

Обзорная статья
УДК 524:631.527.2:633.852.53
DOI: 10.30901/2227-8834-2024-4-252-263



NGS-секвенирование в селекционно-генетических исследованиях сои

М. Т. Меньков^{1,2}, И. В. Розанова¹, А. Я. Евлаш¹, Е. К. Хлесткина^{1,2}

¹ Научно-технологический университет «Сириус», Центр генетики и наук о жизни, Краснодарский край, Россия

² Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова, Санкт-Петербург, Россия

Автор, ответственный за переписку: Михаил Тимофеевич Меньков, menkov.mt@talantiuspeh.ru

Соя (*Glycine max* (L.) Merr.) является одной из важнейших зернобобовых культур, производство которой растет с каждым годом и на 2024 г. составляет около 7 миллионов тонн. Целью данного обзора было обобщение последних достижений в селекции сои, в том числе с применением методов высокопроизводительного секвенирования и геномных технологий. Соя является одним из наиболее изучаемых растений. Ввиду ее важности для сельского хозяйства селекционеры постоянно реализуют в своей работе наиболее передовые методы. Исследования, выполненные в последние годы, показали преимущество подходов, основанных на применении молекулярно-генетических маркеров. Первый вариант последовательности генома сои, геном *G. max* 'Williams 82', представили в 2010 г., и это событие значительно ускорило изучение и развитие генетических исследований культуры. Полученные данные позволили создавать ресурсы, направленные как на изучение функциональной организации генов сои, так и на выведение новых улучшенных сортов. В обзоре обобщены результаты по крупным проектам секвенирования сои, в том числе и пангеномные работы. Описаны методы, применяющиеся для генетического картирования высокого разрешения, такие как анализ SNP с использованием чипов и метод GBS (genotyping-by-sequencing). Представлена информация о генах сои, связанных с хозяйственно ценными и селекционно значимыми признаками, идентификация которых позволила выделить их в качестве мишеней для редактирования.

Ключевые слова: ускоренная селекция, молекулярные маркеры, соя, NGS, *Glycine max*, GBS

Благодарности: результаты получены при финансовой поддержке проекта, реализуемого в рамках государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории «Сириус»» (соглашение № 18-03 от 10.09.2024).

Авторы благодарят рецензентов за их вклад в экспертную оценку этой работы.

Для цитирования: Меньков М.Т., Розанова И.В., Евлаш А.Я., Хлесткина Е.К. NGS-секвенирование в селекционно-генетических исследованиях сои. *Труды по прикладной ботанике, генетике и селекции*. 2024;185(4):252-263. DOI: 10.30901/2227-8834-2024-4-252-263

SURVEYS

Review article

DOI: 10.30901/2227-8834-2024-4-252-263

Next-generation sequencing in soybean breeding and genetic research

Mikhail T. Menkov^{1,2}, Irina V. Rozanova¹, Anastasia Ya. Evlash¹, Elena K. Khlestkina^{1,2}¹ *Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, Krasnodar Territory, Russia*² *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia***Corresponding author:** Mikhail T. Menkov, menkov.mt@talantiuspeh.ru

Soybean (*Glycine max* (L.) Merr.) is one of the most important grain legume crops whose production has been growing every year and by 2024 reached ca. 7 million tons. The objective of this review was to summarize the latest achievements in soybean breeding, including the use of high-throughput sequencing methods and genomic technologies. Soybean is one of the most studied plants. The studies of recent years showed the advantage of approaches based on the use of molecular genetic markers in breeding. The first version of the soybean genome sequence, the *G. max* genome “Williams 82”, was presented in 2010, and this event significantly accelerated the study and development of genetic research on the crop. The data obtained made it possible to develop resources aimed at both studying the functional organization of soybean genes and breeding new improved cultivars. The review summarizes the results of large-scale soybean sequencing projects, including pan-genome works. Methods used for high-resolution genetic mapping, such as the SNP array analysis and the GBS (genotyping-by-sequencing) technique, are described. Information is provided on soybean genes associated with valuable agronomic and breeding-oriented traits whose identification made it possible to single them out as targets for editing.

