development and characterization of a line with substitution of chromosome 4B of wheat *Triticum aestivum* L. on chromosome 4H^{mar} of wild barley *Hordeum marinum* ssp. gussoneanum (4x). Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2023;27(6):545-552. DOI 10.18699/VJGB-23-66]

Швачко Н.А., Семилет Т.В., Тихонова Н.Г. Трихомы у высших растений: гомологические ряды в наследственной изменчивости и молекулярно-генетические механизмы. *Генетика*. 2020;56(11): 1320-1332. DOI 10.31857/S0016675820110089

[Shvachko N.A., Semilet T.V., Tikhonova N.G. Trichomes in higher plants: homological series in hereditary variability and molecular genetic mechanisms. *Russ. J. Genet.* 2020;56(11):1359-1370. DOI 10.1134/S1022795420110083]

Adonina I.G., Timonova E.M., Salina E.A. Introgressive hybridization of common wheat: results and prospects. *Russ. J. Genet.* 2021; 57(4):390-407. DOI 10.1134/S1022795421030029

- Brooks S.A., See D., Brown-Guedira G. SNP-based improvement of a microsatellite marker associated with Karnal bunt resistance in wheat. *Crop Sci.* 2006;46(4):1467-1470. DOI 10.2135/cropsci2005. 05-0065
- Dobrovolskaya O.B., Pshenichnikova T.A., Arbuzova V.S., Lohwasser U., Röder M.S., Börner A. Molecular mapping of genes determining hairy leaf character in common wheat with respect to other species of the Triticeae. *Euphytica*. 2007;155:285-293. DOI 10.1007/s10681-006-9329-7
- Doroshkov A.V., Afonnikov D.A., Dobrovolskaya O.B., Pshenichnikova T.A. Interactions between leaf pubescence genes in bread wheat as assessed by high throughput phenotyping. *Euphytica*. 2016;207: 491-500. DOI 10.1007/s10681-015-1520-2
- Hamaoka N., Yasui H., Yamagata Y., Inoue Y., Furuya N., Araki T., Ueno O., Yoshimura A. A hairy-leaf gene, BLANKET LEAF, of wild *Oryza nivara* increases photosynthetic water use efficiency in rice. *Rice*. 2017;10(1):20. DOI 10.1186/s12284-017-0158-1

- Genaev M.A., Doroshkov A.V., Pshenichnikova T.A., Kolchanov N.A., Afonnikov D.A. Extracting quantitative characteristics of wheat leaf hairiness using image processing technique. *Planta*. 2012;236: 1943-1954. DOI 10.1007/s00425-012-1751-6
- Gordeeva E., Badaeva E., Yudina R., Shchukina L., Shoeva O., Khlestkina E. Marker-assisted development of a blue-grained substitution line carrying the *Thinopyrum ponticum* chromosome 4Th(4D) in the spring bread wheat Saratovskaya 29 background. *Agronomy*. 2019;9:723. DOI 10.3390/agronomy9110723
- Gordeeva E., Shoeva O., Mursalimov S., Adonina I., Khlestkina E. Fine points of marker-assisted pyramiding of anthocyanin biosynthesis regulatory genes for the creation of black-grained bread wheat (*Triticum aestivum* L.) lines. *Agronomy*. 2022;12:2934. DOI 10.3390/ agronomy12122934
- Kaur J., Kariyat R. Role of trichomes in plant stress biology. In: Núñez-Farfán J., Valverde P. (Eds.). Evolutionary Ecology of Plant-Herbivore Interaction. Springer, 2020;15-35. DOI 10.1007/978-3-030-46012-9 2
- Korzun V., Malyshev S., Pickering R.A., Börner A. RFLP mapping of a gene for hairy leaf sheath using a recombinant line from *Hordeum* vulgare L.×*Hordeum bulbosum* L. cross. *Genome*. 1999;42(5):960-963. DOI 10.1139/g99-021
- Leonova I.N., Röder M.S., Kalinina N.P., Budashkina E.B. Genetic analysis and localization of loci controlling leaf rust resistance of *Triticum aestivum* × *Triticum timopheevii* introgression lines. *Russ.* J. Genet. 2008;44:1431-1437. DOI 10.1134/S1022795408120077
- Li H., Wang X. Thinopyrum ponticum and Th. intermedium: the promising source of resistance to fungal and viral diseases of wheat. J. Genet. Genomics. 2009;36(9):557-565. DOI 10.1016/S1673-8527 (08)60147-2
- Li H., Conner R.L., Chen Q., Li H., Laroche A., Graf R.J., Kuzyk A.D. The transfer and characterization of resistance to common root rot from *Thinopyrum ponticum* to wheat. *Genome*. 2004;47(1):215-223. DOI 10.1139/g03-095
- Li M., Wang Y., Liu X., Li X., Wang H., Bao Y. Molecular cytogenetic identification of a novel wheat *Thinopyrum ponticum* 1J^S (1B) substitution line resistant to powdery mildew and leaf rust. *Front. Plant Sci.* 2021;12:727734. DOI 10.3389/fpls.2021.727734
- McIntosh R.A., Devos K.M., Dubcovsky J., Morris C.F., Rogers W.J. Catalogue of Gene Symbols for Wheat. Supplement. 2003. Available at: https://wheat.pw.usda.gov/ggpages/wgc/2003upd.html
- Niu Z., Klindworth D.L., Yu G., Friesen T.L., Chao S., Jin Y., Cai X., Ohm J.-B., Rasmussen J.B., Xu S.S. Development and characterization of wheat lines carrying stem rust resistance gene *Sr43* derived from *Thinopyrum ponticum*. *Theor. Appl. Genet.* 2014;127(4):969-980. DOI 10.1007/s00122-014-2272-4
- Plaschke J., Ganal M.W., Röder M.S. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.* 1995;91(6-7):1001-1007. DOI 10.1007/BF00223912
- Pshenichnikova T.A., Lapochkina I.F., Shchukina L.V. The inheritance of morphological and biochemical traits introgressed into common wheat (*Triticum aestivum* L.) from *Aegilops speltoides* Tausch. *Genet. Resour. Crop Evol.* 2007;54(2):287-293. DOI 10.1007/ s10722-005-4499-z
- Pshenichnikova T.A., Khlestkina E.K., Shchukina L.V., Simonov A.V., Chistyakova A.K., Morozova E.V., Landjeva S., Karceva T., Börner A. Exploitation of Saratovskaya 29 (Janetzkis Probat 4D*7A) substitution and derivate lines for comprehensive phenotyping and molecular mapping of quantitative trait loci (QTL). In: EWAC Newsletter 2012, Proc. of the 15th International EWAC Conference, 7–11 November 2011, Novi Sad, Serbia. 2012;19-22
- Pshenichnikova T.A., Doroshkov A.V., Simonov A.V., Afonnikov D.A., Börner A. Diversity of leaf pubescence in bread wheat and relative species. *Genet. Resour. Crop Evol.* 2017;64(7):1761-1773. DOI 10.1007/s10722-016-0471-3

- Pshenichnikova T.A., Doroshkov A.V., Osipova S.V., Permyakov A.V., Permyakova M.D., Efimov V.M., Afonnikov D.A. Quantitative characteristics of pubescence in wheat (*Triticum aestivum* L.) are associated with photosynthetic parameters under conditions of normal and limited water supply. *Planta*. 2019;249(3):839-847. DOI 10.1007/s00425-018-3049-9
- Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P., Ganal M.W. A microsatellite map of wheat. *Genetics*. 1998;149(4):2007-2023. DOI 10.1093/genetics/149.4.2007
- Saade S., Kutlu B., Draba V., Förster K., Schumann E., Tester M., Pillen K., Maurer A. A donor-specific QTL, exhibiting allelic variation for leaf sheath hairiness in a nested association mapping population, is located on barley chromosome 4H. *PloS One*. 2017;12(12): e0189446. DOI 10.1371/journal.pone.0189446
- Salem K.F.M., Mattar M.Z. Identification of microsatellite alleles for salt tolerance at seedling stage in wheat (*Triticum aestivum* L.). *Life Sci. J.* 2014;11(12s):1064-1073
- Schreuder M.D.J., Brewer C.A., Heine C. Modelled influences of nonexchanging trichomes on leaf boundary layers and gas exchange. *J. Theor. Biol.* 2001;210:23-32. DOI 10.1006/jtbi.2001.2285
- Shchukina L.V., Pshenichnikova T.A., Khlestkina E.K., Misheva S., Kartseva T., Abugalieva A., Börner A. Chromosomal location and mapping of quantitative trait locus determining technological parameters of grain and flour in strong-flour bread wheat cultivar Saratovskaya 29. Cereal Res. Commun. 2018;46(4):628-638. DOI 10.1556/0806.46.2018.047
- Shchukina L.V., Simonov A.V., Demenkova M.A., Klykov A.G., Shamanin V.P., Pozherukova V.E., Lepekhov S.B., Chebatareva M.V., Petin V.A., Börner A., Pshenichnikova T.A. Increased grain protein and gluten contents of bread wheat caused by introgression of a *T. timopheevii* segment into chromosome 2A. *Euphytica*. 2022;218: 170. DOI.10.1007/s10681-022-03121-w
- Simonov A.V., Smirnova O.G., Genaev M.A., Pshenichnikova T.A. The identification of a new gene for leaf pubescence introgressed into bread wheat from *Triticum timopheevii* Zhuk. and its manifestation in a different genotypic background. *Plant Genet. Resour.* 2021; 19(3):238-244. DOI 10.1017/S1479262121000277
- Singh A., Pallavi J.K., Gupta P., Prabhu K.V. Identification of microsatellite markers linked to leaf rust resistance gene *Lr25* in wheat. *J. Appl. Genet.* 2012;53(1):19-25. DOI 10.1007/s13353-011-0070-0
- Sukhwinder-Singh, Brown-Guedira G.L., Grewal T.S., Dhaliwal H.S., Nelson J.C., Singh H., Gill B.S. Mapping of a resistance gene effective against Karnal bunt pathogen of wheat. *Theor. Appl. Genet.* 2003;106(2):287-292. DOI 10.1007/s00122-002-1112-0
- Taketa S., Chang C.L., Ishii M., Takeda K. Chromosome arm location of the gene controlling leaf pubescence of a Chinese local wheat cultivar Hong-mang-mai. *Euphytica*. 2002;125(2):141-147. DOI 10.1023/A:1015812907111
- Wan H., Yang Y., Li J., Zhang Z., Yang W. Mapping a major QTL for hairy leaf sheath introgressed from *Aegilops tauschii* and its association with enhanced grain yield in bread wheat. *Euphytica*. 2015; 205:275-285. DOI 10.1007/s10681-015-1457-5
- Wang S., Wang C., Wang Y., Wang Y., Chen C., Ji W. Molecular cytogenetic identification of two wheat – *Thinopyrum ponticum* substitution lines conferring stripe rust resistance. *Mol. Breed.* 2019;39:143. DOI 10.1007/s11032-019-1053-9
- Yang G., Deng P., Ji W., Fu S., Li H., Li B., Li Zh., Zheng Q. Physical mapping of a new powdery mildew resistance locus from *Thinopyrum ponticum* chromosome 4AgS. *Front. Plant Sci.* 2023;14:1131205. DOI 10.3389/fpls.2023.1131205
- Zeven A.C. Wheats with purple and blue grains: a review. *Euphytica*. 1991;56:243-258
- Zheng Q., Li B., Mu S., Zhou H., Li Z. Physical mapping of the bluegrained gene(s) from *Thinopyrum ponticum* by GISH and FISH in a set of translocation lines with different seed colors in wheat. *Genome*. 2006;49(9):1109-1114. DOI 10.1139/g06-073

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 03.12.2023. После доработки 26.06.2024. Принята к публикации 26.06.2024.

DOI 10.18699/vjgb-24-68

Влияние хромосом 1А и 1D *T. aestivum* на фертильность аллоплазматических рекомбинантных линий (*H. vulgare*)-*T. aestivum* в зависимости от цитоядерной совместимости

Л.А. Першина (^{D¹, 2} ⊠, Н.В. Трубачеева (^{D¹, 2}, В.К. Шумный (^{D¹}

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия ² Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

🖾 pershina@bionet.nsc.ru

Аннотация. Изучено влияние хромосом 1А и 1D T. aestivum L. на фертильность рекомбинантных аллолиний мягкой пшеницы одного происхождения, имеющих цитоплазму ячменя H. vulgare L. и разный уровень цитоядерной совместимости. Аллолиния Л-56 включает преимущественно полностью стерильные (ПС) и частично стерильные (ЧС) растения; аллолиния Л-57 – частично фертильные (ЧФ) растения, а линия Л-58 – фертильные (Ф) растения. Результаты анализа морфобиологических признаков и окраски пыльцы указывают на проявление полной или частичной мужской стерильности у растений аллолиний Л-56 и Л-57. Для разделения генотипов с цитоядерной коадаптацией и генотипов, у которых цитоядерная совместимость нарушена, выполнен ПЦР-анализ 18S/5S митохондриального (мт) повтора. Показано, что ПС, ЧС, ЧФ и часть Ф растений характеризуются гетероплазмией (наличием копий мтДНК ячменя и пшеницы), что ассоциировано с нарушением цитоядерной совместимости. У основной части фертильных растений выявлена гомоплазмия (гм) пшеничного типа, что ассоциировано с цитоядерной коадаптацией. Растения аллолиний, использованные в качестве материнских генотипов, были скрещены с пшенично-ржаными замещенными линиями 1R(1A) и 1R(1D). В F_1 все растения комбинаций ЧФ×1R(1A) и ЧФ×1R(1D) были фертильными, а в F_2 наблюдали расщепление, близкое к 3 (фертильные) : 1 (стерильные). Эти результаты впервые показали, что в хромосомах 1А и 1D локализовано по одному доминантному гену Rf, контролирующему восстановление мужской фертильности мягкой пшеницы, несущей цитоплазму *H. vulgare*. Все растения F₁ комбинаций ПС×1R(1A), ПС×1R(1D), ЧС×1R(1A), ЧС×1R(1D) стерильные, что указывает на то, что одной дозы генов, локализованных в хромосомах пшеницы 1А или 1D, недостаточно для восстановления мужской фертильности у ПС и ЧС растений. Все растения гибридных комбинаций $\Phi_{rm} \times 1R(1A)$ и $\Phi_{rm} \times 1R(1D)$ и в F₁ и в F₂ были фертильными, т.е. у аллолиний с цитоядерной коадаптацией нет зависимости проявления фертильности от влияния хромосом пшеницы 1А и 1D. Ключевые слова: аллолинии (H. vulgare)-T. aestivum; хромосомы 1А и 1D; мтДНК; нарушение цитоядерной совместимости; цитоядерная коадаптация; гены Rf.

Для цитирования: Першина Л.А., Трубачеева Н.В., Шумный В.К. Влияние хромосом 1А и 1D *T. aestivum* на фертильность аллоплазматических рекомбинантных линий (*H. vulgare*)-*T. aestivum* в зависимости от цитоядерной совместимости. *Вавиловский журнал генетики и селекции*. 2024;28(6):610-618. DOI 10.18699/vjgb-24-68

Финансирование. Исследование выполнено в рамках бюджетного проекта FWNR-2022-0017.

Благодарности. Авторы благодарны д.б.н. А.И. Щаповой за предоставленные для работы пшенично-ржаные замещенные линии.

The effect of *T. aestivum* chromosomes 1A and 1D on fertility of alloplasmic recombinant (*H. vulgare*)-*T. aestivum* lines depending on cytonuclear compatibility

L.A. Pershina (D^{1, 2} , N.V. Trubacheeva (D^{1, 2}, V.K. Shumny (D¹

¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia ² Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

🖾 pershina@bionet.nsc.ru

Abstract. The effect of *T. aestivum* L. chromosomes 1A and 1D on fertility of recombinant bread wheat allolines of the same origin carrying the cytoplasm of barley *H. vulgare* L. and different levels of cytonuclear compatibility was studied. Alloline L-56 included mainly fully sterile (FS) and partially sterile (PS) plants, alloline L-57 included partially fertile (PF) plants and line L-58 included fertile (F) ones. Analysis of morphobiological traits and pollen paint-

2024 28•6

ing indicated complete or partial male sterility in plants of allolines L-56 and L-57. To differentiate genotypes with cytonuclear coadaptation and genotypes with cytonuclear incompatibility, PCR analysis of the 185/55 mitochondrial (mt) repeat was performed. Heteroplasmy (simultaneous presence of barley and wheat mtDNA copies) was found in FS, PS, PF and some F plants, which was associated with a violation of cytonuclear compatibility. Wheat-type homoplasmy (hm) was detected in the majority of the fertile plants, which was associated with cytonuclear coadaptation. The allolines used as maternal genotypes were crossed with wheat-rye substitution lines 1R(1A) and 1R(1D). In F₁, all plants of PF×1R(1A) and PF×1R(1D) combinations were fertile, and in F₂, a segregation close to 3 (fertile) : 1 (sterile) was observed. These results showed for the first time that chromosomes 1A and 1D carry one dominant *Rf* gene, which controls the restoration of male fertility of bread wheat carrying the cytoplasm of *H. vulgare*. All plants of F₁ combinations FS×1R(1A), FS×1R(1D), PS×1R(1A), PS×1R(1D) were sterile, which indicates that a single dose of genes localized on wheat chromosomes 1A or 1D is not enough to restore male fertility in FS and PS plants. All plants of hybrid combinations F(hm)×1R(1A) and F(hm)×1R(1D) in both F₁ and F₂ were fertile, that is, fertility of allolines with cytonuclear coadaptation does not depend on wheat chromosomes 1A and 1D. **Key words:** allolines (*H. vulgare*)-*T. aestivum*; chromosomes 1A and 1D; mtDNA; violation of cytonuclear compatibility; cytonuclear coadaptation; *Rf* genes.

For citation: Pershina L.A., Trubacheeva N.V., Shumny V.K. The effect of *T. aestivum* chromosomes 1A and 1D on fertility of alloplasmic recombinant (*H. vulgare*)-*T. aestivum* lines depending on cytonuclear compatibility. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(6):610-618. DOI 10.18699/vjgb-24-68

Введение

Аллоплазматические линии (аллолинии) образуются в результате повторяющихся скрещиваний отдаленных гибридов F₁ с пыльцевым родителем и сочетают цитоплазму материнского вида с ядерным геномом отцовского генотипа (Tsunewaki, 1996). Замещение цитоплазмы приводит к нарушениям регуляции ядерно-митохондриальных и ядерно-хлоропластных взаимодействий (Yang et al., 2008; Crosatti et al., 2013; Soltani et al., 2016), что может вызывать изменения в развитии растений (Бадаева и др., 2006), их устойчивости к стрессовым факторам (Булойчик и др., 2002; Talukder et al., 2015; Takenaka et al., 2019), в проявлении морфологических и агрономических признаков (Liu C.G. et al., 2002; Atienza et al., 2008; Tao et al., 2011; Климушина и др., 2013). Наиболее характерным выражением цитоядерного конфликта, в том числе у аллолиний, является цитоплазматическая мужская стерильность (ЦМС) (Tsunewaki, 1996), связанная с мутациями генов митохондриального генома, оказывающих негативное влияние на развитие органов цветка и пыльцы (Yang et al., 2008).

У ряда экономически важных культур ЦМС-линии в сочетании с линиями, поддерживающими ЦМС, и линиями-носителями гена (генов) восстановления мужской фертильности (Rf, restorer-of-fertility) используют в гибридной селекции (Islam et al., 2014; Bohra et al., 2016; Gupta et al., 2019). В этой технологии ключевая роль принадлежит источникам ЦМС и генов Rf. Кроме того, замещение цитоплазмы способствует увеличению цитоплазматического разнообразия растений, что считают возможным и необходимым для таких культур, как рис (Liu Y. et al., 2016), сахарный тростник (Rafee et al., 2010), мягкая пшеница (Liu C.G. et al., 2002; Климушина и др., 2013; Першина и др., 2018).

В связи с этим изучение особенностей образования аллолиний и генетического контроля восстановления их фертильности считается важной задачей как для поиска новых генетических систем ЦМС-*Rf* в гибридной селекции, так и для получения новых генотипов для традиционных селекционных программ. У мягкой пшеницы внимание уделено восстановлению мужской фертильности генотипов, несущих цитоплазму *T. timopheevii* (Sinha et al., 2013), *H. chilense* (Martin et al., 2010), видов *Aegilops* (Tsunewaki, 2015; Hohn, Lukaszewski, 2016), а в наших работах – генотипов с цитоплазмой культурного ячменя *H. vulgare* (Першина и др., 2012; Trubacheeva et al., 2021). Основная часть генов *Rf* у пшеницы организована в кластеры на хромосомах 1, 2 и 6-й гомеологичных групп, а самой насыщенной является хромосома 1 (Gupta et al., 2019).

В предыдущем исследовании мы впервые установили, что в коротком плече хромосомы 1В мягкой пшеницы локализован доминантный ген, контролирующий восстановление мужской фертильности пшеницы на цитоплазме H. vulgare (Trubacheeva et al., 2021). В настоящей работе мы продолжили исследовать роль первой гомеологичной группы в восстановлении фертильности у аллолиний мягкой пшеницы с цитоплазмой *H. vulgare*. Была поставлена задача изучить влияние хромосом 1A и 1D T. aestivum на проявление мужской фертильности рекомбинантных аллолиний мягкой пшеницы, носителей цитоплазмы культурного ячменя, в зависимости от уровня их фертильности и цитоядерной совместимости. Такой подход дал возможность выявить аллолинии (H. vulgare)-T. aestivum – модели для определения локализации генов Rf в хромосомах 1А и 1D.

Материалы и методы

Растительный материал и особенности его получения. Изучены три рекомбинантные аллолинии (*H. vulgare*)-*T. aestivum*, сформированные от отдельных растений беккроссных (BC) поколений ячменно-пшеничного гибрида *H. vulgare* (Неполегающий) × *T. aestivum* (Саратовская 29), последовательно опыленного сортами пшеницы Саратовская 29, Мироновская 808, Пиротрикс 28, Саратовская 29, Пиротрикс 28 (рис. 1). В предыдущих работах было установлено, что Саратовская 29 является закрепителем стерильности у беккроссных потомков ячменно-пшеничных гибридов (Першина и др., 2012), а Мироновская 808 и Пиротрикс 28 – восстановителями мужской фертильности аллолиний пшеницы с цитоплазмой культурного ячменя (Pershina et al., 1998; Першина и др., 2012). ВС₁–ВС₄-поколения, как и ячменно-пше

H. vulgare (2*n* = 14) (Неполегающий) × *T. aestivum* (2*n* = 42) (Саратовская 29)



Рис. 1. Схема получения аллоплазматических рекомбинантных линий (*H. vulgare*)-*T. aestivum* Л-56, Л-57, Л-58.

ничный гибрид, характеризовались мужской стерильностью, но женской фертильностью, а в поколении BC₅ были выделены отдельные 42-хромосомные растения с частично восстановленной мужской фертильностью. От этих растений получены самоопыленные поколения F_2BC_5 - F_5BC_5 , источники изученных аллолиний. Аллолиния Л-56 выделена из F_2BC_5 поколения, а аллолинии Л-57 и Л-58 – из F_4BC_5 и F_5BC_5 поколений соответственно. Начиная с F_3BC_5 для формирования каждого следующего самоопыленного поколения использовали растения с максимальной продуктивностью.

Методы изучения морфобиологических характеристик аллоплазматических рекомбинантных линий. Растения линий охарактеризованы по уровню фертильности: ПС – полностью стерильные (нет зерен); ЧС – частично стерильные (1–9 зерен); ЧФ – частично фертильные (10–19); Ф – фертильные (более 19 зерен в главном колосе). Оценивали не менее 20 растений каждой линии, выращенных в гидропонной теплице.

У растений с разным уровнем фертильности проанализирована фертильность пыльцы, которая используется в качестве основного критерия оценки мужской фертильности/стерильности. Для этой цели из пыльников, вычлененных во время цветения из трех разных цветков одного колоса, на предметном стекле готовили давленые препараты в растворе Люголя (1% раствор йода в водном растворе йодида калия). У растений каждой аллолинии определяли высоту, число колосьев, длину главного колоса, число колосков в главном колосе, число зерен в главном колосе и растении, массу 1000 зерен. Различия между средними значениями изученных признаков у аллолинии Л-56 по сравнению с линией Л-57, а у аллолинии Л-57 по сравнению с аллолинией Л-58 оценивали с использованием *t*-критерия Стьюдента. Данные анализировали с помощью Statistica v.7.0.61.0.

ПЦР-анализ 18S/5S митохондриального повтора. В качестве ПЦР-маркеров митохондриального (мт) генома были использованы праймеры к 18S/5S повтору. Их структура и условия ПЦР описаны в работе (Coulthart et al., 1993). Разделение продуктов ПЦР проводилось в 1.5 % агарозном геле, содержащем бромистый этидий; визуализацию осуществляли с помощью системы гельдокументирования Gel Doc XR+ (Bio-Rad, CША). Тотальная ДНК была выделена из зеленых листьев, срезанных до колошения, согласно ранее опубликованному протоколу (Current Protocols..., 1987). Анализировали от одного до восьми образцов от отдельных генотипов. В этой части работы контролем при изучении рекомбинантных аллолиний служили линия ячменя сорта Неполегающий, источник цитоплазмы аллолиний, и сорт мягкой пшеницы Пиротрикс 28, источник цитоплазмы пшеницы (один из рекуррентных генотипов).

Оценка фертильности гибридов между аллоплазматическими линиями и пшенично-ржаными замещенными линиями 1A(1R) и 1D(1R) в F_1 и F_2 . Для оценки влияния хромосом пшеницы 1А и 1D на фертильность аллолиний (H. vulgare)-Т. aestivum в зависимости от уровня их цитоядерной совместимости растения этих линий как материнские генотипы скрещивали с пшенично-ржаными замещенными линиями 1A(1R) и 1D(1R), чтобы заместить в F₁ по одной хромосоме 1А и 1D аллолиний на хромосому ржи 1R. Использованные в работе линии 1A(1R) и 1D(1R) получены в результате замещения соответствующих хромосом пшеницы сорта Саратовская 29 на хромосому ржи 1R сорта Онохойская (Shchapova, Kravtzova, 1982). В гибридизацию были включены ПС и ЧС растения аллолинии Л-56, ЧФ растения аллолинии Л-57 и отдельные Ф растения аллолинии Л-58. До цветения колосья материнских растений, а также растений F₁ и F₂, выращиваемых в гидропонной теплице, помещали под пергаментные изоляторы. У индивидуальных растений F₁ и F₂ оценивали завязываемость семян в главном колосе. Разделение на стерильные и фертильные растения в F₂ проводили согласно рекомендациям (Sinha et al., 2013): полностью стерильные растения и растения, завязавшие не более четырех зерен в главном колосе, относили к стерильным; растения, завязавшие в главном колосе от пяти зерен и более, - к фертильным. Для оценки соответствия между наблюдаемым и теоретически ожидаемым расщеплением в F₂ на фертильные и стерильные растения использовали критерий Пирсона (Хи-квадрат) (α = 0.05).

Результаты

Характеристика рекомбинантных аллолиний

Аллолиния Л-56 представлена в основном частично стерильными (60 %) и полностью стерильными растениями (35 %); частота частично фертильных растений составила 5 % (табл. 1).

Большинство растений у аллолинии Л-57 – частично фертильные (85 %), остальные – частично стерильные (5 %) и фертильные (10 %). Аллолинию Л-58 составили фертильные (92 %) и частично фертильные растения (8 %). Образцы колосьев растений с разным уровнем фертильности представлены на рис. 2.

У полностью стерильных растений рыльца нормально развиты, но пыльники отсутствуют. У частично стерильных и частично фертильных растений пыльники по срав-

2024

28•6

Аллолиния	Количество изученных растений	Число и частота (%) растений			
		ПС (0)	ЧС (1–9) [#]	ЧФ (10–19) [#]	Ф (>19) [#]
Л-56	20	7 (35 %)	12 (60 %)	1 (5 %)	0
Л-57	20	0	1 (5 %)	17 (85 %)	2 (10 %)
Л-58	25	0	0	2 (8 %)	23 (92 %)

Таблица	1. Уровень	фертильности растений	рекомбинантных аллолиний (Л	H. vulgare)-T. aestivum Л-56, Л-57, Л-58
---------	------------	-----------------------	-----------------------------	--

Примечание. ПС – полная стерильность; ЧС – частичная стерильность; ЧФ – частичная фертильность; Ф – фертильность. # – число зерен в главном колосе.

нению с фертильными растениями недостаточно развиты и не все пыльцевые зерна окрашены (рис. 3).

Сравнение средних значений показателей изученных признаков у аллолинии Л-56, представленной стерильными и частично стерильными растениями, по сравнению с Л-57, состоящей в основном из частично фертильных растений, показало, что аллолиния Л-56 превосходит Л-57 только по числу колосьев на растение. Значение остальных показателей признаков (высота растений, длина главного колоса, число колосков в главном колосе, число зерен в главном колосе и в растении) у Л-56 достоверно ниже, чем у Л-57 (табл. 2).

У аллолинии Л-57 длина главного колоса, число колосков в главном колосе, число зерен в главном колосе и число зерен в растении достоверно ниже по сравнению с аллолинией Л-58, представленной преимущественно фертильными растениями. По массе 1000 зерен различий между изученными аллолиниями не обнаружено.



Рис. 2. Колосья растений: 1, 2 – фертильных; 3 – частично фертильного; 4 – частично стерильного; 5 – полностью стерильного.



Рис. 3. Рыльце (1) полностью стерильного растения; рыльце и пыльники (2), пыльцевые зерна (3) частично фертильного растения; рыльце и пыльники (4) и пыльцевые зерна (5) фертильного растения.

Таблица 2. Х	арактеристика	рекомбинантных алл	толиний (<i>H. vula</i>	are)-T. aestivum по	о морфобиологическ	ким признакам
				,,		

Признак растений	Л-56	Л-57	Л-58
Высота растений, см	76.38±2.66 ^(**)	87.81±1.97	89.55±2.0
Число колосьев в растении	4.78±0.32*	3.92±0.24	3.85±0.22
Длина главного колоса, см	$6.07 \pm 0.41^{(***)}$	8.12±0.34′*′	9.14±0.23
Число колосков в главном колосе	$13.21 \pm 0.67^{(***)}$	16.50±0.25'*'	18.15±0.65
Число зерен в главном колосе	$3.76 \pm 1.98^{(***)}$	16.35±1.66′***′	33.10±2.56
Число зерен в растении	15.57±6.35 ^(***)	52.50±7.84′***′	115.74±13.67
Масса 1000 зерен, г	35.32±1.18	34.25±1.12	35.67±1.31

Примечание. Разница по сравнению с Л-57 достоверно больше при * p < 0.05; достоверно меньше при ^(**) p < 0.01 и ^(***) p < 0.001; по сравнению с Л-58 – достоверно меньше при ^(**) p < 0.05 и ^(***) p < 0.001.
Результаты анализа последовательности 18S/5S мтДНК у рекомбинантных аллолиний

У всех изученных образцов аллолинии Л-56, включая полностью стерильные, частично стерильные растения и одно частично фертильное, обнаружена гетероплазмия (наличие копий 18S/5S мтДНК ячменя и пшеницы) (рис. 4, табл. 3). У аллолинии Л-57 у образцов шести изученных частично фертильных растений и двух растений, отнесенных к группе фертильных, также выявлена гетероплазмия. У аллолинии Л-58 образцы двух частично фертильных и двух фертильных растений показали гетероплазмию, а шести фертильных растений – гомоплазмию пшеничного типа. Результаты выполненного анализа, согласно данным (Aksyonova et al., 2005; Trubacheeva et al., 2021), послужили для разделения аллоплазматических генотипов на категории с разным уровнем цитоядерной несовместимости (см. табл. 3). У растений с гетероплазмией цитоядерная совместимость нарушена, у растений с гомоплазмией - не нарушена.

Анализ гибридов, полученных между рекомбинантными аллолиниями и пшенично-ржаными замещенными линиями 1R(1A) и 1R(1D)

Отдельные полностью стерильные и частично стерильные растения аллолинии Л-56 были опылены пыльцой пшенично-ржаных замещенных линий 1R(1A) и 1R(1D). Семена завязались во всех комбинациях скрещивания, что определилось проявлением женской фертильности у ПС и ЧС растений. Из завязавшихся семян было выращено 18 растений F_1 комбинации Л-56(ПС) × 1R(1A), 20 растений комбинации Л-56(ПС) × 1R(1A), томбинации Л-56(ЧС) × 1R(1A) и 17 растений комбинации Л-56(ЧС) × 1R(1D). Все растения F_1 этих гибридных комбинаций не завязали семян от самоопыления (табл. 4).

Полная стерильность гибридов F₁, гетерозиготных по хромосомам пшеницы 1A и 1D, указывает на то, что фертильность частично стерильных (ЧС) растений зависит от влияния хромосомы 1A и хромосомы 1D. Однако одной дозы гена, локализованного на каждой из этих хромосом,



Рис. 4. Электрофореграмма ПЦР-продуктов, полученных с использованием маркера мтДНК 185/55.

1 – ячмень *H. vulgare* сорт Неполегающий; 2 – пшеница *T. aestivum* сорт Пиротрикс 28; 3 – полностью стерильное растение аллолинии Л-56; 4 – частично стерильное растение аллолинии Л-56; 5 – частично фертильное растение аллолинии Л-57; 6, 7 – фертильные растения аллолинии Л-58.

недостаточно для восстановления мужской фертильности растений.

Из аллолинии Л-57 в гибридизацию с пшенично-ржаными замещенными линиями были включены частично фертильные растения. Из семян гибридной комбинации Л-57(ЧФ) × 1R(1А) было выращено 15, а комбинации Л-57(ЧФ) × 1R(1D) – 12 растений F₁. Все растения F₁ были фертильными, что говорит о том, что восстановление фертильности у этих аллолиний является доминантным признаком. Анализ завязывания семян в главном колосе 74 растений популяции F₂ гибридной комбинации Л-57(ЧФ) × 1R(1А) выявил 51 растение из группы фертильных и 23 растения из группы стерильных. Наблюдаемое соотношение при расщеплении на фертильные и стерильные растения в F₂ соответствует теоретически ожидаемому 3 (фертильные) : 1 (стерильные) со значением $\chi^2_{05} = 1.46$, что ниже статистического значения $\chi^2_{05} = 3.84$.

Аналогичный результат получен и для гибридной комбинации Л-57(ЧФ) × 1R(1D). Из 61 изученного растения F_2 данной комбинации 45 растений отнесены к группе фертильных, а 16 растений – к группе стерильных: значение $\chi^2_{\phi a \kappa r} = 0.05$ (см. табл. 4). Эти результаты указывают на зависимость фертильности аллолинии Л-57 от влияния хромосом 1А и 1D пшеницы. Соотношение фертильных

Линия	Уровень фертильности	Число изученных растений	18S/5S мтДНК	Цитоядерная совместимость
Л-56	ПС	5	Я + П	Нарушена
	ЧС	8	•	
	ЧФ	1	• •	
Л-57	ЧФ	6	Я + П	Нарушена
	Φ	2	•	
Л-58	ЧФ	2	Я + П	Нарушена
	Φ	2	•	
	Φ	6	Π	Не нарушена
Неполегающий	Φ	2	Я	Не нарушена
Пиротрикс 28	Φ	2	П	Не нарушена

	Таблица 3. Результа	аты изучения 18S/5S	мтДНК у рекомбинантн	ных аллолиний (<i>H. vu</i>	lgare)-T. aestivum
--	---------------------	---------------------	----------------------	------------------------------	--------------------

Примечание. Я – ячмень; П – пшеница; Неполегающий – сорт ячменя; Пиротрикс 28 – сорт мягкой пшеницы.

Гибридная	Поколение	Растений			Ожидаемое	$\chi^2_{\phi a \kappa \tau}$	<i>p</i> -value
їибридная комбинация П-56(ПС) \times 1R(1A) П-56(ПС) \times 1R(1D) П-56(ЧС) \times 1R(1D) П-56(ЧС) \times 1R(1D) П-57(ЧФ) \times 1R(1A) П-57(ЧФ) \times 1R(1D) Л-58(Ф _{ГМ}) \times 1R(1A)		всего	фертильных	стерильных	расщепление в F ₂	Ψ	
Л-56(ПС) × 1R(1A)	F ₁	18	0	18	_		
Л-56(ПС) × 1R(1D)	F ₁	20	0	20	_		
Л-56(ЧС) × 1R(1А)	F ₁	15	0	15	_		
Л-56(ЧС) × 1R(1D)	F ₁	17	0	17	_		
Л-57(ЧФ) × 1R(1A)	F ₁	15	15	_			
JI-57(ΨΦ) × TR(TA)	F ₂	74	51	23	3:1	1.46	0.227
Л-57(ЧФ) × 1R(1D)	F ₁	12	12	_			
	F ₂	61	45	16	3:1	0.05	0.824
Л-58(Ф _{гм}) × 1R(1А)	F ₁	14	14	_			
	F ₂	75	75	_			
Л-58($\Phi_{\rm rm}$) $ imes$ 1R(1D)	F ₁	15	15	_			
	F ₂	86	86	_			

Таблица 4. Завязываемость семян у гибридов F₁ и расщепление по завязыванию семян у гибридов F₂, полученных от скрещивания рекомбинантных аллолиний (*H. vulgare*)-*T. aestivum* с пшенично-ржаными замещенными линиями 18(1A) и 18(1D) сорта пшеницы Саратовская 29

и стерильных растений в F₂ популяциях комбинации Л-57(ЧФ) × 1R(1А) и Л-57(ЧФ) × 1R(1D) показывает, что восстановление мужской фертильности у ЧФ растений аллолинии Л-57 в каждом случае контролируется одним доминантным геном. Один из таких генов локализован в хромосоме 1А, а другой – в хромосоме 1D мягкой пшеницы.

Иной результат получен при скрещивании полностью фертильных растений аллолинии Л-58, у которых выявлена гомоплазмия пшеничного типа, с пшенично-ржаными замещенными линиями. Все растения F_1 и F_2 комбинации Л-58(Φ) × 1R(1A) и комбинации Л-58(Φ) × 1R(1D) были фертильными (см. табл. 4). Из этого следует, что фертильность включенных в скрещивания с пшенично-ржаными замещенными линиями растений рекомбинантной аллолинии Л-58 не зависит от влияния хромосом мягкой пшеницы 1A и 1D.

Обсуждение

Между культурным ячменем *H. vulgare* и мягкой пшеницей *T. aestivum* сильно выражена межтеномная несовместимость, препятствующая как скрещиванию между ними, так и восстановлению фертильности гибридов. Однако благодаря использованию методов преодоления несовместимости и подбору родительских генотипов оказалось возможным получить жизнеспособные ячменно-пшеничные гибриды F_1 с женской фертильностью (Pershina et al., 1998). Это позволило включить гибриды в возвратные скрещивания с разными сортами мягкой пшеницы, что привело к вытеснению хромосом ячменя, образованию рекомбинантного ядерного генома пшеницы и замещению цитоплазмы пшеницы на цитоплазму ячменя у формирующихся аллоплазматических генотипов (Aksyonova et al., 2005; Першина и др., 2012).

Изученные в настоящей работе рекомбинантные аллолинии (*H. vulgare*)-*T. aestivum* Л-56, Л-57 и Л-58 имеют одинаковое происхождение, но характеризуются разной выраженностью морфобиологических признаков и разным уровнем фертильности. Рекомбинантный ядерный геном этих линий сформирован с участием сортов мягкой пшеницы Саратовская 29, Мироновская 808 и Пиротрикс 28. Проявление морфобиологических признаков у линии Л-56 относительно аллолинии Л-57, представленной частично фертильными растениями, подавлено. Аллолиния Л-56 расщепляется на полностью стерильные растения и растения с низким уровнем фертильности. По-видимому, у аллолинии Л-56 в ядерном геноме преобладает генетический материал сорта пшеницы Саратовская 29, который служит закрепителем стерильности мягкой пшеницы на цитоплазме культурного ячменя (Першина и др., 2012).

Отсутствие пыльников у полностью стерильных растений и неполное окрашивание пыльцевых зерен у частично фертильных растений обусловлено ЦМС, которая есть результат нарушения ядерно-митохондриальных взаимодействий (Yang et al., 2008). ПЦР-анализ 18S/5S мт-повтора у аллолиний Л-56 и Л-57 обнаружил гетероплазмию – наличие двух вариантов мтДНК, ячменных и пшеничных. Гетероплазмия мтДНК у ячменно-пшеничных гибридов и полученных на их основе аллолиний является следствием двуродительской передачи мтДНК начиная с F₁ (Aksyonova et al., 2005). Аналогичные данные приведены и для гибридов (Ae. crassa × пшеница сорта Chinese Spring) (Kawaura et al., 2011) и аллолиний (Ae. longissima)-T. turgidum (Noyszewski et al., 2014). Haследование цитоплазматических геномов со стороны обоих родителей по сравнению со строго материнским дает большее разнообразие вариантов мт-и хпДНК у гибридов. Считают, что наследование хлоропластов от двух родителей у покрытосеменных может служить спасением видов с дефектными пластидами (Zhang, Sodmergen, 2010), а также механизмом, который смягчает негативное влияние цитоядерной несовместимости на развитие гибридов F₁ (Barnard-Kubow et al., 2016). Можно предположить, что у изучаемых в нашей работе аллолиний Л-56 и Л-57 наличие пшеничных копий мтДНК, наряду с ячменными, также приводит к нейтрализации цитоядерного конфликта между цитоплазмой ячменя и ядерным геномом пшеницы, обеспечивающим развитие жизнеспособных аллолиний, хотя и с пониженной фертильностью.

При беккроссировании отцовским видом (пшеницей) гибридов *H. vulgare* \times *T. aestivum*, для которых характерна гетероплазмия мтДНК, наблюдается изменчивость не только ядерного генома, но и митохондриального (Aksyonova et al., 2005; Trubacheeva et al., 2012, 2021). При восстановлении фертильности аллолиний происходит увеличение числа копий мтДНК пшеничного (отцовского) типа, и первичное аллоплазматическое состояние утрачивается (Aksyonova et al., 2005; Trubacheeva et al., 2021). Такую же закономерность наблюдали и при образовании аллолиний пшеницы, несущих цитоплазму определенных видов Aegilops (Tsukamoto et al., 2000; Hattori et al., 2002). Аллолиния Л-58 с полным восстановлением фертильности и отсутствием признаков ЦМС выделена при отборе растений с максимальной фертильностью в F₅BC₅ поколении ячменно-пшеничного гибрида (см. рис. 1). Можно предположить, что у Л-58 рекомбинантный ядерный геном, из которого хромосомы ячменя элиминированы, представлен в основном генетическим материалом сортов пшеницы Мироновская 808 и Пиротрикс 28, которые относятся к восстановителям фертильности мягкой пшеницы на цитоплазме культурного ячменя (Першина и др., 2012, 2018). В процессе отбора по фертильности, так же как и при беккроссировании, изменчивость мтДНК от гетеро- к гомоплазмии пшеничного типа коррелирует с изменчивостью хлоропластной ДНК от гомоплазмии ячменного типа к гомоплазмии пшеничного типа (Aksyonova et al., 2005; Trubacheeva et al., 2021).

Как следует из данных, полученных в настоящем исследовании и ранее опубликованных (Aksyonova et al., 2005; Trubacheeva et al., 2012, 2021), гетеро- и гомоплазмия пшеничного типа 18S/5S мтДНК, обнаруженные у аллолиний (*H. vulgare*)-*T. aestivum*, могут служить маркерами для разделения аллолиний с нарушением цитоядерной совместимости и аллолиний с цитоядерной коадаптацией, поскольку не во всех случаях уровень фертильности может быть надежным признаком для такого разделения. Например, и в настоящей, и в более ранней работе (Trubacheeva et al., 2021) у части растений, отнесенных к фертильным, обнаружена гетероплазмия мтДНК, т.е. имелось нарушение цитоядерной совместимости.

Четкие различия между аллолиниями с нарушением цитоядерной совместимости и аллолиниями с цитоядерной коадаптацией обнаружены по результатам изучения влияния хромосом 1А и 1D на фертильность этих линий. У аллолиний Л-56 и Л-57 (с нарушенной цитоядерной совместимостью) мужская фертильность зависит от влияния этих хромосом пшеницы, а у аллолинии Л-58 (без нарушения цитоядерной совместимости) не зависит. Это можно объяснить тем, что у аллолиний Л-56 и Л-57 с гетероплазмией для нейтрализации стерилизующего эффекта цитоплазмы необходимо влияние генов-восстановителей *Rf*, расположенных в хромосомах 1А и 1D. Линия Л-58 (с цитоядерной коадаптацией) имеет вновь сформированную цитоплазму пшеничного типа, поэтому развитие мужскофертильных растений не зависит от наличия генов *Rf* на данных хромосомах.

Аналогичные различия мы наблюдали и в предыдущем исследовании (Trubacheeva et al., 2021): короткое плечо хромосомы 1В оказывает влияние на фертильность аллолиний с нарушением цитоядерной совместимости, но не влияет на фертильность аллолиний, у которых цитоядерная совместимость не нарушена.