Keywords: accelerated breeding, molecular markers, soybean, NGS, *Glycine max*, GBS**Acknowledgments:** the results were obtained with financial support from the state program for the Sirius Federal Territory “Scientific and technological development of the Sirius Federal Territory” (Agreement No. 18-03 dated Sept. 10, 2024). The authors thank the reviewers for their contribution to the peer review of this work.**For citation:** Menkov M.T., Rozanova I.V., Evlash A.Ya., Khlestkina E.K. Next-generation sequencing in soybean breeding and genetic research. *Proceedings on Applied Botany, Genetics and Breeding*. 2024;185(4):252-263. DOI: 10.30901/2227-8834-2024-4-252-263

Введение

Соя является одной из важнейших мировых сельскохозяйственных культур, занимает четвертое место после риса, пшеницы и кукурузы в мировом производстве растительных продуктов питания и служит основным источником ценных пищевых белков и растительных масел. Средняя урожайность сои в России – 15,9 ц/га. Регионами – лидерами по валовому сбору являются дальневосточные и черноземные области России, такие как Амурская, Белгородская и Курская области, Приморский и Краснодарский края, Тамбовская область (<https://rosstat.gov.ru/compendium/document/13277>) (рис. 1). По данным ФАО, производство сои в России началось с 1990-х гг. На данный момент размер занимаемых площадей сои и валовый сбор увеличились в 24 раза (<https://www.fao.org/faostat/en/#data/QCL>) (рис. 2).

Благодаря своим выдающимся характеристикам, в первую очередь высокому содержанию белка (30–52%), соя культивируется во множестве стран как продукт питания, а также используется для кормления животных и в технических целях. По уровню белка она превосходит все другие возделываемые культуры, включая горох (22–24%), подсолнечник (семена) (25%), пшеницу (9–26%), рис (7%), кукурузу (10%) и др. Ценность соевого белка определяется содержанием незаменимых аминокислот, которые в совокупности составляют 20% от общего содержания белка, по сравнению с 18% для пшеничного белка (Gorissen et al., 2018). Соевые бобы занимают второе место в пересчете на сухое вещество после кукурузы по содержанию полиаминов, играющих роль в делении клеток и при окислительном стрессе (Egorov et al., 2021).

Можно выделить два основных направления селекции сои – зерновое и овощное. Овощную сою используют

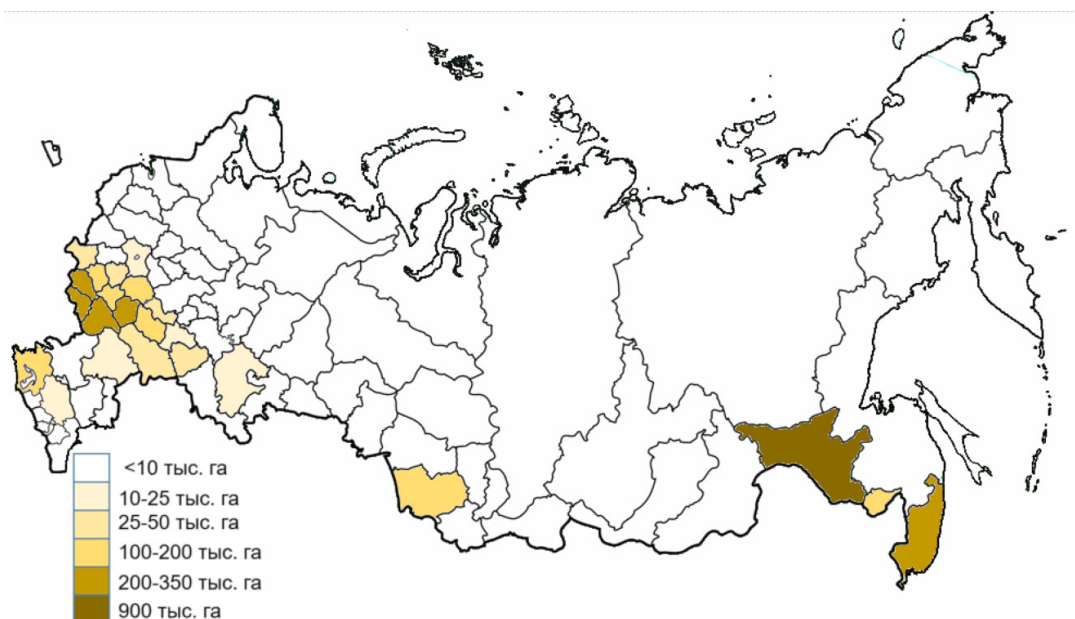


Рис. 1. Топ регионов России по посевным площадям сои (данные за 2023 год, <https://rosstat.gov.ru/compendium/document/13277>)

Fig. 1. Top regions of Russia according to soybean production areas (data for 2023, <https://rosstat.gov.ru/compendium/document/13277>)

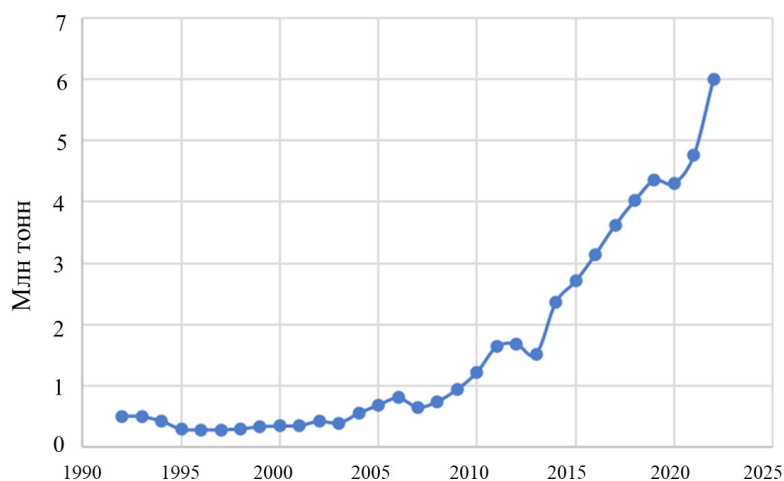


Рис. 2. Увеличение производства сои в России, начиная с 1990-х гг. (<https://www.fao.org/faostat/en/#data/QCL>)

Fig. 2. Increase in soybean production in Russia since the 1990s (<https://www.fao.org/faostat/en/#data/QCL>)

в качестве свежего продукта питания или в виде замороженных бобов. Из зерновой сои получают соевое масло, соевый шрот, использующийся как высокобелковый корм животных. Есть и более нестандартные применения сои, такие как получение фотополимерной смолы из соевого масла (Guit et al., 2020), получение клея из соевого шрота (Luo et al., 2020). Это показывает широкий потенциал применения и актуальность возделывания этой культуры.

Glycine max (L.) Merr. ($2n = 4x = 40$), или соя культурная, и *G. soja* Siebold. & Zucc. ($2n = 4x = 40$), или соя уссурийская, – однолетние растения, принадлежащие к роду *Glycine* Willd. (Fabaceae) (Carter et al., 2004). *G. soja* и *G. max* имеют одинаковое число хромосом, легко гибридизируются, демонстрируют нормальное скрещивание и генерируют жизнеспособные фертильные гибриды. Произрастают в основном в северном Китае, Японии, Корее, восточной части России, Южной и Северной Америке (Carter et al., 2004).