Заключение

Для выполнения настоящей работы среди беккроссных потомков ячменно-пшеничного гибрида *H. vulgare* × *T. aestivum*, последовательно опыленного разными сортами мягкой пшеницы, были выделены три аллолинии мягкой пшеницы, носители цитоплазмы культурного ячменя. Эти аллолинии – одного происхождения, но с разной фертильностью и разным уровнем цитоядерной совместимости – послужили в качестве адекватных моделей для определения локализации генов, контролирующих восстановление фертильности мягкой пшеницы, несущей цитоплазму культурного ячменя.

Основываясь на результатах расщепления в F_2 гибридов, полученных от скрещивания аллолинии, у которой нарушена цитоядерная совместимость, с пшенично-ржаными замещенными линиями 1R(1A) и 1R(1D), мы впервые сделали заключение, что хромосомы 1A и 1D несут по одному доминантному гену *Rf*, контролирующему восстановление мужской фертильности мягкой пшеницы на цитоплазме культурного ячменя. Показано, что одной дозы этих генов недостаточно для восстановления фертильности частично стерильных растений. Данные наших исследований дополнили информацию о локализации генов *Rf* в хромосомах первой гомеологичной группы, соответственно 1A, 1D (в настоящей работе) и 1BS (Trubacheeva et al., 2021).

Получен важный результат, демонстрирующий независимость проявления фертильности линий с цитоядерной совместимостью от влияния хромосом, в которых локализованы гены *Rf*. Это объясняет тот факт, что интрогрессия чужеродного генетического материала у таких линий, в том числе замещение короткого плеча хромосомы пшеницы 1В на короткое плечо хромосомы ржи 1R, не нарушает цитоядерную совместимость и аллолинии сохраняют фертильность (Першина и др., 2018; Pershina et al., 2020). Более того, на основе интрогрессивных аллолиний (*H. vulgare*)-*T. aestivum* получены дигаплоидные линии (Белан и др., 2021), использованные в качестве материнских генотипов при создании коммерческих высокоурожайных сортов яровой мягкой пшеницы Сигма, Уралосибирская 2, Сигма 5.

Список литературы/ References

Бадаева Е.Д., Першина Л.А., Бильданова Л.Л. Цитогенетическое исследование нестабильных по проявлению фертильности и жизнеспособности аллоплазматических рекомбинантных линий (*Hordeum vulgare*)-*Triticum aestivum. Генетика.* 2006;42(2): 198-209

[[]Badaeva E.D., Pershina L.A., Bil'danova L.L. Cytogenetic analysis of alloplasmic recombinant lines (*H. vulgare*)-*T. aestivum* unstable

in fertility and viability. *Russ. J. Genet.* 2006;42(2):140-149. DOI 10.1134/S1022795406020074]

Белан И.А., Россеева Л.П., Блохина Н.П., Мухина Я.В., Трубачеева Н.В., Першина Л.А. Использование дигаплоидных линий – ускорение селекционного процесса в создании сортов яровой мягкой пшеницы. В: Сб. тезисов междунар. конф. «Перспективные технологии в аграрном производстве: человек, "цифра", окружающая среда (AgroProd 2021)». Омск, 2021;128-133

[Belan I.A., Rosseeva L.P., Blokhina N.P., Mukhina Y.V., Trubacheeva N.V., Pershina L.A. The use of double haploid lines is an acceleration of breeding process in creating varieties of spring bread wheat. In: Abstracts from the Int. Conf. "Advanced Technologies in Agricultural Production: People, digital, environment (AgroProd 2021)". Omsk, 2021;128-133 (in Russian)]

Булойчик А.А., Волуевич Е.А., Михно А.М. Эффекты генома и плазмона на экспрессию преодоленных генов устойчивости пшеницы к бурой ржавчине. *Цитология и генетика*. 2002;36(2): 11-19

[Buloychik A.A., Voluevich E.A., Mikhno A.M. Genome and plasmon effects on expression of the defeated genes of resistance to brown rust in wheat. *Tsitologiya i Genetika* = *Cytology and Genetics*. 2002;36(2):11-19 (in Russian)]

Климушина М.В., Дивашук М.Г., Мухаммед Т.А.К., Семенов О.Г., Карлов Г.И. Анализ аллельного состава генов, связанных с хлебопекарными качествами, у аллоцитоплазматических гибридов пшеницы. *Генетика*. 2013;49(5):617-625. DOI 10.7868/S00166 75813050081

[Klimushina M.V., Divashuk M.G., Mokhammed T.A.K., Semenov O.G., Karlov G.I. Analysis of allelic state of genes responsible for baking properties in allocytoplasmic wheat hybrids. *Russ. J. Genet.* 2013;49(5):530-538. DOI 10.1134/S1022795413050074]

Першина Л.А., Девяткина Э.П., Трубачеева Н.В., Кравцова Л.А., Добровольская О.Б. Особенности восстановления фертильности аллоплазматических линий, полученных на основе гибридизации самоопыленного потомка ячменно-пшеничного амфиплоида (*Hordeum vulgare* L. × *Triticum aestivum* L.) с сортами мягкой пшеницы Саратовская 29 и Пиротрикс 28. *Генетика*. 2012;48(12):1372-1379

[Pershina L.A., Devyatkina E.P., Trubacheeva N.V., Kravtsova L.A., Dobrovol'skaya O.B. Characterization of fertility restoration in alloplasmic lines derived from hybridization of self-fertilized of spring of barley–wheat (*Hordeum vulgare* L. × *Triticum aestivum* L.) amphiploid with common wheat varieties Saratovskaya 29 and Pyrotrix 28. *Russ. J. Genet.* 2012;48(12):1184-1190. DOI 10.1134/S1022 795412120101]

Першина Л.А., Белова Л.И., Трубачеева Н.В., Осадчая Т.С., Шумный В.К., Белан И.А., Россеева Л.П., Немченко В.В., Абакумов С.Н. Аллоплазматические рекомбинантные линии (*H. vulgare*)-*T. aestivum* с транслокацией 1RS.1BL: исходные генотипы для создания сортов яровой мягкой пшеницы. *Вавиловский журнал генетики и селекции*. 2018;22(5):544-552. DOI 10.18699/VJ18.393

[Pershina L.A., Belova L.I., Trubacheeva N.V., Osadchaya T.S., Shumny V.K., Belan I.A., Rosseeva L.P., Nemchenko V.V., Abakumov S.N. Alloplasmic recombinant lines (*H. vulgare*)-*T. aestivum* with 1RS.1BL translocation: initial genotypes for production of common wheat varieties. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2018;22(5):544-552. DOI 10.18699/VJ18.393]

- Aksyonova E., Sinyavskaya M., Danilenko N., Pershina L., Nakamura C., Davydenko O. Heteroplasmy and paternally oriented shift of the organellar DNA composition in barley-wheat hybrids during backcrosses with wheat parents. *Genome.* 2005;48(5):761-769. DOI 10.1139/g05-049
- Atienza S.G., Martin A., Peechioni N., Platani C., Cattivelli L. The nuclear-cytoplasmic interaction controls carotenoid content in wheat. *Euphytica*. 2008;159:325-331. DOI 10.1007/s10681-007-9511-6

- Barnard-Kubow K.B., McCoy M.A., Galloway L.F. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytol.* 2016;213(3):1466-1476. DOI 10.1111/nph.14222
- Bohra A., Jha U.C., Adhimoolam P., Bisht D., Sing N.P. Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Rep.* 2016;35(5):967-993. DOI 10.1007/s00299-016-1949-3
- Coulthart M.B., Spencer D.F., Gray M.W. Comparative analysis of a recombining-repeat-sequence family in the mitochondrial genomes of wheat (*Triticum aestivum* L.) and rye (*Secale cereale* L.). *Curr. Genet.* 1993;23(3):255-264. DOI 10.1007/BF00351504
- Crosatti C., Quansah L., Maré C., Giusti L., Roncaglia E., Atienz S.G., Cattivelli L., Fait A. Cytoplasmic genome substitution in wheat affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations. *BMC Genomics*. 2013;14:868-889. DOI 10.1186/1471-2164-14-868
- Current Protocols in Molecular Biology. Greene Publishing Associates. N.Y.: Wiley Interscience, 1987
- Gupta P.K., Balyan H.S., Gahlaut V., Saripalli G., Pal B., Basnet B.R., Joshi A.K. Hybrid wheat: past, present and future. *Theor. Appl. Genet.* 2019;132(9):2463-2483. DOI 10.1007/s00122-019-03397-y
- Hattori N., Kitagawa K., Takumi S., Nakamura C. Mitochondrial DNA heteroplasmy in wheat, *Aegilops* and their nucleus-cytoplasm hybrids. *Genetics*. 2002;160(4):1619-1630. DOI 10.1093/genetics/ 160.4.1619
- Hohn C.E., Lukaszewski A.J. Engineering the 1BS chromosome arm in wheat to remove the *Rf* ^{multi} locus restoring male fertility in cytoplasms of *Aegilops kotschyi*, *Ae. uniaristata* and *Ae. mutica. Theor. Appl. Genet.* 2016;129(9):1769-1774. DOI 10.1007/s00122-016-2738-7
- Islam M.S., Studer B., Møller I.M., Asp T. Genetics and biology of cytoplasmic male sterility and its applications in forage and turf grass breeding. *Plant Breed*. 2014;133(3):299-312. DOI 10.1111/ pbr.12155
- Kawaura K., Saeki A., Masumura T., Morita S., Ogihara Y. Heteroplasmy and expression of mitochondrial genes in alloplasmic and euplasmic wheat. *Genes Genet. Syst.* 2011;86(4):249-255. DOI 10.1266/ggs.86.249
- Liu C.G., Wu Y.W., Hou H., Zhang C., Zhang Y., McIntosh R.A. Value and utilization of alloplasmic common wheats with *Aegilops crassa* cytoplasm. *Plant Breed*. 2002;121(5):407-410. DOI 10.1046/j.1439-0523.2002.755374.x
- Liu Y., Tang L., Xu Q., Ma D., Zha M., Sun J., Chen W. Experimental and genomic evidence for the *indica*-type cytoplasmic effect in *Oryza sativa* L. ssp. *japonica*. J. Integr. Agric. 2016;15(10):2183-2191. DOI 10.1016/S2095-3119(15)61190-X
- Martin A.C., Atienza S.G., Ramírez M.C., Barro F., Martín A. Molecular and cytological characterization of an extra acrocentric chromosome that restores male fertility of wheat in the msH1 CMS system. *Theor. Appl. Genet.* 2010;121(6):1093-1101. DOI 10.1007/s00122-010-1374-x
- Noyszewski A.K., Ghavami F., Alnemer L.M., Soltani A., Gu Y.Q., Huo N., Meinhardt S., Penny M.A., Kianian P.M.A., Kianian S.F. Accelerated evolution of the mitochondrial genome in an alloplasmic line of durum wheat. *BMC Genomics*. 2014;15(1):67. DOI 1471-2164/15/67
- Pershina L.A., Numerova O.M., Belova L.I., Devyatkina E.P. Biotechnological and cytogenetic aspects of producing new wheat genotypes using hybrids. *Euphytica*. 1998;100(1-3):239-244. DOI 10.1023/A:1018312408312
- Pershina L., Trubacheeva N., Badaeva E., Belan I., Rosseeva L. Study of androgenic plant families of alloplasmic introgression lines (*H. vulgare*)–*T. aestivum* and the use of sister DH lines in breeding. *Plants*. 2020;9(6):764-816. DOI 10.3390/plants9060764
- Rafee V.V., Lalitha R., Remadevi A.K., Lekshmi M., Premachandran M.N. Substitution of cytoplasm of sugarcane with that of the wild grass *Erianthus arundinaceus*. *Gregor Mendel Found*. J. 2010; 1(1&2):16-22

- Shchapova A.I., Kravtzova L.A. The production of wheat-rye substitution line by using the Giemsa staining technique. *Cereal Res. Commun.* 1982;10(1):33-39
- Sinha P., Tomar S.M., Vinod, Singh V.K., Balyan H.S. Genetic analysis and molecular mapping of a new fertility restorer gene *Rf8* for *Triticum timopheevi* cytoplasm in wheat (*Triticum aestivum* L.) using SSR markers. *Genetica*. 2013;141(10-12):431-441. DOI 10.1007/ s10709-013-9742-5
- Soltani A., Kumar A., Mergoum M., Pirseyedi S.M., Hegstad J.B., Mazaheri M., Kianian S.F. Novel nuclear-cytoplasmic interaction in wheat (*Triticum aestivum*) induces vigorous plants. *Funct. Integr. Genom.* 2016;16(2):171-182. DOI 10.1007/s10142-016-0475-2
- Takenaka S., Yamamoto R., Nakamura C. Differential and interactive effects of cytoplasmic substitution and seed aging on submergence stress response in wheat (*Triticum aestivum L.*). *Biotechnol. Biotechnol. Equip.* 2019;33(1):75-85. DOI 10.1080/13102818.2018. 1549960
- Talukder S.K., Vara Prasad P.V., Todd T., Babar M.A., Poland J., Bowden R., Fritz A. Effect of cytoplasmic diversity on post anthesis heat tolerance in wheat. *Euphytica*. 2015;204:383-394. DOI 10.1007/s10681-014-1350-7
- Tao D., Xu P., Zhou J., Deng X., Li J., Deng W., Yang J., Yang G., Li Q., Hu F. Cytoplasm affects grain weight and filled-grain ratio in *indica* rice. *BMC Genet.* 2011;12:53. DOI 10.1186/1471-2156-12-53
- Trubacheeva N.V., Kravtsova L.A., Devyatkina E.P., Efremova T.T., Sinyavskaya M.G., Shumny V.K., Pershina L.A. Heteroplasmic and homoplasmic states of mitochondrial and chloroplast DNA regions

in progenies of distant common wheat hybrids of different origins. *Russ. J. Genet. Appl. Res.* 2012;2(6):494-500. DOI 10.1134/S20790 59712060147

- Trubacheeva N.V., Divashuk M.G., Chernook A.G., Belan I.A., Rosseeva L.P., Pershina L.A. The effect of chromosome arm 1BS on the fertility of alloplasmic recombinant lines in bread wheat with the *Hordeum vulgare* cytoplasm. *Plants.* 2021;10(6):1120. DOI 10.3390/plants10061120
- Tsukamoto N., Asakura N., Hattori N., Takumi S., Mori N., Nakamura C. Identification of paternal mitochondrial DNA sequences in the nucleus-cytoplasm hybrid of tetraploidand hexaploid wheat with D and D2 plasmon from *Aegilops* species. *Curr. Genet.* 2000; 38(4):208-217. DOI 10.1007/s002940000153
- Tsunewaki K. Plasmon analysis as the counterpart of genome analysis. In: Jauhar P.P. (Ed.) Methods of Genome Analysis in Plants. Boca Raton: CRC Press, 1996;271-299. http://pi.lib.uchicago.edu/1001/ cat/bib/2607056
- Tsunewaki K. Fine mapping of the first multi-fertility-restoring gene, *Rf*^{multi}, of wheat for three *Aegilops* plasmons, using 1BS-1RS recombinant lines. *Theor. Appl. Genet.* 2015;128(4):723-732. DOI 10.1007/s00122-015-2467-3
- Yang J., Zhang M., Yu J. Mitochondrial retrograde regulation tuning fork in nuclear genes expressions of higher plants. J. Genet. Genomics. 2008;35(2):65-71. DOI 10.1016/S1673-8527(08)60010-7
- Zhang Q., Sodmergen. Why does biparental plastid inheritance revive in angiosperms? J. Plant Res. 2010;123(2):201-206. DOI 10.1007/ s10265-009-0291-z

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 25.03.2024. После доработки 03.05.2024. Принята к публикации 14.05.2024.

DOI 10.18699/vjgb-24-69

Содержание метаболитов и профиль экспрессии генов соответствующих метаболических путей в контрастных по окраске плодах баклажана (*Solanum melongena* L.)

М.А. Филюшин 🔟¹ 🖾, Е.А. Джос^{1, 2}, А.В. Щенникова 🔟¹, Е.З. Кочиева 🔟¹

¹ Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия ² Федеральный научный центр овощеводства, пос. ВНИИССОК, Московская область, Россия

michel7753@mail.ru

Аннотация. Баклажан (Solanum melongena L.) занимает пятое место по значимости среди овощных культур семейства Пасленовых, в том числе благодаря антиоксидантным свойствам плода за счет высокого содержания различных фенольных соединений. Наряду с популярными фиолетовоплодными сортами S. melongena имеются сорта, плоды которых синтезируют фенольные соединения, однако характеризуются белой окраской из-за отсутствия биосинтеза антоцианов. Определение количества антоцианов и других фенольных соединений, а также каротиноидов и сахаров входит в оценку качества плодов баклажана коммерческой (технической) спелости. Кроме антиоксидантных и вкусовых качеств, эти метаболиты связаны с устойчивостью плода к различным стрессовым факторам. В данном исследовании проведен сравнительный анализ содержания антоцианов, каротиноидов и растворимых сахаров (сахарозы, глюкозы, фруктозы) в кожице и мякоти плода как технической, так и биологической спелости у фиолетовоплодного (сорт Влас) и белоплодного (сорт Снежный) образцов баклажана отечественной селекции. Кожица и мякоть плода биологической спелости сортов Влас и Снежный были использованы для сравнительного транскриптомного анализа. Показано, что ключевые гены флавоноидного пути, метаболизма каротиноидов, гидролиза сахарозы, а также транспорта растворимых сахаров дифференциально экспрессируются между тканями плода как внутри каждого сорта, так и между сортами. Подтверждена связь фиолетовой окраски кожицы плода сорта Влас с присутствием значительных количеств антоцианов. Определено, что в сравнении с сортом Снежный спелый плод сорта Влас характеризуется существенно более низким уровнем экспрессии генов биосинтеза флавоноидов. Однако у обоих сортов в спелом плоде не выявлены транскрипты генов биосинтеза антоцианов (DFR, ANS, UFGT). Также показано, что в сравнении с белым плодом сорта Снежный фиолетовый плод сорта Влас накапливает больше каротиноидов и сахарозы и меньше глюкозы и фруктозы. Биохимические данные соответствуют профилю дифференциальной экспрессии ключевых генов, кодирующих структурные белки метаболизма и транспорта анализируемых соединений.

Ключевые слова: сорта баклажана; Solanum melongena L.; каротиноиды; антоцианы; растворимые сахара; экспрессия генов метаболических путей.

Для цитирования: Филюшин М.А., Джос Е.А., Щенникова А.В., Кочиева Е.З. Содержание метаболитов и профиль экспрессии генов соответствующих метаболических путей в контрастных по окраске плодах баклажана (Solanum melongena L.). Вавиловский журнал генетики и селекции. 2024;28(6):619-627. DOI 10.18699/vjgb-24-69

Финансирование. Работа выполнена при поддержке Минобрнауки России в рамках соглашения № 075-15-2022-318 от 20.04.2022 о предоставлении гранта в форме субсидий из федерального бюджета на осуществление государственной поддержки создания и развития научного центра мирового уровня «Агротехнологии будущего».

Metabolite concentrations and the expression profiles of the corresponding metabolic pathway genes in eggplant (*Solanum melongena* L.) fruits of contrasting colors

M.A. Filyushin (D¹ 🖾, E.A. Dzhos^{1, 2}, A.V. Shchennikova (D¹, E.Z. Kochieva (D¹

¹ Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia
 ² Federal Scientific Vegetable Center, VNIISSOK village, Moscow region, Russia
 ² michel7753@mail.ru

Abstract. Eggplant (*Solanum melongena* L.) ranks fifth in importance among vegetable crops of the Solanaceae family, in part due to the high antioxidant properties and polyphenol content of the fruit. Along with the popular purple-fruited varieties of *S. melongena*, there are cultivars, the fruits of which are rich in phenolic compounds, but are white-colored due to the lack of anthocyanin biosynthesis. Determination of the amount of anthocyanins and other phenolic compounds, as well as carotenoids and sugars, is included in the assessment of the quality of eggplant fruits of com-

mercial (technical) ripeness. In addition to antioxidant and taste properties, these metabolites are associated with fruit resistance to various stress factors. In this study, a comparative analysis of the content of anthocyanins, carotenoids and soluble sugars (sucrose, glucose, fructose) in the peel and pulp of the fruit of both technical and biological ripeness was carried out in purple-fruited (cv. Vlas) and white-fruited (cv. Snezhny) eggplant accessions of domestic selection. The peel and pulp of biologically ripe fruits of the cvs Vlas and Snezhny were used for comparative transcriptomic analysis. The key genes of the flavonoid and carotenoid metabolism, sucrose hydrolysis, and soluble sugar transport were shown to be differentially expressed between fruit tissues, both within each cultivar and between them. It has been confirmed that the purple color of the peel of the cv. Vlas fruit is due to substantial amounts of anthocyanins. Flavonoid biosynthesis genes showed a significantly lower expression level in the ripe fruit of the cv. Vlas in comparison with the cv. Snezhny. However, in both cultivars, transcripts of anthocyanin biosynthesis genes (*DFR, ANS, UFGT*) were not detected. Additionally, the purple fruit of the cv. Vlas accumulated more carotenoids and sucrose and less glucose and fructose than the white fruit of the cv. Snezhny. Biochemical data corresponded to the differential expression pattern of the key genes encoding the structural proteins of metabolism and transport of the compounds analyzed. **Key words:** eggplant cultivars; *Solanum melongena* L; carotenoids; anthocyanins; soluble sugars; expression of metabolic pathway genes.

For citation: Filyushin M.A., Dzhos E.A., Shchennikova A.V., Kochieva E.Z. Metabolite concentrations and the expression profiles of the corresponding metabolic pathway genes in eggplant (*Solanum melongena* L.) fruits of contrasting colors. *Vavilovskii Zhurnal Genetiki i Selektsii=Vavilov Journal of Genetics and Breeding*. 2024;28(6):619-627. DOI 10.18699/vjgb-24-69

Введение

Баклажан (Solanum melongena L.) – овощная культура, занимающая пятое место по экономической значимости в семействе Пасленовых (Solanaceae). Несмотря на теплолюбивость, эта культура выращивается не только в тропических и субтропических климатических зонах, но и как тепличная культура в регионах с холодным климатом, включая Российскую Федерацию. Наибольшую известность имеют плоды баклажана с кожицей, окрашенной в разные оттенки фиолетового цвета, что определяется содержанием антоцианов. Данный факт в совокупности с тем, что мякоть плодов обогащена фенольными кислотами, придает баклажанам мощные антиоксидантные свойства и ставит их в один ряд с продуктами с высокой пищевой/ диетической ценностью (Gürbüz et al., 2018; Akhbari et al., 2019; Condurache et al., 2021; Saha et al., 2023).

Кроме сортов с фиолетовоокрашенными плодами, имеются также сорта *S. melongena*, образующие плоды с кожицей белой или зеленой окраски из-за ингибирования биосинтеза антоцианов (Condurache et al., 2021; Yang et al., 2022; You et al., 2022). Окраска белая, зеленая или промежуточных оттенков определяется соотношением двух типов пластид в клетках плода – хлоропластов и лейкопластов (Тао et al., 2023). С точки зрения потребителя важно, что белоплодные сорта могут быть предпочтительней, так как лишены горечи, свойственной плодам с темной окраской, благодаря изменению содержания гликоалкалоидов (Lelario et al., 2019; Saha et al., 2023).

Коммерческим сортам баклажана присуща морфологическая вариативность, и скрининг существующих коллекций по набору характеристик включает группировку по окраске кожицы плода как наиболее важного признака (Martínez-Ispizua et al., 2021). Оценка качества плодов фокусируется на их антиоксидантных свойствах, включая определение содержания фенольных соединений/флавоноидов, каротиноидов и сахаров, и в данном аспекте наблюдается широкое разнообразие (Martínez-Ispizua et al., 2021). Фиолетовоплодные сорта в сравнении с белоплодными характеризуются большей антиоксидантной активностью и повышенным содержанием фенолов и каротиноидов как в кожице, так и в мякоти и слабо отличаются или сходны по общему количеству сахаров (Martínez-Ispizua et al., 2021; Colak et al., 2022).

Корреляций между содержанием флавоноидов, каротиноидов и сахаров в плодах баклажана не наблюдается (Martínez-Ispizua et al., 2021). С другой стороны, имеются косвенные свидетельства существования такой зависимости, опосредованной фитогормонами, в ягодах вишни (Teribia et al., 2016), а именно: обратная корреляция между содержанием растворимых сахаров и транс-зеатина, а также гиббереллина GA4 и антоцианов. Абсцизовая кислота (AБК), напротив, связана положительно с количеством антоцианов и растворимых сахаров (Teribia et al., 2016). При этом накопление антоцианов коррелирует с количеством сахаров у китайского финика *Ziziphus jujube* (Jiang et al., 2020).

Все упомянутые антиоксидантные соединения, а также растворимые сахара тесно связаны с устойчивостью к различным стрессовым факторам как вегетативной части растения (Keunen et al., 2013; Pérez-Torres et al., 2021; Waadt et al., 2022), так и сочного плода (Shi et al., 2019; Jiang et al., 2020). К примеру, показано, что повышенная наработка фенольных соединений определяет устойчивость плода баклажана к низким температурам (Shi et al., 2019). Повышенная температура положительно влияет на содержание сахаров, антоцианов, флавоноидов и каротиноидов в плодах китайского финика *Z. jujube*, однако в сочетании с засухой вызывает обратный эффект (Jiang et al., 2020).

Цель настоящей работы – характеристика плодов двух сортов баклажана, которая включала определение содержания антоцианов, каротиноидов и растворимых сахаров, а также профиля экспрессии ключевых генов соответствующих метаболических путей. Для этого были выбраны два сорта отечественной селекции, имеющих разную окраску плода – белую и фиолетовую соответственно. Существенным отличием от проводимых ранее подобных исследований было то, что для анализа были использованы плоды не только технической (товарной), но и биологической спелости.

Для достижения цели были поставлены следующие задачи: получение растительного материала (плоды двух сортов на стадиях технической и биологической спелости), определение содержания целевых метаболитов в кожице и мякоти плодов технической и биологической спелости, анализ транскриптомов кожицы и мякоти плодов на стадии биологической спелости с фокусировкой на транскриптах генов целевых метаболических путей, валидация транскриптомных данных.

Материалы и методы

Для сравнительного исследования использовали образцы двух раннеспелых сортов баклажана вида *S. melongena* селекции Федерального научного центра овощеводства (ФНЦО, Московская область), различающихся окраской спелого плода. Плоды сорта Снежный (код сорта 9905014, https://gossortrf.ru/registry/) на стадии технической спелости имеют белую окраску кожицы и мякоти. Плоды сорта Влас (код сорта 8057522) в технической спелости имеют темно-фиолетовую кожицу и белую мякоть. На стадии биологической спелости мякоть плодов остается белой у обоих сортов, а кожица приобретает желтоватый (сорт Снежный) или коричневый (сорт Влас) оттенки (рис. 1).

Растения исследуемых сортов выращивали (2023 г.) до стадии плодоношения в пленочной теплице ФНЦО. В августе собирали плоды в технической (commercially mature, CM) и биологической (physiologically ripe, PR) спелости, разделяли их на кожицу (экзокарп) и мякоть (мезокарп), измельчали растиранием в фарфоровой ступке в жидком азоте и использовали для биохимического, метаболомного и транскриптомного анализов.

Содержание антоцианов и каротиноидов определяли спектрофотометрически в хлороформ-метанольных экс-

трактах согласно (Филюшин и др., 2020). Поскольку среди антоцианов, накапливающихся в кожице плода баклажана, доминируют гликозиды дельфинидина (93–98 % от общей суммы) (Condurache et al., 2021; Yang et al., 2022), содержание антоцианов рассчитывали в пересчете на дельфинидин-3-рутинозид.

Содержание сахаров: глюкозы, фруктозы и сахарозы – определяли по данным метаболомов (не опубликовано), которые получали согласно (Filyushin et al., 2023). Вкратце: около 0.2 г тщательно измельченной ткани листьев дважды экстрагировали 200 мкл 80 % метанола. Суммарный экстракт упаривали, растворяли в 30 % метаноле (из расчета 50 мг сырой массы на 100 мкл экстракта) и подвергали масс-спектральному анализу с использованием сверхэффективного метода жидкостной хроматографии с квадрупольной времяпролетной масс-спектрометрией (UPLC-qTOF-MS/MS) согласно протоколу [https://lcms. cz/labrulez-bucket-strapi-h3hsga3/1866243 lcms 148 how potato fights its enemies 02 2019 ebook rev 01 9d3990d6c4/1866243-lcms-148-how-potato-fights-itsenemies-02-2019-ebook-rev-01.pdf]. В качестве относительных показателей по содержанию сахаров использовали уровень сигнала/100 мг аннотированных соединений.

Дифференциально экспрессирующиеся гены (ДЭГ), кодирующие белки, участвующие в гидролизе и транспорте растворимых сахаров (инвертазы и унипортеры сахаров), определяли по данным транскриптомов кожицы и мякоти плода PR (не опубликовано). Для транскриптомного анализа выделяли препараты суммарной PHK (RNeasy Plant Mini Kit, Qiagen, США), послужившие основой для библиотек мPHK (NEBNext[®] mRNA Library Prep Reagent



Рис. 1. Плоды сортов баклажана Снежный (*a*) и Влас (*б*) в технической (СМ; слева) и биологической (PR; справа) спелости.

Сорта различаются по окраске кожицы плода – белой (Снежный) и фиолетовой (Влас). Масштабная линия 5 см.

Set for Illumina; New England BioLabs, CIIIA), которые затем секвенировали (Illumina HiSeq2500; Illumina Inc., США). Для сборки и определения кодирующих последовательностей использовали Trinity v3.5.13 (https://github. com/trinityrnaseq/trinityrnaseq/wiki) и TransDecoder v5.1.0 (https://github.com/TransDecoder/TransDecoder); аннотировали с помощью NCBI-Blast (https://www.ncbi.nlm.nih. gov/). Относительные уровни транскриптов (FPKM; количество фрагментов на 1 кб транскриптов (FPKM; количество фрагментов) оценивали с помощью RSEM (https://github.com/deweylab/RSEM). Для определения ДЭГ как внутри сорта (кожица vs мякоть), так и между сортами (кожица vs кожица; мякоть vs мякоть) данные транскриптомов нормализовали на уровень транскриптов референсного гена *GAPDH*.

Структурный анализ последовательностей ДЭГ выполняли с помощью NCBI-BLAST (https://blast.ncbi.nlm.nih. gov/Blast.cgi) и MEGA 7.0 (https://www.megasoftware.net/) с применением геномных (GCA_000787875.1) (Hirakawa et al., 2014) и транскриптомных данных *S. melongena* (https://www.ncbi.nlm.nih.gov/).

Валидацию транскриптомных данных проводили методом количественной ПЦР в реальном времени (ПЦР-РВ) на амплификаторе CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, CША); программа [95 °С – 5 мин; 40 циклов (95 °С – 15 с, 62 °С – 50 с)]. Для этого на основе имеющихся препаратов суммарной РНК синтезировали кДНК (GoScript[™] Reverse Transcription System, Promega, США) и 3 нг использовали в реакцию. Реакционная смесь включала набор «Реакционная смесь для проведения ПЦР-РВ в присутствии SYBR GreenI и ROX» (ООО «Синтол», Россия) и ген-специфичные праймеры. Реакции проводили в трех технических и двух биологических повторах и нормализовали на уровень транскриптов референсного гена *GAPDH* (Zhang et al., 2014).

Для статистической обработки полученных биохимических и экспрессионных данных использовали GraphPad Prism v. 8 (GraphPad Software Inc., США; https://www.graph pad.com/scientific-software/prism/). Для оценки достоверности различий применяли *t*-test (p < 0.05 указывает на статистическую значимость различий).

Результаты

Исследование было сфокусировано на сравнительной характеристике плода технической и биологической спелости двух сортов баклажана, принадлежащих одному виду, *S. melongena*, и различающихся по окраске кожицы плода, а именно сортов Снежный и Влас соответственно с белой/желтоватой и фиолетовой/коричнево-фиолетовой окраской кожицы плода в технической/биологической спелости (см. рис. 1).

Проведенный биохимический анализ кожицы и мякоти в динамике созревания плода показал, что содержание антоцианов соответствует окраске анализируемых тканей плода биологической спелости. В желтоватой кожице и белой мякоти плода сорта Снежный, а также белой мякоти плода сорта Влас количество антоцианов имело следовые значения, тогда как в коричнево-фиолетовой кожице плода сорта Влас их количество оказалось выше в ~300 раз (рис. 2, *a*).

Плоды обоих сортов как в технической, так и биологической спелости содержали следы каротиноидов в мякоти. В кожице каротиноиды накапливались активнее, в случае сорта Влас количество каротиноидов было выше в ~25 раз, чем у сорта Снежный (см. рис. 2, δ).

Если различие в содержании антоцианов у анализируемых сортов было предсказуемым, то несколько неожиданными оказались значительные расхождения по содержанию растворимых сахаров. По данным проведенного метаболомного профилирования кожицы и мякоти было обнаружено, что в плоде сорта Снежный присутствует в ~2 (кожица) и ~5 (мякоть) раз больше гексоз (глюкоза, фруктоза), а также в ~2 (кожица и мякоть) раза меньше сахарозы, чем у сорта Влас (см. рис. 2, *в*).

Проведенный сравнительный анализ транскриптомов кожицы и мякоти плода сортов Снежный и Влас позволил выявить ряд ДЭГ, куда ожидаемо по результатам биохимического и метаболомного анализа попали гены, связанные с метаболизмом антоцианов, каротиноидов и сахаров (см. таблицу).

Было обнаружено, что основные гены флавоноидного пути (Zhang et al., 2014; Alappat B., Alappat J., 2020) до синтеза антоцианов (*CHS1*, *CHS2*, *F3H*), высоко транскри-



Рис. 2. Содержание суммы антоцианов (*a*), суммы каротиноидов (*б*), гексоз (суммарно глюкозы и фруктозы) (*в*) и сахарозы (*г*) в кожице и мякоти плода технической (CM) и биологической (PR) спелости сортов баклажана Снежный и Влас (*S. melongena*).

В качестве относительных показателей по содержанию сахаров, полученных при нетаргетированном метаболомном профилировании, использовали уровень сигнала/100 мг аннотированных соединений.

Ген	ID, транскриптом S. melongena	Гомолог у S. lycopersicum, NCBI ID
	Семейство GH32	? (кислые инвертазы; гидролиз сахарозы)
VINV1	TRINITY_DN2044_c0_g1_i1.p1	acid vacuolar invertase ASK06213.1
CWINV1	TRINITY_DN7423_c0_g1_i23.p1	beta-fructofuranosidase, insoluble isoenzyme CWINV3-like XP_004241885.1
CWINV2	TRINITY_DN29292_c0_g1_i1.p1	beta-fructofuranosidase, insoluble isoenzyme CWINV1 XP_019068732.1
CWINV3	TRINITY_DN3426_c0_g1_i16.p1	cell-wall invertase AAM22409.1
	Семейство GH100 (нейтр	ально-щелочные инвертазы; гидролиз сахарозы)
N/AINV1	TRINITY_DN5049_c0_g1_i2.p1	neutral/alkaline invertase 3, chloroplastic, XP_004249987.1
N/AINV2	TRINITY_DN5579_c0_g1_i2.p1	probable alkaline/neutral invertase D, XP_004241837.1
N/AINV3	TRINITY_DN5658_c1_g1_i7.p1	alkaline/neutral invertase A, mitochondrial, XP_004230329.1
N/AINV4	TRINITY_DN6542_c0_g1_i11.p1	alkaline/neutral invertase A, mitochondrial, XP_004230329.1
N/AINV5	TRINITY_DN6803_c0_g1_i6.p1	neutral/alkaline invertase 3, chloroplastic, XP_004249987.1
N/AINV6	TRINITY_DN9045_c1_g1_i1.p1	probable alkaline/neutral invertase D, XP_004238357.1
	Семейство SWI	ЕЕТ (унипортеры растворимых сахаров)
SWEET1	TRINITY_DN316_c1_g1_i1.p1	bidirectional sugar transporter SWEET1-like XP_004237723.1
SWEET2	TRINITY_DN2271_c0_g1_i1.p1	bidirectional sugar transporter SWEET1, XP_004242009.1
SWEET3	TRINITY_DN1022_c0_g1_i7.p1	bidirectional sugar transporter N3, XP_019068532.1
SWEET4	TRINITY_DN13252_c0_g1_i6.p1	bidirectional sugar transporter SWEET1-like, XP_004237724.1
SWEET5	TRINITY_DN1022_c0_g1_i4.p1	bidirectional sugar transporter N3, XP_019068532.1
SWEET6	TRINITY_DN10403_c0_g1_i2.p1	bidirectional sugar transporter SWEET2a, XP_004233011.1
	Путь	метаболизма каротиноидов
PSY1	TRINITY_DN59246_c0_g1_i2.p1	phytoene synthase 1, NP_001234812.1
PSY2	TRINITY_DN6268_c0_g1_i3.p1	phytoene synthase 2, NP_001234671.1
NCED1	TRINITY_DN3512_c2_g1_i3.p1	9-cis-epoxycarotenoid dioxygenase, NP_001234455.1
		Флавоноидный путь
CHS1	TRINITY_DN6959_c0_g2_i2.p1	chalcone synthase 1, NP_001234033.2
CHS2	TRINITY_DN6763_c0_g1_i3.p1	chalcone synthase 2, NP_001234036.2
F3H	TRINITY_DN5746_c0_g1_i2.p1	flavanone 3-dioxygenase, NP_001316412.1
DFR	TRINITY_DN49807_c0_q1_i1.p1	dihydroflavonol 4-reductase, NP_001234408.2

Список ДЭГ, связанных с метаболизмом сахаров, каротиноидов и антоцианов

бируются в кожице плода сорта Снежный и детектируются в значительно меньших и сходных между собой количествах в мякоти (оба сорта) и кожице (Влас) (рис. 3). При рассмотрении части пути, относящейся к синтезу антоцианов, оказалось, что уровень экспрессии первого гена ветви, DFR, в кожице и мякоти плода сорта Снежный значительно выше, чем у сорта Влас. Однако количество транскриптов в значении FPKM для DFR предельно низкое во всех четырех образцах (0.49-3.44), поэтому о существенной разнице между сортами говорить нельзя, так как уровень транскриптов гена приближается к нулю. При этом транскрипты последующих генов ветви биосинтеза антоцианов, ANS (антоцианидинсинтаза) и UFGT (UDPглюкозофлавоноид-3-О-глюкозилтрансфераза), не вошли в список ДЭГ и детектировались в следовых количествах (см. рис. 3).

Анализ транскриптов генов фитоинсинтаз – ключевых изоферментов метаболизма каротиноидов (Rosas-Saa-

vedra, Stange, 2016) – показал следовые количества мРНК *PSY1* в кожице (Снежный) и мякоти (оба сорта) плода и их существенный уровень в кожице плода сорта Влас (см. рис. 3). В то же время транскрипты *PSY2* присутствовали в относительно значимых количествах в кожице (Снежный, Влас) и мякоти (Снежный) плода. При этом число транскриптов *PSY2* было существенно больше у сорта Снежный (см. рис. 3). Другой ДЭГ, связанный с катаболизмом каротиноидов, ген 9-*цис*-эпоксикаротиноиддиоксигеназы (*NCED1*), катализирующей синтез АБК из ксантофиллов β , β -ветви пути (Rosas-Saavedra, Stange, 2016), высоко транскрибировался в кожице и мякоти плода сорта Влас, тогда как в плоде сорта Снежный детектировались только следовые количества (см. рис. 3).

В список ДЭГ, связанных с необратимым гидролизом сахарозы и транспортом моно- и дисахаридов, вошли гены вакуолярной инвертазы (*VINV1*), инвертаз клеточной стенки (*CWINV1–3*), нейтральных/щелочных инвер-



Рис. 3. Тепловая карта экспрессии ДЭГ, связанных с метаболизмом антоцианов и каротиноидов, а также с гидролизом сахарозы и транспортом растворимых сахаров в кожице и мякоти биологически спелого плода (PR) сортов Снежный и Влас (*S. melongena*).

Построено по данным транскриптомного анализа.

таз (*N/AINV1-6*) и унипортеров сахаров (*SWEET1-6*) (см. таблицу).

В кожице плода сорта Снежный наиболее высокий уровень экспрессии имели гены четырех инвертаз (VINV1, CWINV1, N/AINV5 и 6) и трех унипортеров сахаров (SWEET1, 3 и 5); в мякоти плода – гены четырех инвертаз (VINV1, N/AINV1, 5 и 6) и двух унипортеров сахаров (SWEET5 и 6) (см. рис. 3).

В целом сорт Влас отличался от сорта Снежный значительными уровнями экспрессии и большим числом активных генов инвертаз и унипортеров сахаров. В кожице плода сорта Влас наиболее высоко транскрибировались гены шести инвертаз (*CWINV3*, *CWINV1* и 2, *N/AINV2–4*) и одного унипортера сахаров (*SWEET6*); в мякоти плода – гены четырех инвертаз (*CWINV1*, *N/AINV1*, 5 и 6) и трех унипортеров сахаров (*SWEET2–4*) (см. рис. 3).

Таким образом, профиль экспрессии генов метаболизма антоцианов, каротиноидов и сахаров различался как



Metabolite concentrations and expression

of metabolic pathway genes in eggplant fruits

CHS1

CHS1

Рис. 4. Профили экспрессии по данным ПЦР-РВ (слева) и транскриптомного анализа (справа) генов *CHS1*, *CHS2*, *F3H*, *PSY1* и *PSY2*.

Отсутствие экспрессии генов *DFR*, *ANS* также подтверждено ПЦР-PB; графики не приведены. Последовательности праймеров для генов *CHS1*, *CHS2*, *F3H*, *DFR*, *ANS* и референсного гена *GAPDH* взяты из статьи (Филюшин и др., 20236); для генов *PSY1* и *PSY2* – из (Кулакова и др., 2023).

внутри каждого сорта (кожица vs мякоть), так и между сортами (кожица vs кожица, мякоть vs мякоть).

Транскриптомные данные были валидированы с помощью ПЩР-РВ: в тех же тканях плода была определена экспрессия генов CHS1, CHS2, F3H, DFR, ANS (флавоноидный путь), PSY1 и PSY2 (каротиногенез) (рис. 4). Было показано, что характер экспрессии этих генов согласуется с транскриптомными данными, за исключением несущественных различий в соотношении уровней экспрессии генов PSY1 и PSY2 в мякоти плода между сортами (см. рис. 4).

Обсуждение

Морфологическое разнообразие сортов баклажана стало предметом многих исследований, что способствует оптимизации селекции новых сортов с улучшенными характеристиками (Martínez-Ispizua et al., 2021). Особое внимание уделяется метаболитам (содержание, регуляция синтеза/накопления), обладающим антиоксидантными свойствами и/или определяющими онтогенез/стрессоустойчивость и вкусовые качества плода (Martínez-Ispizua et al., 2021). Рассматриваемые нутрицевтики главным образом включают полифенолы, аскорбиновую кислоту, каротиноиды и, реже, гликоалкалоиды и сахара (Gürbüz et al., 2018; Akhbari et al., 2019; Condurache et al., 2021; Martínez-Ispizua et al., 2021; Saha et al., 2023).

В нашем исследовании были охарактеризованы образцы двух сортов баклажана *S. melongena*, различающиеся окраской кожицы плода: Снежный (белая окраска) и Влас (фиолетовая окраска) (см. рис. 1). Характеристика включала содержание суммы антоцианов, суммы каротиноидов и растворимых сахаров в кожице и мякоти плода (СМ и PR) в сопровождении анализа экспрессии генов, кодирующих ключевые стадии метаболизма данных соединений в тканях биологически спелого плода (PR).

С помощью биохимического анализа подтверждено, что фиолетовая окраска кожицы плода сорта Влас определяется присутствием антоцианов (см. рис. 2, a). Значительно более высокое содержание каротиноидов в кожице плода сорта Влас в сравнении с мякотью, а также плодом сорта Снежный (см. рис. 2, δ), не сказывается на окраске плода, видимо, в силу наличия большого количества антоцианов.

В случае плода сорта Влас содержание и тех, и других пигментов существенно снижалось при переходе от технической к биологической спелости (см. рис. 2, a, δ). Это может быть связано со снижением экспрессии генов биосинтеза данных метаболитов или с ускоренным катаболизмом пигментных соединений. Уменьшение концентрации наблюдалось и для растворимых сахаров (см. рис. 2, e, e). Эти результаты соответствуют снижению вкусовых и антиоксидантных характеристик плода на стадии биологической спелости и объясняют коммерческое использование плодов технической спелости.

Согласно транскриптомному анализу, гены метаболизма антоцианов, каротиноидов и сахаров дифференциально экспрессируются как между тканями плода внутри одного сорта, так и между сортами (см. таблицу). Это, предположительно, определяет внутри- и межсортовые различия в содержании соответствующих соединений в тканях плода.

В целом полученные данные по экспрессии генов флавоноидного пути соответствуют ранее показанному профилю их экспрессии у сортов баклажана с белой и фиолетовой кожицей (Филюшин и др., 2023б). Согласно этим данным, именно между технической и биологической стадиями созревания плода происходят значительные изменения экспрессии генов флавоноидного пути, за счет чего снижается содержание антоцианов в кожице плода фиолетовоплодного сорта.