Еще Н. И. Вавилов называл сою важнейшим эндемиком умеренной зоны Китайского очага Восточноазиатского центра происхождения культурных растений (Seferova, Vishnyakova, 2014). L. Wang с коллегами в 2016 г. с помощью анализа изменчивости простых повторяющихся последовательностей (SSR) предположили возможный конкретный ареал происхождения сои на основании того, что в его пределах генетическое разнообразие наиболее высокое и снижается по мере отдаления. Один район ареала – в долине среднего и нижнего течения реки Вэй и в долине верхнего течения реки Ханьцзян, другой район – на севере провинции Хэбэй (Wang et al., 2016). Сою одомашнили по меньшей мере пять тысяч лет назад от предка, распространенного по всей Восточной Азии, включая большую часть Китая, Корею, Японию и часть России (Carter et al., 2004). Она была ввезена в Северную Америку в 1765 г. и массово распространилась по Центральной и Южной Америке в первой половине прошлого века. В настоящее время время культивируемая соя широко выращивается по всему миру, в том числе в Азии (Китай, Япония, Корея и Индия), Северной Америке (США и Канада) и Южной Америке (Бразилия, Аргентина и Парагвай).

С развитием генетики начинается новый этап исследований разнообразия генофонда культуры сои. Благодаря современным исследованиям, например, выяснилось, что доместикация сои в районе Японии произошла в период 3000–1000 гг. до н. э. (Jeong et al., 2019). Полученные знания применяются в маркер-ориентированной селекции (МОС) и для внедрения современных стратегий в селекционных программах, а также для разработки технологий, позволяющих создавать начальный селекционный материал с определенными характеристиками с использованием геномного редактирования.

Основная цель данного обзора – исследовать направления научных изысканий в области генетико-селекционных работ по сое, использующих методы высокопроизводительного секвенирования.

Проекты по секвенированию генома сои

В настоящее время данные о геномных последовательностях играют ключевую роль в агрономических программах. Технологический прорыв, произошедший в области генетики, позволил использовать новые возможности в селекции растений. Первый вариант последовательности генома *G. max* представили в 2010 г., и это

событие дало толчок к развитию генетических исследований этой культуры. Геном получили секвенированием с помощью коротких прочтений образца сои “Williams 82”. Размер генома составил 1115 Мб; при представлении первого референсного генома описали и каталогизировали 85% генов. Наиболее значимые открытые исследования представлены в таблице. Исследования показали, что геном сои в ходе своей эволюции пережил два цикла полной дупликации, произошедшие 59 и 13 миллионов лет назад. В результате около 75% генов существуют в нескольких копиях (Schmutz et al., 2010). Это делает сою хорошим объектом исследования для изучения полногеномной дупликации у растений.

Для углубления знаний о процессах, происходящих в живых организмах, в настоящее время недостаточно информации только от одного генома. Для выявления генетического разнообразия секвенируют различные сорта одного вида растений. Такой концепт получил название «пангеном». В целом для сборки пангенома используются в основном три подхода: сравнительная сборка *de novo*, итеративная сборка и стратегия «сопоставление с пангеномом». Соя *G. soja* стала одним из первых культурных растений, для которого начали проводиться пангеномные исследования (Torkamaneh et al., 2021).

Данные по секвенированию перестали быть уникальной информацией. Это стало тем пространством, от которого исследователи отталкиваются, чтобы двигаться дальше. Собранные данные создают полезные ресурсы для будущих исследований, посвященных функциональной организации генов сои, а также для практических разработок, направленных на селекцию новых, усовершенствованных сортов. Это включает в себя применение методов селекции следующего поколения (next-generation breeding), особенно технологий генетического редактирования.

Селекция сои с применением ДНК-маркеров

Применению маркеров в селекции сои уже 40 лет. Первыми стали маркеры RFLP (restriction fragment length polymorphism), которые использовались для разработки анализов профилей ДНК, и их появление создало новое направление в селекции, основанное на привлечении молекулярных методов. С появлением полимеразной цепной реакции (ПЦР) в 1990-х гг. начали широко использовать молекулярные маркеры, основанные на ПЦР, включая простые повторы последовательностей, известные как SSR (simple sequence repeats). Тандемно повторяющиеся 2–5-нуклеотидные последовательности ДНК, такие как (CA)_n, (ATT)_n, (AGAT)_n, хорошо зарекомендовали себя и до сих пор используются в паспортизации генотипов сои (Lavrent'eva et al., 2023). По мере совершенствования геномных технологий появились маркеры на основе однонуклеотидного полиморфизма – SNP. В отличие от других маркеров, обнаружение SNP не ограничивается разделением фрагментов по размеру (как это делается, например, при использовании эндонуклеаз рестрикции для детекции однонуклеотидного полиморфизма ДНК) – именно для анализа SNP-локусов изобрели автоматизированные методы (см. ниже про ДНК-чипы), что позволило повысить производительность и осуществлять анализ полиморфизма ДНК больших выборок организмов. Особенно ценным это стало для анализа количественных признаков.

Улучшение по количественным признакам, таким как урожайность, устойчивость к полеганию, является основ-

Таблица. Крупные проекты по секвенированию генома сои
Table. Major soybean genome sequencing projects

Год	Событие
2010...	Секвенирование первого генома сои <i>G. max</i> сорта 'Williams 82', получение референсного генома (Schmutz et al., 2010). Секвенирование генома дикой сои <i>Glycine soja</i> , выравнивание на референсный геном. Полученные риды общим размером 48,8 Гб были выровнены на референсный геном <i>G. max</i> , и для <i>G. soja</i> определен консенсус. Эта консенсусная последовательность охватывает 915,4 Мб, что составляет 97,7% опубликованной последовательности генома <i>G. max</i> . Показано, что нуклеотидная последовательность генома <i>G. soja</i> содержит 2,5 Мб полиморфных SNP, что составляет миллионы SNP (single nucleotide locus, однонуклеотидный полиморфизм) между <i>G. max</i> и <i>G. soja</i> SNP в негенных и генных регионах. (Kim M.Y. et al., 2010).
2013	Создание чипа SoySNP50K iSelect BeadChip (Song et al., 2013).
2014	Создание первого пангенома сои из семи образцов <i>G. soja</i> (Li et al., 2014). По этой малой выборке обнаружили, что набор переменных генов, не относящихся к обязательным, эволюционировал быстрее и был более изменчивым, чем основной набор. Среди переменного набора генов выявили наличие генов устойчивости (<i>R</i> -генов). Предполагается, что вариации числа копий генов устойчивости (<i>R</i>) могут помочь объяснить различия в устойчивости между дикими и культивируемыми образцами (Li et al., 2014).
2015	Создание пангенома сои из 30 образцов <i>G. soja</i> и <i>G. max</i> . Z. Zhou с коллегами исследовали локусы, ассоциированные с признаками масличности семян сои, и показали, что популяции сои структурированы географически, некоторые признаки и локусы зависят от географического региона (Zhou Z. et al., 2015).
2017	Сравнение китайской и американской популяций сои (Liu Z. et al., 2017).
2018	Секвенирование генома <i>G. latifolia</i> (Liu Q. et al., 2018). Сборка <i>de novo</i> последовательности генома китайского сорта сои 'Zhonghuang 13' с помощью комбинации данных SMRT, Hi-C и оптического картирования (Shen et al., 2018).
2019	Получение эталонного генома образца китайской дикой сои W05 (Shen et al., 2018).
2020	Создание пангенома сои на основании сборки <i>de novo</i> 26 геномов сои. Пангеномный анализ выявил многочисленные (2898) генетические вариации (Liu Y. et al., 2020), что помогло исследователям определить гены-кандидаты, связанные с агрономическими признаками. Сборка таких фенотипически разных растений позволяет совершенствовать платформу для будущих углубленных функциональных геномных исследований сои (Liu Y., Tian, 2020; Lyu, 2020).
2021	Секвенирование 8 сортов сои из разных широт (Chu et al., 2021). Сборка генома популярного корейского сорта сои Hwangkeum (Kim M.S. et al., 2021). Создание пангенома сои (PanSoy) из 204 образцов, различающихся филогенетически и географически (Torkamaneh et al., 2021).
2022	Секвенирование (WGS-Seq) 181 сорта сои (Liu J. et al., 2022). Аннотирование генома овощного сорта сои 'Zhenong 6' (Liu N. et al., 2022).
2023	Создание чипа SoySNP 600K (Ma X. et al., 2023).
2024	Секвенирование генома сои сорта Jack (Huang Y. et al., 2024). Сборка генома <i>de novo</i> сорта сои HJ117 с высоким содержанием белка (Liu Z. et al., 2024).