Неожиданностью стала существенно более высокая экспрессия ключевых генов флавоноидного пути (до антоциановой ветви) в плоде сорта Снежный в сравнении с плодом сорта Влас (см. рис. 3), что говорит о возможности синтеза в плоде сорта Снежный большего количества флавоноидов, исключая антоцианы. Поскольку в плоде сорта Снежный содержание каротиноидов минимально, а экспрессия генов флавоноидного пути сравнительно высока, можно предположить, что желтая окраска зрелого плода (PR) сорта Снежный (см. рис. 1, *a*) связана с накоплением флавоноидов (бесцветных или имеющих желтую окраску). Это отличает плоды баклажана от плодов родственных видов томата (*S. lycopersicum*) и перца (*Capsicum annuum*), окраска которых связана с накоплением каротиноидов (Филюшин и др., 2020).

Кроме того, эти результаты противоположны данным немногочисленных исследований по сравнению содержания фенольных соединений у белых и фиолетовых плодов баклажана, которые свидетельствуют о большем накоплении их фиолетовыми плодами (Martínez-Ispizua et al., 2021; Colak et al., 2022). Обе эти работы включали анализ только одного белоплодного сорта (Martínez-Ispizua et al., 2021; Colak et al., 2022), как и в нашем случае. Таким образом, белоплодные сорта баклажана могут существенно различаться друг с другом по содержанию фенольных соединений и, следовательно, по антиоксидантной активности.

Показанный в работе профиль экспрессии генов, с которых начинается биосинтез каротиноидов, - генов фитоинсинтаз, PSY1, PSY2, соответствует известной специфичности активности каждого из двух изоферментов к определенному типу пластид (Rosas-Saavedra, Stange, 2016). Так, PSY1, кодирующий хромопласт-специфичный фермент, экспрессировался в следовых количествах, тогда как хлоропласт-специфичному PSY2 соответствовало на порядок больше транскриптов (см. рис. 3, 4). В то же время высокий уровень экспрессии гена 9-цисэпоксикаротиноиддиоксигеназы (NCED1), катализирующей превращение каротиноидов β,β-ветви в АБК (Rosas-Saavedra, Stange, 2016), в плоде сорта Влас, и его следовые количества в плоде сорта Снежный (см. рис. 3) предполагают повышенное содержание АБК в фиолетовом плоде. С учетом комплексных функций АБК (Waadt et al., 2022) этот факт может говорить о большей эффективности процессов развития, созревания и ответа на стрессовые факторы у фиолетового плода по сравнению с белым плодом.

Содержание АБК положительно связывают с количеством антоцианов и растворимых сахаров (Teribia et al., 2016), хотя количество последних не коррелирует с накоплением фенольных соединений, а также каротиноидов (Martínez-Ispizua et al., 2021).

Концентрация растворимых сахаров регулируется, в том числе, гидролизом (с помощью инвертаз) и транспортом между тканями (с помощью транспортеров сахаров) (Liu et al., 2022; Ren et al., 2022; Филюшин и др., 2023а). Семейство инвертаз включает нейтральные/щелочные (N/AINV) и кислые (вакуолярные и клеточные стенки; VINV/CWINV) ферменты, участвующие в регуляции онтогенеза и стрессоустойчивости растений (Qian et al., 2016), как и унипортеры сахаров семейства SWEET (Fan et al., 2023; Filyushin et al., 2023).

В сравнении с сортом Влас плоды сорта Снежный содержали больше гексоз и меньше сахарозы (см. рис. 2), что, на первый взгляд, находится в противоречии с более низкой активностью генов инвертаз (см. рис. 3). Однако

2024

эти расхождения могут быть следствием неполного соответствия плодов двух анализируемых сортов по степени биологической спелости. Спелые сочные плоды характеризуются увеличенными клетками с крупными вакуолями, которые активно накапливают и хранят сахара (Hedrich et al., 2015). В кожице и мякоти плода сорта Снежный наблюдается наиболее высокий уровень экспрессии единственного найденного ДЭГ вакуолярной инвертазы VINV1 (см. рис. 3), что соответствует наиболее высокому содержанию там гексоз (см. рис. 2) и, вероятно, является признаком полной биологической спелости плода. В то же время в плоде сорта Влас высоко экспрессируются инвертазы клеточной стенки и нейтральные/щелочные инвертазы (см. рис. 3), работающие в цитоплазме и хлоропластах (Qian et al., 2016), где гексозы активно утилизируются для процессов развития (Hedrich et al., 2015). То есть анализируемый плод данного сорта, возможно, еще не достиг полного созревания и находится на промежуточной стадии, предшествующей биологической спелости. Наблюдаемая межсортовая разница по содержанию сахаров в плоде может быть следствием транспортной регуляции их количества, в том числе с помощью унипортеров семейства SWEET (Filyushin et al., 2023).

Заключение

В настоящей работе проведена сравнительная характеристика спелого плода двух сортов баклажана *S. melongena* с белой (Снежный) и фиолетовой (Влас) окраской кожицы с использованием биохимического и транскриптомного анализов. Показано, что фиолетовая окраска плода сорта Влас связана с присутствием антоцианов и сопровождается повышенным накоплением каротиноидов и сахарозы, что согласуется с профилем экспрессии генов, связанных с ключевыми стадиями метаболизма данных соединений и транспортом растворимых сахаров. В сравнении с сортом Влас плод сорта Снежный характеризуется большим количеством гексоз и, возможно, флавоноидов.

Список литературы / References

Кулакова А.В., Щенникова А.В., Кочиева Е.З. Экспрессия генов биогенеза каротиноидов в процессе длительного холодового хранения клубней картофеля. *Генетика*. 2023;59(8):914-928. DOI 10.31857/S001667582308009X

[Kulakova A.V., Shchennikova A.V., Kochieva E.Z. Expression of carotenoid biosynthesis genes during the long-term cold storage of potato tubers. *Russ. J. Genet.* 2023;59(8):794-807. DOI 10.1134/S1022795423080094]

Филюшин М.А., Джос Е.А., Щенникова А.В., Кочиева Е.З. Зависимость окраски плодов перца от соотношения основных пигментов и профиля экспрессии генов биосинтеза каротиноидов и антоцианов. Физиология растений. 2020;67(6):644-653. DOI 10.31857/S0015330320050048

[Filyushin M.A., Dzhos E.A., Shchennikova A.V., Kochieva E.Z. Dependence of pepper fruit colour on basic pigments ratio and expression pattern of carotenoid and anthocyanin biosynthesis genes. *Russ. J. Plant Physiol.* 2020;67(6):1054-1062. DOI 10.1134/S1021 443720050040]

Филюшин М.А., Слугина М.А., Щенникова А.В., Кочиева Е.З. Дифференциальный профиль экспрессии генов унипортеров сахаров семейства *SWEET* в регуляции качественных признаков плода у видов томата (*Solanum* секция Lycopersicon). *Физиолосия растений*. 2023а;70(4):354-364. DOI 10.31857/S001533032 360002X [Filyushin M.A., Slugina M.A., Shchennikova A.V., Kochieva E.Z. Differential expression of sugar uniporter genes of the *SWEET* family in the regulation of qualitative fruit traits in tomato species (*Solanum* section Lycopersicon). *Russ. J. Plant Physiol.* 2023a; 70(4):70. DOI 10.1134/S102144372360023X]

Филюпин М.А., Щенникова А.В., Кочиева Е.З. Коэкспрессия структурных и регуляторных генов флавоноидного пути выявляет особенности биосинтеза антоцианов в органах баклажана (*Solanum melongena* L.). *Физиология растений*. 20236;70(3):241-250. DOI 10.31857/S0015330322600747

[Filyushin M.A., Shchennikova A.V., Kochieva E.Z. Coexpression of structural and regulatory genes of the flavonoid pathway reveals the characteristics of anthocyanin biosynthesis in eggplant organs (*Solanum melongena* L.). *Russ. J. Plant Physiol.* 2023b;70:27. DOI 10.1134/S1021443722603147]

- Akhbari M., Hamedi S., Aghamiri Z.S. Optimization of total phenol and anthocyanin extraction from the peels of eggplant (*Solanum melongena* L.) and biological activity of the extracts. J. Food Measure. Character. 2019;13:3183-3197. DOI 10.1007/s11694-019-00241-1
- Alappat B., Alappat J. Anthocyanin pigments: beyond aesthetics. Molecules. 2020;25(23):5500. DOI 10.3390/molecules25235500
- Colak N., Kurt-Celebi A., Gruz J., Strnad M., Hayirlioglu-Ayaz S., Choung M.G., Esatbeyoglu T., Ayaz F.A. The phenolics and antioxidant properties of black and purple versus white eggplant cultivars. *Molecules*. 2022;27(8):2410. DOI 10.3390/molecules27082410
- Condurache N.N., Croitoru C., Enachi E., Bahrim G.E., Stanciuc N., Rapeanu G. Eggplant peels as a valuable source of anthocyanins: extraction, thermal stability and biological activities. *Plants.* 2021; 10:577. DOI 10.3390/Plants10030577
- Fan X.W., Sun J.L., Cai Z., Zhang F., Li Y.Z., Palta J.A. MeSWEET15a/b genes play a role in the resistance of cassava (Manihot esculenta Crantz) to water and salt stress by modulating sugar distribution. Plant Physiol. Biochem. 2023;194:394-405. DOI 10.1016/j.plaphy. 2022.11.027
- Filyushin M.A., Anisimova O.K., Shchennikova A.V., Kochieva E.Z. Genome-wide identification, expression, and response to *Fusarium* infection of the *SWEET* gene family in garlic (*Allium sativum* L.). *Int. J. Mol. Sci.* 2023;24(8):7533. DOI 10.3390/ijms24087533
- Gürbüz N., Uluişikb S., Frarya A., Fraryc A., Doğanlara S. Health benefits and bioactive compounds of eggplant. *Food Chem.* 2018; 268:602. DOI 10.1016/j.foodchem.2018.06.093
- Hedrich R., Sauer N., Neuhaus H.E. Sugar transport across the plant vacuolar membrane: nature and regulation of carrier proteins. *Curr. Opin. Plant Biol.* 2015;25:63-70. DOI 10.1016/j.pbi.2015.04.008
- Hirakawa H., Shirasawa K., Miyatake K., Nunome T., Negoro S., Ohyama A., Yamaguchi H., Sato S., Isobe S., Tabata S., Fukuoka H. Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. *DNA Res.* 2014;21:649. DOI 10.1093/dnares/dsu027
- Jiang W., Li N., Zhang D., Meinhardt L., Cao B., Li Y., Song L. Elevated temperature and drought stress significantly affect fruit quality and activity of anthocyanin-related enzymes in jujube (*Ziziphus jujuba* Mill. cv. 'Lingwuchangzao'). *PLoS One.* 2020;15(11):e0241491. DOI 10.1371/journal.pone.0241491
- Keunen E., Peshev D., Vangronsveld J., Van Den Ende W., Cuypers A. Plant sugars are crucial players in the oxidative challenge during abiotic stress: extending the traditional concept. *Plant Cell Environ*. 2013;36(7):1242-1255. DOI 10.1111/pce.12061
- Lelario F., De Maria S., Rivelli A.R., Russo D., Milella L., Bufo S.A., Scrano L. A complete survey of glycoalkaloids using LC-FTICR-MS and IRMPD in a commercial variety and a local landrace of eggplant (*Solanum melongena* L.) and their anticholinesterase and antioxidant activities. *Toxins* (*Basel*). 2019;11(4):230. DOI 10.3390/ toxins11040230
- Liu Y.H., Song Y.H., Ruan Y.L. Sugar conundrum in plant-pathogen interactions: roles of invertase and sugar transporters depend on pathosystems. J. Exp. Bot. 2022;73(7):1910-1925. DOI 10.1093/jxb/ erab562

- Martínez-Ispizua E., Calatayud Á., Marsal J.I., Mateos-Fernández R., Díez M.J., Soler S., Valcárcel J.V., Martínez-Cuenca M.R. Phenotyping local eggplant varieties: commitment to biodiversity and nutritional quality preservation. *Front. Plant Sci.* 2021;12:696272. DOI 10.3389/fpls.2021.696272
- Pérez-Torres I., Castrejón-Téllez V., Soto M.E., Rubio-Ruiz M.E., Manzano-Pech L., Guarner-Lans V. Oxidative stress, plant natural antioxidants, and obesity. *Int. J. Mol. Sci.* 2021;22(4):1786. DOI 10.3390/ijms22041786
- Qian W., Yue C., Wang Y., Cao H., Li N., Wang L., Hao X., Wang X., Xiao B., Yang Y. Identification of the invertase gene family (INVs) in tea plant and their expression analysis under abiotic stress. *Plant Cell Rep.* 2016;35(11):2269-2283. DOI 10.1007/s00299-016-2033-8
- Ren R., Wan Z., Chen H., Zhang Z. The effect of inter-varietal variation in sugar hydrolysis and transport on sugar content and photosynthesis in *Vitis vinifera* L. leaves. *Plant Physiol. Biochem.* 2022;189: 1-13. DOI 10.1016/j.plaphy.2022.07.031
- Rosas-Saavedra C., Stange C. Biosynthesis of carotenoids in plants: enzymes and color. Subcell. Biochem. 2016;79:35-69. DOI 10.1007/ 978-3-319-39126-7 2
- Saha P., Singh J., Bhanushree N., Harisha S.M., Tomar B.S., Rathinasabapathi B. Eggplant (*Solanum melongena* L.) nutritional and health promoting phytochemicals. In: Kole C. (Ed.). Compendium of Crop Genome Designing for Nutraceuticals. Singapore: Springer Nature Singapore, 2023;1463-1493. DOI 10.1007/978-981-19-4169-6_53
- Shi J., Zuo J., Xu D., Gao L., Wang Q. Effect of low-temperature conditioning combined with methyl jasmonate treatment on the chil-

ling resistance of eggplant (*Solanum melongena* L.) fruit. J. Food Sci. Technol. 2019;56(10):4658-4666. DOI 10.1007/s13197-019-03917-0

- Tao T., Hu W., Yang Y., Zou M., Zhou S., Tian S., Wang Y. Transcriptomics reveals the molecular mechanisms of flesh colour differences in eggplant (*Solanum melongena*). *BMC Plant Biol.* 2023;23(1):5. DOI 10.1186/s12870-022-04002-z
- Teribia N., Tijero V., Munné-Bosch S. Linking hormonal profiles with variations in sugar and anthocyanin contents during the natural development and ripening of sweet cherries. *Nat. Biotechnol.* 2016; 33(6):824-833. DOI 10.1016/j.nbt.2016.07.015
- Waadt R., Seller C.A., Hsu P.K., Takahashi Y., Munemasa S., Schroeder J.I. Plant hormone regulation of abiotic stress responses. *Nat. Rev. Mol. Cell Biol.* 2022;23(10):680-694. DOI 10.1038/s41580-022-00479-6
- Yang G., Li L., Wei M., Li J., Yang F. SmMYB113 is a key transcription factor responsible for compositional variation of anthocyanin and color diversity among eggplant peels. *Front. Plant Sci.* 2022;13: 843996. DOI 10.3389/fpls.2022.843996
- You Q., Li H., Wu J., Li T., Wang Y., Sun G., Li Z., Sun B. Mapping and validation of the epistatic *D* and *P* genes controlling anthocyanin biosynthesis in the peel of eggplant (*Solanum melongena* L.) fruit. *Hortic. Res.* 2022;10(2):uhac268. DOI 10.1093/hr/uhac268
- Zhang Y., Hu Z., Chu G., Huang C., Tian S., Zhao Z., Chen G. Anthocyanin accumulation and molecular analysis of anthocyanin biosynthesis-associated genes in eggplant (*Solanum melongena* L.). *J. Agric. Food Chem.* 2014;62:2906. DOI 10.1021/jf404574c

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 13.05.2024. После доработки 27.06.2024. Принята к публикации 28.06.2024.

DOI 10.18699/vjgb-24-70

Импутация генотипов в геномных исследованиях человека

А.А. Бердникова (D^{1, 2}, И.В. Зоркольцева (D¹, Я.А. Цепилов (D¹, Е.Е. Елгаева (D^{1, 2} 🖂

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия ² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия © elizabeth.elgaeva@gmail.com

> Аннотация. Импутация – это метод, позволяющий восстанавливать недостающую информацию о генетических вариантах, которые не удалось генотипировать напрямую с помощью ДНК-микрочипов или секвенирования с низким покрытием. Импутация играет важнейшую роль в полногеномном анализе ассоциаций (genome wide associations study, GWAS). Она приводит к существенному увеличению количества изучаемых вариантов, что повышает разрешающую способность метода и увеличивает сопоставимость данных, полученных в разных когортах и/или с помощью разных технологий, что важно при проведении метаанализов. При ее выполнении информацию о генотипах в исследуемой выборке, у которой известна только часть генетических вариантов, дополняют за счет эталонной (референсной) выборки, имеющей более полные данные о генотипах (чаше всего это результаты полногеномного секвенирования). Импутация стала неотъемлемой частью геномных исследований человека благодаря преимуществам, которые она дает, а также увеличению доступности инструментов для импутации и данных референсных выборок. Обзор посвящен импутации в геномных исследованиях человека. В первом разделе приводятся описание технологий получения информации о генотипах человека и характеристика получаемых типов данных. Во втором разделе представлена методология импутации, перечисляются этапы ее проведения и соответствующие программы, дается описание наиболее популярных референсных панелей и способов оценки качества импутации. В заключении представлены примеры использования импутации в геномных исследованиях выборок из России. Настояций обзор показывает важность проведения импутации, дает информацию о том, как ее выполнять, и систематизирует результаты ее применения на примере российских выборок.

> Ключевые слова: импутация; генотипирование; секвенирование; полногеномный анализ ассоциаций; человек; ДНК-микрочип.

Для цитирования: Бердникова А.А., Зоркольцева И.В., Цепилов Я.А., Елгаева Е.Е. Импутация генотипов в геномных исследованиях человека. *Вавиловский журнал генетики и селекции*. 2024;28(6):628-639. DOI 10.18699/vjgb-24-70

Финансирование. Работа БАА и ЕЕЕ поддержана грантом Российского научного фонда № 22-15-20037 и Правительством Новосибирской области. Работа ЗИВ и ЦЯА финансирована за счет бюджетного проекта № FWNR-2022–0020.

Благодарности. Авторы выражают благодарность В.С. Фишману за рекомендации по улучшению текста статьи.

Genotype imputation in human genomic studies

A.A. Berdnikova (D^{1, 2}, I.V. Zorkoltseva (D¹, Y.A. Tsepilov (D¹, E.E. Elgaeva (D^{1, 2}

¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
 ² Novosibirsk State University, Novosibirsk, Russia
 ² elizabeth.elgaeva@gmail.com

Abstract. Imputation is a method that supplies missing information about genetic variants that could not be directly genotyped with DNA microarrays or low-coverage sequencing. Imputation plays a critical role in genome-wide association studies (GWAS). It leads to a significant increase in the number of studied variants, which improves the resolution of the method and enhances the comparability of data obtained in different cohorts and/or by using different technologies, which is important for conducting meta-analyses. When performing imputation, genotype information from the study sample, in which only part of the genetic variants are known, is complemented using the standard (reference) sample, which has more complete genotype data (most often the results of whole-genome sequencing). Imputation has become an integral part of human genomic research due to the benefits it provides and the increasing availability of imputation tools and reference sample data. This review focuses on imputation in human genomic research. The first section of the review provides a description of technologies for obtaining information about human genotypes and characteristics of these types of data. The second section describes the imputation methodology, lists the stages of its implementation and the corresponding programs, provides a description of the most popular reference panels and methods for assessing the quality of imputation. The review concludes with examples of the use of imputation in genomic studies of samples from Russia. This review shows the importance of imputation, provides information on how to carry it out, and systematizes the results of its application using Russian samples.

Key words: imputation; genotyping; sequencing; genome-wide association study; human; DNA-microarray.

For citation: Berdnikova A.A., Zorkoltseva I.V., Tsepilov Y.A., Elgaeva E.E. Genotype imputation in human genomic studies. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(6):628-639. DOI 10.18699/vjgb-24-70

Технологии получения данных о генотипе человека и их особенности

Данные генотипа человека являются ключевым аспектом для многих генетических исследований. Существует несколько технологий, разработанных для прочтения, анализа и интерпретации генетической информации. Наиболее часто используемые методы включают в себя секвенирование по Сэнгеру, секвенирование следующего поколения (next generation sequencing, NGS) и ДНК-микрочипы.

Генотипирование с помощью ДНК-микрочипов

ДНК-микрочип (или просто «микрочип», «чип», однако не следует путать с РНК-микрочипом, что является другой технологией), представляет собой небольшую стеклянную или кремниевую подложку, к которой в заданном порядке с большой плотностью прикреплены десятки тысяч зондов (коротких фрагментов одноцепочечной ДНК, комплементарных определенным нуклеотидным последовательностям). Эти зонды расположены на чипе так, чтобы каждый фрагмент можно было идентифицировать по его местоположению (рис. 1).

В ходе анализа к исследуемым молекулам ДНК, разрезанным на фрагменты эндонуклеазами рестрикции, присоединяют флуоресцентные маркеры и помещают их на чип. Целевые фрагменты ДНК связываются с комплементарными ДНК-зондами, а все оставшиеся удаляют с чипа. Для детекции флуоресценции фрагментов, записи картины эмиссии (излучения) и последующей идентификации последовательностей используют лазерные лучи и компьютерную обработку. Это очень быстрый метод, позволяющий одновременно определять нуклеотидную последовательность сразу нескольких фрагментов ДНК (Govindarajan et al., 2012).

Альтернативный подход к решению задачи генотипирования был реализован академиком А.Д. Мирзабековым в отечественных разработках по созданию гелевых микрочипов (Мирзабеков, 2003). Они представляют собой подложку из стекла, пластика или силикона с фиксированными на ее поверхности полусферическими каплями гидрогеля. Отличием этого метода является то, что фрагменты ДНК оказываются иммобилизованными в трехмерном пространстве, что обеспечивает большую чувствительность и емкость микрочипа. Данная технология также нашла свое применение в анализе РНК, белковых и клеточных биочипах.

Существует несколько стратегий идентификации однонуклеотидного полиморфизма (single nucleotide polymorphism, SNP) для микрочипов (рис. 2).

Аллель-специфичная гибридизация (allele discrimination by hybridization) (рис. 2, *a*). Меченая целевая ДНК гибридизуется с зондами, содержащими полиморфный сайт в центре. Правильно спаренные олигонуклеотиды получаются стабильнее (имеют большую температуру плавления), по сравнению с дуплексами с некомплементарным основанием. Поэтому после промывки чипа в жестких температурных условиях на нем остаются лишь верно спаренные цепочки. Принято использовать несколько фрагментов для каждого аллеля, чтобы улучшить качество сигнала по отношению к шуму (Wang D.G. et al., 1998).

GoldenGate-анализ компании Illumina (рис. 2, б) – два аллель-специфических олигонуклеотида, каждый из которых имеет 5'-конец с разными универсальными праймерами, P1 и P2 (праймеры метят уникальным флуорофором для последующего различия сайтов), гибридизуются в растворе с геномной ДНК. Третий олигонуклеотид, помимо универсального праймера, P3, имеет хвостовую часть с последовательностью «штрих-код», комплементарную фрагменту на чипе. Удлиненные полимеразой аллель-специфичные праймеры лигируются с третьим олигонуклеотидом, после чего полученные фрагменты



Псевдоцвет (красный, желтый или зеленый) определяется количеством связавшихся с зондом молекул, помеченных разными красителями. Дальнейшие пояснения к рисунку см. в тексте.



Рис. 2. Стратегии обнаружения SNP для ДНК-микрочипов.

a – аллель-специфичная гибридизация; б – GoldenGate-анализ компании Illumina; в – матричное удлинение праймера.

амплифицируют с помощью полимеразной цепной реакции и гибридизуют на чип. Использование нескольких штрих-кодов (по одному для каждого интересующего локуса) позволяет проводить анализ сразу для нескольких участков генома (Fan et al., 2003).

Матричное удлинение праймера (arrayed primer extension, APEX) (рис. 2, *в*). Здесь чип содержит фрагмент ДНК, 5'-конец которого закреплен на подложке, а 3'-конец заканчивается нуклеотидом, предшествующим определяемому SNP. Фрагменты геномной ДНК гибридизуются с чипом, при этом искомый SNP остается неспаренным. В ходе реакции секвенирования прикрепленная к подложке последовательность нуклеотидов удлиняется на один терминирующий нуклеотид, помеченный красителем (Kurg et al., 2000). Этот нуклеотид предотвращает дальнейший рост цепи ДНК, а цвет его красителя позволяет определить, какой из нуклеотидов (А, Т, Г или Ц) находится в данной позиции.

Одно из основных преимуществ ДНК-микрочипов – высокая пропускная способность (Hayat, 2002; Brown et al., 2024). Микрочип обеспечивает основу для одновременного генотипирования тысяч различных локусов и выявления однонуклеотидных замен. Так, микрочипы используются для анализа выборок больших объемов с целью генотипировать часто встречаемые генетические варианты (с частотой минорного аллеля в популяции >0.01).

Однако существуют и некоторые ограничения в интерпретации результатов. Данные микрочипов обычно являются бинарными (указывают на наличие или отсутствие определенного аллеля), высокопроизводительными (позволяют проанализировать тысячи и миллионы SNP) и требующими специальных методов анализа для извлечения значимой информации. Речь в данном случае идет о программном обеспечении (например, GenomeStudio (Illumina Inc., San Diego, CA, США)), которое включает в себя инструменты для контроля качества, определения генотипа, визуализации и анализа данных. Кроме того, микрочипы могут давать как ложноположительные, так и ложноотрицательные результаты. Эти проблемы подчеркивают важность тщательной интерпретации данных и необходимость использования соответствующих статистических методов для контроля качества и подтверждения результатов.

Секвенирование генома

В этом подразделе описаны различные технологии секвенирования. Примерно в 1976 г. было разработано два метода, которые могли декодировать сотни оснований за полдня - обрыв цепи Сэнгера и Коулсона и химическое расщепление Максама и Гилберта (Maxam, Gilbert, 1977; Sanger et al., 1977). В обоих методах исследуемую ДНК распределяют по четырем пробиркам с разным составом реакционной смеси для конкретного типа азотистого основания (А, Т, Г или Ц). В методе Гилберта используются ДНК, радиоактивно меченная с одного конца, и смесь ферментов, специфично ее разрезающих перед нуклеотидом определенного вида. Секвенирование по Сэнгеру, напротив, содержит праймеры и дидезоксинуклеотиды, приводящие к прекращению синтеза цепи при включении радиоактивно-меченого дидезоксинуклеозидтрифосфата (ddNTP; в каждой пробирке свой). Таким образом, в результате применения обоих методов в каждой пробирке образуются меченые фрагменты ДНК различной длины, заканчивающиеся одинаковым основанием. Последовательности разделяют по длине с помощью электрофореза на полиакриламидных пластинчатых гелях (по одной дорожке на каждый тип основания) с разрешением в один нуклеотид. Изображение, полученное на рентгеновской пленке после электрофореза, позволяет восстановить исходную последовательность. Описанные методы сразу же вошли в обиход, а к 1987 г. автоматизированные флуоресцентные секвенаторы Сэнгера могли прочитывать уже



Рис. 3. Секвенирование третьего поколения. *a* – Pacific Biosciences; *6* – Oxford Nanopore Technology. Пояснения см. в тексте ниже.

около 1000 оснований в день (Smith et al., 1986; Connell et al., 1987).

В 2005 г. впервые были представлены технологии секвенирования второго поколения (next generation sequencing, NGS), в основе которых лежат два подхода. Первый из них – секвенирование путем гибридизации (sequencing by hybridization, SBH). Суть метода состоит в следующем: сначала короткие участки ДНК фиксируют на стеклянной подложке (ДНК-чипе). Затем фрагменты, подлежащие идентификации, метят флуорофором и наносят на чип для гибридизации с зафиксированными участками. Одноцепочечную ДНК смывают, и картина гибридизации считывается по цветным меткам и их яркости. Альтернативным подходом в NGS является секвенирование путем синтеза (sequencing by synthesis, SBS) (Shendure et al., 2017).

В технологиях, использующих методику SBS, как правило, заранее фрагментированные последовательности закрепляют в проточной ячейке, где происходит циклический синтез новой цепи. Путем последовательного добавления одного из четырех дезоксинуклеотидов в смесь, при удалении заранее из нее предыдущих, можно считать сигналы из тех ячеек, где реакция синтеза прошла успешно. Тем самым на выходе получают информацию о том, где и какой нуклеотид находится.

Технологии секвенирования с подходом, отличным от NGS, были впервые описаны в 2008–2009 гг. и получили название «секвенирование третьего поколения» (Check Hayden, 2009). Они включают два основных подхода (рис. 3).

Первая технология, Pacific Biosciences (PacBio) (Rhoads, Au, 2015), заключается в оптическом наблюдении синтеза ДНК с помощью полимеразы в режиме реального времени. В конструкции имеется отверстие меньше половины длины световой волны, ограничивающее флуоресцентное возбуждение небольшим объемом, в котором находятся только полимераза и ее матрица (см. рис. 3, *a*). При таком устройстве только флуоресцентно-меченые нуклеотиды, включенные в растущую цепь ДНК, излучают сигналы достаточной длительности, чтобы быть считанными. Частота ошибок при этом методе секвенирования очень высока (около 10 %), но они распределены случайным образом. Благодаря длинным прочтениям и толерантности к высокому содержанию GC и случайным ошибкам, технология РасВіо позволяет получить сборки *de novo* беспрецедентного качества в отношении точности и непрерывности.

Вторая из основных технологий секвенирования третьего поколения - Oxford Nanopore (ONT) (Deamer et al., 2016). Данная методика впервые была предложена в 1980-х гг. Специальная камера, где происходит процесс секвенирования, заполнена электролитическим раствором и поделена двухслойной мембраной с нанопорой (ее размеры находятся в нанодиапазоне). После подключения напряжения ионы электролита и молекула ДНК начинают двигаться через пору. Нуклеиновая кислота физически мешает миграции ионов, что приводит к колебаниям силы тока, по которым можно определять нуклеотидную последовательность (см. рис. 3, б). Основным отличием от других технологий секвенирования является исключительная портативность устройств с нанопорами, которые могут быть размером с карту памяти (USB), поскольку они основаны на обнаружении электронных, а не оптических сигналов.

Сравнение технологий и их применения для решения разных задач

Наиболее часто для крупномасштабных проектов (полногеномного секвенирования, анализа транскриптома и эпигенетического профилирования) используют NGS технологии Illumina, однако для проведения сборки *de novo* больше соответствует РасВіо, а для портативного секвенирования – ONT. Метод Сэнгера подходит для секвенирования коротких фрагментов ДНК, таких как отдельные гены, плазмиды или вирусные геномы.

Следует также упомянуть конкурирующий с Illumina секвенатор, разработанный компаниями Complete Genomics и MGI Tech, DNBSEQ-T7 (ранее известный как MGISEQ-T7). У DNBSEQ-T7 процесс клональной амплификации происходит по типу катящегося кольца, т.е. всегда с исходной матрицы, что позволяет исключить накопление ошибок ДНК-полимеразы (Drmanac et al., 2010). Основные преимущества MGI включают меньшую в сравнении с Illumina стоимость и возможность обработки большего объема образцов за короткое время. Как показывают последние исследования, новый секвенатор MGISEQ-2000 может быть использован в качестве полноценной альтернативы секвенаторам Illumina при проведении полногеномных исследований (поиск вариантов, выявление инделов); различия между двумя платформами незначительны (Korostin et al., 2020; Jeon et al., 2021; Feng et al., 2024).

Недавно была показана эффективность применения полногеномного секвенирования (whole-genome sequencing, WGS) для проведения GWAS (DePristo et al., 2011; Chat et al., 2022). Этот подход – многообещающая альтернатива для генотипирования с помощью ДНК-микрочипов, поскольку дает возможность получить информацию о большей доле генетических вариаций, повышая мощность тестов ассоциации и последующего анализа тонкого картирования (fine-mapping) (Wang Q.S., Huang, 2022). Однако, несмотря на снижение стоимости технологий на основе NGS, для GWAS используются главным образом высокопроизводительные и относительно дешевые ДНКмикрочипы, содержащие от сотен тысяч до миллионов общих генетических маркеров, которые дают возможность протестировать практически весь геном на наличие ассоциаций с изучаемым признаком.

Генотипирование SNP с помощью ДНК-микрочипов может содержать до 5 % ошибок в зависимости от производителя (Lamy et al., 2006; Yang et al., 2011; Guo et al., 2014). Однако существующие протоколы по контролю качества полученных данных позволяют заметно снизить количество ошибок (в среднем на 1.7 %) (Zhao et al., 2018). Таким образом, микрочипы дают возможность достаточно точно генотипировать образцы даже для видов с высокой гетерозиготностью (т.е. с большей генетической изменчивостью по сравнению с ожидаемой при равновесии Харди-Вайнберга) (Bourke et al., 2018). Более того, на конец 2023 г. цена генотипирования образца на микрочипе была на порядок меньше стоимости секвенирования NGS, что позволяет при одном и том же бюджете проекта охватить больший размер выборки. Основной их недостаток при проведении GWAS - они не дают обнаружить ассоциацию между SNP и признаком в том случае, если данный генетический вариант не представлен на микрочипе.

Дополнительные трудности в использовании ДНКмикрочипов могут быть вызваны тем, что информация (например, расположение SNP на хромосоме), использованная для дизайна чипа, устарела или отличается у разных производителей. Вышеперечисленные проблемы могут быть решены импутацией данных генотипирования (Pasaniuc et al., 2012). Этот подход позволяет увеличить плотность охвата изучаемых генетических вариантов (общее число маркеров) и долю общих вариантов при проведении метаанализа (объединения данных разных исследований и/или платформ генотипирования) (Li Y. et al., 2009).

Заменой ДНК-микрочипов может стать WGS с низким покрытием (low-coverage, lcWGS), в котором секвенируются случайные области генома (Chat et al., 2022). Исследования показывают, что lcWGS значительно превосходит микрочипы по плотности распределения маркеров, что также дает возможность более тщательно оценивать ассоциации с менее распространенными вариантами. Такие данные также требуют импутации с использованием гаплотипов (например, из проекта «1000 геномов») (Auton et al., 2015). Затраты WGS со сверхнизким покрытием (глубина секвенирования $\leq 0.5x$) могут быть сравнимы или ниже, чем при использовании ДНК-микрочипов, однако его потенциал в качестве альтернативы еще тщательно не оценен (Martin et al., 2021).

Секвенирование ДНК и генотипирование решают задачу анализа генетической информации по-разному. Так, секвенирование позволяет читать фрагменты ДНК целиком и потому применяется для выявления редких (частота минорного аллеля < 0.01 %) и *de novo* мутаций, широко используется для изучения структуры отдельных генов или участков генома. Генотипирование, с другой стороны, - более быстрый и экономичный метод анализа генетической изменчивости, что особенно полезно для крупномасштабных геномных исследований, включающих тысячи или даже миллионы образцов. Таким образом, если цель исследования состоит в том, чтобы всесторонне исследовать генетическую архитектуру признака или заболевания, секвенирование, вероятно, является лучшим подходом. Однако если фокус исследования сосредоточен на частых генетических вариантах или анализе популяционной или родственной структуры выборки, то генотипирование часто бывает достаточным и более эффективным (Gresham et al., 2008).

Импутация данных генотипирования

Хотя секвенирование всего генома у сотен тысяч людей пока трудноосуществимо, можно добиться значительного прогресса, определяя лишь относительно небольшое количество генетических вариантов у каждого человека. Такой тип «неполной» информации все равно полезен, потому что данные о любом наборе SNP в группе людей дают возможность делать выводы о многих других вариантах, не наблюдаемых у тех же людей. Подход, это осуществляющий, называется импутацией.

Методология

Процедура импутации предполагает следующие этапы: контроль качества данных генотипирования, фазирование, собственно импутация и на завершающем этапе – контроль качества импутированных генотипов (рис. 4).

Генетические варианты, расположенные рядом на хромосоме, имеют большую вероятность наследоваться вместе, что происходит из-за того, что на каждую хромосому приходится всего несколько рекомбинаций. Данный принцип получил название «неравновесие по сцеплению» (linkage disequilibrium, LD). Благодаря этому принципу мы наблюдаем блоки гаплотипов (гаплоблоки) – наборы близко расположенных генетических вариантов, которые в ходе эволюции наследовались вместе.

В импутации гаплоблоки используются для выявления общих коротких участков ДНК на хромосомах, которые индивиды в случайно выбранной популяции могли унаследовать от общего предка. Путем сравнения гаплотипов в двух выборках (исследуемой и референсной) по набору общих генетических вариантов алгоритмы импутации предоставляют вывод о генотипах исследуемых индивидов. Обе эти выборки должны относиться к одной этнической группе, чтобы импутация давала точные результаты (Mills et al., 2020).

Хотя данные генотипирования не содержат информации о гаплотипах, их можно вывести и реконструировать с помощью поэтапного анализа. Фазирование – это процесс статистической оценки гаплотипов. Импутация может быть реализована как на необработанных данных нефазированного генотипирования, так и на реконструированных смешанных гаплотипах, хотя известно, что фазирование приводит к повышению точности импутации (Anderson et al., 2010). Кроме того, фазирование зачастую необходимо ввиду того, что стандартные алгоритмы для импутации (подробнее ниже) работают именно с гаплоблоками.

Контроль качества данных генотипирования

Важным шагом любого геномного исследования является проведение контроля качества данных. Значимость этого этапа иллюстрирует пример со статьей, опубликованной в "Science", которая была отозвана из-за недостаточного учета технических ошибок при генотипировании на чипе Illumina (Marees et al., 2018).

Контроль качества данных ДНК-микрочипового генотипирования подразделяется на два основных шага: контроль на уровне индивидов и контроль на уровне маркеров. Контроль на уровне индивидов заключается в удалении из выборки образца в следующих случаях (Anderson et al., 2010):

- наблюдается несоответствие между фенотипом и генотипом (в частности, фенотипический пол отличается от генетического);
- количество гетерозиготных локусов в геноме отклоняется от ожидаемого значения (завышение или занижение этого показателя может свидетельствовать о контаминации образца или об инбридинге соответственно);
- в выборке присутствуют дубликаты, родственники первой или второй степени родства (схожие генотипы будут представлены избыточно, вследствие чего частоты аллелей в популяции могут быть отображены недостоверно);
- имеет другое этническое происхождение, т.е. наблюдается стратификация популяции (наиболее распространенный подход для выявления таких индивидов – анализ главных компонент (principal component analysis, PCA) на матрице родства).

Контроль качества данных на уровне отдельных маркеров также состоит из нескольких пунктов, которые предполагают удаление SNP в случае, если:

- частота минорного аллеля (minor allele frequency, MAF)
 < 0.01;
- они отсутствуют у большой части индивидов в выборке;
- значительно отклоняются от равновесия Харди–Вайнберга.

Для проведения контроля качества используют ряд готовых программ, находящихся в открытом доступе: PLINK 1.9/PLINK 2 (Purcell et al., 2007; Chang et al., 2015), RICOLI (Lam et al., 2020), SMARTPCA (Price et al., 2006) и FlashPCA (Abraham et al., 2017).

Инструменты для импутации

За последние 20 лет несколько различных исследовательских групп разработали и опубликовали ряд инструментов для фазирования и последующей импутации, большинство из которых основано на скрытой марковской модели (Hidden Markov Model, HMM) Ли и Стивенса (LS) (Li N., Stephens, 2003). Эта статистическая модель, впервые описанная в 2003 г., предполагает, что гаплотипы наследуются в виде гаплоблоков и события рекомбинации происходят на их границах. Модель вероятностно реконструирует ис-



Рис. 4. Импутация данных генотипирования.

1 – фазирование; 2 – собственно импутация. Пояснения см. в тексте.



Рис. 5. Визуализация работы алгоритмов, основанных на HMM, для четырех индивидов из референсной выборки.

Каждая колонка – отдельный SNP с двумя аллелями (пустые и зачеркнутые квадраты обозначают разные аллели одного SNP), а каждая пара строк представляет собой две копии ДНК (от каждого из родителей). Тесно связанные SNP объединены в группы по цветам; каждый гаплотип моделируется как мозаика цветных комбинаций (Scheet, Stephens, 2006).

следуемые гаплотипы в виде мозаики, составленной из гаплотипов небольшой референсной выборки (рис. 5). Было показано, что методы, основанные на HMM Ли и Стивенса, точнее и эффективнее (Weale, 2004) таких подходов, как, например, алгоритм Кларка (Clark, 1990) или EM-алгоритм (expectation-maximization – максимизация ожидания) (Dempster et al., 1977) (Browning S.R., Browning B.L., 2011). В настоящее время наиболее часто применяемыми программами, использующими HMM Ли и Стивенса, являются Beagle 5 (Browning et al., 2021), Eagle 2 (Loh et al., 2016) и ShapeIT (Delaneau et al., 2012) для фазирования, а также Beagle 5 (Browning et al., 2018), Impute 5 (Rubinacci et al., 2020) и Minimac 4 (Das et al., 2016) – для импутации. Beagle 5 и ShapeIT 2 при этом позволяют проводить обе эти процедуры.

Сравнительный анализ текущего программного обеспечения (ПО) для фазирования и импутации показал, что в целом Beagle5.4 несколько лучше, чем Impute5 и Minimac4. Эта программа отличается более высоким коэффициентом конкордации и высокой производительностью даже на больших наборах данных (De Marino et al., 2022). Однако Minimac4 и Impute5, как правило, лучше работают с редкими вариантами, потому что, в отличие от Beagle 5.4, которая вычисляет кластеры гаплотипов и выполняет вычисления на их основе, эти программы выполняют поиск во всем пространстве гаплотипов. Программа Minimac4 требует наименьшего объема памяти, но вычисления занимают больше времени. Если использование памяти ограничено, а потеря точности приемлема, тогда Minimac4 может быть оптимальным выбором программного обеспечения для импутации.

Вышеперечисленные программы предполагают запуск с локального сервера и требуют наличия референсных гаплотипов. Но большинства таких крупномасштабных наборов данных нет в открытом доступе. По этой причине наиболее часто для импутации данных человека используются специальные серверы, содержащие информацию о разных референсных панелях, такие как Michigan Imputation Server¹ (Das et al., 2016) и TOPMed Imputation Server²

¹ https://imputationserver.sph.umich.edu/index.html#!pages/home

² https://imputation.biodatacatalyst.nhlbi.nih.gov/#!pages/home

(Das et al., 2016). Исследователи могут загружать туда свои наборы данных, настраивать параметры через пользовательский веб-интерфейс (выбирать инструменты, референсные панели и т.д.), выполнять фазирование и импутацию генотипов на сервере и загружать выходные файлы.

В качестве недостатков данного подхода нужно отметить необходимость отправки своих данных за пределы локального сервера (хотя и с использованием безопасных протоколов соединения) и возможные очереди. Кроме того, пользователи часто ограничены выбором программ или референсных панелей, не могут объединять несколько панелей или интегрировать свои. Однако есть возможность обойти данные ограничения, например с помощью ПО Docker (Das et al., 2016), и запускать импутацию на своем сервере. Проблема автономного запуска заключается в чуть большей сложности из-за настроек вручную, где пользователю нужно устанавливать дополнительные программы для конвейера и учитывать конфликты библиотек. В Приложении 1³ можно сравнить инструментарий, доступный на двух описанных выше серверах.

Референсные панели для импутации данных генотипирования человека

Один из важных вопросов при импутации генотипов заключается в том, как выбрать референсную панель, обеспечивающую высокую точность импутации в интересующей популяции. Так, было показано (Huang, Tseng, 2014), что на качество импутации влияет не только размер панели, но и этнический состав референсной выборки. Наиболее часто используемыми панелями для европейских популяций на сегодняшний день являются: 1000 Genomes (Sudmant et al., 2015), Haplotype Reference Consortium (HRC) (Haplotype Reference Consortium, 2016) и Trans-Omics for Precision Medicine (TOPMed) (Taliun et al., 2021).

Референсная панель 1000 Genomes Phase 3 Version 5 была подготовлена в рамках проекта «1000 геномов» в 2008 г. (Auton et al., 2015). Всего в этом проекте путем использования комбинации полногеномного секвенирования с низким покрытием, экзомного секвенирования с высоким покрытием и генотипирования на микрочипах удалось охарактеризовать 88 млн генетических вариантов (84.7 млн SNP, 3.6 млн коротких вставок/делеций и 60000 структурных вариантов). Данная версия референсной панели включает в себя 49 млн маркеров от 2504 индивидов из смешанной популяции.

Референсная панель HRC r1.1 2016 набрана консорциумом HRC (The Haplotype Reference Consortium) с целью создания крупной справочной панели гаплотипов. Панель HRC объединяет наборы данных из 20 различных исследований, большинство из которых получены с помощью полногеномного секвенирования с низким покрытием (4–8x) и состоят из выборок индивидов преимущественно европейского происхождения. Референсная панель состоит из 64976 гаплотипов 32 тыс. индивидов с 39235157 SNP, делеций и вставок не содержит.

Проект TOPMed (The Trans-Omics for Precision Medicine) был инициирован в 2010 г. с целью сбора и анализа

³ Приложения 1–3 см. по адресу:

https://vavilovj-icg.ru/download/pict-2024-28/appx22.pdf

данных полногеномного секвенирования. По состоянию на сентябрь 2021 г. ТОРМеd насчитывает около 180 тыс. участников, преимущественно неевропейского происхождения из более 85 различных исследований. На базе данных ТОРМеd была создана референсная панель, которая включает 286068980 SNP; 5815513 вставок и 16222592 делеций в генотипах 97256 индивидов. Эти генетические варианты распределены по 22 аутосомам и X-хромосоме. ТОРМеd (Version r3) – первая панель, которая основана исключительно на данных глубокого полногеномного секвенирования и при этом значительно превосходит ранее опубликованные альтернативы.

Несмотря на то что большинство генетических исследований и референсных панелей ориентированы на выборки индивидов европейского происхождения, нужно отметить существование различных проектов, направленных на изучение генетического разнообразия других популяций. В их число входят ChinaMAP (10588 образцов и 136.7 млн SNP) (Li L. et al., 2021), NARD (1779 индивидов, 40.6 млн SNP) (Yoo et al., 2019), GAsP (1739 образцов, 1 млн аутосомных SNP) (Wall et al., 2019), SG10K (4810 образцов, 89.1 млн SNP) (Wu et al., 2019) для выборок людей азиатского происхождения; AFAM (2269 образцов, 45 млн SNP) (O'Connell et al., 2021) и UGR (4778 образцов, 2.2 млн маркеров) (Fatumo et al., 2022) для афроамериканцев. Панель TOPMed также может быть использована для импутации неевропейских выборок индивидов как африканского, так и азиатского происхождения.

Идеальным решением при выборе панели для импутации стало объединение данных от нескольких референсных выборок для построения комбинированной эталонной панели. Однако в различных исследованиях, как правило, используются разные стратегии контроля качества и фильтрации вариантов, что может затруднить объединение результатов.

Еще одна важная проблема – ограничения на совместное использование данных. Например, информация о генотипе на индивидуальном уровне во многих референсных панелях не является общедоступной; поэтому может быть невозможно напрямую объединить ее с результатами секвенирования других выборок. В связи с этим был предложен метод метаимпутации (Yu et al., 2022). Вместо того чтобы объединять эталонные панели, сначала импутируют генотипы с использованием нескольких эталонных панелей по отдельности, а затем объединяют импутированные результаты в согласованный набор данных.

Оценка качества импутации

Качество импутации данных генотипирования можно оценить: 1) с помощью стандартных метрик качества импутации; 2) опытным путем (например, провести GWAS на исследуемом признаке и проверить воспроизводимость известных из литературы сигналов ассоциации или вычислить полигенную оценку признака и сравнить ее с реальными фенотипами).

Метрики качества импутации также можно разделить на две большие группы (Stahl et al., 2021): 1) те, что оценивают качество импутации без использования генотипированных напрямую SNP и вычисляются автоматически при запуске соответствующего ПО для импутации, и 2) те, что позволяют сравнить импутированные SNP с генотипами и вычисляются вручную.

Метрики качества из первой группы специфичны для каждой отдельной программы. Так, например, для программ Minimac4 и Beagle 5 оценивается показатель R^2 (Marchini, Howie, 2010), который при этом считается поразному для каждой из программ, в то время как Impute 5 вычисляет параметр Info (Marchini, Howie, 2010). Из-за своей специфичности они не подходят для сравнения качества данных, импутированных разными методами. С этой задачей успешно справляются метрики из второй группы, к которым относятся: коэффициент конкордации (concordation rate, CR), показатель качества импутации IQS (imputation quality score) (Lin et al., 2010), показатель Хеллингера (Hellinger score) (Roshyara et al., 2014), квадрат евклидовой нормы (SEN, squared Euclidean norm score) (Roshyara et al., 2014) и другие. На практике чаще всего используют стандартные метрики первой группы.

При импутации оцениваются апостериорные вероятности генотипа. Так, для биаллельных SNP в аддитивной модели (где генотип закодирован как 0, 1 и 2, а референсный и альтернативный аллель – 0 и 1 соответственно) оцененная вероятность индивиду *i* иметь генотип *j* в конкретном локусе обозначается G_j^i (j = 0, 1, 2). Этот показатель рассчитывается соответствующим ПО для импутации по данным референсной и целевой выборок с помощью встроенных алгоритмов (например, скрытой марковской модели, как описано выше). Доза альтернативного аллеля рассчитывается как $D_i = G_1^i + 2G_2^i$.

Метрика *R*² представляет собой аппроксимацию квадрата корреляции между дозой импутированного аллеля и ожидаемым генотипом и считается как отношение дисперсии дозы аллеля и ожидаемой дисперсии при равновесии Харди–Вайнберга

$$\hat{R}_{d}^{2} = \frac{\frac{1}{N} \sum_{i=1}^{N} (D_{i} - 2\hat{p})^{2}}{2\hat{p}(1 - \hat{p})},$$

$$(1)$$

$$\hat{p} = \sum_{i=1}^{N} \frac{D_{i}}{2N},$$

где N – количество индивидов в выборке; D_i – доза импутированного аллеля для *i*-го индивида; \hat{p} – оценка частоты аллеля.

Многие современные алгоритмы (как, например, Minimac), проводят импутацию на предварительно фазированных генотипах, т. е. работают с гаплотипами. Формула претерпевает небольшие изменения, так как теперь множество генотипов описывается как пул из 2N бинарно закодированных аллелей

$$\hat{R}_{h}^{2} = \frac{\frac{1}{2N} \sum_{i=1}^{2N} (H_{i} - \hat{p})^{2}}{\hat{p}(1 - \hat{p})},$$

$$\hat{p} = \sum_{i=1}^{2N} \frac{H_{i}}{2N},$$
(2)

где H_i – вероятность импутированного аллеля в *i*-м гаплотипе (варьируется от 0 до 1 и оценивается встроенными

алгоритмами скрытых марковских моделей); N – размер выборки; \hat{p} – оценка частоты аллеля (вывод формул представлен в Приложениях 2 и 3).

При расчете метрик второго типа часть информации о генотипах в исследуемой выборке искусственно «маскируется» (удаляется из общего набора данных с сохранением информации об этих SNP). После чего полученные пропуски импутируют и сравнивают с реальными генотипами. Например, CR представляет собой долю правильно рассчитанных SNP ко всем SNP. Показатель Хеллингера является мерой расстояния между двумя распределениями вероятностей генотипа, в его основе лежит коэффициент Бхаттачарии (Bhattacharyya, 1946), который измеряет степень перекрытия между двумя распределениями. Показатель SEN представляет собой масштабированное евклидово расстояние между распределением истинной дозы и импутированной. Как показатель Хеллингера, так и показатель SEN рассчитываются для отдельных SNP каждого индивида. Метрика IQS основана на каппа-статистике Коэна и позволяет учитывать случайные совпадения импутированных и реальных SNP.

Как было сказано выше, помимо перечисленных метрик, для контроля качества импутации можно использовать полигенную оценку признака (polygenic score, PGS) (Choi et al., 2020). Она представляет собой показатель индивидуального генетического риска в отношении некоторого признака, полученный путем суммирования количественной оценки эффекта многих распространенных вариантов (обычно с частотой минорных аллелей ≥ 1 %) в геноме, каждый из которых может делать небольшой вклад в генетический риск человека для данного признака или заболевания. Показатель PGS обычно вычисляется как взвешенная сумма набора генетических вариантов, как правило, SNP, определяемых как вариации одной пары оснований из референсного генома. Полученная оценка имеет распределение, близкое к нормальному в общей популяции, причем высокие оценки указывают на более высокий риск.

В общем виде уравнение для расчета взвешенной оценки полигенного риска для отдельного индивида выглядит следующим образом (Collister et al., 2022):

$$PGS_i = \sum_{j=1}^{M} \hat{\beta} * dosage_{ij},$$

где M – количество SNP в модели; $\hat{\beta}_j$ – оценка размера эффекта *j*-го варианта; dosage_{*ij*} – генотип в кодировке 0, 1, 2 для *j*-го варианта в генотипе *i*-го индивида. Величину эффекта SNP (β) часто получают из результатов GWAS.

После расчета PGS оценки признака ее значения сравнивают со значениями реальных фенотипов. В случае если наблюдается достоверная корреляция между этими двумя наборами данных, мы можем сделать вывод о высоком качестве данных после импутации.

Примеры импутации в геномных исследованиях на российских выборках

Несмотря на описанные выше преимущества импутации и фазирования, существует крайне мало упоминаний их использования в исследованиях российских выборок. Так, в работе (Pinakhina et al., 2022), посвященной изучению депрессии в выборке из 4520 индивидов из различных регионов России, проведена импутация с помощью референсных панелей HRC и 1000G с использованием программы Beagle 5.1. Аналогично в исследовании генетической структуры западнорусской популяции (выборка из 4145 индивидов) в качестве панели для импутации была выбрана панель HRC, сама процедура проводилась с помощью Beagle 4.0 и позволила рассматривать в анализе, помимо 623249 генотипированных вариантов, еще 10454514 импутированных (Usoltsev et al., 2023). А в работе (Moreland et al., 2022) по изучению маркеров, ассоциированных с мышечной силой и мощностью у 292 россиян (из них 83 профессиональные атлеты), проведены не только импутация на панели 1000G, но и фазирование с помощью SHAPEIT.

Один из важнейших факторов для осуществления качественной импутации – правильный выбор референсной панели. N. Kolosov с коллегами (2022) оценивали надежность импутации генотипов выборки из 230 пожилых людей из Санкт-Петербурга (501100 SNP) такими панелями как HRC, 1000G, HGDP (Human Genome Diversity Project (Cann et al., 2002) – референсная панель на основе 929 человек различной этнической принадлежности). Им удалось увеличить с помощью Beagle 5.1 (данные были заранее фазированы) общее количество исследуемых вариантов до 37.6, 37.5 и 26.6 млн SNP для каждой из панелей соответственно. Помимо этого, HRC, по сравнению с двумя другими панелями, показала наибольшую точность импутации (метрики IQS и CR).

Все эти работы в качестве референсных панелей используют HRC или 1000G, однако данный подход несколько устарел и подлежит пересмотру в связи с появлением более крупного набора данных TOPMed, применение которой служит своего рода золотым стандартом в зарубежных исследованиях на сегодняшний день. Из программного обеспечения в рассмотренных работах применяют различные версии Beagle.

В упомянутых исследованиях на российских выборках не проводились метаанализы или тонкое картирование генов, однако, как показывают примеры других работ (Barton et al., 2021), именно благодаря импутации и фазированию такие анализы могут быть выполнены существенно более качественно.

Заключение

В настоящее время проведение импутации данных генотипирования – неотъемлемая часть многих геномных исследований человека, в частности GWAS. Она обеспечивает увеличение числа анализируемых SNP и дает возможность объединять результаты разных исследований. Импутация также существенно улучшает результаты тонкого картирования, позволяя наиболее точно определять конкретные генетические варианты и гены, обусловливающие ассоциацию всего участка генома с изучаемым признаком (Chundru et al., 2019).

Следует отметить, что для крупномасштабных исследований, где важны размер выборки и покрытие генотипирования, комбинация ДНК-микрочипов/секвенирования с низким покрытием и дальнейшей импутацией является наиболее оптимальной и дешевой стратегией получения данных, подходящих для большинства дизайнов геномных исследований. Эта комбинация используется во всех крупных национальных биобанках, как, например, UK Biobank (Sudlow et al., 2015), AllOfUs (Ramirez et al., 2022) и другие.

Наряду с перечисленными достоинствами метод импутации имеет ряд недостатков и ограничений. В частности, ошибки чтения при низком покрытии, а также некорректный выбор параметров для импутации наравне с неподходящей референсной панелью часто приводят к низкой точности импутированных данных, что может отрицательно влиять на результаты дальнейших этапов анализа. Необходимо также помнить, что импутация использует информацию о гаплотипах из референсной выборки, поэтому при ее устаревании генетические варианты, которые стали частыми в популяции сравнительно недавно, могут импутироваться хуже (Ali et al., 2022). Помимо этого, высокий уровень рекомбинации снижает точность фазирования и последующей импутации генотипов, в связи с чем в некоторых случаях необходим дополнительный анализ рекомбинаций (Weng et al., 2014).

Кроме того, импутация может сгладить генетические различия между индивидами в выборках типа «случай– контроль» (Lau et al., 2024): в импутированных данных могут возникнуть неточные генотипы в регионах, где ожидаются различия между случаем и контролем. При этом такой эффект проявляется независимо от того, насколько велика и разнообразна референсная панель. Наконец, при использовании метода важно помнить, что то, что верно для популяции в целом, не всегда может быть верно для конкретного индивида.

Существует большое разнообразие программ и референсных панелей для проведения импутации геномных данных человека, и, как следствие, множество их комбинаций. Благодаря этому у исследователей есть возможность подобрать оптимальный набор инструментов для импутации, наиболее подходящий под особенности выборки и задачи конкретного исследования. Обзор работ на российских выборках показал, что наиболее популярное ПО для импутации – Beagle различных версий, а среди референсных панелей чаще всего используют HRC и 1000G, что несколько отличается от зарубежной практики, где лидером среди референсных панелей является TOPMed.

Большая осведомленность о тонкостях проведения импутации и осознанный подход к подбору инструментов позволят улучшить качество геномных данных без увеличения стоимости их получения, облегчат их интеграцию с результатами других исследований и дадут более точную информацию о генетическом контроле признаков человека.

Список литературы / References

- Мирзабеков А.Д. Биочипы в биологии и медицине XXI века. Вестн. Рос. акад. наук. 2003;73(5):412.
 - [Mirzabekov A.D. Biochips in the biology and medicine of the XXI century. *Vestnik Rossiyskoj Akademii Nauk* = *Herald of the Russian Academy of Sciences*. 2003;73(5):412 (in Russian)]
- Abraham G., Qiu Y., Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics*. 2017; 33(17):2776-2778. DOI 10.1093/bioinformatics/btx299

- Ali A.T., Liebert A., Lau W., Maniatis N., Swallow D.M. The hazards of genotype imputation in chromosomal regions under selection: A case study using the lactase gene region. *Ann. Hum. Genet.* 2022; 86(1):24-33. DOI 10.1111/ahg.12444
- Anderson C.A., Pettersson F.H., Clarke G.M., Cardon L.R., Morris A.P., Zondervan K.T. Data quality control in genetic case-control association studies. *Nat. Protoc.* 2010;5(9):1564-1573. DOI 10.1038/nprot.2010.116
- Auton A., Abecasis G.R., Altshuler D.M., Durbin R.M., Abecasis G.R., Bentley D.R., ... Min Kang H., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A., Abecasis G.R. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. DOI 10.1038/ nature15393
- Barton A.R., Sherman M.A., Mukamel R.E., Loh P.-R. Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses. *Nat. Genet.* 2021;53(8):1260-1269. DOI 10.1038/s41588-021-00892-1
- Bhattacharyya A. On a measure of divergence between two multinomial populations. *Sankhyā: Ind. J. Stat.* 1946;7(4):401-406
- Bourke P.M., Voorrips R.E., Visser R.G.F., Maliepaard C. Tools for genetic studies in experimental populations of polyploids. *Front. Plant. Sci.* 2018;9:513. DOI 10.3389/fpls.2018.00513
- Brown A., Ampratwum P.O., Ray S.D. Microarray analysis. In: Encyclopedia of Toxicology. 4 ed. 2024;6:385-392. DOI 10.1016/B978-0-12-824315-2.00210-4
- Browning B.L., Zhou Y., Browning S.R. A One-penny imputed genome from next-generation reference panels. Am. J. Hum. Genet. 2018;103(3):338-348. DOI 10.1016/j.ajhg.2018.07.015
- Browning B.L., Tian X., Zhou Y., Browning S.R. Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. 2021;108(10): 1880-1890. DOI 10.1016/j.ajhg.2021.08.005
- Browning S.R., Browning B.L. Haplotype phasing: existing methods and new developments. *Nat. Rev. Genet.* 2011;12(10):703-714. DOI 10.1038/nrg3054
- Cann H.M., de Toma C., Cazes L., Legrand M.F., Morel V., Piouffre L., Bodmer J., ... Zhu S., Weber J.L., Greely H.T., Feldman M.W., Thomas G., Dausset J., Cavalli-Sforza L.L. A human genome diversity cell line panel. *Science*. 2002;296(5566):261-262. DOI 10.1126/ science.296.5566.261b
- Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., Lee J.J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4(1):7. DOI 10.1186/s13742-015-0047-8
- Chat V., Ferguson R., Morales L., Kirchhoff T. Ultra low-coverage whole-genome sequencing as an alternative to genotyping arrays in genome-wide association studies. *Front. Genet.* 2022;12:790445. DOI 10.3389/fgene.2021.790445
- Check Hayden E. Genome sequencing: the third generation. *Nature*. 2009;457(7231):768-769. DOI 10.1038/news.2009.86
- Choi S.W., Mak T.S.-H., O'Reilly P.F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* 2020;15(9):2759-2772. DOI 10.1038/s41596-020-0353-1
- Chundru V.K., Marioni R.E., Prendergast J.G.D., Vallerga C.L., Lin T., Beveridge A.J., Gratten J., Hume D.A., Deary I.J., Wray N.R., Visscher P.M., McRae A.F. Examining the impact of imputation errors on fine-mapping using DNA methylation QTL as a model trait. *Genetics*. 2019;212(3):577-586. DOI 10.1534/genetics.118.301861
- Clark A.G. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol. Biol. Evol.* 1990;7(2):111-122. DOI 10.1093/oxfordjournals.molbev.a040591
- Collister J.A., Liu X., Clifton L. Calculating polygenic risk scores (PRS) in UK biobank: A practical guide for epidemiologists. *Front. Genet.* 2022;13:818574. DOI 10.3389/fgene.2022.818574
- Connell C., Fung S., Heiner C., Bridgham J., Chakerian V., Heron E., Jones B., Menchen S., Mordan W., Raff M., Recknor M., Smith L.M., Springer J., Woo S., Hunkapiller M. Automated DNA-sequence analysis. *Biotechniques*. 1987;5:342-348

- Das S., Forer L., Schönherr S., Sidore C., Locke A.E., Kwong A., Vrieze S.I., Chew E.Y., Levy S., McGue M., Schlessinger D., Stambolian D., Loh P.-R., Iacono W.G., Swaroop A., Scott L.J., Cucca F., Kronenberg F., Boehnke M., Abecasis G.R., Fuchsberger C. Nextgeneration genotype imputation service and methods. *Nat. Genet.* 2016;48(10):1284-1287. DOI 10.1038/ng.3656
- De Marino A., Mahmoud A.A., Bose M., Bircan K.O., Terpolovsky A., Bamunusinghe V., Bohn S., Khan U., Novković B., Yazdi P.G. A comparative analysis of current phasing and imputation software. *PLoS One.* 2022;17(10):e0260177. DOI 10.1371/journal. pone.0260177
- Deamer D., Akeson M., Branton D. Three decades of nanopore sequencing. *Nat. Biotechnol.* 2016;34(5):518-524. DOI 10.1038/nbt. 3423
- Delaneau O., Marchini J., Zagury J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods*. 2012;9(2):179-181. DOI 10.1038/nmeth.1785
- Dempster A.P., Laird N.M., Rubin D.B. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statist. Society. 1977;39(1):1-38. DOI 10.1111/j.2517-6161.1977.tb01600.x
- DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., del Angel G., Rivas M.A., Hanna M., McKenna A., Fennell T.J., Kernytsky A.M., Sivachenko A.Y., Cibulskis K., Gabriel S.B., Altshuler D., Daly M.J. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 2011;43(5):491-498. DOI 10.1038/ ng.806
- Drmanac R., Sparks A.B., Callow M.J., Halpern A.L., Burns N.L., Kermani B.G., Carnevali P., ... Drmanac S., Oliphant A.R., Banyai W.C., Martin B., Ballinger D.G., Church G.M., Reid C.A. Human genome sequencing using unchained base reads on selfassembling DNA nanoarrays. *Science*. 2010;327(5961):78-81. DOI 10.1126/science.1181498
- Fan J.B., Oliphant A., Shen R., Kermani B.G., Garcia F., Gunderson K.L., Hansen M., ... Kruglyak S., Bentley D., Haas J., Rigault P., Zhou L., Stuelpnagel J., Chee M.S. Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* 2003;68:69-78. DOI 10.1101/ sqb.2003.68.69
- Fatumo S., Mugisha J., Soremekun O.S., Kalungi A., Mayanja R., Kintu C., Makanga R., Kakande A., Abaasa A., Asiki G., Kalyesubula R., Newton R., Nyirenda M., Sandhu M.S., Kaleebu P. Uganda genome resource: A rich research database for genomic studies of communicable and non-communicable diseases in Africa. *Cell Genom.* 2022;2(11):100209. DOI 10.1016/j.xgen.2022.100209
- Feng Z., Peng F., Xie F., Liu Y., Zhang H., Ma J., Xing J., Guo X. Comparison of capture-based mtDNA sequencing performance between MGI and illumina sequencing platforms in various sample types. *BMC Genomics*. 2024;25(1):41. DOI 10.1186/s12864-023-09938-6
- Govindarajan R., Duraiyan J., Kaliyappan K., Palanisamy M. Microarray and its applications. J. Pharm. Bioallied Sci. 2012;4(6):310. DOI 10.4103/0975-7406.100283
- Gresham D., Dunham M.J., Botstein D. Comparing whole genomes using DNA microarrays. *Nat. Rev. Genet.* 2008;9(4):291-302. DOI 10.1038/nrg2335
- Guo Y., He J., Zhao S., Wu H., Zhong X., Sheng Q., Samuels D.C., Shyr Y., Long J. Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* 2014;9(11):2643-2662. DOI 10.1038/nprot.2014.174
- Hayat M.A. DNA microarrays technology. In: Handbook of Immunohistochemistry and *in situ* Hybridization of Human Carcinomas. 2002;49-55. DOI 10.1016/S1874-5784(04)80015-1
- Huang G.-H., Tseng Y.-C. Genotype imputation accuracy with different reference panels in admixed populations. *BMC Proc.* 2014;8(S1): S64. DOI 10.1186/1753-6561-8-S1-S64
- Jeon S.A., Park J.L., Park S.-J., Kim J.H., Goh S.-H., Han J.-Y., Kim S.-Y. Comparison between MGI and illumina sequencing platforms for whole genome sequencing. *Genes Genom.* 2021;43(7): 713-724. DOI 10.1007/s13258-021-01096-x

- Kolosov N., Rezapova V., Rotar O., Loboda A., Freylikhman O., Melnik O., Sergushichev A., Stevens C., Voortman T., Kostareva A., Konradi A., Daly M.J., Artomov M. Genotype imputation and polygenic score estimation in northwestern Russian population. *PLoS One.* 2022;17(6):e0269434. DOI 10.1371/journal.pone.0269434
- Korostin D., Kulemin N., Naumov V., Belova V., Kwon D., Gorbachev A. Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing. *PLoS One.* 2020;15(3):e0230301. DOI 10.1371/journal.pone. 0230301
- Kurg A., Tõnisson N., Georgiou I., Shumaker J., Tollett J., Metspalu A. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genet. Test.* 2000;4(1):1-7. DOI 10.1089/109065700316408
- Lam M., Awasthi S., Watson H.J., Goldstein J., Panagiotaropoulou G., Trubetskoy V., Karlsson R., Frei O., Fan C.-C., De Witte W., Mota N.R., Mullins N., Brügger K., Lee S.H., Wray N.R., Skarabis N., Huang H., Neale B., Daly M.J., Mattheisen M., Walters R., Ripke S. RICOPILI: rapid imputation for COnsortias PIpeLIne. *Bioinformatics*. 2020;36(3):930-933. DOI 10.1093/bioinformatics/ btz633
- Lamy P., Andersen C.L., Wikman F.P., Wiuf C. Genotyping and annotation of Affymetrix SNP arrays. *Nucleic Acids Res.* 2006;34(14):e100. DOI 10.1093/nar/gkl475
- Lau W., Ali A., Maude H., Andrew T., Swallow D.M., Maniatis N. The hazards of genotype imputation when mapping disease susceptibility variants. *Genome Biol.* 2024;25(1):7. DOI 10.1186/s13059-023-03140-3
- Li L., Huang P., Sun X., Wang S., Xu M., Liu S., Feng Z., Zhang Q., Wang X., Zheng X., Dai M., Bi Y., Ning G., Cao Y., Wang W. The ChinaMAP reference panel for the accurate genotype imputation in Chinese populations. *Cell Res.* 2021;31(12):1308-1310. DOI 10.1038/s41422-021-00564-z
- Li N., Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*. 2003;165(4):2213-2233. DOI 10.1093/genetics/165.4.2213
- Li Y., Willer C., Sanna S., Abecasis G. Genotype imputation. Annu. Rev. Genomics Hum. Genet. 2009;10(1):387-406. DOI 10.1146/ annurev.genom.9.081307.164242
- Lin P., Hartz S.M., Zhang Z., Saccone S.F., Wang J., Tischfield J.A., Edenberg H.J., Kramer J.R., Goate A.M., Bierut L.J., Rice J.P. A new statistic to evaluate imputation reliability. *PLoS One.* 2010; 5(3):e9697. DOI 10.1371/journal.pone.0009697
- Loh P.-R., Danecek P., Palamara P.F., Fuchsberger C., Reshef Y.A., Finucane H.K., Schoenherr S., Forer L., McCarthy S., Abecasis G.R., Durbin R., L Price A. Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* 2016;48(11):1443-1448. DOI 10.1038/ng.3679
- Marchini J., Howie B. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 2010;11(7):499-511. DOI 10.1038/ nrg2796
- Marees A.T., de Kluiver H., Stringer S., Vorspan F., Curis E., Marie Claire C., Derks E.M. A tutorial on conducting genome wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 2018;27(2). DOI 10.1002/mpr.1608
- Martin A.R., Atkinson E.G., Chapman S.B., Stevenson A., Stroud R.E., Abebe T., Akena D., ... Ramesar R., Shiferaw W., Stein D.J., Teferra S., van der Merwe C., Zingela Z. Low-coverage sequencing cost-effectively detects known and novel variation in underrepresented populations. *Am. J. Hum. Genet.* 2021;108(4):656-668. DOI 10.1016/j.ajhg.2021.03.012
- Maxam A.M., Gilbert W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* 1977;74(2):560-564. DOI 10.1073/pnas.74. 2.560
- Mills M.C., Barban N., Tropf F.C. An Introduction to Statistical Genetic Data Analysis. Cambridge, MA: MIT Press, 2020
- Moreland E., Borisov O.V., Semenova E.A., Larin A.K., Andryushchenko O.N., Andryushchenko L.B., Generozov E.V., Wil-

2024 28•6

liams A.G., Ahmetov I.I. Polygenic profile of elite strength athletes. *J. Strength. Cond. Res.* 2022;36(9):2509-2514. DOI 10.1519/JSC. 00000000003901

- O'Connell J., Yun T., Moreno M., Li H., Litterman N., Kolesnikov A., Noblin E., ... Wang W., Weldon C.H., Wilton P., Wong C., Auton A., Carroll A., McLean C.Y. A population-specific reference panel for improved genotype imputation in African Americans. *Commun. Biol.* 2021;4(1):1269. DOI 10.1038/s42003-021-02777-9
- Pasaniuc B., Rohland N., McLaren P.J., Garimella K., Zaitlen N., Li H., Gupta N., ... Haas D.W., Liang L., Sunyaev S., Patterson N., de Bakker P.I.W., Reich D., Price A.L. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat. Genet.* 2012;44(6):631-635. DOI 10.1038/ng.2283
- Pinakhina D., Yermakovich D., Vergasova E., Kasyanov E., Rukavishnikov G., Rezapova V., Kolosov, ... Plotnikov N., Ilinsky V., Neznanov N., Mazo G., Kibitov A., Rakitko A., Artomov M. GWAS of depression in 4,520 individuals from the Russian population highlights the role of MAGI2 (S-SCAM) in the gut-brain axis. *Front. Genet.* 2022;13:972196. DOI 10.3389/fgene.2022.972196
- Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 2006;38(8):904-909. DOI 10.1038/ng1847
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. PLINK: a tool set for whole-genome association and populationbased linkage analyses. *Am. J. Hum. Genet.* 2007;81(3):559-575. DOI 10.1086/519795
- Ramirez A.H., Sulieman L., Schlueter D.J., Halvorson A., Qian J., Ratsimbazafy F., Loperena R., ... Denny J.C., Carroll R.J., Glazer D., Harris P.A., Hripcsak G., Philippakis A., Roden D.M.; All of Us research program. The *All of Us* research program: Data quality, utility, and diversity. *Patterns (N Y)*. 2022;3(8):100570. DOI 10.1016/j.patter.2022.100570
- Rhoads A., Au K.F. PacBio Sequencing and its applications. *Genomics Proteomics Bioinformatics*. 2015;13(5):278-289. DOI 10.1016/j.gpb. 2015.08.002
- Roshyara N.R., Kirsten H., Horn K., Ahnert P., Scholz M. Impact of pre-imputation SNP-filtering on genotype imputation results. *BMC Genet*. 2014;15(1):88. DOI 10.1186/s12863-014-0088-5
- Rubinacci S., Delaneau O., Marchini J. Genotype imputation using the Positional Burrows Wheeler Transform. *PLoS Genet*. 2020;16(11): e1009049. DOI 10.1371/journal.pgen.1009049
- Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 1977;74(12):5463-5467. DOI 10.1073/pnas.74.12.5463
- Scheet P., Stephens M. A fast and flexible statistical model for largescale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 2006;78(4): 629-644. DOI 10.1086/502802
- Shendure J., Balasubramanian S., Church G.M., Gilbert W., Rogers J., Schloss J.A., Waterston R.H. DNA sequencing at 40: past, present and future. *Nature*. 2017;550(7676):345-353. DOI 10.1038/nature 24286
- Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Heiner C., Kent S.B.H., Hood L.E. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986;321(6071): 674-679. DOI 10.1038/321674a0
- Stahl K., Gola D., König I.R. Assessment of imputation quality: comparison of phasing and imputation algorithms in real data. *Front. Genet.* 2021;12:724037. DOI 10.3389/fgene.2021.724037
- Sudlow C., Gallacher J., Allen N., Beral V., Burton P., Danesh J., Downey P., Elliott P., Green J., Landray M., Liu B., Matthews P., Ong G., Pell J., Silman A., Young A., Sprosen T., Peakman T., Collins R. UK Biobank: An open access resource for identifying the causes

of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779. DOI 10.1371/journal.pmed.1001779

- Sudmant P.H., Rausch T., Gardner E.J., Handsaker R.E., Abyzov A., Huddleston J., Zhang Y., ... Gerstein M.B., Bashir A., Stegle O., Devine S.E., Lee C., Eichler E.E., Korbel J.O. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015; 526(7571):75-81. DOI 10.1038/nature15394
- Taliun D., Harris D.N., Kessler M.D., Carlson J., Szpiech Z.A., Torres R., ... Cupples L.A., Laurie C.C., Jaquish C.E., Hernandez R.D., O'Connor T.D., Abecasis G.R. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021; 590(7845):290-299. DOI 10.1038/s41586-021-03205-y
- The Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 2016;48:1279-1283. DOI 10.1038/ng.3643
- Usoltsev D., Kolosov N., Rotar O., Loboda A., Boyarinova M., Moguchaya E., Kolesova E., ... Laiho P., Kostareva A., Konradi A., Shlyakhto E., Palotie A., Daly M.J., Artomov M. Understanding complex trait susceptibilities and ethnical diversity in a sample of 4,145 Russians through analysis of clinical and genetic data. *bioRxiv*. 2023. DOI 10.1101/2023.03.23.534000
- Wall J.D., Stawiski E.W., Ratan A., Kim H.L., Kim C., Gupta R., Suryamohan K., ... Radha V., Mohan V., Majumder P.P., Seshagiri S., Seo J.-S., Schuster S.C., Peterson A.S. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019; 576(7785):106-111. DOI 10.1038/s41586-019-1793-z
- Wang D.G., Fan J.-B., Siao C.-J., Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., Spencer J., Kruglyak L., Stein L., Hsie L., Topaloglou T., Hubbell E., Robinson E., Mittmann M., Morris M.S., Shen N., Kilburn D., Rioux J., Nusbaum C., Rozen S., Hudson T.J., Lipshutz R., Chee M., Lander E.S. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*. 1998;280(5366):1077-1082. DOI 10.1126/science.280.5366.1077
- Wang Q.S., Huang H. Methods for statistical fine-mapping and their applications to auto-immune diseases. *Semin. Immunopathol.* 2022; 44(1):101-113. DOI 10.1007/s00281-021-00902-8
- Weale M.E. A survey of current software for haplotype phase inference. *Hum. Genomics.* 2004;1(2):141. DOI 10.1186/1479-7364-1-2-141
- Weng Z.-Q., Saatchi M., Schnabel R.D., Taylor J.F., Garrick D.J. Recombination locations and rates in beef cattle assessed from parentoffspring pairs. *Gen. Select. Evol.* 2014;46(1):34. DOI 10.1186/ 1297-9686-46-34
- Wu D., Dou J., Chai X., Bellis C., Wilm A., Shih C.C., ... Wong W.-C., Xie Z., Yeo K.K., Zhang L., Zhai W., Zhao Y. Large-scale wholegenome sequencing of three diverse Asian populations in Singapore. *Cell*. 2019;179(3):736-749.e15. DOI 10.1016/j.cell.2019.09.019
- Yang H.-C., Lin H.-C., Kang M., Chen C.-H., Lin C.-W., Li L.-H., Wu J.-Y., Chen Y.-T., Pan W.-H. SAQC: SNP array quality control. *BMC Bioinformatics*. 2011;12(1):100. DOI 10.1186/1471-2105-12-100
- Yoo S.-K., Kim C.-U., Kim H.L., Kim S., Shin J.-Y., Kim N., Yang J.S.W., Lo K.-W., Cho B., Matsuda F., Schuster S.C., Kim C., Kim J.-I., Seo J.-S. NARD: whole-genome reference panel of 1779 Northeast Asians improves imputation accuracy of rare and lowfrequency variants. *Genome Med.* 2019;11(1):64. DOI 10.1186/ s13073-019-0677-z
- Yu K., Das S., LeFaive J., Kwong A., Pleiness J., Forer L., Schönherr S., Fuchsberger C., Smith A.V., Abecasis G.R. Meta-imputation: An efficient method to combine genotype data after imputation with multiple reference panels. *Am. J. Hum. Genet.* 2022;109(6):1007-1015. DOI 10.1016/j.ajhg.2022.04.002
- Zhao S., Jing W., Samuels D.C., Sheng Q., Shyr Y., Guo Y. Strategies for processing and quality control of Illumina genotyping arrays. *Brief. Bioinform.* 2018;19(5):765-775. DOI 10.1093/bib/bbx012

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 12.12.2023. После доработки 23.04.2024. Принята к публикации 04.07.2024.

DOI 10.18699/vjgb-24-71

Поиск сигналов положительного отбора генов циркадных ритмов *PER1*, *PER2*, *PER3* в различных популяциях людей

А.И. Мишина (D 🔄, С.Ю. Бакоев (D, А.Ю. Ооржак, А.А. Кескинов (D, Ш.Ш. Кабиева (D, А.В. Коробейникова (D, В.С. Юдин (D, М.М. Боброва (D, Д.А. Шестаков, В.В. Макаров (D, А.В. Гетманцева (D)

Центр стратегического планирования и управления медико-биологическими рисками здоровью Федерального медико-биологического агентства, Москва, Россия arinamishina32@yandex.ru

> Аннотация. Разнообразие географически распределенных человеческих популяций демонстрирует большую вариацию внешних и внутренних признаков индивидов. Такие различия в значительной степени объясняются генетической адаптацией к различным воздействиям окружающей среды, к которым относят изменения климатических условий, колебания условий сна и бодрствования, вариации рациона и другие. Полногеномные данные, полученные от людей различных популяций, дают возможность идентифицировать конкретные генетические участки, ответственные за эти адаптации, и глубже понимать генетическую структуру, лежащую в основе адаптивных характеристик человека. В данной работе проведен поиск сигналов однонуклеотидных полиморфизмов (SNP), находящихся под давлением отбора у людей различных популяций. Для выявления сигналов отбора в различных популяционных группах были исследованы гены PER1, PER2 и PER3, играющие важнейшую роль в координации термогенных функций и регуляции циркадных ритмов, что напрямую отражается на адаптационных способностях организма. Анализ данных осуществляли на основе общедоступных данных из проекта «1000 геномов» (1000 Genomes Project) по 23 популяциям. Для поиска следов отбора был выбран статистический метод XP-EHH (expanded haplotype homozygosity score). Проведенный сравнительный анализ позволил идентифицировать точки, подверженные давлению отбора. Найденные SNP были аннотированы через каталог GWAS, а также вручную, путем анализа интернет-ресурсов и публикаций. Исследование позволяет сделать вывод о том, что условия проживания, климат и другие внешние факторы напрямую влияют на генетическую структуру популяций и варьируют в зависимости от расы и географического местоположения. Кроме того, многие из вариантов отбора в генах PER1, PER2, PER3, по-видимому, регулируют биологические процессы, связанные с основными современными заболеваниями, включая ожирение, онкологию, метаболический синдром, биполярное расстройство личности, депрессию, ревматоидный артрит, сахарный диабет, красную волчанку, инсульт и болезнь Альцгеймера, что делает их крайне интересными объектами для дальнейших исследований, направленных на идентификацию генетически обусловленных причин заболеваний человека.

Ключевые слова: популяции; SNP; адаптация; PER1; PER2; PER3.

Для цитирования: Мишина А.И., Бакоев С.Ю., Ооржак А.Ю., Кескинов А.А., Кабиева Ш.Ш., Коробейникова А.В., Юдин В.С., Боброва М.М., Шестаков Д.А., Макаров В.В., Гетманцева Л.В. Поиск сигналов положительного отбора генов циркадных ритмов *PER1*, *PER2*, *PER3* в различных популяциях людей. *Вавиловский журнал генетики и селекции*. 2024;28(6):640-649. DOI 10.18699/vjgb-24-71

Search for signals of positive selection of circadian rhythm genes *PER1*, *PER2*, *PER3* in different human populations

A.I. Mishina (D) (D), S.Y. Bakoev (D), A.Y. Oorzhak, A.A. Keskinov (D), Sh.Sh. Kabieva (D), A.V. Korobeinikova (D), V.S. Yudin (D), M.M. Bobrova (D), D.A. Shestakov, V.V. Makarov (D), L.V. Getmantseva (D)

Centre for Strategic Planning and Management of Biomedical Health Risks of the Federal Medical Biological Agency, Moscow, Russia arinamishina32@yandex.ru

Abstract. The diversity of geographically distributed human populations shows considerable variation in external and internal traits of individuals. Such differences are largely attributed to genetic adaptation to various environmental influences, which include changes in climatic conditions, variations in sleep and wakefulness, dietary variations, and others. Whole-genome data from individuals of different populations make it possible to determine the specific genetic sites responsible for adaptations and to further understand the genetic structure underlying human adaptive characteristics. In this article, we searched for signals of single nucleotide polymorphisms (SNPs) under

© Мишина А.И., Бакоев С.Ю., Ооржак А.Ю., Кескинов А.А., Кабиева Ш.Ш., Коробейникова А.В., Юдин В.С., Боброва М.М., Шестаков Д.А., Макаров В.В., Гетманцева Л.В., 2024

Контент доступен под лицензией Creative Commons Attribution 4.0

selection pressure in people of different populations. To identify selection signals in different population groups, the *PER1*, *PER2* and *PER3* genes that are involved in the coordination of thermogenic functions and regulation of circadian rhythms, which is directly reflected in the adaptive abilities of the organism, were investigated. Data were analyzed using publicly available data from the 1000 Genomes Project for 23 populations. The Extended Haplotype Homozygosity Score statistical method was chosen to search for traces of selection. The comparative analysis performed identified points subject to selection pressure. The SNPs were annotated through the GWAS catalog and manually by analyzing Internet resources. This study suggests that living conditions, climate, and other external factors directly influence the genetic structure of populations and vary across races and geographic locations. In addition, many of the selection variants in the *PER1*, *PER2*, *PER3* genes appear to regulate biological processes that are associated with major modern diseases, including obesity, cancer, metabolic syndrome, bipolar personality disorder, depression, rheumatoid arthritis, diabetes mellitus, lupus erythematosus, stroke and Alzheimer's disease, making them extremely interesting targets for further research aimed at identifying the genetic causes of human disease. **Key words:** populations; SNP; adaptation; *PER1*; *PER2*; *PER3*.

For citation: Mishina A.I., Bakoev S.Y., Oorzhak A.Y., Keskinov A.A., Kabieva Sh.Sh., Korobeinikova A.V., Yudin V.S., Bobrova M.M., Shestakov D.A., Makarov V.V., Getmantseva L.V. Search for signals of positive selection of circadian rhythm genes *PER1*, *PER2*, *PER3* in different human populations. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(6):640-649. DOI 10.18699/vjgb-24-71

Введение

Достижения в методах генотипирования SNP привели к быстрому переходу от исследований, сосредоточенных на пространственно явных нейтральных генетических процессах, к исследованиям, посвященным адаптивным генетическим процессам (Ahrens et al., 2018). Одним из инструментов отслеживания этих процессов является поиск локусов под давлением отбора (Carlson et al., 2005). Уникальные генетические паттерны или следы, оставленные в геномных регионах, подвергшихся отбору, называются «подписи отбора» (selection signatures) (Nielsen, 2005; Jensen et al., 2016; Bakoev et al., 2021). Подписи отбора представляют собой геномные области, содержащие последовательности ДНК, функционально участвующие в генетической изменчивости признаков, подлежащих отбору (Lopez et al., 2015; Bakoev et al., 2023). Такие участки представляют интерес в связи с их значимостью для отслеживания эволюционной биологии и потенциальной связи с генами, которые контролируют фенотипы в диких и домашних популяциях (Xu et al., 2015).

Для идентификации локусов под давлением отбора применяют различные статистические подходы, один из них – анализ гомозиготности по расширенному гаплотипу (ЕНН – extended haplotype homozygosity). Следует отметить, что слово «гомозиготность» как часть термина ЕНН означает вероятность того, что две случайно выбранные хромосомы из популяции идентичны (в определенном локусе или регионе) (Klassmann, Gautier, 2022). Результат интерпретируется из теории о том, что основные гаплотипы с необычно высоким ЕНН и высокой популяционной частотой указывают на наличие мутации, которая стала заметной в генофонде человека быстрее, чем ожидалось при нейтральной эволюции (Sabeti et al., 2002).

Для изучения генетического разнообразия и эволюции человеческих популяций хорошо зарекомендовал себя метод ХР-ЕНН (expanded haplotype homozygosity score), позволяющий выявить потенциальные места генетических изменений, которые могут быть связаны с адаптацией к изменяющимся условиям окружающей среды (Voight et al., 2006).

Гены *PER1*, *PER2* и *PER3* вовлечены в координацию циркадных ритмов, регуляцию адаптивных способностей

организма, а также связаны с различными заболеваниями (Liberman et al., 2017; Rijo-Ferreira, Takahashi, 2019). Haпример, в некоторых исследованиях обнаружена высокая связь экспрессии гена PER2 с адаптацией организмов к низким температурам. S. Chappuis с коллегами (2013) доказали, что мыши с выключенным геном Period2 (PER2) чувствительны к холоду, так как их система адаптивного термогенеза становится менее эффективной. Что касается гена PER1, то Y. Shi с коллегами (2021) утверждают, что световая адаптация, генерируемая путем CRTC1-SIK1, в которую вовлечен ген *PER1*, в супрахиазматическом ядре обеспечивает надежный механизм, позволяющий циркадной системе поддерживать гомеостаз в присутствии световых возмущений. Этот механизм, по-видимому, важен для быстрой адаптации к изменяющимся условиям окружающей среды. Опираясь на данные L. Zhang с коллегами (2013), удалось установить, что полиморфизм в гене PER3 связан с уровнем адаптации к сменному графику работы и чередованию фаз сна у медсестер, работающих посменно.

Таким образом, гены из группы PER являются перспективным объектом для поиска сигналов положительного отбора в различных популяциях. Кроме того, интерес вызывает наличие связи между адаптационными способностями, сигналами отбора и основными современными заболеваниями.

Материалы и методы

Для анализа использовали публичные данные из проекта «1000 геномов» (1000 Genomes, 2008), представляющие 23 популяции, сгруппированные в соответствующие кластеры (см. таблицу).

С помощью Plink 1.9 (Purcell et al., 2007) были объединены все данные. С использованием bcftools были удалены SNP-дубликаты и SNP с одинаковыми позициями, а также нормализованы все данные в соответствии с референсом GRCh38. Координаты начального и конечного положений для областей генов *PER1*, *PER2* и *PER3* (сборка GRCh 38) получены из NCBI, США (National Library of Medicine).