ной задачей селекционных программ. Значительное преимущество по сравнению с традиционной имеет селекция с использованием молекулярных маркеров, благодаря повышению эффективности отбора и сокращению времени селекции. При разработке маркеров для селекционных программ в первых работах исследователи столкнулись с такой проблемой, как неравновесное сцепление (LD, linkage disequilibrium). Это неслучайная связь между двумя аллелями, в силу которой определенные комбинации аллелей встречаются наиболее часто. То есть при разработке маркеров необходимо учитывать генетическую структуру популяции, происхождение выбранных генотипов и родственные связи между ними. Двуродительская популяция F_2 имеет максимальное количество неравновесного сцепления между несвязанными геномными районами и геномными районами на

разных хромосомах, что приводит к получению ложноположительных ассоциаций. Влияние неравновесия по сцеплению продемонстрировала компания Monsanto при исследовании трансгенного сорта сои 'Roundup Ready' с встроенным геном устойчивости к гербицидам глифосата Roundup (Eathington et al., 2007). Всего провели генотипирование 750 линий сои, содержащих множество однонуклеотидных полиморфизмов (SNP). Примерно половина этих линий произошла от сорта сои 'Roundup Ready'. Линии распределили по категориям устойчивости и восприимчивости в зависимости от их фенотипической реакции на гербицид глифосат. Выявили 49 молекулярных маркеров на 15 различных хромосомах, ассоциированных с устойчивостью. Однако локализация трансгена 40-3-2, в результате которого появилась устойчивость к гербициду, была известна (Padgett et al., 1995). Она пред-

ставляет собой одну вставку и находилась в группе сцепления D1b (U19) генетической карты сои USDA. Поскольку эта область была единственно значимой, то, следовательно, 48 молекулярных маркеров, ассоциированных с устойчивостью, оказались ложноположительными из-за LD. В настоящее время разработано несколько алгоритмов для выявления сцепленных SNP и уменьшения степеней LD. Метод, разработанный Дж. К. Причардом с соавторами (Pritchard et al., 2000), использует информацию о молекулярных маркерах для определения структуры популяции и учета этой структуры в анализе. Общедоступную программу под названием Structure (Pritchard et al., 2000) разработали для анализа ассоциативных исследований. Среди используемых программ можно отметить Arlequin (Excoffier, Lischer, 2010), MIDAS (Gaunt et al., 2006). Популярными стратегиями реализованы в PLINK 1.07/1.9 (Purcell et al., 2007). Данные программы позволяют сокращать количество SNP среди высококоррелирующих, оставляя один.

При повышении технологического уровня и удешевлении методов секвенирования ДНК появилось много возможностей изучения взаимосвязей между генетическим и фенотипическим разнообразием. Высокопроизводительное секвенирование и высокопроизводительное фенотипирование легли в основу геномной селекции (Varabaschi et al., 2016).

Геномной селекцией предшествовал отбор с помощью маркеров (МОС). К настоящему времени работы по МОС выявили большое количество QTL, ассоциированных с различными количественными признаками, включая физиологические, морфологические, а также связанные с урожайностью и качеством урожая. Многие исследования продемонстрировали успешное использование МОС в селекции сои: при направленном отборе сои упор делали на укрупнение боба, укрупнение семян, их твердость, что включает в себя водопроницаемость сухих семян и твердость приготовленных, архитектуру стебля, потерю чувствительности к фотопериоду и зависимости от длины дня, устойчивость к патогенам, вкусовые качества, препятствование его раннего растрескивания и др. МОС учитывает основные гены и локусы количественных признаков с большим эффектом (QTL) для различных признаков (Bhat, Yu, 2021). Однако этот подход зависит от доступности маркеров, связанных с интересующим признаком, которые можно выявить либо с помощью анализа двуродительских популяций, либо посредством исследований ассоциаций по всему геному (GWAS).

Геномная селекция (ГС), в отличие от традиционной МОС, использует полногеномный маркерный профиль селекционных линий для прогнозирования геномной селекционной ценности (genomically estimated breeding value, GEV). Таким образом учитываются геномные вариации, которые регулируются в том числе и незначительными эффектами QTL/генов (Kim J.M. et al., 2020; Ukhatoва et al., 2023).

По мере развития селекционных программ постоянно расширяется число исследуемых признаков. В числе прочих приступили к решению таких задач, как уменьшение содержания олигосахаридов семейства рафинозы, получение гипоаллергенных сортов, увеличение содержания олеиновой кислоты и уменьшение содержания линолевой и линоленовой кислот, увеличение содержания белка, увеличение числа семян на боб, увеличение размера листьев, увеличение веса семян, ранний переход к цветению в условиях длинного дня, повышение засухо- и солеустойчивости, увеличение содержания изофлаво-

ноидов и изменение профилей жирных кислот (Cai et al., 2023), сопротивление вирусу мозаики за счет сверхэкспрессии гена *GmMLR1* (Che et al., 2023), улучшение клубенькообразования, устойчивость к полеганию (Ukhatoва et al., 2023), изменение архитектуры стебля с увеличением числа узлов и контролем высоты растений. Расширение спектра исследуемых признаков продолжается по настоящее время.

Все это невозможно без современных быстрых, простых и сравнительно недорогих методов секвенирования и генотипирования растений.

Методы, используемые для генотипирования сои

Генотипирование при помощи ДНК-чипов

Изучение полиморфизма ДНК и выявление полиморфных локусов, связанных с экономически важными характеристиками, представляет собой ключевую задачу в области генетики и селекции растений. Предыдущие методы анализа полиморфизма ДНК у растений включали отбор праймеров, ПЦР-амплификацию и секвенирование ампликонов из небольшого набора различных генотипов. Внутривидовое разнообразие может быть изучено за счет сравнения считываний коротких последовательностей референсного генома и между собой (Kim M.Y. et al., 2010). С появлением доступного качественного полного генома сои (Schmutz et al., 2010) разработка чипа для поиска SNP стало лишь вопросом времени. В 2013 г., благодаря предшествующим работам по секвенированию генома сои и ресеквенированию его у разных сортов с применением NGS, разработали SoySNP50K iSelect BeadChip (Song et al., 2013), который на тот момент стал самым плотным по численности SNP среди аналогичных чипов для растений (Song et al., 2015). Он нашел свое применение в оценке генетического разнообразия, построении карт генетических связей, установлении связи «генотип – фенотип» (genome-wide association studies, GWAS). Данный чип долгое время не терял своей актуальности: так, в 2021 г. с его помощью охарактеризовали генетическое разнообразие и структуру 481 сорта сои, из которых 52 относились к диким (*G. soja*) образцам и 429 – к культурным (*G. max*) сортам. Среди культивируемых сортов обнаружили большое количество высококонсервативных регионов, которые, видимо, селективались в процессе доместикации культуры. В исследовании также выявили, что большое количество высококонсервативных регионов коррелирует с определенным географическим положением (Valliyodan et al., 2021).

В 2023 г. X. Ma et al. (2023) собрали информацию обо всех известных на тот момент геномах, провели фильтрацию по принципу наиболее часто встречаемых и значимых SNP и на основе этого спроектировали чип SoySNP 600K с более чем 600 тысячами SNP-маркеров (см. таблицу). Интервал между каждой парой маркеров составил ~1,6 Кб, а расстояние интервала более 5 Кб приходилось только на 2,4% пар. Примерно 47% SNP были расположены во внутривидовых регионах, особенно в экзонных регионах и сайтах сплайсинга, в результате чего было охвачено большинство генов, кодирующих белок в геноме (Ma X. et al., 2023). Разработав чип, исследователи сконструировали и генотипировали популяцию RIL (recombinant inbred lines – рекомбинантные инбредные линии) сои и смоделировали генетическую карту высокого разрешения (Ma X. et al., 2023).