Для идентификации сигналов отбора использовали метод ХР-ЕНН, реализованный в программе selscan (Szpiech, 2021). Нестандартизированные оценки нормализовали с использованием скрипта "norm", предоставленного в

Популяции из проекта «1000 геномов», выбранные для анализа Группа Ν Место проживания/этническая самоидентификация Африканцы (AFR) ESN 97 Юг Нигерии (Esan in Nigeria)/эсан (ишан) GWD Западный округ Гамбии (Gambian in Western Division – Mandinka)/мандинка 113 83 Сьерра-Леоне (Mende in Sierra Leone)/менде MSL 108 YRI Ибадан, Нигерия (Yoruba in Ibadan, Nigeria)/йоруба 108 LWK Вебуйе, округ Бунгома в западной Кении (Luhya in Webuye, Kenya)/лухья Европейцы (EUR) GBR 89 Великобритания (British from England and Scotland)/контрольная популяция Великобритании FIN 96 Финляндия (Finnish in Finland)/финны TSI 107 Тоскана, Италия (Toscani in Italia)/тосканцы 107 IBS Испания (Iberian Populations in Spain)/испанцы Смешанные американцы (AMR) CLM 94 Столичный район Медельин, Колумбия (Colombian in Medellín, Colombia)/колумбийцы MXI 64 Лос-Анджелес, штат Калифорния, США (Mexican Ancestry in Los Angeles, CA, USA)/мексиканцы PEL 85 Столичный регион Лима-Кальяо, Перу (Peruvian in Lima, Peru)/перуанцы PUR 104 Пуэрто-Рико (Puerto Rican in Puerto Rico)/пуэрториканцы Жители Восточной Азии (EAS) GIH 103 Столичный район Хьюстона, штат Техас, США (Gujarati Indians in Houston, Texas, USA)/гуджарати STU 102 Великобритания (Sri Lankan Tamil in the UK)/шри-ланкийские тамилы ITU 102 Великобритания (Indian Telugu in the UK)/телугу PJL 92 Лахор, Пакистан (Punjabi in Lahore, Pakistan)/пенджабцы BEB 84 Бангладеш (Bengali in Bangladesh)/бенгальцы Жители Южной Азии (SAS) CHS 105 Провинция Хунань и Фуцзянь Южного Китая (Han Chinese South, China)/ханьцы CHB 103 Жилой район Пекинского педагогического университета (Han Chinese in Beijing, China)/ханьцы CDX 93 Сообщество Школы здоровья Сишуанбаньна в Сишуанбаньне, Юньнань, Китай (Chinese Dai in Xishuangbanna, China)/дайцы KHV 96 Хошимин, Вьетнам (Kinh in Ho Chi Minh City, Vietnam)/вьетнамцы

программе selscan. SNP со значениями crit = 1/-1 рассматривались как генетические варианты, находящиеся под давлением отбора (выбросы) (crit = 1 – предковый аллель под давлением отбора, crit = -1 – производный аллель).

JPT

103

Сигналы отбора определяли путем межпопуляционных сравнений, в качестве группы сравнения использовали YRI из африканского кластера. Это позволило определить выбросы между африканской популяцией йоруба (YRI) из Ибадана и другими группами (в том числе и африканского кластера, а именно ESN, GWD, MSL и LWK). Кроме того, интерес представляли и варианты отбора, связанные с внутрикластерной изменчивостью. Для этого в каждом кластере была выбрана группа сравнения, на основе которой проводили анализ с другими группами этого же кластера. Так, в кластере EUR в качестве группы для сравнения приняли GBR и соответственно провели анализ между GBR&FIN, GBR&IBS и GBR&TSI. В кластере AMR в качестве группы для сравнения приняли группу PUR и провели анализ между PUR&CLM, PUR&MXL и PUR&PEL. В кластерах EAS и SAS были определены группы CHB и BEB соответственно и проведен анализ между CHB&CDX, CHB&CHS, CHB&KHV, CHB&JPT и BEB&PJL, BEB&ITU, BEB&STU, BEB&GIH соответственно.

Результаты и обсуждение

Геномика и молекулярная биология существенно повлияли на исследования «отбора и адаптации» посредством идентификации генетических основ различных признаков, связанных с поддержанием жизнедеятельности и здоровья у людей и животных (Hancock et al., 2010; Gintis et al., 2012). Наряду с этим результаты генетической и геномной «революции» обеспечили возможность секвенирования геномов и предоставили новые инструменты измерения как прошлых, так и, возможно, продолжающихся адаптаций (Zheng et al., 2023).

Столичный район Токио (Japanese in Tokyo, Japan)/японцы



Рис. 1. Генетические варианты под давлением отбора в гене *PER1*. Здесь и на рис. 2 и 3 роз – позиция.

Предполагают, что поведение человека определяется взаимодействием между природой и развитием общества (Saravanan et al., 2020). Можно допустить, что особенности генетической структуры в различных популяциях людей, в том числе связанные с миграцией людей из Африки, в дальнейшем легли в основу индивидуальных особенностей их развития (Benton et al., 2021). Так, рассматриваемые нами гены *PER1*, *PER2* и *PER3* показали сигналы положительного отбора, часть из которых прослеживалась в нескольких популяциях (эти варианты в основном локализованы в гене *PER2*), а другая часть встречалась только в одной популяции.

В результате анализа полногеномных профилей исследуемых популяционных групп выявлено 110 локусов (78 точек в гене PER2, 25 – в PER3 и 7 – в PER1), находящихся под давлением отбора. При анализе гена PER1 обнаружено 8 выбросов при межгрупповом сравнении жителей Южной Азии, проживающих в Пекине (СНВ), с жителями Южной Азии, проживающими в Юньнани (CDX) (рис. 1). Кроме того, при межгрупповом сравнении жителей Восточной Азии, проживающих на территории Бангладеша (ВЕВ), с жителями Восточной Азии, проживающими на территории Шри-Ланки (STU), и жителей Восточной Азии, проживающих на территории Бангладеша (ВЕВ), с жителями Восточной Азии, проживающими на территории Хьюстона, штат Техас (GIH), также идентифицированы четыре участка под давлением отбора (см. рис. 1).

Обнаруженные участки вовлечены в процессы, способствующие предрасположенности к развитию большого депрессивного расстройства, болезни Паркинсона, болезни Альцгеймера, зависимости от употребления алкогольных напитков, рака груди и долголетию (см. Приложение)¹.

Хотелось бы уделить внимание точкам, находящимся под давлением отбора у нескольких сравниваемых групп. Такие SNP идентифицированы при межгрупповом сравнении жителей Восточной и Южной Азии. Найденные точки дают основание предполагать, что на сравниваемые группы оказывали воздействие схожие внешние факторы, которые, в свою очередь, имели одинаковое влияние, необходимое для адаптации исследуемых этносов. Нужно обратить внимание на выбросы в позициях chr17:8149097 (предрасположенность к развитию рака груди). Возможно, закрепление аллелей в группах сравнения жителей Вос-

https://vavilovj-icg.ru/download/pict-2024-28/appx23.pdf

точной Азии из Бангладеш (ВЕВ) и жителей Восточной Азии из Шри-Ланки (STU), а также жителей Южной Азии из Пекина (СНВ) и жителей Южной Азии из Юньнани (CDX) могло произойти вследствие преобладания влажного климата на территориях, где проживали исследуемые этносы. Согласно мнению некоторых ученых, влажный климат может стать фактором риска при развитии ряда онкологических заболеваний (Maryanaji, 2020; Guo et al., 2021; Pan et al., 2023).

При анализе гена PER2 идентифицировано 78 точек, находящихся под давлением отбора. Выбросы были идентифицированы при сравнении всех исследуемых этносов с африканцами, а также в единичных участках при межгрупповом сравнении жителей Южной Азии и европейцев (рис. 2). При анализе точек в сравниваемых популяционных группах можно сделать вывод о сильной дифференциации африканской популяции по сравнению с другими изучаемыми этносами. Аннотация участков, находящихся под давлением отбора у нескольких сравниваемых групп, выявила участие SNP в формировании хронотипов, координации сна, предрасположенности к развитию сахарного диабета, инсульта, красной волчанки и биполярного расстройства, абсорбции холестерина в кишечнике, а также в формировании метаболического фенотипа. Более подробно связи SNP с различными заболеваниями и фенотипами у людей представлены в Приложении.

Наличие общего количества выбросов при сравнении популяционной группы африканцев и других этносов указывает на длительный период воздействия определенных внешних факторов на все изучаемые популяции. Интересно отметить, что все аллели, находящиеся под давлением отбора, оказались производными вариантами. Так как ведущая функция гена *PER2* – формирование хронотипов, можно предположить, что выявление общего массива точек, находящихся под давлением отбора, также объясняется действием внешних факторов, присущих местности, на которой проживали изучаемые этносы. К таким факторам можно отнести общее количество часов светового дня, действие магнитного поля, особенности климата и другие.

Идентификация локусов под давлением отбора у нескольких групп сравнения между жителями Южной Азии и африканцами, а также внутри групп жителей Южной Азии, проживающих на территории Пекина (СНВ) и Юньнани (CDX), обнаружила SNP, ответственные за предрасположенность к развитию ряда заболеваний желу-

¹ Приложение см. по адресу:

	z	ND	SL	¥	XQ	HS	Ν	ΡT	HS	×	≥	н	ΓW	1XL	Ш	Ŋ	Ч	Ŀ.	JR	_		D	Ξ	В		۔		Т	ŝS	SI	z	~	æ	S	_
	L_ES	6	ž				BK		그	5	국	ľ	2	2	R P	5	Ň	<u>–</u> PE	<u>I</u> PL	B_P		B_S	B	<u>BE</u>		Ē	L_ST	5	2	E H	н Ц		۳ ۳	LIB!	L_TS
pos	ΥR	ΥR	ΥR	ΥR	£	£	£	Ð	ΥR	ΥR	ΥR	ΥR	Ы	Ы	Ы	ΥR	ΥR	ΥR	ΥR	BE	BE	BE	BE	ΥR	ΥR	ΥR	ΥR	ΥR	8	B	B	ΥR	ΥR	ΥR	K
chr2:238243478																•			•					•	•			•				•			
chr2:238245124					-								-	-		•			•					•	•			•	-	-	-	•			
chr2:238246412									•	•	•	•				•	•		•					•	•	•	•	•		\vdash		•	•		•
chr2:238247001									•	•	•	•				•	٠		•					•	•	•	•	•				•	•		•
chr2:238247091									•	•	•	•				•	•		•					•	•	•	•	•				•	•		•
chr2:238247330				-	-				•	•	•	•	-	\vdash	-	•	•		•					•	•	•	•	•	-	\vdash	-	•	•	•	•
chr2:238249900									•	•	•	•				•	•		•					•	•	•	•	•				•	•		•
chr2:238249929																															•				
chr2:238251128																															•				
chr2:238251323					-									-															-	\vdash	•				
chr2:238252450														\vdash																\vdash	•				
chr2:238254093																															•				
chr2:238256995																															•				
chr2:238257022					-								-	-	-														-	-	•				
chr2:238258845																															•				
chr2:238259807																															•				
chr2:238266065												•																							
chr2:238266159									•		0	•										H	-						-	\vdash					
chr2:238266693																															•				
chr2:238267745																															•				
chr2:238269574																														-	•				
chr2:238270268																														-					
chr2:238270777																															•				
chr2:238271002																															•				
chr2:238271146					-										_															-	•				
chr2:238271273					•				•			•	-	-								\vdash							-	-	•			-	
chr2:238272107																															•				
chr2:238272850																															٠				
chr2:238273219														-															-	-	•				
chr2:238273283				-										\vdash								\vdash							-	\vdash	•				
chr2:238273760																														\vdash	•				
chr2:238273793																															٠				
chr2:238274081														-																-	•				
chr2:238274048					-									\vdash								\vdash							-	\vdash	•				
chr2:238275483																															•				
chr2:238275562																															•				
chr2:238275940									•		•		_																	-		•			
chr2:238270803				-	-									\vdash								\vdash							-	\vdash	•				
chr2:238277948																															•				
chr2:238278317																															•				
chr2:238278487													_																		•				
chr2:238278568																														-	•				
chr2:238279349																															•				
chr2:238279838																															•				
chr2:238279852																														-	•				
chr2:238280642																															•				
chr2:238282057																															•				
chr2:238282265																															•				
chr2:238282966																													-	-	•				
chr2:238284937																															•				
chr2:238285197																															•				
chr2:238285454																															•				
chr2:238286196																													-	-	•				
chr2:238287253																															•				
chr2:238287359																															•				
chr2:238289506																															•				
chr2:238289684									•	•	•	•																	-	-					
chr2:238289924																													-	-	•				
chr2:238294347																															•				
chr2:238294521																															•				
chr2:238294650											•	•																							
chr2:238299790									•	•	•	•																		-					
									-								-	-	-	-				100	-	-						-		-	

Рис. 2. Генетические варианты под давлением отбора в гене PER2.

Точками обозначены варианты под давлением отбора; черный цвет точек означает, что под давлением отбора оказался предковый аллель, синий цвет – производный аллель.



Рис. 3. Генетические варианты под давлением отбора в гене PER3.

Точками обозначены варианты под давлением отбора; черный цвет точек означает, что под давлением отбора оказался предковый аллель, синий цвет – производный аллель.

дочно-кишечного тракта и сердечно-сосудистой системы. При этом в группах сравнения жителей Южной Азии с африканцами были идентифицированы производные аллели, а при сравнении жителей Южной Азии внутри подгрупп (в зависимости от региона проживания) – предковый аллель. Наибольший интерес представляет выброс в chr2:238289684, обнаруженный при сравнении жителей Южной Азии с африканцами: он ассоциируется с системной красной волчанкой, которую вызывают такие нарушения, как гормональный дисбаланс в период полового созревания, стресс, а также внешние средовые факторы, а именно воздействие солнца и вирусные инфекции (Quaglia et al., 2021; Kim et al., 2022; Molina et al., 2022).

На наш взгляд, различный уровень вирусной нагрузки, а также аутентичные климатические условия могли сыграть ключевую роль в развитии адаптивных способностей этих этносов, таким образом, закрепив данные аллели за исследуемыми популяциями. Такая же гипотеза может объяснить закрепление локусов, ассоциированных с заболеваниями желудочно-кишечного тракта. В результате миграции людей с африканского континента в другие местности их гастрономические предпочтения менялись, модифицируя микробную экосистему кишечника (Clemente et al., 2015; Syromyatnikov et al., 2022). Исходя из этого, можно предположить, что болезни желудочнокишечного тракта среди жителей Южной Азии и Африки различались вследствие отличий в микробиоме кишечника (Donin et al., 2010; Porras et al., 2021).

Проведенное исследование выявило 42 точки, находящиеся под давлением отбора в гене *PER3*, при сравнении жителей Восточной Азии, проживающих на территории

Бангладеш (ВЕВ), с жителями Восточной Азии, проживающими в Хьюстоне, штат Texac (GIH); смешанных американцев, проживающих на территории Пуэрто-Рико (PUR), со смешанными американцами, проживающими на территории Перу (PEL); африканцев с колумбийцами (YRI CLM) и африканцев с пуэрториканцами (YRI PUR) (рис. 3). После проведения аннотации были обнаружены следующие ассоциации с найденными SNP: реакция на применение препаратов лития при лечении биполярного расстройства, формирование хронотипов разных типов, предрасположенность к депрессивным расстройствам, предрасположенность к развитию метаболического синдрома, вероятность развития колоректального рака, предрасположенность к развитию ожирения. Более подробно связи SNP с различными заболеваниями и фенотипами у людей представлены в Приложении.

Основная функция участков, идентифицированных нами как находящиеся под давлением отбора, – формирование хронотипа утреннего типа. Нужно отметить, что при внутреннем сравнении групп смешанных американцев и жителей Восточной Азии были идентифицированы предковые аллели, а при сравнении африканцев со смешанными американцами – производные. Возможно, ключевым отличием африканцев от смешанных американцев является особенность режима сна у представителей данных популяций. Так, среди смешанных американцев чаще преобладает вечерний хронотип, в то время как у африканцев наиболее часто встречается утренний (Egan et al., 2017). Вероятно, это связано с влиянием различных внешних факторов, таких как широта, долгота, действие магнитного поля или активность солнца. Кроме того, полученные данные позволяют сделать предположение о влиянии внешних факторов на формирование исследуемых популяций, которое в результате привело к различному действию механизмов их адаптационных способностей. Например, обособленность популяции смешанных американцев из Лима-Кальяо (PEL) от африканцев может быть обусловлена отдаленностью нахождения данной группы людей по сравнению с другими изучаемыми этносами. Достоверно известно, что народности из других районов Латинской Америки подвергались большей ассимиляции с европейцами, по сравнению с жителями из Перу (Chacón-Duque et al., 2018). Таким образом, самобытность образовавшейся популяции сформировала наиболее обособленный генетический кластер.

Прогрессивные статистические методы, направленные на поиск локусов, находящихся под давлением отбора, позволяют ученым из разных стран проводить исследования в этой области. В их работах встречаются упоминания об отдельных SNP, идентифицированных нами в данном исследовании как находящиеся под давлением отбора. В общей сложности нами аннотировано 35 таких участков.

Исследователями выполнена работа по аннотации SNP, нахождению связи с полиморфизмами в данных точках и корреляции с некоторыми заболеваниями и физиологическими особенностями. Например, S.E. Jones с коллегами провели анализ поведенческих показателей циркадных ритмов путем анализа полногеномных данных у 697828 жителей Соединенного Королевства. Исследование позволило обнаружить новые локусы, связанные с хронотипом утреннего типа. Среди таких локусов был идентифицирован rs58574366 (2:238286196). Проведенные нами анализы продемонстрировали, что данный SNP находится под давлением отбора у групп сравнения европейцев из Соединенного Королевства (GBR) с европейцами из Финляндии (FIN). Отрицательные значения показателей расчета xpehh дали возможность сделать вывод об выявлении производных аллелей в двух сравниваемых выборках (Jones et al., 2019).

Еще одна точка, представляющая интерес, – rs74508725 (2:238278568). Такой выброс встречается при сравнении групп европейцев из Соединенного Королевства (GBR) с европейцами из Финляндии (FIN) и несет в себе отрицательные значения, что может свидетельствовать о дифференциации данного участка внутри исследуемых групп. В работе (Kichaev et al., 2019) обнаружены ассоциации данного участка с фенотипом роста.

Локус rs2585399 (17:8151441) был идентифицирован нами как находящийся под давлением отбора при сопоставлении сразу нескольких исследуемых групп людей: жителей Восточной Азии из Бангладеш (ВЕВ) с жителями Восточной Азии из штата Техас (GIH) и жителей Южной Азии из Пекина (СНВ) с жителями Южной Азии из Юньнани (CDX). Интересным фактом является то, что в исследовании авторов были найдены ассоциации данного SNP с большим депрессивным расстройством. Анализ транскриптомных ассоциаций позволил обнаружить значительные ассоциации с экспрессией NEGR1 в гипоталамусе и DRD2 в прилежащем ядре (Levey et al., 2021).

Еше олним сигналом отбора, изученным ранее, оказался rs228654 (1:7837168). Однако необходимо отметить, что в результате сравнения популяционных групп африканцев и смешанных американцев YRI (Ибадане, Нигерия) с CLM (Медельин, Колумбия) и YRI (Ибадане, Нигерия) с PUR (Пуэрто-Рико) были выявлены отрицательные значения показателей ЕНН, что говорит о наличии производного аллеля между группами. Напротив, между группами жителей Восточной Азии, проживающих в Texace (BEB), и жителей Восточной Азии, проживающих в Бангладеш (GIH), был обнаружен положительный отбор, что свидетельствует о наличии предкового аллеля. Группа исследователей во главе с P.R. Jansen провела анализ генома человека с целью получения информации о путях, тканях и типах клеток, участвующих в регуляции бессонницы (Jansen et al., 2019). Однонуклеотидный полиморфизм rs228654 вошел в число локусов, ассоциированных с развитием данного заболевания.

Заключение

Настоящее исследование позволяет сделать вывод о том, что условия проживания, климат и другие внешние факторы напрямую влияют на генетическую структуру популяций и варьируют в зависимости от расы и географического положения. Кроме того, многие из вариантов отбора в генах PER1, PER2, PER3, по-видимому, регулируют биологические процессы, которые связаны с основными современными заболеваниями, включая ожирение, онкологические заболевания, метаболический синдром, биполярное расстройство личности, депрессию, ревматоидный артрит, сахарный диабет, красную волчанку, инсульт и болезнь Альцгеймера, что делает их крайне интересными объектами для дальнейших исследований, направленных на идентификацию генетически обусловленных причин болезней человека, в том числе кардиометаболических и психических расстройств, а также онкологических заболеваний.

Список литературы / References

- Гафаров В.В., Гагулин И.В., Громова Е.А., Гафарова А.В., Панов Д.О. Ассоциация полиморфизма rs934945 гена *Per2* с нарушениями сна в мужской новосибирской популяции 25-44 лет. *Мир науки, культуры, образования.* 2016;5:283-287
 - [Gafarov V.V., Gagulin I.V., Gromova E.A., Gafarova A.V., Panov D.O. Association of polymorphism rs934945 gene Per2 with sleep disorders in the male population of Novosibirsk 25-44. *Mir Nauki, Kul'tury, Obrazovaniya = The World of Science, Culture and Education.* 2016;5:283-287 (in Russian)]
- 1000 Genomes. [WWW Document]. 2008. URL: https://www.ncbi. nlm.nih.gov/projects/faspftp/1000genomes/ (accessed 9.6.23)
- Ahrens C.W., Rymer P.D., Stow A., Bragg J., Dillon S., Umbers K.D.L., Dudaniec R.Y. The search for loci under selection: trends, biases and progress. *Mol. Ecol.* 2018;27(6):1342-1356. DOI 10.1111/mec. 14549
- Azevedo P.G., Miranda L.R., Nicolau E.S., Alves R.B., Bicalho M.A.C., Couto P.P., Ramos A.V., Souza R.P., Longhi R., Friedman E., Marco L., Bastos-Rodrigues L. Genetic association of the *PERIOD3 (PER3)* Clock gene with extreme obesity. *Obes. Res. Clin. Pract.* 2021;15(4):334-338. DOI 10.1016/j.orcp.2021.06.006
- Bacalini M.G., Palombo F., Garagnani P., Giuliani C., Fiorini C., Caporali L., Stanzani Maserati M., Capellari S., Romagnoli M., De Fanti S., Benussi L., Binetti G., Ghidoni R., Galimberti D., Scarpini E.,

Arcaro M., Bonanni E., Siciliano G., Maestri M., Guarnieri B.; Italian Multicentric Group on clock genes, actigraphy in AD; Martucci M., Monti D., Carelli V., Franceschi C., La Morgia C., Santoro A. Association of rs3027178 polymorphism in the circadian clock gene *PER1* with susceptibility to Alzheimer's disease and longevity in an Italian population. *GeroScience*. 2022;44(2):881-896. DOI 10.1007/ s11357-021-00477-0

- Bakoev S., Getmantseva L., Kostyunina O., Bakoev N., Prytkov Y., Usatov A., Tatarinova T.V. Genome-wide analysis of genetic diversity and artificial selection in Large White pigs in Russia. *PeerJ*. 2021;9:e11595. DOI 10.7717/peerj.11595
- Bakoev S.Y., Korobeinikova A.V., Mishina A.I., Kabieva S.S., Mitrofanov S.I., Ivashechkin A.A., Akinshina A.I., Snigir E.A., Yudin S.M., Yudin V.S., Getmantseva L.V., Anderzhanova E.A. Genomic signatures of positive selection in human populations of the OXT, OXTR, AVP, AVPR1A and AVR1B gene variants related to the regulation of psychoemotional response. Genes (Basel). 2023;14(11): 2053. DOI 10.3390/genes14112053
- Baranger D.A.A., Ifrah C., Prather A.A., Carey C.E., Corral-Frías N.S., Drabant Conley E., Hariri A.R., Bogdan R. PER1 rs3027172 genotype interacts with early life stress to predict problematic alcohol use, but not reward-related ventral striatum activity. *Front. Psychol.* 2016;7:464. DOI 10.3389/fpsyg.2016.00464
- Benton M.L., Abraham A., LaBella A.L., Abbot P., Rokas A., Capra J.A. The influence of evolutionary history on human health and disease. *Nat. Rev. Genet.* 2021;22(5):269-283. DOI 10.1038/s41576-020-00305-9
- Biscontin A., Zarantonello L., Russo A., Costa R., Montagnese S. Toward a molecular approach to chronotype assessment. *J. Biol. Rhythms*. 2022;37(3):272-282. DOI 10.1177/07487304221099365
- Blomeyer D., Buchmann A.F., Lascorz J., Zimmermann U.S., Esser G., Desrivieres S., Schmidt M.H., Banaschewski T., Schumann G., Laucht M. Association of *PER2* genotype and stressful life events with alcohol drinking in young adults. *PLoS One*. 2013;8(3):e59136. DOI 10.1371/journal.pone.0059136
- Bondarenko E.A., Shadrina M.I., Druzhkova T.A., Akzhigitov R.G., Gulyaeva N.V., Gekht A.B., Slominsky P.A. An association study of rs10462021 polymorphism in the clock gene *PERIOD3* and different clinical types of depression. *Mol. Genet. Microbiol. Virol.* 2018; 33(1):26-29. DOI 10.3103/S0891416818010056
- Cade B.E. Variation and selection in human circadian clock genes. Doctoral Thesis. University of Surrey, 2010
- Carlson C.S., Thomas D.J., Eberle M.A., Swanson J.E., Livingston R.J., Rieder M.J., Nickerson D.A. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 2005;15(11):1553-1565. DOI 10.1101/gr.4326505
- Carpen J.D., Archer S.N., Skene D.J., Smits M., von Schantz M. A single-nucleotide polymorphism in the 5'-untranslated region of the *hPER2* gene is associated with diurnal preference. J. Sleep Res. 2005;14(3):293-297. DOI 10.1111/j.1365-2869.2005.00471.x
- Chacón-Duque J.-C., Adhikari K., Fuentes-Guajardo M., Mendoza-Revilla J., Acuña-Alonzo V., Barquera R., Quinto-Sánchez M., ... Poletti G., Gallo C., Bedoya G., Rothhammer F., Balding D., Hellenthal G., Ruiz-Linares A. Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nat. Commun.* 2018;9(1):5388. DOI 10.1038/s41467-018-07748-z
- Chang A.-M., Bjonnes A.C., Aeschbach D., Buxton O.M., Gooley J.J., Anderson C., Van Reen E., Cain S.W., Czeisler C.A., Duffy J.F., Lockley S.W., Shea S.A., Scheer F.A.J.L., Saxena R. Circadian gene variants influence sleep and the sleep electroencephalogram in humans. *Chronobiol. Int.* 2016;33(5):561-573. DOI 10.3109/0742 0528.2016.1167078
- Chang Y.-C., Chiu Y.-F., Liu P.-H., Hee S.W., Chang T.-J., Jiang Y.-D., Lee W.-J., Lee P.-C., Kao H.-Y., Hwang J.-J., Chuang L.-M. Genetic variation in the *NOC* gene is associated with body mass index in chinese subjects. *PLoS One*. 2013;8(7):e69622. DOI 10.1371/journal. pone.0069622

- Chappuis S., Ripperger J.A., Schnell A., Rando G., Jud C., Wahli W., Albrecht U. Role of the circadian clock gene *Per2* in adaptation to cold temperature. *Mol. Metab.* 2013;2(3):184-193. DOI 10.1016/ j.molmet.2013.05.002
- Clemente J.C., Pehrsson E.C., Blaser M.J., Sandhu K., Gao Z., Wang B., Magris M., Hidalgo G., Contreras M., Noya-Alarcón Ó., Lander O., McDonald J., Cox M., Walter J., Oh P.L., Ruiz J.F., Rodriguez S., Shen N., Song S.J., Metcalf J., Knight R., Dantas G., Dominguez-Bello M.G. The microbiome of uncontacted Amerindians. *Sci. Adv.* 2015;1(3):e1500183. DOI 10.1126/sciadv.1500183
- Dan Y.-L., Zhao C.-N., Mao Y.-M., Wu Q., He Y.-S., Hu Y.-Q., Xiang K., Yang X.-K., Sam N.B., Wu G.-C., Pan H.-F. Association of PER2 gene single nucleotide polymorphisms with genetic susceptibility to systemic lupus erythematosus. *Lupus*. 2021;30(5): 734-740. DOI 10.1177/0961203321989794
- Donin A.S., Nightingale C.M., Owen C.G., Rudnicka A.R., McNamara M.C., Prynne C.J., Stephen A.M., Cook D.G., Whincup P.H. Nutritional composition of the diets of South Asian, black African-Caribbean and white European children in the United Kingdom: The Child Heart and Health Study in England (CHASE). *Br. J. Nutr.* 2010;104(2):276-285. DOI 10.1017/S000711451000070X
- Egan K.J., Knutson K.L., Pereira A.C., von Schantz M. The role of race and ethnicity in sleep, circadian rhythms and cardiovascular health. *Sleep Med. Rev.* 2017;33:70-78. DOI 10.1016/j.smrv.2016.05.004
- Forbes E.E., Dahl R.E., Almeida J.R.C., Ferrell R.E., Nimgaonkar V.L., Mansour H., Sciarrillo S.R., Holm S.M., Rodriguez E.E., Phillips M.L. *PER2 rs2304672* polymorphism moderates circadianrelevant reward circuitry activity in adolescents. *Biol. Psychiatry*. 2012;71(5):451-457. DOI 10.1016/j.biopsych.2011.10.012
- Gintis H., Doebeli M., Flack J. The evolution of human cooperation. *Cliodynamics*. 2012;3(1):172-190. DOI 10.21237/C7CLIO3112928
- Gu Z., Wang B., Zhang Y.-B., Ding H., Zhang Y., Yu J., Gu M., Chan P., Cai Y. Association of *ARNTL* and *PER1* genes with Parkinson's disease: a case-control study of Han Chinese. *Sci. Rep.* 2015;5(1): 15891. DOI 10.1038/srep15891
- Guo H., Li X., Li W., Wu J., Wang S., Wei J. Climatic modification effects on the association between PM1 and lung cancer incidence in China. *BMC Public Health*. 2021;21(1):880. DOI 10.1186/s12889-021-10912-8
- Hancock A.M., Alkorta-Aranburu G., Witonsky D.B., Di Rienzo A. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philos. Trans. R. Soc. B Biol. Sci.* 2010; 365(1552):2459-2468. DOI 10.1098/rstb.2010.0032
- Holipah Hinoura T., Kozaka N., Kuroda Y. The correlation between PER3 rs2640908 polymorphism and colorectal Cancer in the Japanese population. *Appl. Cancer Res.* 2019;39(1):3. DOI 10.1186/ s41241-019-0072-5
- Jansen P.R., Watanabe K., Stringer S., Skene N., Bryois J., Hammerschlag A.R., de Leeuw C.A., Benjamins J.S., Muñoz-Manchado A.B., Nagel M., Savage J.E., Tiemeier H., White T., Tung J.Y., Hinds D.A., Vacic V., Wang X., Sullivan P.F., van der Sluis S., Polderman T.J.C., Smit A.B., Hjerling-Leffler J., Van Someren E.J.W., Posthuma D. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* 2019;51(3):394-403. DOI 10.1038/s41588-018-0333-3
- Jensen J.D., Foll M., Bernatchez L. The past, present and future of genomic scans for selection. *Mol. Ecol.* 2016;25(1):1-4. DOI 10.1111/ mec.13493
- Jones S.E., Lane J.M., Wood A.R., van Hees V.T., Tyrrell J., Beaumont R.N., Jeffries A.R., ... Gehrman P.R., Lawlor D.A., Frayling T.M., Rutter M.K., Hinds D.A., Saxena R., Weedon M.N. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* 2019;10(1):343. DOI 10.1038/s41467-018-08259-7
- Kichaev G., Bhatia G., Loh P.-R., Gazal S., Burch K., Freund M.K., Schoech A., Pasaniuc B., Price A.L. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* 2019;104(1):65-75. DOI 10.1016/j.ajhg.2018.11.008

- Kim J.-W., Kim H.-A., Suh C.-H., Jung J.-Y. Sex hormones affect the pathogenesis and clinical characteristics of systemic lupus erythematosus. *Front. Med.* 2022;9:906475. DOI 10.3389/fmed.2022. 906475
- Klassmann A., Gautier M. Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data. *PLoS One.* 2022;17(1):e0262024. DOI 10.1371/journal. pone.0262024
- Kripke D.F., Nievergelt C.M., Joo E., Shekhtman T., Kelsoe J.R. Circadian polymorphisms associated with affective disorders. J. Circadian Rhythms. 2009;7:2. DOI 10.1186/1740-3391-7-2
- Lee H., Nah S.-S., Chang S.-H., Kim H.-K., Kwon J.-T., Lee S., Cho I.-H., Lee S.W., Kim Y.O., Hong S.-J., Kim H.-J. PER2 is downregulated by the LPS-induced inflammatory response in synoviocytes in rheumatoid arthritis and is implicated in disease susceptibility. *Mol. Med. Rep.* 2017;16(1):422-428. DOI 10.3892/mmr. 2017.6578
- Lesicka M., Jabłońska E., Wieczorek E., Pepłońska B., Gromadzińska J., Seroczyńska B., Kalinowski L., Skokowski J., Reszka E. Circadian gene polymorphisms associated with breast cancer susceptibility. *Int. J. Mol. Sci.* 2019;20(22):5704. DOI 10.3390/ijms20225704
- LeVan T.D., Xiao P., Kumar G., Kupzyk K., Qiu F., Klinkebiel D., Eudy J., Cowan K., Berger A.M. Genetic variants in circadian rhythm genes and self-reported sleep quality in women with breast cancer. J. Circadian Rhythms. 2019;17(1):184. DOI 10.5334/jcr.184
- Levey D.F., Stein M.B., Wendt F.R., Pathak G.A., Zhou H., Aslan M., Quaden R., Harrington K.M., Nuñez Y.Z., Overstreet C., Radhakrishnan K., Sanacora G., McIntosh A.M., Shi J., Shringarpure S.S., Concato J., Polimanti R., Gelernter J. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat. Neurosci.* 2021; 24(7):954-963. DOI 10.1038/s41593-021-00860-2
- Levran O., Randesi M., Rotrosen J., Ott J., Adelson M., Kreek M.J. A 3' UTR SNP rs885863, a *cis*-eQTL for the circadian gene *VIPR2* and lincRNA 689, is associated with opioid addiction. *PLoS One*. 2019;14(11):e0224399. DOI 10.1371/journal.pone.0224399
- Liberman A.R., Kwon S.B., Vu H.T., Filipowicz A., Ay A., Ingram K.K. Circadian clock model supports molecular link between *PER3* and human anxiety. *Sci. Rep.* 2017;7(1):9893. DOI 10.1038/s41598-017-07957-4
- Lin E., Kuo P.-H., Liu Y.-L., Yang A.C., Kao C.-F., Tsai S.-J. Effects of circadian clock genes and health-related behavior on metabolic syndrome in a Taiwanese population: Evidence from association and interaction analysis. *PLoS One*. 2017;12(3):e0173861. DOI 10.1371/ journal.pone.0173861
- Lopez M.E., Neira R., Yáñez J.M. Applications in the search for genomic selection signatures in fish. *Front. Genet.* 2015;5:458. DOI 10.3389/fgene.2014.00458
- Maryanaji Z. The effect of climatic and geographical factors on breast cancer in Iran. *BMC Res. Notes.* 2020;13(1):519. DOI 10.1186/ s13104-020-05368-9
- McCarthy M.J., Welsh D.K. Cellular circadian clocks in mood disorders. J. Biol. Rhythms. 2012;27(5):339-352. DOI 10.1177/0748730 412456367
- Melhuish Beaupre L.M., Gonçalves V.F., Zai C.C., Tiwari A.K., Harripaul R.S., Herbert D., Freeman N., Müller D.J., Kennedy J.L. Genome-wide association study of sleep disturbances in depressive disorders. *Mol. Neuropsychiatry*. 2020;5(Suppl. 1):34-43. DOI 10.1159/000505804
- Min W., Tang N., Zou Z., Chen Y., Zhang X., Huang Y., Wang J., Zhang Y., Zhou B., Sun X. A panel of rhythm gene polymorphisms is involved in susceptibility to type 2 diabetes mellitus and bipolar disorder. *Ann. Transl. Med.* 2021;9(20):1555. DOI 10.21037/atm-21-4803
- Miranda A., Shekhtman T., McCarthy M., DeModena A., Leckband S.G., Kelsoe J.R. Study of 45 candidate genes suggests *CACNG2* may be associated with lithium response in bipolar disor-

der. J. Affect. Disord. 2019;248:175-179. DOI 10.1016/j.jad.2019. 01.010

- Molina E., Gould N., Lee K., Krimins R., Hardenbergh D., Timlin H. Stress, mindfulness, and systemic lupus erythematosus: An overview and directions for future research. *Lupus*. 2022;31(13):1549-1562. DOI 10.1177/09612033221122980
- National Library of Medicine (US) [WWW Document]. URL https:// www.ncbi.nlm.nih.gov/(accessed 9.17.23)
- Nielsen R. Molecular signatures of natural selection. *Annu. Rev. Genet.* 2005;39(1):197-218. DOI 10.1146/annurev.genet.39.073003. 112420
- Pan Z., Yu L., Shao M., Ma Y., Cheng Y., Wu Y., Xu S., Zhang C., Zhu J., Pan F., Sun G. The influence of meteorological factors and total malignant tumor health risk in Wuhu city in the context of climate change. *BMC Public Health*. 2023;23(1):346. DOI 10.1186/ s12889-023-15200-1
- Porras A.M., Shi Q., Zhou H., Callahan R., Montenegro-Bethancourt G., Solomons N., Brito I.L. Geographic differences in gut microbiota composition impact susceptibility to enteric infection. *Cell Rep.* 2021;36(4):109457. DOI 10.1016/j.celrep.2021.109457
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. PLINK: A tool set for whole-genome association and populationbased linkage analyses. *Am. J. Hum. Genet.* 2007;81(3):559-575. DOI 10.1086/519795
- Qu F., Qiao Q., Wang N., Ji G., Zhao H., He L., Wang H., Bao G. Genetic polymorphisms in circadian negative feedback regulation genes predict overall survival and response to chemotherapy in gastric cancer patients. *Sci. Rep.* 2016;6(1):22424. DOI 10.1038/srep22424
- Quaglia M., Merlotti G., De Andrea M., Borgogna C., Cantaluppi V. Viral infections and systemic lupus erythematosus: new players in an old story. *Viruses*. 2021;13(2):277. DOI 10.3390/v13020277
- Rijo-Ferreira F., Takahashi J.S. Genomics of circadian rhythms in health and disease. *Genome Med.* 2019;11(1):82. DOI 10.1186/ s13073-019-0704-0
- Sabeti P.C., Reich D.E., Higgins J.M., Levine H.Z.P., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V., Patterson N.J., McDonald G.J., Ackerman H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E.S. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419(6909):832-837. DOI 10.1038/nature01140
- Sakurada K., Konta T., Takahashi S., Murakami N., Sato H., Murakami R., Watanabe M., Ishizawa K., Ueno Y., Yamashita H., Kayama T. Circadian clock gene polymorphisms and sleep-onset problems in a population-based cohort study: The yamagata study. *Tohoku J. Exp. Med.* 2021;255(4):325-331. DOI 10.1620/tjem.255.325
- Saravanan K.A., Panigrahi M., Kumar H., Bhushan B., Dutt T., Mishra B.P. Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livest. Sci.* 2020;241: 104257. DOI 10.1016/j.livsci.2020.104257
- Schroor M.M., Plat J., Mensink R.P. Relation between single nucleotide polymorphisms in circadian clock relevant genes and cholesterol metabolism. *Mol. Genet. Metab.* 2023;138(4):107561. DOI 10.1016/j.ymgme.2023.107561
- Shareefa D. Genetic analysis of bipolar disorder and alcohol use disorder. Doctoral Thesis. University of Cape Town, 2015
- Shi Y., Liu Y., Yang L., Yan J. A mathematical model to characterize the role of light adaptation in mammalian circadian clock. *Front. Mol. Biosci.* 2021;8:681696. DOI 10.3389/fmolb.2021.681696
- Soria V., Martínez-Amorós È., Escaramís G., Valero J., Pérez-Egea R., García C., Gutiérrez-Zotes A., Puigdemont D., Bayés M., Crespo J.M., Martorell L., Vilella E., Labad A., Vallejo J., Pérez V., Menchón J.M., Estivill X., Gratacòs M., Urretavizcaya M. Differential association of circadian genes with mood disorders: CRY1 and NPAS2 are associated with unipolar major depression and CLOCK and VIP with bipolar disorder. *Neuropsychopharmacology*. 2010; 35(6):1279-1289. DOI 10.1038/npp.2009.230

- Syromyatnikov M., Nesterova E., Gladkikh M., Smirnova Y., Gryaznova M., Popov V. Characteristics of the gut bacterial composition in people of different nationalities and religions. *Microorganisms*. 2022;10(9):1866. DOI 10.3390/microorganisms10091866
- Szpiech Z.A. Selscan 2.0: scanning for sweeps in unphased data. *bioRxiv*. 2021. DOI 10.1101/2021.10.22.465497
- Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4(3):e72. DOI 10.1371/journal.pbio.0040072
- Wang W.M., Yuan P., Wang J.Y., Ma F., Fan Y., Li Q., Zhang P., Xu B.H. Association of genetic variantions of circadian clock genes and risk of breast cancer. *Zhonghua Zhong Liu Za Zhi*. 2013;35(3):236-239. DOI 10.3760/cma.j.issn.0253-3766.2013.03.017
- Wen M., Jiang X., She H., Han C., Pei Z., Cai Y., Zhang T. The *Per2* polymorphism rs10462023 is associated with the risk of stroke in

a Chinese population. *Biol. Rhythm Res.* 2015;46(4):545-551. DOI 10.1080/09291016.2015.1026675

- Xu L., Bickhart D.M., Cole J.B., Schroeder S.G., Song J., Tassell C.P., Sonstegard T.S., Liu G.E. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol. Biol. Evol.* 2015;32(3):711-725. DOI 10.1093/molbev/msu333
- Zhang L., Ptáček L.J., Fu Y.-H. Diversity of human clock genotypes and consequences. 2013;119:51-81. DOI 10.1016/B978-0-12-396971-2.00003-8
- Zheng W., He Y., Guo Y., Yue T., Zhang H., Li J., Zhou B., Zeng X., Li L., Wang B., Cao J., Chen L., Li C., Li H., Cui C., Bai C., Baimakangzhuo, Qi X., Ouzhuluobu, Su B. Large-scale genome sequencing redefines the genetic footprints of high-altitude adaptation in Tibetans. *Genome Biol.* 2023;24(1):73. DOI 10.1186/s13059-023-02912-1

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 31.01.2024. После доработки 25.06.2024. Принята к публикации 26.06.2024.
DOI 10.18699/vjgb-24-72

Генетические аспекты лактазной недостаточности у коренного населения Сибири

Б.А. Малярчук 🕩

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия 🖾 malbor@mail.ru

Аннотация. Способность метаболизировать лактозу во взрослом состоянии связана с сохранением активности фермента лактазы. В европейских популяциях персистенция лактазы детерминируется главным образом наличием варианта rs4988235-Т в гене МСМ6, который увеличивает экспрессию гена LCT, кодирующего лактазу. Наиболее высокие показатели персистенции лактазы характерны для европейцев, а самые низкие – для населения Восточной Азии. Анализ опубликованных данных о распределении варианта rs4988235-С, связанного с гиполактазией, у населения Центральной Азии и Сибири выявил, что частота этого варианта увеличивается в северо-восточном направлении. В Центральной Азии частота этого аллеля составляет 87 %, на юге Сибири – 90.6 % и на северо-востоке Сибири – 92.9 %. Соответственно, в таком же географическом направлении убывает способность населения метаболизировать лактозу. Анализ палеогеномных данных показал, что более высокая частота аллеля rs4988235-Т в популяциях Центральной Азии и Южной Сибири связана с распространением на восток древнего населения восточноевропейских степей начиная с эпохи бронзового века. Результаты анализа полиморфизма экзонов и прилегающих к ним интронов генов MCM6 и LCT у коренного населения Сибири свидетельствуют о возможности существования в восточноазиатских популяциях вариантов полиморфизма, потенциально связанных с метаболизмом лактозы. В популяциях Восточной Азии, в том числе в сибирских этнических группах, обнаружен участок гена МСМ6 длиной ~26.5 тыс. пар нуклеотидов, включающий комбинацию аллелей rs4988285-A, rs2070069-G, rs3087353-T, rs2070068-A. Локусы rs4988285 и rs2070069 находятся в области энхансера, регулирующего активность гена LCT. Анализ палеогеномных последовательностей показал, что указанной выше комбинацией аллелей гена МСМ6 характеризуются геномы денисовцев и неандертальцев. Таким образом, обнаруженный гаплотип, по всей видимости, является архаичным. Он мог быть унаследован от общего предка современных людей, неандертальцев и денисовцев, или же был приобретен в результате гибридизации с денисовцами или неандертальцами. Полученные данные свидетельствуют о возможной функциональной значимости архаичных вариантов полиморфизма гена МСМб.

Ключевые слова: генетический полиморфизм; персистенция лактазы; ген *MCM6*; ген *LCT*; популяции человека; Сибирь; архаичные варианты полиморфизма.

Для цитирования: Малярчук Б.А. Генетические аспекты лактазной недостаточности у коренного населения Сибири. *Вавиловский журнал генетики и селекции*. 2024;28(6):650-658. DOI 10.18699/vjgb-24-72

Genetic aspects of lactase deficiency in indigenous populations of Siberia

B.A. Malyarchuk 🕩

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

Abstract. The ability to metabolize lactose in adulthood is associated with the persistence of lactase enzyme activity. In European populations, lactase persistence is determined mainly by the presence of the rs4988235-T variant in the *MCM6* gene, which increases the expression of the *LCT* gene, encoding lactase. The highest rates of lactase persistence are characteristic of Europeans, and the lowest rates are found in East Asian populations. Analysis of published data on the distribution of the hypolactasia-associated variant rs4988235-C in the populations of Central Asia and Siberia showed that the frequency of this variant increases in the northeastern direction. The frequency of this allele is 87 % in Central Asia, 90.6 % in Southern Siberia, and 92.9 % in Northeastern Siberia. Consequently, the ability of the population to metabolize lactose decreases in the same geographical direction. The analysis of paleogenomic data has shown that the higher frequency of the rs4988235-T allele in populations of Central Asia and Southern Siberia is associated with the eastward spread of ancient populations of the Eastern European steppes, starting from the Bronze Age. The results of polymorphism analysis of exons and adjacent introns of the *MCM6* and *LCT* genes in indigenous populations of Siberia indicate the possibility that polymorphic variants may potentially be related to lactose metabolism exist in East Asian populations. In East Asian populations, including Siberian ethnic groups, a ~26.5 thousand nucleotide

pairs long region of the *MCM6* gene, including a combination of the rs4988285-A, rs2070069-G, rs3087353-T, and rs2070068-A alleles, was found. The rs4988285 and rs2070069 loci are located in the enhancer region that regulates the activity of the *LCT* gene. Analysis of paleogenomic sequences showed that the genomes of Denisovans and Neanderthals are characterized by the above combination of alleles of the *MCM6* gene. Thus, the haplotype discovered appears to be archaic. It could have been inherited from a common ancestor of modern humans, Neanderthals, and Denisovans, or it could have been acquired by hybridization with Denisovans or Neanderthals. The data obtained indicate a possible functional significance of archaic variants of the *MCM6* gene.

Key words: genetic polymorphism; lactase persistence; *MCM6* gene; LCT gene; human populations; Siberia; archaic variants of polymorphism.

For citation: Malyarchuk B.A. Genetic aspects of lactase deficiency in indigenous populations of Siberia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(6):650-658. DOI 10.18699/vjgb-24-72

Введение

Лактоза (молочный сахар) – основной дисахарид молока различных млекопитающих, для гидролиза которого необходим фермент лактаза, кодируемый геном LCT, который экспрессируется преимущественно в тонком кишечнике. Активность лактазы снижается в процессе онтогенеза таким образом, что многие взрослые не могут усваивать лактозу (Ségurel, Bon, 2017). Первичная гиполактазия (OMIM: 223100) проявляется целым рядом симптомов (вздутием кишечника, тошнотой, диареей) после употребления молока и молочных продуктов. Вместе с тем обнаружено, что этнорегиональные группы населения мира различаются по способности метаболизировать лактозу (Evershed et al., 2022). Эта способность, или персистенция лактазы (ПЛ), имеет наследственную природу. Одним из основных вариантов генетического полиморфизма, строго ассоциированного с ПЛ, является вариант Т в локусе rs4988235 гена MCM6, который регулирует экспрессию гена LCT (Enattah et al., 2002; Olds, Sibley, 2003; Troelsen et al., 2003). Несмотря на то что этот генетический вариант находится на расстоянии примерно в 14 тыс. пар нуклеотидов (п.н.) от гена LCT (из-за чего часто обозначается как -13910*T), он ответственен за повышение ферментативной активности лактазы, расщепляющей лактозу на молекулы глюкозы и галактозы.

Самые низкие показатели ПЛ характерны для населения Восточной Азии, а самые высокие – для европейцев (Liebert et al., 2017). Это связано с тем, что, согласно археологическим данным, молочное животноводство возникло, по всей видимости, в степной зоне Северного Кавказа и Причерноморья примерно 4–5 тыс. лет назад (л. н.) (Scott et al., 2022). Палеогеномные данные свидетельствуют, что частота связанного с ПЛ варианта rs4988235-Т начала увеличиваться примерно 6 тыс. л. н. в составе предковых геномных компонентов ЕНG и CHG, характерных для восточноевропейских и кавказских охотников-собирателей соответственно (Segurel et al., 2020; Irving-Pease et al., 2024). Было выявлено также сцепление между вариантами полиморфизма в локусах rs4988235 и rs1438307, причем рост частоты аллеля rs1438307-Т начался намного раньше – примерно 12 тыс. л. н. (Irving-Pease et al., 2024). В отношении варианта rs1438307-Т высказано предположение о том, что он возник в результате адаптации древних людей к голоданию и воздействию патогенов, поскольку отмечалось его участие в регуляции энергетических затрат организма и развитии метаболических заболеваний (Evershed et al., 2022).

Несмотря на высокий интерес к генетическим аспектам проблемы гиполактазии в популяциях человека со стороны генетического и медицинского научного сообщества, многие регионы мира остаются малоисследованными (Liebert et al., 2017; Anguita-Ruiz et al., 2020). Цель настоящей работы – попытка обобщить результаты исследований полиморфизма генов *LCT* и *MCM6*, имеющих непосредственное отношение к персистенции лактазы, в популяциях коренного населения Сибири – одного из малоизученных регионов.

Распространенность вариантов полиморфизма локуса rs4988235 в современных и древних популяциях Северной Азии

Генетико-эпидемиологические исследования показали, что в популяциях европейской части России первичная гиполактазия детерминируется преимущественно или исключительно аллелем rs4988235-С гена МСМ6 (Боринская и др., 2006; Kovalenko et al., 2023), и, соответственно, ПЛ определяется аллелем rs4988235-T. Однако в восточноазиатских популяциях, включая сибирские группы, эта взаимосвязь не столь очевидна – некоторые популяции, например буряты и уйгуры, демонстрируют высокие (на уровне 95 %) частоты варианта rs4988235-С на фоне пониженной распространенности гиполактазии (Боринская и др., 2006; Соколова и др., 2007). В связи с этим высказывалось предположение о том, что меньшая частота гиполактазии у некоторых народов Сибири и Центральной Азии связана с присутствием не только варианта rs4988235-T, но и каких-то других генетических ПЛмаркеров (Соколова и др., 2007).

К настоящему времени установлено, что в этнических группах Африки и Ближнего Востока распространены, помимо аллеля rs4988235-T, некоторые другие варианты генетического полиморфизма, определяющие способность расщепления лактозы, например в локусах rs41525747, rs41380347, rs145946881 и rs182549 гена *MCM6* (Ingram et al., 2007; Tishkoff et al., 2007). Между тем сведения об ассоциативных связях между вариантами генетического полиморфизма и ПЛ у населения Восточной Азии довольно противоречивы. Так, в популяциях Центральной Азии – в смешанной выборке таджиков и узбеков, а также у казахов – случаи обнаружения варианта rs4988235-T (с частотой 10 и 16.5 % соответственно) довольно хорошо коррелировали со способностью усваивать лактозу (11–30 % у таджиков и узбеков и 25–32 % у казахов) (Неуег et al., 2011). Также на уровне 30 % усваивают лактозу и тибетцы, среди которых распространена традиция употребления, например, молока яка, однако у них отсутствуют варианты полиморфизма rs4988235-Т и rs182549-T, выявленные в соседних популяциях Северного Китая с частотой 3.8 и 6.9 % соответственно (Xu et al., 2010; Peng et al., 2012). У тибетцев обнаружен свой собственный спектр аллелей энхансерного участка гена *MCM6*, предположительно, связанных с ПЛ, среди которых преобладает вариант –13838*А с частотой около 6.5 % (Peng et al., 2012).

Более детальное исследование генетической адаптации к потреблению молока у народов Центральной Азии, характеризующихся различным хозяйственным укладом, выявило, что скотоводы (казахи, киргизы, каракалпаки, буряты, монголы, алтайцы), рационы питания которых в значительной степени зависят от молочных продуктов, в большинстве своем не обладают высокой способностью метаболизировать лактозу в сравнении с земледельцами (туркменами, таджиками, узбеками), которые показывают более высокую распространенность варианта rs4988235-Т (Sequrel et al., 2020). Относительно низкой частотой, ~10 %, этого генетического варианта также характеризуются этнические группы жителей Южной Сибири (хакасы, шорцы, тубалары), ведущих полукочевой образ жизни и занимающихся лесными промыслами и таежной охотой (Sequrel et al., 2020). Таким образом, из полученных данных следует, что частота встречаемости варианта rs4988235-Т в популяциях Центральной Азии и Сибири не особо зависит от хозяйственного уклада и уровня потребления молока.

В табл. 1 приводятся данные о распределении варианта rs4988235-С в различных выборках коренного населения Северо-Восточного Китая, Центральной Азии и Сибири. Частота этого варианта в популяциях изменяется от 70 до 100 %, однако если выборки разделить на три региональные группы, то заметен рост частоты варианта rs4988235-С с юга на северо-восток Сибири (см. рисунок).

В Центральной Азии частота аллеля rs4988235-С составляет 87.0±2.0 %, в Южной Сибири – 90.6±1.7 % и в Северо-Восточной Сибири – 92.9±2.3 %. Соответственно, в таком же географическом направлении убывает способность населения метаболизировать лактозу. Вместе с тем необходимо отметить, что различия по частоте аллелей локуса rs4988235 статистически значимы только между популяциями Центральной Азии и Сибири (и южной, и северовосточной ее частей; $P < 10^{-5}$, точный тест Фишера), а две сибирские группы популяций между собой не различаются (P = 0.09). Тем не менее нужно заметить, что частота аллелей может варьировать в разных выборках в пределах одной этнической группы - как, например, в случае казахов, бурят и особенно чукчей (см. табл. 1). Помимо случайных факторов (преимущественно для малых выборок), вклад в такого рода гетерогенность частот может вносить метисация с представителями этнических групп, для которых характерна более высокая встречаемость варианта rs4988235-T.

Очень наглядно это продемонстрировано в исследовании полиморфизма локуса rs4988235 у ненцев, основным занятием которых является оленеводство, однако ненцы практически не пьют молоко (Khabarova et al., 2012). По



Распределение варианта rs4988235-С в региональных группах Сибири и Центральной Азии.

Показана средняя частота генетического варианта (в %) и границы стандартного отклонения частоты.

всей видимости, главным образом это связано с очень высокой непереносимостью лактозы - частота варианта rs4988235-С у ненцев, имеющих всех четырех бабушек и дедушек ненецкого происхождения, составляет 92.7 % (частота генотипа rs4988235-CC - 90 %). Между тем частота аллеля rs4988235-С в группе ненцев, имеющих одного и более родственника не ненецкого происхождения, снижается до 73 % (Khabarova et al., 2012). В межэтнических браках с ненцами в основном участвуют коми и северные русские, частота аллеля rs4988235-Т у которых находится на уровне 35-42 % (Khabarova et al., 2011). Таким образом, некоторое снижение гиполактазии у народов Крайнего Севера Европы и Сибири можно объяснить брачными контактами с пришлым населением восточноевропейского происхождения начиная с XVII в. в связи с продвижением русских первопроходцев и особенно интенсивно в советское время.

Длительность и интенсивность контактов с восточными европейцами, очевидно, были выше на территории Юго-Западной Сибири и Центральной Азии, принимая во внимание миграции населения восточноевропейских степей в эпоху бронзового века. Как отмечалось выше, согласно палеогеномным данным, появление мутации rs4988235-T регистрируется примерно 6 тыс. л. н., предположительно, у древнего населения Северного Причерноморья, а позже этот вариант полиморфизма распространился по всей Северной Евразии - от Испании до Казахстана (Segurel et al., 2020; Irving-Pease et al., 2024). Причем в Европе увеличению частоты варианта rs4988235-T, определяющего стабильную активность лактазы, необходимой для усвоения молока взрослыми, способствовал положительный отбор, обусловленный недостатком витамина D в высоких широтах и необходимостью повышенного поступления кальция из молока (Козлов, Вершубская, 2017).

Анализ палеогеномных данных, находящихся в базе данных AADR (Allen Ancient DNA Resource, https://reich. hms.harvard.edu/), показал, что первые случаи появления аллеля rs4988235-Т регистрируются у древнего населения Европы – на территории Украины (~6 тыс. л. н.), Ирландии (5.5 тыс. л. н.), а от 4.5 тыс. л. н. и позже – в Литве, Германии, Чехии, Эстонии, России. В Восточной Азии вариант rs4988235-T впервые обнаружен у представителя ботайской археологической культуры Северного Казахстана (5.3 тыс. л. н.). В Центральной Азии (у предков казахов, киргизов, монголов, туркмен, узбеков, таджиков) в интервале от 0.5 до 5.3 тыс. л. н. частота варианта rs4988235-T составляла 4.2 %, а, по современным данным (см. табл. 1), ее величина оценивается примерно в 13 %. Предполагается, что в Центральной Азии эта мутация уже с железного века была распространена с частотой около 5 % (Segurel et al., 2020). Таким образом, если бы этот вариант полиморфизма находился под сильным селективным давлением, то у него должно было быть достаточно времени, чтобы

достичь высоких частот в современных популяциях – 51 %, по оценкам L. Segurel с коллегами (2020). Однако этого не произошло, в связи с чем правомочен вывод о том, что вариант rs4988235-Т не испытал существенного селективного давления у населения Центральной Азии, в отличие от европейцев и некоторых популяций Африки и Ближнего Востока (Segurel et al., 2020).

Согласно сведениям базы данных AADR, у древнего населения Сибири и Урала (в интервале 0.5–10 тыс. л. н.) вариант rs4988235-Т был распространен с частотой 1.8 %, но только на западе этой территории. Все случаи этого

Таблица 1. Частота аллеля rs4988235-С в популяциях Северной Азии

Этническая группа	Размер выборки, N	Частота аллеля rs4988235-С	Литературный источник
Таджики (Узбекистан, Таджикистан)	254	0.82	Segurel et al., 2020
Узбеки (Узбекистан)	45	0.76	Segurel et al., 2020
Туркмены (Узбекистан)	50	0.80	Segurel et al., 2020
Казахи (Китай)	94	0.95	Sun et al., 2007
Казахи (Узбекистан)	83	0.83	Heyer et al., 2011
Казахи (Узбекистан)	159	0.79	Segurel et al., 2020
Казахи (Казахстан)	34	0.88	Соколова и др., 2007
Алтайские казахи	128	0.91	Пилипенко и др., 2016
Киргизы (Киргизия)	201	0.88	Segurel et al., 2020
Каракалпаки (Узбекистан)	45	0.93	Segurel et al., 2020
Уйгуры (Казахстан)	30	0.95	Соколова и др., 2007
Монголы (Китай)	82	0.98	Sun et al., 2007
Монголы (Монголия)	32	0.88	Segurel et al., 2020
Маньчжуры (Китай)	75	1.0	Sun et al., 2007
Орочены (Китай)	45	0.99	Sun et al., 2007
Нанайцы (Китай)	77	1.0	Sun et al., 2007
Алтайцы южные	24	0.85	Cardona et al., 2014
Алтайцы южные	62	0.92	Segurel et al., 2020
Алтайцы северные	29	0.93	Segurel et al., 2020
Шорцы	24	0.94	Cardona et al., 2014
Шорцы	29	0.90	Segurel et al., 2020
Хакасы	29	0.86	Segurel et al., 2020
Хакасы	64	0.92	Пилипенко и др., 2016
Буряты	78	0.95	Соколова и др., 2007
Буряты	24	0.98	Cardona et al., 2014
Буряты	28	0.82	Segurel et al., 2020
Якуты	22	0.93	Cardona et al., 2014
Якуты	55	0.95	Liebert et al., 2017
Якуты	25	0.94	Bersaglieri et al., 2004
Эвенки западные	24	0.96	Cardona et al., 2014
Эвены	24	0.96	Cardona et al., 2014
Коряки	25	0.96	Cardona et al., 2014
Чукчи	35	0.94	Боринская и др., 2006
Чукчи	14	0.75	Cardona et al., 2014
Эскимосы	19	0.97	Cardona et al., 2014

Этническая группа	Размер выборки, N	Частота аллеля			
		rs4988235-T	rs182549-A		
Алтайцы южные	24	0.15	0.15		
Шорцы	24	0.06	0.06		
Буряты	24	0.02	0.02		
Якуты	22	0.07	0.07		
Эвенки западные	24	0.04	0.04		
Эвены	24	0.04	0.04		
Коряки	25	0.04	0.04		
Чукчи	14	0.25	0.25		
Эскимосы	19	0.03	0.05		

Таблица 2. Частоты аллелей rs4988235-Т и rs182549-А в популяциях Сибири, по данным (Cardona et al., 2014)

аллеля зарегистрированы примерно 3.1–3.8 тыс. л. н. у представителей карасукской (Юго-Западная Сибирь) и синташтинской (Южный Урал) археологических культур. У современного коренного населения Сибири средняя частота варианта rs4988235-T составляет 0.8 % (см. табл. 1). Таким образом, у сибирских народов за прошедшие 3 тыс. лет частота аллеля, способствующего усилению ферментативной активности лактазы, не увеличилась, несмотря на произошедшие изменения в диете и рост потребления молочных продуктов. Все это свидетельствует о том, что в популяциях Сибири аллель rs4988235-T ведет себя как нейтральный вариант генетического полиморфизма.

Возвращаясь к вопросу о возможных дополнительных вариантах полиморфизма энхансерной области гена МСМб у населения Восточной Азии, необходимо отметить, что такого рода скрининг проводился в отношении нескольких локусов, например rs41525747, rs41380347, rs869051967, rs145946881, rs182549 (Xu et al., 2010; Liebert et al., 2017; Anguita-Ruiz et al., 2020). Полиморфизм энхансерного элемента гена LCT изучался также в двух популяциях коренного населения Южной Сибири – у алтайских казахов и хакасов (Пилипенко и др., 2016). Однако частоты аллелей, потенциально связанных с ПЛ, как правило, были очень низкими. Исключение составляет только локус rs182549, в отношении которого сообщалось, что в некоторых восточноазиатских популяциях аллель rs182549-А даже более информативен, чем rs4988235-Т, поскольку аллель rs182549-А отмечался даже в отсутствие аллеля rs4988235-T (Sun et al., 2007; Mattar et al., 2010; Xu et al., 2010). Аналогичный вывод предлагался и для некоторых популяций африканского, европейского и западноазиатского происхождения (Bersaglieri et al., 2004; Coelho et al., 2005; Raz et al., 2013). Однако это противоречит предложенному ранее заключению о существовании полного неравновесия по сцеплению между вариантами полиморфизма rs4988235-Т и rs182549-А (Enattah et al., 2002; Troelsen et al., 2003), в том числе в восточноазиатских популяциях (Kato et al., 2018).

Анализ данных о распределении частот вариантов rs4988235-Т и rs182549-А в этнических группах Сибири, по данным (Cardona et al., 2014), также показывает наличие сцепления между указанными аллелями – из девяти исследованных популяций лишь в одном случае (у эскимосов) частота аллеля rs182549-А была выше, чем у аллеля rs4988235-Т (табл. 2). Таким образом, представляется маловероятным, что персистенция лактазы у коренного населения Сибири может быть обусловлена аллелем rs182549-А.

Полиморфные варианты, в том числе архаичные, генов *LCT* и *MCM6* у коренного населения Сибири

Результаты анализа полиморфизма экзонов и прилегающих к ним интронов генов MCM6 и LCT у коренного населения Сибири свидетельствуют о возможности существования в восточноазиатских популяциях вариантов полиморфизма, потенциально связанных с метаболизмом лактозы. В табл. 3 приведены данные о распространенности вариантов полиморфизма гена LCT и MCM6 у 102 представителей коренного населения Северо-Восточной (эскимосы, чукчи, коряки), Центральной (эвены, эвенки, якуты), Южной (тувинцы, шорцы, алтайцы, буряты) и Западной (кеты, ханты, манси, селькупы, ненцы) Сибири по результатам исследования полногеномной изменчивости (Pagani et al., 2016). В гене LCT выявлен 21 полиморфный локус, в гене МСМ6 – 7 локусов. Бо́льшая часть полиморфных вариантов, обнаруженных у коренного населения Сибири, относится к числу полиморфизмов, распространенных как в восточноазиатских, так и в европейских популяциях. Редкие варианты, характерные только для населения Восточной Азии, найдены в локусах rs201668742, rs144864087 и rs3739021; характерные только для европейцев – в локусе rs34307240.

Однако обращает на себя внимание группа полиморфных вариантов в локусах гs79023654 гена *LCT* и rs4988285, rs2070069, rs3087353 и rs2070068 – гена *MCM6* (отмечены полужирным в табл. 3). Локусы гs79023654, rs4988285 и rs2070069 присутствуют в некодирующей области генов; локусы rs3087353 и rs2070068 – в экзонах, но нуклеотидные замены в них не приводят к заменам аминокислот. Указанные в табл. 3 аллели находятся в сцепленном виде как у представителей коренного населения Сибири, так и в других популяциях Восточной Азии – у японцев, корейцев, вьетнамцев (см. табл. 3 и 4). В отношении сибир-

в ненулация двравни								
Вариант полиморфизма	Ген	CBC (<i>N</i> = 25)	ЦС (N = 29)	ЮС (N = 28)	3C (<i>N</i> = 20)	BA3	EBP	
rs62170085-G	LCT	6.0	0	0	2.5	0.26	2.84	
rs1042712-C	LCT	0	6.9	14.3	10.0	21.7	19.4	
rs2278544-G	LCT	60.0	51.7	64.3	67.5	43.3	68.6	
rs3213890-A	LCT	0	6.9	14.3	10.0	20.1	19.5	
rs2322659-C	LCT	62.0	58.6	64.3	70.0	45.6	66.5	
rs2304371-G	LCT	0	8.6	21.4	12.5	22.6	23.3	
rs3739022-A	LCT	36.0	32.8	12.5	20.0	21.5	13.9	
rs201668742-T	LCT	0	5.2	0	0	0.03	0	
rs144864087-C	LCT	4.0	6.9	7.1	5.0	1.03	0	
rs79023654-A	LCT	4.0	10.3	10.7	7.5	16.2	0	
rs35093754-C	LCT	0	1.7	7.1	2.5	5.04	2.92	
rs6719488-T	LCT	60.0	48.3	57.1	62.5	39.7	62.6	
rs2322812-G	LCT	36.0	32.8	12.5	17.5	21.5	13.9	
rs2874874-C	LCT	36.0	32.8	12.5	17.5	21.5	13.9	
rs7579771-A	LCT	40.0	51.7	44.6	37.5	60.4	37.4	
rs2164210-C	LCT	60.0	48.3	55.4	62.5	39.7	62.6	
rs60376570-A	LCT	36.0	32.8	12.5	17.5	21.5	13.9	
rs3816088-C	LCT	0	1.7	7.1	2.5	5.04	3.0	
rs3754689-T	LCT	4.0	19.0	26.8	20.0	37.7	20.2	
rs2236783-A	LCT	54.0	48.3	51.8	57.5	37.01	62.7	
rs34307240-A	LCT	2.0	0	1.8	0	0	0.95	
rs4988285-A	МСМ6	4.0	10.3	10.7	7.5	16.2	0	
rs3739021-A	МСМ6	0	1.7	3.6	2.5	0.17	0	
rs3087350-T	МСМ6	0	1.7	7.1	2.5	5.2	3.0	
rs2070069-G	МСМ6	4.0	10.3	10.7	7.5	16.2	0	••••
rs3087353-T	МСМ6	4.0	10.3	12.5	7.5	15.7	0	•••••
rs2070068-A	МСМ6	4.0	10.3	12.5	7.5	15.8	0	•••••
rs1057031-A	МСМ6	0	8.6	14.3	7.5	21.3	20.5	•••••

Таблица 3. Варианты полиморфизма экзонов и прилегающих к ним интронов генов *LCT* и *MCM6* и их частота (в %) в популяциях Евразии

Примечание. СВС – Северо-Восточная Сибирь; ЦС – Центральная Сибирь; ЮС – Южная Сибирь; ЗС – Западная Сибирь; ВАЗ – Восточная Азия; ЕВР – Европа. Для популяций Сибири частоты приводятся по работе (Pagani et al., 2016), для Восточной Азии и Европы – по базе данных dbSNP.

Таблица 4.	Частота (в %) вариантов	rs79023654-A, rs498	8285-A, rs2070069-	G, rs3087353-T и rs20)70068-А в популяциях мира
------------	-------------------------	---------------------	--------------------	-----------------------	----------------------------

Регион/страна	rs79023654-A	rs4988285-A	rs2070069-G	rs3087353-T	rs2070068-A
Европа	0	0	0	0	0
Сибирь	8.3	8.3	8.3	8.8	8.8
Восточная Азия	16.2	16.2	16.2	15.7	15.8
Япония	13.9	13.9	13.9	13.9	13.9
Вьетнам	17.1	17.1	11.2	12.8	13.1
Южная Корея	18.4	18.5	18.6	18.5	18.5
Южная Азия	1.1	0.58	0.58	0.58	0.58
Африка	0	0	0	0	24.7

Примечание. Популяционные частоты приводятся по базе данных dbSNP, для популяций Сибири – по работе (Pagani et al., 2016).

ских индивидуумов известно, что все эти аллели входят в состав генотипа rs4988235-CC, определяющего состояние первичной гиполактазии. Локус rs79023654 гена *LCT* находится на расстоянии примерно 29.7 тыс. п. н. от локуса rs4988285 гена *MCM6*, а сами полиморфные локусы внутри гена *MCM6* расположены друг от друга на расстоянии примерно 26.5 тыс. п. н. Причем локусы rs4988285 и rs2070069 находятся в области энхансера, регулирующего активность гена *LCT*. Все это свидетельствует о возможной функциональной значимости вариантов полиморфизма выявленного гаплотипа.

Анализ сведений из базы генетических данных dbSNP (https://www.ncbi.nlm.nih.gov/snp/) показал, что варианты rs79023654-A, rs4988285-A, rs2070069-G и rs3087353-T характерны только для населения Восточной Азии и с небольшой частотой (около 1 %) отмечались в популяциях Южной Азии (табл. 4). Однако пятый из этой группы аллель - rs2070068-А - обнаружен с высокой частотой (24.7%) в африканских популяциях (см. табл. 4). Из такого распределения можно сделать вывод, если основываться на сценарии заселения человеком Евразии из Африки, о том, что восточноазиатский гаплотип rs79023654-А, rs4988285-A, rs2070069-G, rs3087353-Т сформировался на основе предковых (африканских) гаплотипов, которые характеризовались вариантом rs2070068-А. Однако анализ палеогеномных данных (база данных AADR) показал, что вариант rs2070068-А появился в Африке намного позже, чем в Евразии. Самая ранняя находка этого аллеля в Африке связана с севером континента (на территории Марокко) примерно 14.5 тыс. л. н., остальные случаи обнаружены примерно 9 тыс. л. н. и позже. Зато оказалось, что в Евразии этим вариантом полиморфизма гена МСМ6 характеризовались денисовцы и неандертальцы (особи, жившие в интервале примерно от 40 до 110 тыс. л.н.), а также многие наиболее древние представители Ното sapiens в Европе и Восточной Азии (возрастом примерно от 34 до 44 тыс. л.н.).

Более детальный анализ, выполненный с помощью информации из баз данных, включающих палеогеномные последовательности (Denisova Variants Track Settings; https:// genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=dhcVcf DenisovaPinky), показал, что у денисовцев и неандертальцев был распространен MCM6-гаплотип: rs4988285-A, rs2070069-G, rs3087353-T, rs2070068-A. Локус rs79023654 гена LCT попал в область секвенирования с очень низким покрытием, поэтому наличие полиморфизма в этом локусе у денисовцев и неандертальцев остается под вопросом.

Таким образом, полученные результаты свидетельствуют о том, что обнаруженный у населения Восточной Азии (и в значительно меньшей мере Южной Азии) гаплотип гена *MCM6* является архаичным. Он мог быть либо унаследованным от общего предка *H. sapiens*, неандертальцев и денисовцев (примерно 600 тыс. л. н. от момента дивергенции предка *H. sapiens* от предков неандертальцев и денисовцев (Zeberg et al., 2024)), либо приобретенным в результате гибридизации с неандертальцами или денисовцами. Учитывая восточноазиатское распространение архаичного гаплотипа, более вероятной представляется интрогрессия от денисовцев. В свою очередь, установле-

но, что неандертальцы и денисовцы тоже обменивались генами, например, примерно 80–90 тыс. л. н. на юге Сибири (Slon et al., 2018). Поэтому переход вариантов полиморфизма от денисовцев к неандертальцам тоже вполне вероятен.

В последние годы проводится работа по каталогизации архаичных вариантов генетического полиморфизма, найденных в генофонде современных людей (https://bioinf. eva.mpg.de/catalogbrowser), однако неполнота такого рода сведений зависит от степени изученности популяций (Zeberg et al., 2024). Очевидно, что по мере географического расширения геномных исследований эта база данных станет намного полнее. Уже сейчас есть интересные находки редких предковых вариантов полиморфизма в далеко разобщенных популяциях, например идентичных аллелей ряда генов у южноафриканских койсанов и филиппинских аэта (Zeberg et al., 2024).

Имеется много информации о генетических вариантах, унаследованных современными людьми от неандертальцев, особенно в связи с преимуществами, которые люди получили в результате смешения, в отношении метаболизма, работы органов чувств, в частности ощущения боли, иммунитета (в том числе SARS-CoV-2), экспрессии некоторых генов (Telis et al., 2020; Zeberg et al., 2020, 2024; Pairo-Castineira et al., 2021; Haeggström et al., 2022).

Значительно меньше известно о функциональном проявлении генетического влияния денисовцев. Основные примеры такого влияния связаны с адаптацией к условиям высокогорья и холоду. Так, у жителей Тибета обнаружен участок ДНК денисовцев длиной примерно 33 тыс. п. н., который кодирует индуцируемый гипоксией фактор транскрипции EPAS1, участвующий в адаптации к низкому уровню кислорода (Zhang et al., 2021). У эскимосов Гренландии с высокой частотой найден участок ДНК денисовцев длиной примерно 28 тыс. п. н., включающий в состав гены WARS и TBX15; предполагается, что эти варианты полиморфизма способствуют адаптации арктических аборигенов к низким температурам (Racimo et al., 2017). По всей видимости, и в случае архаичного гаплотипа гена МСМ6, обнаруженного у населения Восточной Азии, представляется вполне вероятным, что с его помощью реализуется некая программа регуляции ферментативной активности лактазы, которая не потеряла актуальности и в наше время. Для прояснения деталей ее работы необходимо проведение комплексных медико-генетических, биохимических и физиологических исследований носителей ланного гаплотипа.

Заключение

Результаты обзора данных об изменчивости генов *LCT* и *MCM6* свидетельствуют о том, что коренное население Сибири издревле характеризовалось низкой частотой варианта rs4988235-T, способствующего усилению ферментативной активности лактазы. Некоторое увеличение со временем частоты этого аллеля в популяциях Центральной Азии и Юго-Западной Сибири связано с распространением на восток древнего населения степей Восточной Европы начиная с эпохи бронзового века (Heyer et al., 2011; Пилипенко и др., 2016; Segurel et al., 2020). Однако вариант rs4988235-T не достиг высоких частот в по-

пуляциях Центральной Азии, в отличие от Европы, что свидетельствует об отсутствии существенного селективного давления на этот вариант полиморфизма у населения Центральной Азии (Segurel et al., 2020). До сих пор неясно, почему в различных группах восточноазиатского населения, традиционно употребляющего молочные продукты, не выработались специфические варианты генетического полиморфизма, позволяющие метаболизировать лактозу. Одним из объяснений является гипотеза о культурной адаптации населения Центральной Азии, включающей в себя формирование культуры использования бактерий для сбраживания лактозы во время ферментации, что способствовало заселению кишечника специфической микрофлорой (Segurel et al., 2020).

Вместе с тем в регуляцию экспрессии генов метаболизма лактозы могут быть вовлечены и некоторые эпигенетические механизмы (в основном метилирование ДНК) (Labrie et al., 2016). Продемонстрировано также, что характер метилирования ДНК в энхансерных и промоторных участках гена LCT достаточно хорошо предсказывает фенотипы лактазы и, по всей видимости, эпигенетические изменения играют большую роль в регуляции лактазной недостаточности (Leseva et al., 2018). Таким образом, оба подхода - и генетический, и эпигенетический - должны быть использованы для исследований функциональной значимости вариантов полиморфизма, потенциально связанных с ПЛ, – в том числе архаичных генетических вариантов, которые, как показало настоящее исследование, все еще имеют некоторое распространение в популяциях человека.

Список литературы / References

Боринская С.А., Ребриков Д.В., Нефёдова В.В., Кофиади И.А., Соколова М.В., Колчина Е.В., Куликова Е.А., Чернышов В.Н., Куцев С.И., Полоников А.В., Иванов В.П., Козлов А.И., Янковский Н.К. Молекулярная диагностика и распространенность первичной гиполактазии в популяциях России и сопредельных стран. Молекуляр. биология. 2006;40(6):1031-1036

[Borinskaya S.A., Rebrikov D.V., Nefedova V.V., Kofiadi I.A., Sokolova M.V., Kolchina E.V., Kulikova E.A., Chernyshov V.N., Kutsev S.I., Polonikov A.V., Ivanov V.P., Kozlov A.I., Yankovsky N.K. Molecular diagnosis and frequencies of primary hypolactasia in populations of Russia and neighboring countries. *Mol. Biol.* 2006; 40(6):931-935. DOI 10.1134/S0026893306060124]

Козлов А.И., Вершубская Г.Г. D-витаминный статус и персистенция лактазы в европейских популяциях (обзор литературы с элементами мета-анализа). Вестн. Моск. ун-та. Серия XXIII. Антропология. 2017;3:68-75

[Kozlov A.I., Vershubskaya G.G. D-vitamin status and lactase persistence in European populations (review with the elements of metaanalysis. *Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Moscow University Anthropology Bulletin.* 2017;3:68-75 (in Russian)]

Пилипенко И.В., Пристяжнюк М.С., Кобзев В.Ф., Воевода М.И., Пилипенко А.С. Полиморфизм регуляторной области гена *LCT* в некоторых тюркоязычных популяциях Алтае-Саянского региона. *Вавиловский журнал генетики и селекции*. 2016;20(6): 887-893. DOI 10.18699/VJ16.209

[Pilipenko I.V., Pristyazhnyuk M.S., Kobzev V.F., Voevoda M.I., Pilipenko A.S. Polymorphism of the *LCT* gene regulatory region in Turkic-speaking populations of the Altay-Sayan region (southern Siberia). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2016;20(6):887-893. DOI 10.18699/ VJ16.209 (in Russian)] Соколова М.В., Васильев Е.В., Козлов А.И., Ребриков Д.В., Сенкеева С.С., Кожекбаева Ж.М., Люндуп А.В., Свечникова Н.С., Огурцов П.П., Хуснутдинова Э.К., Янковский Н.К., Боринская С.А. Полиморфизм С/Т-13910 регуляторного участка гена лактазы *LCT* и распространенность гиполактазии в популяциях Евразии. Экол. генетика. 2007;5:25-34. DOI 10.17816/ecogen 5325-34

[Sokolova M.V., Vasilyev E.V., Kozlov A.I., Rebrikov D.V., Senkeeva S.S., Kozhekbaeva Zh.M., Lyundup A.V., Svechnikova N.S., Ogurtsov P.P., Khusnutdinova E.K., Yankovsky N.K., Borinskaya S.A. Polymorphism C/T-13910 of the *LCT* gene regulatory region and lactase deficiency in Eurasian populations. *Ekologicheskaya Genetika* = *Ecological Genetics*. 2007;5:25-34. DOI 10.17816/ ecogen5325-34 (in Russian)]

- Anguita-Ruiz A., Aguilera C.M., Gil Á. Genetics of lactose intolerance: an updated review and online interactive world maps of phenotype and genotype frequencies. *Nutrients*. 2020;12(9):2689. DOI 10.3390/nu12092689
- Bersaglieri T., Sabeti P.C., Patterson N., Vanderploeg T., Schaffner S.F., Drake J.A., Rhodes M., Reich D.E., Hirschhorn J.N. Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. 2004;74:1111-1120. DOI 10.1086/421051
- Cardona A., Pagani L., Antao T., Lawson D.J., Eichstaedt C.A., Yngvadottir B., Shwe M.T.T., Wee J., Romero I.G., Raj S., Metspalu M., Villems R., Willerslev E., Tyler-Smith C., Malyarchuk B.A., Derenko M.V., Kivisild T. Genome-wide analysis of cold adaption in indigenous Siberian populations. *PLoS One.* 2014;9:e98076. DOI 10.1371/journal.pone.0098076
- Coelho M., Luiselli D., Bertorelle G., Lopes A.I., Seixas S., Destro-Bisol G., Rocha J. Microsatellite variation and evolution of human lactase persistence. *Hum. Genet.* 2005;117(4):329-339. DOI 10.1007/ s00439-005-1322-z
- Enattah N.S., Sahi T., Savilahti E., Terwilliger J.D., Peltonen L., Järvelä I. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 2002;30(2):233-237. DOI 10.1038/ng826
- Evershed R.P., Davey Smith G., Roffet-Salque M., Timpson A., Diekmann Y., Lyon M.S., Cramp L.J.E., ... Tasić N., van Wijk I., Vostrovská I., Vuković J., Wolfram S., Zeeb-Lanz A., Thomas M.G. Dairying, diseases and the evolution of lactase persistence in Europe. *Nature*. 2022;608(7922):336-345. DOI 10.1038/s41586-022-05010-7
- Haeggström S., Ingelman-Sundberg M., Pääbo S., Zeberg H. The clinically relevant CYP2C8*3 and CYP2C9*2 haplotype is inherited from Neandertals. *Pharmacogenomics J.* 2022;22(4):247-249. DOI 10.1038/s41397-022-00284-6
- Heyer E., Brazier L., Ségurel L., Hegay T., Austerlitz F., Quintana-Murci L., Georges M., Pasquet P., Veuille M. Lactase persistence in Central Asia: phenotype, genotype, and evolution. *Hum. Biol.* 2011; 83(3):379-392. DOI 10.3378/027.083.0304
- Ingram C.J., Elamin M.F., Mulcare C.A., Weale M.E., Tarekegn A., Raga T.O., Bekele E., Elamin F.M., Thomas M.G., Bradman N., Swallow D.M. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum. Genet.* 2007;120(6):779-788. DOI 10.1007/s00439-006-0291-1
- Irving-Pease E.K., Refoyo-Martínez A., Barrie W., Ingason A., Pearson A., Fischer A., Sjögren K.G., ... Korneliussen T., Werge T., Allentoft M.E., Sikora M., Nielsen R., Racimo F., Willerslev E. The selection landscape and genetic legacy of ancient Eurasians. *Nature*. 2024;625(7994):312-320. DOI 10.1038/s41586-023-06705-1
- Kato K., Ishida S., Tanaka M., Mitsuyama E., Xiao J.Z., Odamaki T. Association between functional lactase variants and a high abundance of *Bifidobacterium* in the gut of healthy Japanese people. *PLoS One.* 2018;13(10):e0206189. DOI 10.1371/journal.pone. 0206189
- Khabarova Y., Tornianen S., Tuomisto S., Järvelä I., Karhunen P., Isokoski M., Mattila K. Lactase non-persistent genotype influences milk consumption and gastrointestinal symptoms in Northern Russians. *BMC Gastroenterol.* 2011;11:124. DOI 10.1186/1471-230X-11-124

- Khabarova Y., Grigoryeva V., Tuomisto S., Karhunen P.J., Mattila K., Isokoski M. High prevalence of lactase non-persistence among indigenous nomadic Nenets, north-west Russia. *Int. J. Circumpolar Health.* 2012;71(1):1-6. DOI 10.3402/ijch.v71i0.17898
- Kovalenko E., Vergasova E., Shoshina O., Popov I., Ilinskaya A., Kim A., Plotnikov N., Barenbaum I., Elmuratov A., Ilinsky V., Volokh O., Rakitko A. Lactase deficiency in Russia: multiethnic genetic study. *Eur. J. Clin. Nutr.* 2023;77(8):803-810. DOI 10.1038/ s41430-023-01294-8
- Labrie V., Buske O.J., Oh E., Jeremian R., Ptak C., Gasiunas G., Maleckas A., Petereit R., Žvirbliene A., Adamonis K., Kriukienė E., Koncevičius K., Gordevičius J., Nair A., Zhang A., Ebrahimi S., Oh G., Šikšnys V., Kupčinskas L., Brudno M., Petronis A. Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat. Struct. Mol. Biol.* 2016;23(6):566-573. DOI 10.1038/ nsmb.3227
- Leseva M.N., Grand R.J., Klett H., Boerries M., Busch H., Binder A.M., Michels K.B. Differences in DNA methylation and functional expression in lactase persistent and non-persistent individuals. *Sci. Rep.* 2018;8(1):5649. DOI 10.1038/s41598-018-23957-4
- Liebert A., López S., Jones B.L., Montalva N., Gerbault P., Lau W., Thomas M.G., Bradman N., Maniatis N., Swallow D.M. World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection. *Hum. Genet.* 2017;136(11-12):1445-1453. DOI 10.1007/s00439-017-1847-y
- Mattar R., Monteiro M., Silva J.M., Carrilho F.J. LCT-22018G>A single nucleotide polymorphism is a better predictor of adult-type hypolactasia/lactase persistence in Japanese-Brazilians than LCT-13910C>T. *Clinics (Sao Paulo).* 2010;65(12):1399. DOI 10.1590/s1807-59322010001200030
- Olds L.C., Sibley E. Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a *cis* regulatory element. *Hum. Mol. Genet.* 2003;12(18):2333-2340. DOI 10.1093/ hmg/ddg244
- Pagani L., Lawson D.J., Jagoda E., Mörseburg A., Eriksson A., Mitt M., Clemente F., ... Thomas M.G., Manica A., Nielsen R., Villems R., Willerslev E., Kivisild T., Metspalu M. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature*. 2016; 538(7624):238-242. DOI 10.1038/nature19792
- Pairo-Castineira E., Clohisey S., Klaric L., Bretherick A.D., Rawlik K., Pasko D., Walker S., ... Maslove D., Ling L., McAuley D., Montgomery H., Walsh T., Pereira A.C., Renieri A.; GenOMICC Investigators; ISARIC4C Investigators; COVID-19 Human Genetics Initiative; 23andMe Investigators; BRACOVID Investigators; Gen-COVID Investigators; Shen X., Ponting C.P., Fawkes A., Tenesa A., Caulfield M., Scott R., Rowan K., Murphy L., Openshaw P.J.M., Semple M.G., Law A., Vitart V., Wilson J.F., Baillie J.K. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021; 591(7848):92-98. DOI 10.1038/s41586-020-03065-y
- Peng M.S., He J.D., Zhu C.L., Wu S.F., Jin J.Q., Zhang Y.P. Lactase persistence may have an independent origin in Tibetan populations from Tibet, China. J. Hum. Genet. 2012;57(6):394-397. DOI 10.1038/jhg.2012.41
- Racimo F., Gokhman D., Fumagalli M., Ko A., Hansen T., Moltke I., Albrechtsen A., Carmel L., Huerta-Sanchez E., Nielsen R. Archaic adaptive introgression in TBX15/WARS2. *Mol. Biol. Evol.* 2017; 34(3):509-524. DOI 10.1093/molbev/msw283

- Raz M., Sharon Y., Yerushalmi B., Birk R. Frequency of LCT-13910C/T and LCT-22018G/A single nucleotide polymorphisms associated with adult-type hypolactasia/lactase persistence among Israelis of different ethnic groups. *Gene.* 2013;519(1):67-70. DOI 10.1016/ j.gene.2013.01.049
- Scott A., Reinhold S., Hermes T., Kalmykov A.A., Belinskiy A., Buzhilova A., Berezina N., ... Krause R., Karapetian M., Stolarczyk E., Krause J., Hansen S., Haak W., Warinner C. Emergence and intensification of dairying in the Caucasus and Eurasian steppes. *Nat. Ecol. Evol.* 2022;6(6):813-822. DOI 10.1038/s41559-022-01701-6
- Ségurel L., Bon C. On the evolution of lactase persistence in humans. Annu. Rev. Genomics Hum. Genet. 2017;8:297-319. DOI 10.1146/ annurev-genom-091416-035340
- Segurel L., Guarino-Vignon P., Marchi N., Lafosse S., Laurent R., Bon C., Fabre A., Hegay T., Heyer E. Why and when was lactase persistence selected for? Insights from Central Asian herders and ancient DNA. *PLoS Biol.* 2020;18(6):e3000742. DOI 10.1371/journal. pbio.3000742
- Slon V., Mafessoni F., Vernot B., de Filippo C., Grote S., Viola B., Hajdinjak M., Peyregne S., Nagel S., Brown S., Douka K., Higham T., Kozlikin M.B., Shunkov M.V., Derevianko A.P., Kelso J., Meyer M., Prüfer K., Pääbo S. The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature*. 2018;561(7721):113-116. DOI 10.1038/s41586-018-0455-x
- Sun H.M., Qiao Y.D., Chen F., Xu L.D., Bai J., Fu S.B. The lactase gene -13910T allele can not predict the lactase-persistence phenotype in north China. Asia Pac. J. Clin. Nutr. 2007;16(4):598-601
- Telis N., Aguilar R., Harris K. Selection against archaic hominin genetic variation in regulatory regions. *Nat. Ecol. Evol.* 2020;4(11): 1558-1566. DOI 10.1038/s41559-020-01284-0
- Tishkoff S., Reed F., Ranciaro A., Voight B.F., Babbitt C.C., Silverman J.S., Powell K., Mortensen H.M., Hirbo J.B., Osman M., Ibrahim M., Omar S.A., Lema G., Nyambo T.B., Ghori J., Bumpstead S., Pritchard J.K., Wray G.A., Deloukas P. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 2007; 39(1):31-40. DOI 10.1038/ng1946
- Troelsen J.T., Olsen J., Møller J., Sjöström H. An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology*. 2003;125(6):1686-1694. DOI 10.1053/j.gastro.2003.09.031
- Xu L., Sun H., Zhang X., Wang J., Sun D., Chen F., Bai J., Fu S. The -22018A allele matches the lactase persistence phenotype in northern Chinese populations. *Scand. J. Gastroenterol.* 2010;45(2):168-174. DOI 10.3109/00365520903414176
- Zeberg H., Dannemann M., Sahlholm K., Tsuo K., Maricic T., Wiebe V., Hevers W., Robinson H.P.C., Kelso J., Pääbo S. A Neanderthal sodium channel increases pain sensitivity in present-day humans. *Curr. Biol.* 2020;30(17):3465-3469.e4. DOI 10.1016/j.cub.2020. 06.045
- Zeberg H., Jakobsson M., Pääbo S. The genetic changes that shaped Neandertals, Denisovans, and modern humans. *Cell*. 2024;187(5): 1047-1058. DOI 10.1016/j.cell.2023.12.029
- Zhang X., Witt K.E., Banuelos M.M., Ko A., Yuan K., Xu S., Nielsen R., Huerta-Sanchez E. The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proc. Natl. Acad. Sci. USA*. 2021; 118(22):e2020803118. DOI 10.1073/pnas.2020803118

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию 26.04.2024. После доработки 31.05.2024. Принята к публикации 03.06.2024.

DOI 10.18699/vjgb-24-73

Следы палеолитической экспансии в генофонде нивхов по данным о полиморфизме аутосомных SNP и Y-хромосомы

В.Н. Харьков ^[]¹ ⊠, Н.А. Колесников ^[], Л.В. Валихова¹, А.А. Зарубин ^[], А.Л. Сухомясова ^[], И.Ю. Хитринская ^[], В.А. Степанов ^[]

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

² Северо-Восточный федеральный университет им. М.К. Аммосова, Якутск, Россия

vladimir.kharkov@medgenetics.ru

Аннотация. Нивхи – малочисленный коренной народ Дальнего Востока, проживающий на территории Хабаровского края и острова Сахалин, который относится к потомкам древнего населения этих территорий. У нивхов преобладает специфичный сахалино-амурский антропологический тип. Они являются достаточно обособленными за счет длительной изоляции от контактов с другими народами. Генофонд нивхов охарактеризован по полногеномной панели аутосомных однонуклеотидных полиморфных маркеров и гаплогруппам У-хромосомы в сравнении с другими дальневосточными и сибирскими популяциями. Биоинформатическая обработка частот аутосомных SNP, гаплогрупп Y-хромосомы и YSTR-гаплотипов показала, что генофонд нивхов существенно отличается от генофондов других популяций. При анализе частот SNP методом РСА дальневосточные популяции располагаются в полном соответствии с территориями их проживания и делятся на северную группу чукчей и коряков и южную, включающую нивхов и удэгейцев. Удаленность нивхов совпадает с их географической локализацией, при этом нивхи и удэгейцы демонстрируют наибольшее родство. У нивхов выделяется специфичный для них компонент генофонда, который с гораздо меньшей частотой присутствует у удэгейцев и забайкальских эвенков и бурятов-А. По IBD-блокам генотипы нивхов демонстрируют очень небольшую долю совпадения с удэгейцами, коряками, эвенками и чукчами, значение которых является самым низким по сравнению с IBD-блокам между другими сибирскими популяциями. Показан специфичный для нивхов состав гаплогрупп и YSTRгаплотипов. Гаплогруппа С2а1 у нивхов разделена на три сублинии, которые имеют достаточно древнее происхождение и связаны с предками современных северных монголоидов. Нивхская гаплогруппа O2a1b1a2a-F238 есть у жителей Китая и Мьянмы. Линия Q1a1a1-M120 в исследованных в данной работе выборках представлена у нивхов, коряков, эвенков и юкагиров. Филогенетический анализ отдельных Ү-хромосомных гаплогрупп демонстрирует близость генофонда нивхов с коряками и тунгусскими народами, а также родство в меньшей степени с древним населением Приамурья и Приохотья и населением Юго-Восточной Азии. Генофонд нивхов подтверждает относительную малочисленность их предковой группы без смешения с другими популяциями. Ключевые слова: генофонд; популяции человека; генетическое разнообразие; генетические компоненты; Ү-хромосома; нивхи.

Для цитирования: Харьков В.Н., Колесников Н.А., Валихова Л.В., Зарубин А.А., Сухомясова А.Л., Хитринская И.Ю., Степанов В.А. Следы палеолитической экспансии в генофонде нивхов по данным о полиморфизме аутосомных SNP и Y-хромосомы. *Вавиловский журнал генетики и селекции*. 2024;28(6):659-666. DOI 10.18699/vjgb-24-73

Финансирование. Исследование выполнено за счет гранта Российского научного фонда № 22-64-00060, https://rscf.ru/project/22-64-00060/.

Traces of Paleolithic expansion in the Nivkh gene pool based on data on autosomal SNP and Y chromosome polymorphism

V.N. Kharkov (D¹ 🗟, N.A. Kolesnikov (D¹, L.V. Valikhova¹, A.A. Zarubin (D¹, A.L. Sukhomyasova (D², I.Yu. Khitrinskaya (D¹, V.A. Stepanov (D¹)

¹ Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia ² M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia

vladimir.kharkov@medgenetics.ru

Abstract. The Nivkhs are a small ethnic group indigenous of the Russian Far East, living in the Khabarovsk Territory and on Sakhalin Island, descending from the ancient inhabitants of these territories. In the Nivkhs, a specific Sakhalin-Amur anthropological type is prevalent. They are quite isolated, due to long isolation from contacts with other peoples. The gene pool of the Nivkhs and other Far Eastern and Siberian populations was characterized using a genome-wide panel of autosomal single-nucleotide polymorphic markers and Y chromosome haplogroups. Bioinformatic processing of

frequencies of autosomal SNPs, Y chromosome haplogroups and YSTR haplotypes showed that the Nivkh gene pool is very different from the other populations'. Analysis of the SNP frequencies using the PCA method divided the Far Eastern populations in full accordance with the territories of their residence into the northern group of the Chukchi and Koryaks and the southern group, including the Nivkhs and Udege. The remoteness of the Nivkhs coincides with their geographic localization, with the Nivkhs and Udege demonstrating the greatest kinship. The Nivkhs have a specific component of their gene pool, which is present with much less frequency in the Udege and Transbaikal Evenks. According to the IBD blocks, the genotypes of the Nivkhs show a very small percentage of coincidence with the Udege, Koryaks, Evenks and Chukchi, the value of which is the lowest compared to the IBD blocks among all other Siberian populations. The Nivkh-specific composition of haplogroups and YSTR haplotypes was shown. In the Nivkhs, the C2a1 haplogroup is divided into three sublines, which have a fairly ancient origin and are associated with the ancestors of modern northern Mongoloids. The Nivkh haplogroup O2a1b1a2a-F238 is found among residents of China and Myanmar. The Q1a1a1-M120 line is represented among the Nivkhs, Koryaks, Evenks and Yukaghirs. Phylogenetic analysis of individual Y chromosomal haplogroups demonstrated the closeness of the Nivkh gene pool with the ancient population of the Amur and Okhotsk regions, the Koryaks, the Tungus peoples and the population of Southeast Asia. The Nivkh gene pool confirms the relative smallness of their ancestral groups without mixing with other populations. Key words: gene pool; human populations; genetic diversity; genetic components; Y chromosome; Nivkhs.

For citation: Kharkov V.N., Kolesnikov N.A., Valikhova L.V., Zarubin A.A., Sukhomyasova A.L., Khitrinskaya I.Yu., Stepanov V.A. Traces of Paleolithic expansion in the Nivkh gene pool based on data on autosomal SNP and Y chromosome polymorphism. *Vavilovskii Zhurnal Genetiki i Selektsii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(6):659-666. DOI 10.18699/vjgb-24-73

Введение

Нивхи – это небольшой народ, населяющий районы Дальнего Востока: остров Сахалин и бассейн нижнего Амура. Численность нивхов на 2022 г. составила 3842 человека. Самоназвание – нивхгу (человек). Соседние народы называли их гиляха, гильями. Это название заимствовали русские, придав ему форму «гиляки». Слово «гиляки» на тунгусо-маньчжурских языках означает «люди, передвигающиеся при помощи парных весел на больших лодках».

По территориальному признаку нивхи делятся на две группы – островную (сахалинскую) и материковую. В прошлом нивхи занимали весьма обширную территорию. Ареал их расселения на материке простирался от Амура до бассейна Уды, на Сахалине - по всему западному и восточному побережью и в устье реки Поронай. В настоящее время нивхи Сахалина расселены в северной части острова и в бассейне реки Тымь. На материке они концентрируются в двух районах Хабаровского края – Николаевском и Ульчском. Говорят на нивхском языке, имеющем два диалекта: амурский и восточно-сахалинский. Нивхский язык вместе с кетским и юкагирским принадлежит к изолированным языкам. Ранее его относили к группе палеоазиатских языков в связи с нечеткой выраженностью генеалогического начала. Довольно сильная взаимосвязь между нивхским и чукотско-камчатскими языками была показана в работе (Fortescue, 2011).

Нивхи – прямые потомки древнейшего населения Сахалина и низовья Амура, расселенного в прошлом значительно шире. Они входят в палеоазиатский тип монголоидной расы. У нивхов преобладает сахалино-амурский антропологический тип, который частично есть и у ульчей. Вместе с чукчами, коряками и другими народами Северо-Востока Сибири они входят в группу палеоазиатов. Имеется точка зрения, что предки современных нивхов, северо-восточных палеоазиатов, эскимосов и индейцев Америки – звенья одной этнической цепи, охватывавшей в далеком прошлом северо-западные берега Тихого океана. На современный этнический облик нивхов большое влияние оказали их этнокультурные контакты с тунгусоманьчжурскими народами, айнами и японцами (Народы России, 1994; Суляндзига и др., 2003; Народы Северо-Востока Сибири, 2010).

Полученные в результате генотипирования высокоплотных микрочипов данные по аутосомным SNP в выборках нивхов и других коренных дальневосточных и сибирских народов позволяют точнее описать компонентный состав их генофондов, выявить общие по происхождению блоки сцепления и блоки гомозиготности по сравнению с ограниченными наборами различных ДНК-маркеров. Генотипирование расширенного набора специфичных SNP Y-хромосомы дает возможность гораздо более подробно охарактеризовать молекулярно-филогенетическую структуру отдельных гаплогрупп Y-хромосомы. Современные биоинформатические методы анализа генотипов на уровне отдельных индивидов позволяют максимально детально охарактеризовать генофонд исследованных выборок с применением различных методик.

В геноме человека насчитывается огромное количество SNP, что позволяет использовать их в качестве эффективного инструмента в структуре анализа генетических взаимоотношений между популяциями. Современная популяционная генетика человека имеет широкий выбор различных маркерных систем: аутосомных и однородительских ДНК-маркеров (определяющих филогению гаплогрупп У-хромосомы и митохондриальной ДНК).

Специфичной особенностью митохондриальных генофондов всех популяций Приморья является присутствие линий мтДНК гаплогруппы Ү, максимальные частоты которой отмечены у нивхов Сахалина (66.1 %) и ульчей (37.9 %) (Starikovskaya et al., 2005). Высока ее частота у айнов (25.5 %), негидальцев (21.2 %) (Starikovskaya et al., 2005), коряков (5.7 %), эвенов (8.1 %), восточных эвенков (8.9 %) (Деренко, Малярчук, 2010). В других азиатских популяциях частота этой линии значительно ниже и постепенно падает, по мере удаления от территории проживания основных ее носителей. Именно с территориями нижнего течения Амура и Сахалина связывают происхождение специфичных линий мтДНК данной гаплогруппы.

2024

28•6

Географическое распространение подгруппы Y1a1 ограничено Северо-Восточной Азией. Все линии, присутствующие в популяциях коряков, эвенов, ительменов, негидальцев, нивхов, ороков и айнов, относятся исключительно к подгруппе Y1a1 (Horai et al., 1996; Schurr et al., 1999; Бермишева и др., 2005; Starikovskaya et al., 2005; Деренко, Малярчук, 2010). Основной ареал этой митохондриальной линии и частот сублиний У-хромосомной гаплогруппы C2a1 хорошо соотносятся друг с другом, являясь примером параллельной экспансии У-хромосомной и митохондриальной гаплогрупп в пределах одного и того же региона. Эти результаты согласуются с полученными ранее данными по древним геномам коренного населения бассейна Амура, которые формируют отдельный генетический кластер, включающий древние и современные популяции Приамурья (народы тунгусской языковой группы и нивхи) (Wang et al., 2021).

Целью настоящего исследования был комплексный анализ структуры генофонда нивхов в сравнении с другими популяциями коренного населения Сибири и Дальнего Востока. Для решения вопросов генетической близости нивхов с другими коренными народами выполнено генотипирование широкого геномного набора аутосомных маркеров на высокоплотных микроматрицах ДНК, а также расширенного набора SNP и STR-маркеров Y-хромосомы у различных этнических групп: удэгейцев, чукчей, коряков, якутов, эвенков, бурятов, тувинцев, хакасов, южных алтайцев, кетов, чулымцев и хантов.

Материалы и методы

Материал исследования составили образцы ДНК мужчин и женщин популяции нивхов (N = 155) из поселков Некрасовка и Москальво Охинского района Сахалинской области. Забор первичного биологического материала (венозной крови) у доноров производили с соблюдением процедуры письменного информированного согласия на проведение исследования (протокол № 10 Комитета по биомедицинской этике НИИ медицинской генетики от 15.02.2021). На каждого донора составляли анкету с краткой родословной, указанием этнической принадлежности и мест рождения предков. Индивида относили к данной этнической группе на основании его собственной этнической идентификации, его родителей и места рождения.

Для анализа Y-хромосомных гаплогрупп и гаплотипов мужчин из выборки нивхов были использованы 52 образца ДНК. Для генотипирования на высокоплотных микроматрицах были выбраны неродственные образцы нивхов (N = 13), не имеющих метисации с представителями других народов. Такое небольшое количество образцов для генотипирования связано с тем, что существенная доля собранных индивидов являются потомками межэтнических браков за последние несколько поколений, а также с относительной малочисленностью этих народов и наличием в выборках кровных родственников по материнской и отцовской линии.

Другие популяции коренного населения Сибири для генотипирования представлены удэгейцами (N = 15, села Красный Яр Пожарского района и Агзу Тернейского района на северо-востоке Приморского края), коряками (N = 20, Корякский автономный округ Камчатской области), чукча-

ми (N=25, поселки Лорино, Сиреники, Янарыкот и Новое Чаплино Чукотского автономного округа – относятся к береговой группе), южными алтайцами (N=24, с. Бешпельтир Чемальского района; N = 25, с. Кулада Онгудайского района), кетами (N=15, пос. Келлог Туруханского района Красноярского края), томскими татарами (N = 20, поселки Черная Речка, Эушта и Тахтамышево Томского района), тувинцами (N = 28, с. Тээли Бай-Тайгинского кожууна), бурятами (N = 23, пос. Агинское Агинского района; N = 28, с. Курумкан Курумканского района), хантами (N = 30, с. Казым Белоярского района; N = 26, дер. Русскинская Сургутского района), хакасами (сагайцами Таштыпского района, N = 29, и качинцами Ширинского района, N=26), чулымцами (N=22), эвенками (3-забайкальские, пос. Чара Каларского района, села Моклакан и Тупик Тунгиро-Олёкминского района, N = 25; Я – якутские, N = 28) и якутами (N = 26, с. Чериктей в Усть-Алданском улусе). Материал депонирован в биоресурсной коллекции «Биобанк населения Северной Евразии».

Данные широкогеномного генотипирования получены с использованием микрочипов Infinium Multi-Ethnic Global 8 (Illumina) для SNP-генотипирования, включающего свыше 1.7 млн маркеров. Кластеризацию массива генотипов аутосомных SNP (single nucleotide polymorphism) и контроль качества выполняли с помощью протокола, разработанного Y. Guo с коллегами (2014), с использованием GenomeStudio (Illumina, GenomeStudio, модуль генотипирования v2.0.3). Для фильтрации, нормализации и расчета стандартных геномных статистик и показателей оптимальным был стандартный набор программ, включающий vcftools, bcftools и plink.

Идентичные по происхождению блоки сцепления анализировали с использованием алгоритма Refined IBD (Browning B.L., Browning S.R., 2013), показывающего более точные результаты по сравнению со встроенными в plink алгоритмами. Предварительно генотипы были фазированы с помощью программного обеспечения Beagle 5.1 (Browning S.R., Browning B.L., 2007). Для сравнения популяций были получены суммы средних длин идентичных по происхождению блоков (сегментов IBD – identical by descent) между парами индивидов.

Анализ генетических взаимоотношений между популяциями осуществляли методом PCA. Компонентный состав и количество примесей у отдельных индивидов и популяций определяли с использованием методики NGSadmix (Skotte et al., 2013) и программы Admixture (Alexander et al., 2009; Alexander, Lange, 2011).

Для изучения состава и структуры гаплогрупп Y-хромосомы в исследование были включены две системы генетических маркеров: диаллельные локусы, представленные SNP, и полиаллельные высоковариабельные микросателлиты (YSTR). С помощью 589 SNP-маркеров определяли принадлежность мужчин к различным гаплогруппам. Генотипирование SNP-маркеров проводили с применением полимеразной цепной реакции и последующего анализа фрагментов ДНК с помощью ПДРФ-анализа (полиморфизм длин рестрикционных фрагментов). Для специфичных терминальных SNP по отдельным сублиниям проводилось генотипирование небольшого количества образцов по данным их YSTR-гаплотипов и результатам NGS-секвенирования Y-хромосомы. Обозначение гаплогрупп приводится с привязкой к ISOGG 2019 Y-DNA Нарlogroup Tree. Анализ STR-гаплотипов внутри гаплогрупп выполняли с использованием 44 STR-маркеров нерекомбинирующей части Y-хромосомы (DYS19, 385a, 385b, 388, 389I, 389II, 390, 391, 392, 393, 426, 434, 435, 436, 437, 438, 439, 442, 444, 445, 448, 449, 456, 458, 460, 461, 481, 504, 505, 518, 525, 531, 533, 537, 552, 570, 576, 635, 643, YCAIIa, YCAIIb, GATA H4.1, Y-GATA-A10, GGAAT1B07). STR-маркеры генотипировали с помощью капиллярного электрофореза на приборах ABI Prism 3730 и Нанофор-05.

Экспериментальные исследования осуществлены на базе Центра коллективного пользования научно-исследовательским оборудованием «Медицинская геномика» (НИИ медицинской генетики Томского НИМЦ). Построение медианных сетей гаплотипов Y-хромосомы выполняли в программе Network v.10.2.0.0 (Fluxus Technology Ltd; www.fluxus-engineering.com) по методу медианных сетей Бандельта (Bandelt et al., 1999). Возраст генерации наблюдаемого разнообразия гаплотипов в гаплогруппах оценивали методом ASD (Zhivotovsky et al., 2004) на основании среднеквадратичных отличий в числе повторов между всеми маркерами.

Результаты и обсуждение

После обработки данных по результатам микрочипового исследования для фильтрации генотипированных образцов и проведения дальнейших расчетов был выполнен поиск среди нивхов метисов с использованием программы NGSadmix. Метод NGSadmix при запуске на сформированном нами массиве данных показал, что все образцы чистых нивхов не имеют метисации с другими народами, что совпадает с результатами их анкетирования.

Генетические взаимоотношения нивхов с народами Восточной и Северо-Восточной Сибири

При анализе данных по частотам аутосомных SNP с помощью метода PCA на уровне отдельных образцов (рис. 1) видно, что нивхи наиболее близки удэгейцам, а также к эвенкам из Забайкалья и Якутии. Чукчи и коряки значительно удалены от остальных популяций, что

согласуется с их сильной географической изоляшией на территории Северо-Восточной Сибири. Именно РС2 отделяет их от всех проанализированных выборок, но по РС1 чукчи и коряки очень близки к нивхам и удэгейцам. Их удаление от более южных народов говорит о наличии в чукотско-корякском генофонде более древнего, специфичного для них генетического компонента, связанного с аборигенным палеолитическим населением территорий их проживания. Дальневосточные образцы разделены в полном соответствии с территориями их проживания на северную группу чукчей и коряков и южную, включающую нивхов и удэгейцев. Эвенки из Забайкалья и Якутии также близки друг к другу. Немного более удалены якуты и буряты. Удаленность нивхов от остальных популяций совпадает с их географической локализацией (см. рис. 1). Нивхи, удэгейцы, чукчи и коряки составляют дальневосточную группу популяций, при этом нивхи и удэгейцы демонстрируют наибольшее родство.

Практически все образцы отдельных этносов формируют специфичные кластеры (см. рис. 1), которые могут частично пересекаться, за исключением томских татар, имеющих достаточно гетерогенный состав генофонда (Валихова и др., 2022). При трехкомпонентном анализе и на графике методом t-SNE все этноспецифичные кластеры намного больше удалены друг от друга. Отдельные образцы разных выборок, выделяющиеся из этих общих групп, показывают метисацию при их анализе методом NGSadmix, что как раз и влияет на их расположение на графике.

Компонентный состав генофонда популяций

Для определения генетических компонент в генофонде изучаемых популяций была использована программа Admixture, позволяющая выявлять гетерогенность компонентного состава генома индивидов на основе данных о генотипах и точно определять их распределение на уровне популяций и отдельных образцов. При задании числа предковых компонент больше четырех почти во всех популяциях выявляется специфичный для нивхов генетический компонент, особенно отчетливо проявляющийся на анализируемом массиве популяционных выборок при



Рис. 1. Дифференциация геномов населения Дальнего Востока и Сибири по двум компонентам РСА.

2024

28•6



Рис. 2. Упорядоченная картина компонент Admixture при ранжировании популяций Сибири с запада на восток, К = 8. Здесь и на рис. 3: хакасы С – сагайцы Аскизского района; хакасы Т – сагайцы Таштыпского района; хакасы К – качинцы Ширинского района; чулымцы К – Красноярский край; чулымцы Т, татары Т – Томская область.



Рис. 3. Упорядоченная картина компонент Admixture при ранжировании популяций Сибири с запада на восток, К = 10.

K = 8, который можно интерпретировать как «сахалиноамурский» генетический пласт в генофонде современных популяций (рис. 2).

При К = 8 этот компонент полностью доминирует у нивхов (0.92) и удэгейцев (0.61), встречается у бурят (0.50), алтай-кижи (0.34), хакасов-качинцев, тувинцев (0.30), алтайцев пос. Бешпельтир (0.23), томских татар (0.11), эвенков (0.02–0.09), хакасов-сагайцев (0.07) и якутов (0.01). Возможно, этот генетический пласт связан с древним субстратом в рассматриваемых популяциях.

При К = 10 происходит более детальное разделение (рис. 3): у нивхов выделяется специфичный для них компонент (0.98), выделенный голубым цветом на рис. 3, который присутствует у удэгейцев (0.22) и в небольшой степени у забайкальских эвенков (0.05), эвенков Якутии, хакасов, томских татар и бурят (0.02). Доминирование по частоте этого компонента у всех образцов нивхов подтверждает, что их предки долгое время не контактировали с другими народами и проживали в изоляции на острове Сахалин. Полученные данные доказывают, что коренное население Сахалина длительное время не смешивалось с другими этносами.

Идентичные по происхождению блоки сцепления

Для оценки общих по происхождению блоков ДНК был проведен анализ совпадения на уровне отдельных индивидов и популяций. Фрагмент, имеющий идентичные нуклеотидные последовательности, у разных людей является наследием их общего предка. Размер такого сегмента IBD сопоставим с числом поколений за счет рекомбинации хромосом при формировании половых клеток. Использование информации об общих по происхождению участках генома на уровне отдельных индивидов и популяций позволяет количественно оценить степень генетического родства между людьми и дает дополнительную информацию о генетических связях популяций (Gusev et al., 2012).

Генотипы нивхов показали совпадение по IBD-блокам между собой >1.5 сМ (11 %), далее – с удэгейцами (0.58 %), коряками (0.47 %), эвенками (0.28 %) и чукчами (0.18 %). С другими сибирскими популяциями доля их совпадений намного ниже (рис. 4). Совпадения между нивхами и другими исследованными популяциями являются самыми низкими по сравнению с другими этносами. Это подтверждает их очень долгую по времени изоляцию и отсутствие контактов с другими народами. Доля межпо-



Рис. 4. Сумма длин сегментов IBD > 1.5 сМ между парами индивидов нивхов и сибирскими популяциями.

пуляционных IBD-блоков между нивхами, удэгейцами, коряками, чукчами и эвенками согласуется с результатами PCA и Admixture. Анализ IBD внутри популяций нивхов, коряков и чукчей показал, что они имеют больше общих IBD, чем люди из других выборок. При этом у чукчей (55 %), коряков (57 %) и нивхов (59 %) наибольший вклад вносят короткие IBD-фрагменты, что может говорить о «бутылочном горлышке» в прошлом, во время миграций на север и северо-восток или об изоляции от остальных популяций, населяющих территорию Сибири.

Коэффициент геномного инбридинга

При оценке коэффициента геномного инбридинга для длин ROH > 1.5 м.п.о. у нивхов выявлен относительно невысокий уровень кровного родства (FROH = 0.0268). У коряков (FROH = 0.0446) и чукчей (FROH = 0.0431) он максимален для сибирских популяций и почти вдвое выше их среднестатистического значения на территории Сибири и Дальнего Востока. Для нивхов, чукчей и коряков показано значительное увеличение суммарной длины среднего класса ROH на индивида по сравнению с остальными популяциями. Эти данные вносят дополнение к сравнению с классом коротких ROH в сибирских популяциях. Полученные результаты свидетельствуют об относительно небольшой численности предковых групп указанных народов на протяжении многих поколений и брачных контактах между родственниками, а о также возможном эффекте «бутылочного горлышка». Уровень гомозиготности в геномах представителей дальневосточных народов отличается наиболее высокими показателями инбридинга среди всех коренных сибирских народов. Они имеют длинные гомозиготные участки для всех категорий длины ROH у большинства обследованных образцов. Эти данные подтверждают относительную малочисленность их предковых групп на протяжении длительного времени и территориальную изолированность, которая исключала смешение с другими популяциями.

Гаплогруппы Ү-хромосомы

В результате генотипирования SNP и YSTR-маркеров и определения гаплогрупп Y-хромосомы у всех образцов мужчин нивхов было показано совпадение данных их анкет по отцовской линии. Мужчины, являющиеся метисами с восточно-европейскими народами по отцовской линии, относятся к специфичным европейским сублиниям гаплогрупп E, I1, N1a1, N1a2 и R1a1. Гаплогруппы метисов с корейцами и орочонами принадлежат к восточно-азиатским вариантам клад C2, O1 и O2. Все остальные образцы нивхов относятся к специфичным для них сублиниям трех гаплогрупп.

Наиболее частой гаплогруппой у нивхов является C2a1 (86%). Такая высокая частота этой Y-хромосомной линии не зафиксирована ни у одного из проанализированных этносов. Гаплогруппа C2a1 максимальна у чистокровных по мужской линии нивхов по сравнению с другими народами. Она является субстратным элементом их генофонда, связанным с автохтонными группами населения Приохотья.

Из 37 мужчин нивхов без метисации по отцовской линии 16 человек относятся к сублинии C2a1a (B90, Z32902, Z32912, Z32919, Z32926, Z32937 (xB93, Z32958)) (см. таблицу). Ее возраст ранее был определен в 4216 лет (3700-4667) (Liu et al., 2021). Эта ветвь формирует особый кластер YSTR-гаплотипов, специфичный для нивхов и коряков, который характеризуется сокращением до десяти числа тандемных повторов в локусе DYS389I. Параллельная ей линия С2а1а-В93 присутствует также у эвенков, эвенов, коряков, юкагиров и якутов. У якутских эвенков и юкагиров она составляет 15-20 %. В очень большой выборке якутов к ней принадлежат всего четыре образца; один образец обнаружен также у забайкальских эвенков. Наличие у этих популяций специфичной ветви C2a1a2b (В93) связано с древними коренными популяциями Приамурья и Приохотья, которые очень давно отделились от азиатских предков из более южных регионов. По данным

Частоты встречаемости гаплогрупп	Ү-хромосомы у нивхов
----------------------------------	----------------------

Гаплогруппы	% (<i>N</i> = 37)
C2a1a1b1a~ – B473, F10085, F13958 (xZ32848, FGC28920, BY186309)	32.4 % (12)
C2a1a – B90, Z32902, Z32912, Z32919, Z32926, Z32937 (xB93, Z32958)	43.2 % (16)
C2a1 – F3447, ACT1932, ACT1942	10.8 % (4)
O2a1b1a2a – F238	5.4 % (2)
Q1a1a1 – M120, F746, Y34108, Y34449	8.1 % (3)

научного коллектива из Тарту (Karmin et al., 2015), у трех образцов мужчин коряков, принадлежащих к гаплогруппе C3c2, гаплотипы почти полностью совпадают с нашими образцами из этой линии. У двух эвенков из Монголии (Liu et al., 2021) и одного из России тоже была обнаружена сублиния C2a1a2b-B90 (Karmin et al., 2015). Эта ветвь и родственная ей C2a1a2b-M86 разделились с ветвью C2a-M48 около 11.6 тыс. лет назад (Liu et al., 2021). Распространение C2a1a2b-M86 на территории Восточной Сибири связано с относительно недавней миграцией тунгусских племен из Приамурья и Маньчжурии. Специфичная для нивхов сублиния C2a1a2b (хВ93) отделилась от общего предка еще до формирования мутации В93 у тунгусских народов.

Второй по частоте у нивхов является линия C2a1a1b1a~ F13958 (32.4 %). Эта линия обнаружена у одного казаха и трех киргизов, но по гаплотипам они значительно отличаются от нивхов. По данным сайта YFull, возраст ее общего предка составляет 4300 лет (ДИ = 5200–3500). К линии C2a1 (F3447, ACT1932, ACT1942) относятся четыре нивха. Согласно данным сайта YFull, возраст ее общего предка 16 000 лет (ДИ = 17 300–14 800). Эта древняя по происхождению линия обнаружена у двух китайцев из провинции Ляонин, корейца и японца.

Большое разнообразие линий C2a1 у нивхов и их возраст свидетельствуют об очень раннем появлении данной гаплогруппы на указанной территории. Распространение этой линии при формировании генофонда древнего населения Северо-Восточной Азии связано с ранними миграциями монголоидных племен. Таким образом, C2a1 является маркерной для расселения предков современных северных континентальных монголоидов и их дальнейшей дифференциации в Сибири, а также второй волны заселения Америки, представители которой сохранили морфологические особенности древних протомонголоидов Азии.

В целом генофонд нивхов по аутосомным SNP и гаплогруппам Y-хромосомы, с одной стороны, занимает промежуточное положение между генофондами коряков и удэгейцев, с другой – менее разнообразен по составу и отличается наличием трех специфичных вариантов. Самая высокая среди сибирских популяций частота гаплогрупп C2a1a-B90 (хB93) делает его уникальным объектом для исследования палеолитических пластов суммарного дальневосточного генофонда и реконструкции наиболее ранних этапов заселения человеком Северо-Востока Азии.

Общая медианная сеть гаплотипов гаплогруппы C2a1 очень разветвленная и состоит из трех кластеров гаплотипов, совпадающих по генотипам терминальных SNP для этих сублиний (рис. 5). Это соответствует оценке времени их разделения. Все три кластера демонстрируют наличие общих предков по мужской линии, потомками которых являются все проанализированные образцы нивхов.

Таким образом, популяции, принесшие гаплогруппу C2a1 на территорию Приамурья и Камчатки, по-видимому, мигрировали на север вдоль тихоокеанского побережья. Высокое гаплотипическое разнообразие C2a1 на Дальнем Востоке говорит о значительно более раннем появлении этой гаплогруппы на указанной территории в сравнении с Южной Сибирью. Распространение данной линии при формировании генофонда древнего населения Северной Азии связано, вероятно, с миграциями монголоидных племен, сложившихся в центральноазиатскую, байкальскую и арктическую группы антропологических типов.

Два нивха имеют гаплогруппу O2a1b1a2a-F238 (см. таблицу). Она представлена у жителей Китая и одного из Мьянмы. Возраст ее общего предка 7500 лет (ДИ = 8600–6400). Еще три нивха относятся к редкой линии Q1a1a1 – M120, F746, Y34108, Y34449, к которой принадлежит один коряк, эвенк из Якутии и четыре юкагира.

Заключение

Распространение носителей C2a1 происходило, несомненно, с ассимиляцией более древнего местного населения. Таким образом, генофонд нивхов является достаточно специфичным по составу гаплогрупп Y-хромосомы и мтДНК, но очень близким по аутосомным маркерам между всеми исследованными нивхами. Анализ образцов свидетельствует о близком генетическом родстве нивхов с ко-



Рис. 5. Медианная сеть YSTR-гаплотипов гаплогруппы C2a1 у нивхов.

ряками, чукчами, удэгейцами и эвенками. Специфичность сублиний Y-хромосомы и YSTR-гаплотипов доказывает, что нивхи долго не контактировали с другими этносами и проживали в относительной изоляции на протяжении многих столетий. Это подтверждается и результатами микрочипового анализа. Данные о генофонде нивхов дополняют информацию палеогенетических, лингвистических, антропологических и этнологических направлений исследований. По данным этногенеза нивхи являются палеоазиатами. Именно на их генетическом субстрате позднее формировались и другие амурские народы, что хорошо согласуется с результатами настоящего исследования их генофондов.

Список литературы / References

- Бермишева М.А., Кутуев И.А., Спицын В.А., Виллемс Р., Батырова А.З., Коршунова Т.Ю., Хуснутдинова Э.К. Анализ изменчивости мтДНК в популяции ороков. *Генетика*. 2005;41(1): 78-84
 - [Bermisheva M.A., Kutuev I.A., Spitsyn V.A., Villems R., Batyrova A.Z., Korshunova T.Yu., Khusnutdinova E.K. Analysis of mitochondrial DNA variation in the population of Oroks. *Russ. J. Genet.* 2005;41(1):66-71. DOI 10.1007/pl00022112]
- Валихова Л.В., Харьков В.Н., Зарубин А.А., Колесников Н.А., Сваровская М.Г., Хитринская И.Ю., Штыгашева О.В., Волков В.Г., Степанов В.А. Генетическая взаимосвязь чулымских тюрков с хакасами и кетами по данным аутосомных SNP и гаплогруппам Y-хромосомы. *Генетика*. 2022;58(10):1177-1184. DOI 10.31857/ S0016675822100113

[Valikhova L.V., Kharkov V.N., Zarubin A.A., Kolesnikov N.A., Svarovskaya M.G., Khitrinskaya I.Yu., Shtygasheva O.V., Volkov V.G., Stepanov V.A. Genetic interrelation of the Chulym Turks with Khakass and Kets according to autosomal SNP data and Y-chromosome haplogroups. *Russ. J. Genet.* 2022;58(10):1228-1234. DOI 10.1134/S1022795422100118]

Деренко М.В., Малярчук Б.А. Молекулярная филогеография населения Северной Евразии по данным об изменчивости митохондриальной ДНК. Магадан, 2010

[Derenko M.V., Malyarchuk B.A. Molecular Phylogeography of the Population of Northern Eurasia Based on Data on Mitochondrial DNA Variability. Magadan, 2010 (in Russian)]

Народы России: Энциклопедия. М.: Большая Рос. энцикл., 1994 [The Peoples of Russia: Encyclopedia. Moscow: Bolshaya Rossiyskaya Entsiklopediya Publ., 1994 (in Russian)]

Народы Северо-Востока Сибири. М.: Наука, 2010

[Peoples of North-East Siberia. Moscow: Nauka Publ., 2010 (in Russian)]

Суляндзига Р.В., Кудряшова Д.А., Суляндзига П.В. Коренные малочисленные народы Севера, Сибири и Дальнего Востока Российской Федерации. Обзор современного положения. М., 2003

[Sulyandziga R.V., Kudryashova D.A., Sulyandziga P.V. Indigenous Peoples of the North, Siberia, and the Far East of the Russian Federation. Review of the current situation. Moscow, 2003 (in Russian)]

- Alexander D.H., Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform*. 2011;12: 246. DOI 10.1186/1471-2105-12-246
- Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655-1664. DOI 10.1101/gr.094052

Bandelt H.J., Forster P., Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 1999;16:37-48. DOI 10.1093/oxfordjournals.molbev.a026036

- Browning B.L., Browning S.R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 2013; 194(2):459-471. DOI 10.1534/genetics.113.150029
- Browning S.R., Browning B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 2007; 81(5):1084-1097. DOI 10.1086/521987
- Fortescue M.D. The relationship of Nivkh to Chukotko-Kamchatkan revisited. *Lingua*. 2011;121:1359-1376. DOI 10.1016/j.lingua.2011. 03.001
- Guo Y., He J., Zhao S., Wu H., Zhong X., Sheng Q., Samuels D.C., Shyr Y., Long J. Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* 2014;9:2643-2662. DOI 10.1038/ nprot.2014.174
- Gusev A., Palamara P.F., Aponte G., Zhuang Z., Darvasi A., Gregersen P., Pe'er I. The architecture of long-range haplotypes shared within and across populations. *Mol. Biol. Evol.* 2012;29(2):473-486. DOI 10.1093/molbev/msr133
- Horai S., Murayama K., Hayasaka K., Matsubayashi S., Hattori Y., Fucharoen G., Harihara S., Park K.S., Omoto K., Pan I.H. mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* 1996;59(3):579-590
- Karmin M., Saag L., Vicente M., Wilson Sayres M.A., Järve M., Talas U.G., Rootsi S., Ilumäe A.M., Mägi R., Mitt M., ... Tyler-Smith K., Underhill P.A., Willerslev E., Nielsen R., Metspalu M., Villems R., Kivisild T. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 2015; 25(4):459-466. DOI 10.1101/gr.186684.114
- Liu B.L., Ma P.C. Wang C.Z., Yan S., Yao H.B., Li H.L., Xie Y.M., Meng S.L., Sun J., Cai J.H., Sarengaowa S., Li H., Cheng H.Z., Wei L.H. Paternal origin of Tungusic-speaking populations: insights from the updated phylogenetic tree of Y-chromosome haplogroup C2a-M86. *Am. J. Hum. Biol.* 2021;33(2):e23462. DOI 10.1002/ajhb. 23462
- Schurr T., Sukernik R., Starikovskaya Y., Wallace D. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea–Bering Sea region during the Neolithic. *Am. J. Phys. Anthropol.* 1999;108:1-39. DOI 10.1002/(SICI)1096-8644(199901)108:1<1::AID-AJPA1>3.0.CO;2-1
- Skotte L., Korneliussen T., Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics*. 2013;195(3):693-702. DOI 10.1534/genetics.113.154138
- Starikovskaya E.B., Sukernik R.I., Derbeneva O.A., Volodko N.V., Ruiz-Pesini E., Torroni A., Brown M.D., Lott M.T., Hosseini S.H., Huoponen K., Wallace D.C. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann. Hum. Genet.* 2005;69:67-89. DOI 10.1046/j.1529-8817.2003.00127.x
- Wang C.C., Yeh H.Y., Popov A.N., Zhang H.Q., Matsumura H., Sirak K., Cheronet O., Kovalev A., Rohland N., Kim A.M., ... Schiffels S., Kennett D.J., Jin L., Li H., Krause J., Pinhasi R., Reich D. The genomic formation of human populations in East Asia. *Nature*. 2021;591(7850):413-419. DOI 10.1038/s41586-021-03336-2
- Zhivotovsky L.A., Underhill P.A., Cinnioğlu C., Kayser M., Morar B., Kivisild T., Scozzari R., Cruciani F., Destro-Bisol G., Spedini G., Chambers G.K., Herrera R.J., Yong K.K., Gresham D., Tournev I., Feldman M.W., Kalaydjieva L. On the effective mutation rate at Y-chromosome STRs with application to human population divergence time. *Am. J. Hum. Genet.* 2004;74(1):50-61. DOI 10.1086/ 380911

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 03.04.2024. После доработки 21.06.2024. Принята к публикации 22.06.2024.

DOI 10.18699/vjgb-24-74

Полиморфные варианты гена рецептора дофамина DRD2 (rs6277, rs1800497) у подростков с проблемным использованием компьютерных видеоигр

С.Ю. Терещенко 🕕, К.В. Афоничева 🕕, И.В. Марченко 🕕, М.В. Шубина 🕕, М.В. Смольникова 🔟 🜌

Научно-исследовательский институт медицинских проблем Севера – обособленное подразделение Федерального исследовательского центра «Красноярский научный центр Сибирского отделения Российской академии наук», Красноярск, Россия 🖾 smarinv@yandex.ru

Аннотация. Проблемное использование видеоигр как специфическая форма проблемного использования Интернета широко распространено среди подростков и может оказывать негативный эффект на их психическое и соматическое благополучие. Рост зависимости от пользования видеоиграми, как и Интернетом, среди молодого населения делает актуальным изучение факторов подверженности к ним, в том числе генетической составляющей. Существует ряд исследований, посвященных изучению вовлеченности полиморфных вариантов генов системы нейромедиаторов в развитие Интернет-зависимости, результаты которых различаются в разных этнических группах. Ген рецептора дофамина второго типа DRD2 является одним из кандидатных генов подверженности к патологической зависимости от использования видеоигр. Целью работы было исследование полиморфных вариантов гена рецептора дофамина DRD2 (rs6277, rs1800497) у русских подростков с проблемным использованием компьютерных видеоигр. Протестирована выборка из 407 подростков в возрасте 14.1 ± 1.8 года, у 56 (13.8 %) из которых на основании результатов оценки шкалы GASA было выявлено проблемное использование видеоигр. Мальчики в выборке чаще были зависимы от видеоигр, чем девочки (p = 0.041). В результате сравнения частоты аллелей DRD2 rs6277 обнаружена тенденция к большей частоте минорного аллеля Т в группе подростков с проблемным использованием видеоигр по сравнению с подростками без проблемного использования видеоигр (0.563 и 0.466 соответственно, p = 0.06). В доминантной модели наследования у подростков с проблемным использованием видеоигр статистически значимо чаще встречалось носительство аллеля T (CT+TT) (p = 0.04, OR 2.14, CI = 1.01–4.53). Носительство аллеля T DRD2 rs6277 ассоциировано с низкой экспрессией дофаминового рецептора D2 и приводит к снижению плотности и аффинности экстрастриарных дофаминовых рецепторов второго типа, что сопряжено в том числе с нарушением социальной коммуникации. Мы полагаем, что наличие генотипов СТ и ТТ rs6277 гена DRD2 может выступать потенциальным фактором риска развития проблемного использования видеоигр у подростков.

Ключевые слова: полиморфизм генов; дофамин; подростки; проблемное использование компьютерных видеоигр; игровая зависимость; интернет-зависимость.

Для цитирования: Терещенко С.Ю., Афоничева К.В., Марченко И.В., Шубина М.В., Смольникова М.В. Полиморфные варианты гена рецептора дофамина *DRD2* (rs6277, rs1800497) у подростков с проблемным использованием компьютерных видеоигр. *Вавиловский журнал генетики и селекции*. 2024;28(6):667-674. DOI 10.18699/vjgb-24-74

Финансирование. Исследование выполнено в рамках темы государственного задания № 124020100064-6 «Психосоматические расстройства у подростков Центральной Сибири: распространенность, структура, психологические факторы риска и нейрогенетические предикторы».

Polymorphic variants of the dopamine receptor gene *DRD2* (rs6277, rs1800497) in adolescents with problematic video game use

S.Yu. Tereshchenko (D, K.V. Afonicheva (D, I.V. Marchenko (D, M.V. Shubina (D, M.V. Smolnikova (D) 🛛

Scientific Research Institute of Medical Problems of the North – a separate division of the Federal Research Center "Krasnoyarsk Science Center" of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia Smarinv@yandex.ru

Abstract. Problematic video games use, as a specific form of problematic Internet use, is widespread among adolescents and can have negative effects on their mental and somatic well-being. An increasing incidence of addictive video gaming, as well as the overuse of the Internet, among the young population makes the current study of susceptibility factors, including the genetic component, relevant. There has been a number of investigations related to the involvement of gene variants of the neurotransmitter system in the development of Internet addiction, with the results being different for various ethnic groups. The dopamine type 2 receptor gene (*DRD2*) is one of the candidate genes for susceptibility to video game addiction. The aim of the work was to study polymorphic variants of the dopamine receptor gene *DRD2* (rs6277, rs1800497) in Russian adolescents with problematic use of computer video games. A sampling of 407 adolescents aged 14.1 ± 1.8 years was tested, of which 56 (13.8 %) were identified as having problems with the pathological use of video games use based on the GASA scale results. Boys in the sample proved to be addicted to video games more than girls (p = 0.041). As a result of comparing the allele frequency of *DRD2* (rs6277), a tendency to a higher frequency of the minor allele T was revealed in the group of adolescents with problematic video game use compared with adolescents without problematic video game use (i.e. 0.563 and 0.466, respectively, p = 0.06). When using the dominant inheritance model, it was revealed that adolescents with problematic use of video games were statistically significantly more likely to carry the T (CT+TT) allele (p = 0.04, OR = 2.14, CI = 1.01–4.53). The T allele *DRD2* (rs6277) is associated with low expression of the dopamine receptor D2 and leads to decreasing the density and affinity of extrastriatal dopamine type 2 receptors, which is associated with impaired social communication as well. We suggest that the presence of CT and TT genotypes of rs6277 *DRD2* may be a potential risk factor for developing problematic video game use in adolescents.

Key words: gene polymorphism; dopamine; teenagers; problematic video game use; game addiction; Internet addiction.

For citation: Tereshchenko S.Yu., Afonicheva K.V., Marchenko I.V., Shubina M.V., Smolnikova M.V. Polymorphic variants of the dopamine receptor gene *DRD2* (rs6277, rs1800497) in adolescents with problematic video game use. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(6):667-674. DOI 10.18699/vjgb-24-74

Введение

Проблемное использование компьютерных видеоигр среди подростков является актуальной проблемой современного общества и характеризуется чрезмерным увлечением видеоиграми, что приводит к негативным последствиям в различных сферах жизни – социальной, образовательной, соматической и психологической (Griffiths et al., 2012; Paulus et al., 2018; Männikkö et al., 2020).

Поскольку в настоящее время в большинстве случаев использование видеоигр сопряжено с активным использованием Интернета, большая часть экспертов рассматривают проблемное использование компьютерных видеоигр как специфическую форму проблемного использования Интернета, или Интернет-зависимость. В литературе можно встретить несколько синонимичных терминов, по существу описывающих единый психологический конструкт: game addiction (МКБ-11), Internet gaming disorder (DSM-5), gaming disorder, pathological video gaming, excessive video game use, compulsive gaming, problematic digital gaming, problematic online gaming, problematic video game use (PVGU). В научных публикациях эти термины часто применяют как взаимозаменяемые, однако они могут иметь некоторые смысловые нюансы в зависимости от контекста и теоретической основы исследования. Европейская исследовательская группа рекомендует использовать термин "problematic use of the Internet" для генерализованной формы Интернет-зависимости и ее специфических форм, таких, например, как "problematic social media use" и PVGU (Fineberg et al., 2022). Только один из множества специфических видов аддиктивного Интернетповедения, а именно PVGU, сейчас официально рассматривается как ментальное расстройство (Internet gaming disorder, DSM-5; American Psychiatric Association, 2013; gaming disorder, MKE-11, 2019).

Как показано в систематическом обзоре (Mihara, Higuchi, 2017), распространенность PVGU варьирует от 0.7 до 27.5 % и, так же как и генерализованная Интернет-зависимость, сильно зависит от использованных опросников и оценочных критериев. Распространенность PVGU, как и в случае генерализованной Интернет-зависимости, демонстрирует большие значения в азиатских странах с преимущественно монголоидным населением в сравнении с другими регионами (Sussman et al., 2018).

В отличие от других видов аддикций (например, злоупотребления психоактивными веществами или азартными играми) очень небольшое количество исследований было посвящено поиску генетических основ Интернетзависимости. Так, первое близнецовое исследование было проведено в 2014 г.: авторам на основании обследования 825 детей 10-12 лет китайской популяции удалось оценить долю общей изменчивости, обусловленную генетическими эффектами, которая в зависимости от пола варьировала от 58 до 66 % (Li M. et al., 2014). Подобные результаты немного позднее были получены при исследовании турецкой (19-86 %) (Deryakulu, Ursavaş, 2014), нидерландской (48 %) (Vink et al., 2016), австралийской (41 %) (Long et al., 2016) и немецкой (21-44 %) (Hahn et al., 2017) близнецовых когорт. Хотя указанные данные лимитированы объемом выборок и различными этногеографическими условиями, вероятной является тенденция к большему вкладу генетических факторов у лиц мужского пола. Таким образом, наличие генетического компонента формирования Интернет-зависимости было убедительно показано близнецовыми исследованиями на примере различных популяций, однако к настоящему времени конкретные гены, вовлеченные в механизмы такой наследуемости, точно не идентифицированы.

В связи с этим активно изучаются кандидатные гены, полиморфные варианты которых могут нарушать функционирование систем нейромедиаторов и обусловливать психические и поведенческие расстройства. Одним из таких является ген рецептора дофамина второго типа *DRD2* (Kim et al., 2022). Дофамин – гормон, отвечающий за мотивацию, стремление и пристрастия, функционально связан с «центром удовольствия». Дофаминергические нейроны в головном мозге формируют нигростриарный, мезолимбический, мезокортикальный, тубероинфундибулярный пути (Колотилова и др., 2014). Рецептор D2, классифицируемый как тормозящий, в высокой концентрации присутствует в полосатом теле, обонятельном бугорке, миндалевидном теле, прилежащем ядре, гипоталамусе, черной субстанции и вентральной области покрышки (Ford, 2014; Arnsten et al., 2015). Ген дофаминового рецептора человека DRD2 расположен на хромосоме 11 (q22-q23) и является полиморфным, различные генетические варианты изменяют доступность и экспрессию гена рецептора дофамина D2, что влияет на чувствительность и плотность рецепторов (Magistrelli et al., 2021). Полиморфизм rs6277 в экзоне 7 гена DRD2 представляет собой замену аминокислоты серина на цистеин (Ser311Cys). Гомозиготный генотип CC rs6277 DRD2 обусловливает низкую чувствительность нейронов к дофамину в полосатом теле (стриатуме) (Hänninen et al., 2006). В то же время вне полосатого тела (экстрастриатумной области) данный генотип определяет высокую аффинность рецепторов D2 к дофамину (Liu et al., 2014; Smith et al., 2017; Della Torre et al., 2018). Потенциал связывания дофамина рецепторами D2 в стриатуме выше у носителей TT rs6277 DRD2, в экстрастриатуме наблюдается обратный эффект (Hänninen et al., 2006). Известно, что снижение плотности DRD2 в области стриатума и влияние факторов окружающей среды приводят к формированию зависимостей, например, от алкоголя, наркотиков, компьютерных игр (Hill et al., 2008; Bhaskar, Kumar, 2014; Gao et al., 2017; Анохин et al., 2019; Picci et al., 2022). Однако наблюдается неоднозначность опубликованных данных в том, какой аллель (С или Т) rs6277 DRD2 ассоциирован с зависимостью от психоактивных веществ (Hill et al., 2013). В ряде исследований указывается, что аллель T rs6277 DRD2 ассоциирован с повышенной склонностью к патологической зависимости от видеоигр (Kim et al., 2022).

Надо отметить, что генетические факторы представляют лишь один из аспектов склонности к зависимому поведению; важную роль играет также влияние окружающей среды и социокультурных факторов. Так, известно, что стрессовая среда в сочетании с аллелем T rs6277 DRD2 способствует снижению подавления тяги к компьютерным играм (Kim et al., 2022). Выявлено, что лица с гомозиготным генотипом TT rs6277 DRD2 воспринимают никотиновую заместительную терапию лучше относительно носителей аллеля С (Hill et al., 2008). Аллельный вариант С rs6277 DRD2 обусловливает гиподофаминергическое состояние, проявляющееся в сниженной способности к подавлению реакции на стимулы, связанные с вознаграждением (Machulska et al., 2016; Richter et al., 2017; Rył et al., 2024). Показано, что носители гомозиготного генотипа СС rs6277 DRD2, которые в детском возрасте подвергались жестокому обращению или испытывали травмирующие жизненные события, во взрослом возрасте обладают высокой степенью импульсивности и отличаются более частым употреблением алкоголя (Klaus et al., 2021). Риск развития такой зависимости выше у взрослых носителей аллеля C rs6277 DRD2, тогда как у подростков 11-13 лет данный аллельный вариант может быть протективным в отношении формирования зависимости от психоактивных веществ, а также предрасполагать к более позднему началу употребления алкоголя (Picci et al., 2022).

Полиморфизм rs1800497 гена *DRD2* вызывает аминокислотную замену глицина на лизин (Glu713Lys), что приводит к изменению специфичности связывания рецептора с дофамином. По некоторым данным этот полиморфизм называется *DRD2/ANKK1 Taq1A*, поскольку расположен

в пределах гена протеинкиназы PKK2 (Ankyrin Repeat And Kinase Domain Containing 1, ANKK1), белка системы передачи пострецепторных внутриклеточных сигналов (Гафаров и др., 2019). Полиморфизм rs1800497 DRD2 часто исследуют также в контексте нервно-психических расстройств и зависимостей (Volkow et al., 1996; Pohjalainen et al., 1998). Установлено, что у носителей А1-аллеля (Т) наблюдается снижение плотности дофаминовых рецепторов D2 в стриатуме головного мозга на 30 %, что приводит к снижению внимания и обучаемости, повышению тревожности, наблюдается ассоциация с синдромом «недостатка вознаграждения» и «поиска новизны» (Klein et al., 2007; Кушнарев, 2022). В работе (Pohjalainen et al., 1998) аналогично показано, что наличие минорного аллеля T rs1800497 связано с уменьшенным количеством сайтов связывания дофамина в мозге. Существует связь генотипов A1/A1 (TT) и A1/A2 (TC) rs1800497 гена DRD2 с синдромом дефицита вознаграждения (Klein et al., 2007). Синдром дефицита вознаграждения вызывает различные психические и поведенческие расстройства: никотиновую и наркотическую зависимость, игровую зависимость, СДВГ, расстройства аутистического спектра, расстройства пищевого поведения с компульсивным перееданием (Pohjalainen et al., 1998). Выявлено, что мужчины-носители аллельного варианта Т rs1800497 чаще страдают зависимостью от онлайн-игр (Paik et al., 2017). Этот аллельный вариант также чаще встречается у людей с пристрастием к видеоиграм для удовлетворения стремления к вознаграждению (Werling, Grünblatt, 2022). Таким образом, люди с пониженным количеством дофаминовых рецепторов D2 склонны к поиску экстремальных способов получения удовольствия от жизни. Нарушение чувствительности дофаминовых рецепторов обусловливает снижение способности людей делать правильные выводы из отрицательного опыта, поскольку дофамин участвует в процессах обучения и обеспечивает возможность эффективно учиться на своих ошибках.

Целью данной работы было исследование полиморфных вариантов гена рецептора дофамина *DRD2* (rs6277, rs1800497) у подростков с проблемным использованием компьютерных видеоигр для выявления возможной связи между генетическими вариантами и поведенческими аспектами игровой зависимости.

Материал и методы

Проведено психологическое и генетическое тестирование 407 подростков в возрасте 12–18 лет. Национальность всех включенных в исследование подростков – русские (верифицировано по национальности матери и отца одновременно). От подростков или их родителей (законных представителей) получены информированные согласия. После получения информированного согласия ученики были уведомлены о добровольности и конфиденциальности исследования. Участникам было предложено заполнить анкету, которая включала демографические данные (пол, возраст, национальность матери и отца), и переводную версию опросника для оценки игровой зависимости "Game Addiction Scale for Adolescents" (GASA) (Lemmens et al., 2009). Опросник GASA состоит из семи вопросов, касающихся нарушений поведения у подростка, вызванных чрезмерным увлечением Интернет-играми. Каждый из вопросов оценивается по 5-балльной шкале: «никогда» (0 баллов), «редко» (1 балл), «иногда» (2 балла), «часто» (3 балла), «очень часто» (4 балла). Согласно предложенным авторами опросника критериям (Lemmens et al., 2009) определяли наличие PVGU (если на любые четыре или более из семи вопросов подросток ответил «иногда», «часто» или «очень часто»).

После заполнения опросника подросткам предлагалось сдать образцы слюны в специальные контейнеры. Образцы слюны были забраны с использованием «Устройств для сбора и сохранения ДНК слюны» (Кат. № RU 49080, Norgen Biotek Corp., Канада). ДНК выделяли из образцов слюны с помощью набора DIAtom DNA Prep (Isogene Lab, Россия). Генотипирование полиморфных вариантов rs6277 и rs1800497 *DRD2* проводили по технологии TaqMan с использованием зондов и праймеров («ДНК-Синтез», Россия) и реакционной смеси («Синтол», Россия) на приборе Rotor-Gene 6000 (Qiagen, Германия). Исследование одобрено этическим комитетом ФИЦ «Красноярский научный центр Сибирского отделения Российской академии наук» (протокол № 12 от 18.12.2018).

Статистический анализ выполнен с использованием программного обеспечения Statistica v.10 (StatSoft Inc., США). Различия категориальных данных оценивали с помощью критерия χ^2 Пирсона с поправкой Йейтса, различия количественных данных оценивали с помощью *t*-критерия Стьюдента.

Результаты

Описательная статистика основных переменных представлена в табл. 1. Средний возраст 407 протестированных подростков составил 14.1±1.8 года, соотношение мальчики/девочки – 174 (42.8%)/233 (57.2%). У 56 подростков

(13.8%) выявлено PVGU на основании результатов оценки шкалы GASA (см. табл. 1). У мальчиков средние показатели балла по шкале игровой зависимости значимо выше, чем у девочек. Помимо этого, среди мальчиков выявлено больше подростков с PVGU по сравнению с девочками.

Частота распределения генотипов полиморфных вариантов гена *DRD2* rs6277 и rs1800497 у обследованных в работе подростков соответствует их распределению в европеоидных популяциях (согласно данным ресурса ensembl.org). Распределение частот генотипов соответствовало равновесию Харди–Вайнберга как для случаев PVGU, так и для группы без PVGU. Таким образом, частоты аллелей выбранных полиморфных вариантов в изученной популяции сбалансированы и применимы для ассоциативных исследований.

Распределение частот генотипов и аллелей полиморфных участков rs6277 и rs1800497 гена *DRD2* в зависимости от наличия и отсутствия PVGU представлены в табл. 2 и 3. Для полиморфных вариантов *DRD2* rs6277 частота генотипов существенно не отличалась между группой с PVGU и ее отсутствием (p = 0.12) (см. табл. 2). В то же время при сравнении частот аллелей *DRD2* rs6277 нами выявлена отчетливая тенденция к большей частоте минорного аллеля T в группе подростков с PVGU в сравнении с группой без PVGU (p = 0.06). Анализ полиморфных вариантов rs1800497 гена *DRD2* не показал значимых различий в частотах генотипов и аллелей между группами с отсутствием и наличием PVGU (см. табл. 3).

Далее нами был проведен анализ распределения частот генотипов полиморфного участка rs6277 гена *DRD2* с использованием доминантной модели наследования, где гетерозиготы и гомозиготы по минорному аллелю rs6277 *DRD2* (СТ и ТТ соответственно) были объединены (табл. 4).

Параметр	Общее количество	Мальчики	Девочки	<i>р</i> (мальчики–девочки)
Возраст 12–14 лет	241	95 (39.4 %)	146 (60.6 %)	-
Возраст 15–18 лет	166	79 (47.6 %)	87 (52.4 %)	-
Общее количество	407	174 (42.8 %)	233 (57.2 %)	-
	GASA результат	(<i>n</i> = 407)		
Шкала игровой зависимости для подростков (GASA), балл	10.8 ± 6.8	12.2 ± 6.7	9.8 ± 6.9	0.0005 <i>t</i> = 3.5
Проблемное использование видеоигр (PVGU)	56 (13.8 %)	31 (17.8 %)	25 (10.7 %)	p = 0.041 $\chi^2 = 4.21$, df = 1

Примечание. Данные представлены в виде n (%) и среднего ± стандартное отклонение.

Таблица 2. Распределение	астот генотипов и аллелей rs6	277 гена DRD2 у подростков	с наличием и отсутствием PVGU

Генотипы и аллели rs6277	Отсутствие PVGU	Наличие PVGU	χ ²	р	OR	95 % CI
	n = 351	n = 56				
Генотип СС	0.291 (102)	0.161 (9)	4.26	0.12	0.47	0.22–0.99
Генотип СТ	0.487 (171)	0.554 (31)			1.31	0.74–2.30
Генотип TT	0.222 (78)	0.285 (16)			1.40	0.74–2.30
Аллель С	0.534	0.437	3.62	0.06	0.68	0.45–1.01
Аллель Т	0.466	0.563			1.47	0.99–2.20

Генотипы и аллели rs1800497	Отсутствие PVGU	Наличие PVGU	χ ²	p	OR	95 % CI
	n = 351	n = 56	•			
Генотип СС	0.638 (224)	0.714 (40)	1.24	0.54	1.42	0.76–2.63
Генотип СТ	0.342 (120)	0.268 (15)			0.70	0.37–1.32
Генотип TT	0.020 (7)	0.018 (1)	-		0.89	0.11–7.40
Аллель С	0.809	0.848	0.98	0.32	1.32	0.76–2.28
Аллель Т	0.191	0.152			0.76	0.44–1.31

Таблица 3. Распределение частот генотипов rs1800497 гена DRD2 у подростков с наличием и отсутствием PVGU

Таблица 4. Распределение частот генотипов полиморфного участка rs6277 гена *DRD2* у подростков с наличием и отсутствием PVGU

Генотип	Отсутствие PVGU	Наличие PVGU	χ²	p	OR	95 % Cl
	n = 351	n = 56				
CC	0.291	0.161	4.11	0.04	0.47	0.22–0.99
CT+TT	0.709	0.839			2.14	1.01–4.53

Согласно полученным результатам, в группе подростков с PVGU статистически значимо чаще встречалось носительство аллеля T (генотип CT+TT) в сравнении с подростками без PVGU. Расчет показателя отношения шансов показал значимую ассоциацию носительства аллеля T и наличия у подростков PVGU.

Обсуждение

Общая частота PVGU в обследованной выборке русских подростков составила 13.8 %, что существенно не отличается от более ранних наших данных распространенности зависимости от компьютерных игр, полученных в результате крупномасштабного эпидемиологического проекта (n = 4514, распространенность PVGU – 10.4 %) (Tereshchenko et al., 2022). Нами показано, что мальчики чаще зависимы от видеоигр, чем девочки (p = 0.041), что соответствует данным упомянутого проекта и результатам других эпидемиологических работ с использованием опросника GASA (Mihara, Higuchi, 2017; Tereshchenko et al., 2022). Распределение генотипов и аллелей в исследуемой выборке схоже с их частотой в глобальной европеоидной популяции в соответствии с базами данных "1000 Genomes Project" и НарМар (pecypc ensemble.org), как для rs6277, так для rs1800497. Таким образом, с точки зрения распространенности основных переменных изученная нами популяция является достаточно типичной, и данные анализа могут быть успешно экстраполированы на другие европеоидные подростковые популяции.

Нами установлено, что носители генотипов СТ и ТТ полиморфного участка rs6277 гена *DRD2*, т. е. носители аллеля Т, согласно результатам, полученным при использовании доминантной модели наследования, значимо чаще проявляют признаки PVGU относительно подростков с генотипом CC.

Известно, что носители аллеля T rs6277 гена *DRD2* имеют более низкую, в сравнении с носителями аллеля C, плотность и аффинность дофаминовых рецепторов второго типа во всех областях головного мозга (включая префронтальную корковую область), за исключением по-

лосатого тела (стриатума) – C/C > C/T > T/T (Hirvonen et al., 2009; Smith et al., 2017). Низкая плотность DRD2 в экстрастриарных областях головного мозга может приводить к определенным психофизиологическим последствиям. В частности, в статье, посвященной роли экстрастриарных DRD2, рассматриваются функциональные последствия наличия этих рецепторов в экстрастриарных областях, включая кору и таламус (Takahashi et al., 2006). Обзор включает посмертные данные, а также исследования in vivo на людях и животных, фокусируясь на роли низкой функциональной активности экстрастриарных DRD2 при шизофрении (Takahashi et al., 2006). С. Мигауата с коллегами в своей работе указывают на то, что низкая доступность рецепторов D2/3 в экстрастриарных областях у взрослых мужчин с социокоммуникативным дефицитом при аутизме была связана с уменьшением плотности дофаминовых рецепторов (Murayama et al., 2022).

Носители аллеля T rs6277 гена *DRD2* менее эффективно подавляют импульсивные тенденции к нежелательным действиям по сравнению с носителями аллеля C (Colzato et al., 2010). Установлено, что носители аллеля T rs6277 гена *DRD2* (6–18 лет) характеризуются трудностями в контроле импульсов, самоконтроле эмоций и волевой корректировке поведения (Della Torre et al., 2018). В качестве теоретической модели, подтверждающей генетические данные, авторы этого исследования приводят мнение G.S. Dichter с коллегами (2012) о том, что снижение активности дофаминергической системы ассоциировано с проблемами в обучении и недостаточным самоконтролем поведения.

Наши данные об ассоциативной связи носительства аллеля T rs6277 гена DRD2 с PVGU у подростков корреспондируют с данными работы (Kim et al., 2022), в которой с помощью регрессионного анализа установлена прямая связь выраженности PVGU с носительством аллеля T у студентов колледжа (b = 19.58, p = 0.04). Два других исследования, проведенных на выборках взрослых лиц, не подтверждают подобной ассоциативной связи (Paik et al., 2017; Rył et al., 2024). Противоречивость полученных результатов может быть объяснена различием возрастного, полового, этнического и количественного состава выборок. В частности, влияние генетического компонента аддикций может по-разному проявляться у подростков и взрослых лиц. Подростковый период в развитии головного мозга характеризуется различными по времени траекториями формирования лимбической системы и префронтальных корковых отделов (Casey et al., 2008). Затянувшееся развитие префронтальной коры по сравнению с лимбической системой в течение подросткового периода приводит к ослабленному торможению со стороны корковых отделов в отношении нижележащих подкорковых структур и повышенной импульсивности, что способствует высокому риску формирования аддиктивного поведения (He, Crews, 2007).

Мы считаем, что установленная нами и Е. Кіт с коллегами (2022) ассоциативная связь носительства аллеля Т rs6277 гена *DRD2* с PVGU у подростков и студентов имеет определенное теоретическое и эмпирическое обоснование. Носительство аллеля Т rs6277 приводит к снижению плотности и аффинности экстрастриарных дофаминовых рецепторов второго типа (Hirvonen et al., 2009; Smith et al., 2017) и своеобразному феномену «дофаминовой десентизации», что сопряжено со снижением чувствительности к вознаграждению, повышенной импульсивностью, недостаточным самоконтролем поведения (Colzato et al., 2010; Della Torre et al., 2018; Weinstein, Lejoyeux, 2020; Kim et al., 2022), а также с возможным нарушением социальной коммуникации (Takahashi et al., 2006; Murayama et al., 2022).

Гипореактивность орбитофронтальной коры и снижение дофаминергической функции в этой области головного мозга связаны с гипочувствительностью системы вознаграждения, способствующей трансгрессивному поведению, делинквентности и злоупотреблению психоактивными веществами (Matthys et al., 2013). Исследования показывают, что некоторые варианты гена DRD2 могут способствовать развитию гиподофаминергического состояния, при котором частичная доступность дофаминовых рецепторов определяет снижение чувствительности к вознаграждению (Alcaro et al., 2021). Последнее может привести к тому, что подросток будет стремиться к дополнительной стимуляции дофаминергической системы, которая может проявляться в виде аддиктивного поведения, включающего компонент активного персистирующего вознаграждения, например компульсивного использования компьютерных видеоигр (Weinstein, Lejoyeux, 2020; Kim et al., 2022) или азартных игр.

В свою очередь, повышенная импульсивность и нарушение социальных связей являются важнейшими предикторами формирования генерализованной Интернетзависимости и ее специфической формы – зависимости от компьютерных видеоигр (PVGU). Импульсивность и самоконтроль связаны с широким спектром особенностей поведения. Эмпирические исследования показывают, что люди с высоким самоконтролем лучше контролируют свои мысли, регулируют свои эмоции и подавляют свои импульсы, чем люди с низким самоконтроля и высокая импульсы, чем люди с низким самоконтроля и высокая импульсивность ассоциируются с делинквентностью, преступностью, антисоциальным поведением, экстернализирующим поведением, виктимизацией и аддиктивными расстройствами. Известно, что одним из наиболее часто встречающихся коморбидных с Интернет-зависимостью психиатрических расстройств является синдром дефицита внимания с гиперактивностью, для которого характерна высокая импульсивность в действиях и поступках (Wang et al., 2017). Во многих психологических исследованиях указывается, что Интернет-зависимое поведение тесно ассоциировано с низким уровнем самоконтроля/высоким уровнем импульсивности (Li W. et al., 2016; Li S. et al., 2021; Yu et al., 2021). Метаанализ 40 нейрофизиологических исследований проблемного использования Интернета показал, что независимо от контента Интернет-зависимое поведение характеризуется существенным нарушением процессов ингибиторного контроля, принятия решений и рабочей памяти (Ioannidis et al., 2019). Проведенный в работе (Zhang et al., 2021) метаанализ выявил наличие общего паттерна структурных изменений головного мозга при химических и поведенческих зависимостях - изменения в префронтальной и инсулярной областях, которые связаны с повышенной импульсивностью.

Редкий аллельный вариант T rs1800497 гена DRD2 тоже ассоциирован с низкой экспрессией дофаминового рецептора D2 в префронтальной коре и, согласно (Paik et al., 2017), чаще встречается у корейских мужчин (19-47 лет) с зависимостью от Интернет-игр. Этот вариант также чаще встречается у корейских молодых людей (учеников старшей школы и студентов) с PVGU и высокой зависимостью от вознаграждений (Han et al., 2007). Однако результаты нашего исследования не показали статистически значимых отличий между разными генотипами и аллелями полиморфизма rs1800497 гена DRD2 в группах с признаками проблемного использования видеоигр и без них. Противоречия в результатах анализа могут быть обусловлены этническими и половыми особенностями выборок, а также применением разных психометрических инструментов для верификации PVGU. В частности, существенное влияние может оказывать факт выраженных этнических различий в частотах генотипов и аллелей rs1800497 гена DRD2 у представителей европеоидных и монголоидных популяций.

Заключение

Результаты настоящего исследования полиморфных вариантов гена рецептора дофамина *DRD2* у подростков с проблемным использованием компьютерных видеоигр позволяют сделать вывод о важности генетических факторов в развитии данного поведенческого расстройства. Наличие генотипов СТ и ТТ полиморфного локуса rs6277 гена *DRD2* может выступать потенциальным предиктором риска развития проблемного использования видеоигр у подростков. Изучение генетических основ поведенческих расстройств в дальнейшем даст возможность индивидуально подходить к профилактике и лечению игровой зависимости с учетом генетического профиля пациента.

Список литературы / References

Анохин П.К., Веретинская А.Г., Давыдова Т.В., Шамакина И.Ю. Агонисты дофаминовых D2 рецепторов в фармакотерапии экспериментального алкоголизма. *Патол. физиология и эксперим. терапия.* 2019;63(1):33-39. DOI 10.25557/0031-2991.2019.01. 33-39 [Anokhin P.K., Veretinskaya A.G., Davidova T.V., Shamakina I.Yu. Dopamine D2 agonists in the treatment of experimental alcoholism. *Patologicheskaya Fiziologiya i Eksperimental 'naya Terapiya. = Pathological Physiology and Experimental Therapy.* 2019;63(1):33-39. DOI 10.25557/0031-2991.2019.01.33-39 (in Russian)]

- Гафаров В.В., Громова Е.А., Панов Д.О., Максимов В.Н., Гагулин И.В., Гафарова А.В. Ассоциация полиморфизма гена *DRD2/ ANKK1 Taq1A* с депрессией в открытой популяции мужчин 45–64 лет (международные эпидемиологические программы НАРІЕЕ и ВОЗ MONICA). *Неврология, нейропсихиатрия, психосоматика.* 2019;11(2):37-41. DOI 10.14412/2074-2711-2019-2-37-41
 - [Gafarov V.V., Gromova E.A., Panov D.O., Maximov V.N., Gagulin I.V., Gafarova A.V. Association of *DRD2/ANKK1 Taq1A* polymorphism with depression in an open 45–64 year-old male population (international epidemiological HAPIEE and WHO MONICA programs). *Nevrologiya, Neiropsikhiatriya, Psikhosomatika = Neurology, Neuropsychiatry, Psychosomatics.* 2019;11(2):37-41. DOI 10.14412/2074-2711-2019-2-37-41]
- Колотилова О.И., Коренюк И.И., Хусаинов Д.Р., Черетаев И.В. Дофаминергическая система мозга. Вестн. Брян. гос. ун-та. 2014; 4:97-106

[Kolotilova O.I., Koreniuk I.I., Khusainov D.R., Cheretaev I.V. Dopaminergic brain system. *Vestnik Bryanskogo Gosudarstvennogo Universiteta = The Bryansk State University Herald*. 2014;4:97-106 (in Russian)]

Кушнарев А.П. Изучение полиморфизма генов рецептора дофамина 4-го типа (DRD4) и дофаминового транспортера (DAT) у лиц с антисоциальным поведением и представителей экстремальных профессий. Медицина. Социология. Философия. Прикл. исследования. 2022;3:40-47

[Kushnarev A.P. The study of polymorphism of the dopamine receptor type 4 (DRD4) and dopamine transporter (DAT) genes in individuals with antisocial behavior and representatives of extreme professions. *Meditsina. Sotsiologiya. Filosofiya. Prikladnyye Issledovaniya = Medicine. Sociology. Philosophy. Applied Research.* 2022;3:40-47 (in Russian)]

- Alcaro A., Brennan A., Conversi D. The SEEKING drive and its fixation: a neuro-psycho-evolutionary approach to the pathology of addiction. *Front. Hum. Neurosci.* 2021;15:635932. DOI 10.3389/ fnhum.2021.635932
- Arnsten A.F.T., Wang M., Paspalas C.D. Dopamine's actions in primate prefrontal cortex: challenges for treating cognitive disorders. *Pharmacol. Rev.* 2015;67(3):681-696. DOI 10.1124/pr.115.010512
- Bhaskar L.V.K.S., Kumar S.A. Polymorphisms in genes encoding dopamine signalling pathway and risk of alcohol dependence: a systematic review. Acta Neuropsychiatrica. 2014;26(2):69-80. DOI 10.1017/neu.2013.27
- Casey B.J., Jones R.M., Hare T.A. The adolescent brain. *Ann. N.Y. Acad. Sci.* 2008;1124:111-126. DOI 10.1196/annals.1440.010
- Colzato L.S., van den Wildenberg W.P.M., Van der Does A.J.W., Hommel B. Genetic markers of striatal dopamine predict individual differences in dysfunctional, but not functional impulsivity. *Neuroscience*. 2010;170(3):782-788. DOI 10.1016/j.neuroscience.2010. 07.050
- de Ridder D.T.D., Lensvelt-Mulders G., Finkenauer C., Stok F.M., Baumeister R.F. Taking stock of self-control: a meta-analysis of how trait self-control relates to a wide range of behaviors. *Pers. Soc. Psychol. Rev.* 2012;16(1):76-99. DOI 10.1177/1088868311418749
- Della Torre O.H., Paes L.A., Henriques T.B., de Mello M.P., Celeri E.H.R.V., Dalgalarrondo P., Guerra-Júnior G., Santos-Júnior A.D. Dopamine D2 receptor gene polymorphisms and externalizing behaviors in children and adolescents. *BMC Med. Genet.* 2018;19(1): 65. DOI 10.1186/s12881-018-0586-9
- Deryakulu D., Ursavaş Ö.F. Genetic and environmental influences on problematic Internet use: a twin study. *Comput. Hum. Behav.* 2014; 39:331-338. DOI 10.1016/j.chb.2014.07.038

- Dichter G.S., Damiano C.A., Allen J.A. Reward circuitry dysfunction in psychiatric and neurodevelopmental disorders and genetic syndromes: animal models and clinical findings. J. Neurodev. Disord. 2012;4(1):19. DOI 10.1186/1866-1955-4-19
- Fineberg N.A., Menchón J.M., Hall N., Dell'Osso B., Brand M., Potenza M.N., Chamberlain S.R., Cirnigliaro G., Lochner C., Billieux J., ... Cataldo I., Riva G.M., Yücel M., Flayelle M., Hall T., Griffiths M., Zohar J. Advances in problematic usage of the internet research – a narrative review by experts from the European network for problematic usage of the internet. *Compr. Psychiatry*. 2022;118: 152346. DOI 10.1016/j.comppsych.2022.152346
- Ford C.P. The role of D2-autoreceptors in regulating dopamine neuron activity and transmission. *Neuroscience*. 2014;282:13-22. DOI 10.1016/j.neuroscience.2014.01.025
- Gao X., Wang Y., Lang M., Yuan L., Reece A.S., Wang W. Contribution of genetic polymorphisms and haplotypes in *DRD2*, *BDNF*, and opioid receptors to heroin dependence and endophenotypes among the Han Chinese. *OMICS*. 2017;21(7):404-412. DOI 10.1089/omi. 2017.0057
- Griffiths M.D., Kuss D.J., King D.L. Video game addiction: past, present and future. *Curr. Psychiatry Rev.* 2012;8(4):308-318. DOI 10.2174/157340012803520414
- Hahn E., Reuter M., Spinath F.M., Montag C. Internet addiction and its facets: the role of genetics and the relation to self-directedness. *Addict. Behav.* 2017;65:137-146. DOI 10.1016/j.addbeh.2016. 10.018
- Han D.H., Lee Y.S., Yang K.C., Kim E.Y., Lyoo I.K., Renshaw P.F. Dopamine genes and reward dependence in adolescents with excessive internet video game play. J. Addict. Med. 2007;1(3):133-138. DOI 10.1097/ADM.0b013e31811f465f
- Hänninen K., Katila H., Kampman O., Anttila S., Illi A., Rontu R., Mattila K.M., Hietala J., Hurme M., Leinonen E., Lehtimäki T. Association between the C957T polymorphism of the dopamine D2 receptor gene and schizophrenia. *Neurosci. Lett.* 2006;407(3):195-198. DOI 10.1016/j.neulet.2006.08.041
- He J., Crews F.T. Neurogenesis decreases during brain maturation from adolescence to adulthood. *Pharmacol. Biochem. Behav.* 2007;86(2): 327-333. DOI 10.1016/j.pbb.2006.11.003
- Hill S.Y., Hoffman E.K., Zezza N., Thalamuthu A., Weeks D.E., Matthews A.G., Mukhopadhyay I. Dopaminergic mutations: within-family association and linkage in multiplex alcohol dependence families. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 2008;147B(4): 517-526. DOI 10.1002/ajmg.b.30630
- Hill S.Y., Lichenstein S., Wang S., Carter H., McDermott M. Caudate volume in offspring at ultra high risk for alcohol dependence: COMT Val158Met, DRD2, externalizing disorders, and working memory. *Adv. J. Mol. Imaging.* 2013;3(4):43-54. DOI 10.4236/ami. 2013.34007
- Hirvonen M.M., Lumme V., Hirvonen J., Pesonen U., Någren K., Vahlberg T., Scheinin H., Hietala J. C957T polymorphism of the human dopamine D2 receptor gene predicts extrastriatal dopamine receptor availability *in vivo. Prog. Neuropsychopharmacol. Biol. Psychiatry.* 2009;33(4):630-636. DOI 10.1016/j.pnpbp.2009.02.021
- Ioannidis K., Hook R., Goudriaan A.E., Vlies S., Fineberg N.A., Grant J.E., Chamberlain S.R. Cognitive deficits in problematic internet use: meta-analysis of 40 studies. *Br. J. Psychiatry*. 2019;215(5): 639-646. DOI 10.1192/bjp.2019.3
- Kim E., Lee D., Do K., Kim J. Interaction effects of DRD2 genetic polymorphism and interpersonal stress on problematic gaming in college students. *Genes (Basel)*. 2022;13(3):449. DOI 10.3390/ genes13030449
- Klaus K., Vaht M., Pennington K., Harro J. Interactive effects of *DRD2* rs6277 polymorphism, environment and sex on impulsivity in a population-representative study. *Behav. Brain Res.* 2021;403:113131. DOI 10.1016/j.bbr.2021.113131
- Klein T.A., Neumann J., Reuter M., Hennig J., von Cramon D.Y., Ullsperger M. Genetically determined differences in learning from er-

rors. *Science*. 2007;318(5856):1642-1645. DOI 10.1126/science. 1145044

- Lemmens J.S., Valkenburg P.M., Peter J. Development and validation of a game addiction scale for adolescents. *Media Psychology*. 2009;12(1):77-95. DOI 10.1080/15213260802669458
- Li M., Chen J., Li N., Li X. A twin study of problematic internet use: its heritability and genetic association with effortful control. *Twin Res. Hum. Genet.* 2014;17(4):279-287. DOI 10.1017/thg.2014.32
- Li S., Ren P., Chiu M.M., Wang C., Lei H. The relationship between self-control and internet addiction among students: a meta-analysis. *Front. Psychol.* 2021;12:735755. DOI 10.3389/fpsyg.2021.735755
- Li W., Zhang W., Xiao L., Nie J. The association of Internet addiction symptoms with impulsiveness, loneliness, novelty seeking and behavioral inhibition system among adults with attention-deficit/ hyperactivity disorder (ADHD). *Psychiatry Res.* 2016;243:357-364. DOI 10.1016/j.psychres.2016.02.020
- Liu L., Fan D., Ding N., Hu Y., Cai G., Wang L., Xin L., Xia Q., Li X., Xu S., Xu J., Yang X., Zou Y., Pan F. The relationship between DRD2 gene polymorphisms (C957T and C939T) and schizophrenia: a meta-analysis. *Neurosci. Lett.* 2014;583:43-48. DOI 10.1016/ j.neulet.2014.09.024
- Long E.C., Verhulst B., Neale M.C., Lind P.A., Hickie I.B., Martin N.G., Gillespie N.A. The genetic and environmental contributions to Internet use and associations with psychopathology: a twin study. *Twin Res. Hum. Genet.* 2016;19(1):1-9. DOI 10.1017/thg.2015.91
- Machulska A., Zlomuzica A., Rinck M., Assion H.-J., Margraf J. Approach bias modification in inpatient psychiatric smokers. J. Psychiatric Res. 2016;76:44-51. DOI 10.1016/j.jpsychires.2015.11.015
- Magistrelli L., Ferrari M., Furgiuele A., Milner A.V., Contaldi E., Comi C., Cosentino M., Marino F. Polymorphisms of dopamine receptor genes and Parkinson's disease: clinical relevance and future perspectives. *Int. J. Mol. Sci.* 2021;22(7):3781. DOI 10.3390/ijms 22073781
- Männikkö N., Ruotsalainen H., Miettunen J., Pontes H.M., Kääriäinen M. Problematic gaming behaviour and health-related outcomes: a systematic review and meta-analysis. J. Health. Psychol. 2020; 25(1):67-81. DOI 10.1177/1359105317740414
- Matthys W., Vanderschuren L.J.M.J., Schutter D.J.L.G. The neurobiology of oppositional defiant disorder and conduct disorder: altered functioning in three mental domains. *Dev. Psychopathol.* 2013; 25(1):193-207. DOI 10.1017/S0954579412000272
- Mihara S., Higuchi S. Cross-sectional and longitudinal epidemiological studies of Internet gaming disorder: a systematic review of the literature. *Psychiatry Clin. Neurosci.* 2017;71(7):425-444. DOI 10.1111/ pcn.12532
- Murayama C., Iwabuchi T., Kato Y., Yokokura M., Harada T., Goto T., Tamayama T., Kameno Y., Wakuda T., Kuwabara H., Senju A., Nishizawa S., Ouchi Y., Yamasue H. Extrastriatal dopamine D2/3 receptor binding, functional connectivity, and autism socio-communicational deficits: a PET and fMRI study. *Mol. Psychiatry*. 2022; 27(4):2106-2113. DOI 10.1038/s41380-022-01464-3
- Paik S.-H., Choi M.R., Kwak S.M., Bang S.H., Chun J.-W., Kim J.-Y., Choi J., Cho H., Jeong J.-E., Kim D.-J. An association study of *Taq1A ANKK1* and *C957T* and *-141C DRD2* polymorphisms in adults with internet gaming disorder: a pilot study. *Ann. Gen. Psychiatry*. 2017;16:45. DOI 10.1186/s12991-017-0168-9
- Paulus F.W., Ohmann S., von Gontard A., Popow C. Internet gaming disorder in children and adolescents: a systematic review. *Dev. Med. Child Neurol.* 2018;60(7):645-659. DOI 10.1111/dmcn.13754
- Picci G., Fishbein D.H., VanMeter J.W., Rose E.J. Effects of OPRM1 and DRD2 on brain structure in drug-naïve adolescents: genetic and neural vulnerabilities to substance use. *Psychopharmacology* (*Berl.*). 2022;239(1):141-152. DOI 10.1007/s00213-021-06030-3

- Pohjalainen T., Rinne J.O., Någren K., Lehikoinen P., Anttila K., Syvälahti E.K., Hietala J. The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Mol. Psychiatry*. 1998;3(3):256-260. DOI 10.1038/sj.mp. 4000350
- Richter A., Barman A., Wüstenberg T., Soch J., Schanze D., Deibele A., Behnisch G., Assmann A., Klein M., Zenker M., Seidenbecher C., Schott B.H. Behavioral and neural manifestations of reward memory in carriers of low-expressing versus high-expressing genetic variants of the dopamine D2 receptor. *Front. Psychol.* 2017;8:654. DOI 10.3389/fpsyg.2017.00654
- Rył A., Tomska N., Jakubowska A., Ogrodniczak A., Palma J., Rotter I. Genetic aspects of problematic and risky Internet use in young men – analysis of ANKK1, DRD2 and NTRK3 gene polymorphism. *Genes (Basel)*. 2024;15(2):169. DOI 10.3390/genes15020169
- Smith C.T., Dang L.C., Buckholtz J.W., Tetreault A.M., Cowan R.L., Kessler R.M., Zald D.H. The impact of common dopamine D2 receptor gene polymorphisms on D2/3 receptor availability: C957T as a key determinant in putamen and ventral striatum. *Transl. Psychiatry*. 2017;7(4):e1091-e1091. DOI 10.1038/tp.2017.45
- Sussman C.J., Harper J.M., Stahl J.L., Weigle P. Internet and video game addictions: diagnosis, epidemiology, and neurobiology. *Child Adolesc. Psychiatr. Clin. N. Am.* 2018;27(2):307-326. DOI 10.1016/ j.chc.2017.11.015
- Takahashi H., Higuchi M., Suhara T. The role of extrastriatal dopamine D2 receptors in schizophrenia. *Biol. Psychiatry*. 2006;59(10):919-928. DOI 10.1016/j.biopsych.2006.01.022
- Tereshchenko S., Kasparov E., Semenova N., Shubina M., Gorbacheva N., Novitckii I., Moskalenko O., Lapteva L. Generalized and specific problematic internet use in Central Siberia adolescents: a school-based study of prevalence, age-sex depending content structure, and comorbidity with psychosocial problems. *Int. J. Environ. Res. Public Health.* 2022;19(13):7593. DOI 10.3390/ijerph 19137593
- Vink J.M., van Beijsterveldt T.C.E.M., Huppertz C., Bartels M., Boomsma D.I. Heritability of compulsive Internet use in adolescents. *Addict. Biol.* 2016;21(2):460-468. DOI 10.1111/adb.12218
- Volkow N.D., Wang G.J., Fowler J.S., Logan J., Hitzemann R., Ding Y.S., Pappas N., Shea C., Piscani K. Decreases in dopamine receptors but not in dopamine transporters in alcoholics. *Alcohol Clin. Exp. Res.* 1996;20(9):1594-1598. DOI 10.1111/j.1530-0277.1996. tb05936.x
- Wang B.-Q., Yao N.-Q., Zhou X., Liu J., Lv Z.-T. The association between attention deficit/hyperactivity disorder and internet addiction: a systematic review and meta-analysis. *BMC Psychiatry*. 2017; 17(1):260. DOI 10.1186/s12888-017-1408-x
- Weinstein A., Lejoyeux M. Neurobiological mechanisms underlying internet gaming disorder. *Dialogues Clin. Neurosci.* 2020;22(2): 113-126. DOI 10.31887/DCNS.2020.22.2/aweinstein
- Werling A.M., Grünblatt E. A review of the genetic basis of problematic Internet use. *Curr. Opin. Behav. Sci.* 2022;46:101149. DOI 10.1016/j.cobeha.2022.101149
- Yu Y., Mo P.K.-H., Zhang J., Li J., Lau J.T.-F. Impulsivity, self-control, interpersonal influences, and maladaptive cognitions as factors of internet gaming disorder among adolescents in China: cross-sectional mediation study. J. Med. Internet. Res. 2021;23(10):e26810. DOI 10.2196/26810
- Zhang M., Gao X., Yang Z., Wen M., Huang H., Zheng R., Wang W., Wei Y., Cheng J., Han S., Zhang Y. Shared gray matter alterations in subtypes of addiction: a voxel-wise meta-analysis. *Psychopharmacology (Berl.*). 2021;238(9):2365-2379. DOI 10.1007/s00213-021-05920-w

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 06.04.2024. После доработки 30.06.2024. Принята к публикации 15.07.2024.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции»/"Vavilov Journal of Genetics and Breeding" до 2011 г. выходил под названием «Информационный вестник ВОГиС»/ "The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Сетевое издание «Вавиловский журнал генетики и селекции» – реестровая запись СМИ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Russian Science Citation Index на платформе Web of Science, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:

русскоязычная версия – на сайте https://vavilovj-icg.ru/ и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440 англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour и платформе PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

При перепечатке материалов ссылка обязательна.

email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», проспект Академика Лаврентьева, 10, Новосибирск, 630090. Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090. Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977. Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218. Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич. Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева. Дата публикации 30.09.2024. Формат 60 × 84 ¹/₈. Уч.-изд. л. 13.3.