Поиск на чипах значимых SNP, ставших золотым стандартом в исследованиях генетического разнообразия культурных растений, несмотря на развитие других методов изучения генома, еще долго будет применяться и совершенствоваться.

Генотипирование на основе NGS-секвенирования

Для разработки SNP-чипов маркеры необходимо предварительно обнаружить. Среди платформ генотипирования, представленных на основе NGS, GBS (genotyping-by-sequencing) является наиболее мощным, простым, высокопроизводительным и эффективным методом получения SNP. В основу метода GBS положено снижение сложности генома с помощью ферментов рестрикции (restriction enzymes, RE) у видов с большим и сложным геномом, таких как растения. Метод позволяет одновременно обнаруживать и оценивать миллионы SNP в больших коллекциях линий, которые можно использовать для анализа генетического разнообразия, картирования сцепления, GWAS, геномной селекции и в эволюционных исследованиях. Обнаруженные участки секвенируются с небольшим покрытием. Полученные последовательности используют для определения аллельного разнообразия для каждого образца. Подход объединяет обнаружение и генотипирование больших популяций, что делает его превосходной маркерной платформой для селекционных приложений даже в отсутствие референсной последовательности генома или ранее открытого полиморфизма (Bhat et al., 2016). GBS на основе рестрикционных ферментов (RE-based) может быть как одноферментным, так и двухферментным с контролем размера рестрикционных фрагментов. Существуют вариации метода: Ё. Суяма и Ю. Мацуки предложили заменить RE на ПЦП (Suyama, Matsuki, 2015). Также интересен такой подход, как Hyper-seq (Zou, Xia, 2022), когда разрабатывают специальные праймеры, богатые GC, нацеленные на генные области генома. Затем проводят ПЦП и секвенируют большой набор проб с применением баркодов для их различия. Или же такой подход, как SNP на основе транскриптома, который предложили в качестве альтернативы для полного секвенирования большого и сложного генома (Telfer et al., 2019). Этот метод фокусируется на кодирующих и регуляторных областях генома, которые представляют основной интерес для исследований функциональной геномики (Reyes et al., 2022).

Было продемонстрировано, что GBS является эффективным методом для генетического картирования высокого разрешения и высокоточной селекции сои, в том числе основанной на редактировании генома (прецизионной) (He et al., 2014). GBS охватывает гораздо большую долю генома по сравнению даже с самыми плотными чипами, доступными в настоящее время для растений рода Соя (Bhat et al., 2016). Массивы SNP для создания чипа обычно разрабатываются из ограниченной выборки, тогда как GBS может захватывать генетическую изменчивость, специфичную для каждой конкретной популяции. Преимущество GBS в том, что маркеры обнаруживаются с использованием популяции, подлежащей генотипированию. Метод GBS, как и любой инструмент, имеет свои ограничения: большое количество пропущенных данных, ошибки и недооцененные гетерозиготные генотипы (Reyes et al., 2022). Однако гибкость, низкая стоимость и точность прогнозирования GEBV с помощью GBS делают этот подход привлекательным для современных

исследований растений, в том числе и сои (Bhat et al., 2016).

Применение методов NGS в идентификации и характеризации генов-мишеней для редактирования

При помощи вышеописанных методов выявили и охарактеризовали целый ряд генов сои, связанных с хозяйственно ценными (влияющими на урожайность и качество) и селекционно значимыми (ценными для развития гибридной селекции сои) признаками сои. Их идентификация и характеризация позволила выделить их в качестве мишеней для редактирования с целью улучшения сои. Среди них – группы генов, значимых для повышения урожайности и улучшения качества.

Гены-мишени, значимые для повышения урожайности

Ген *Pdh* кодирует продукт, относящийся к семейству диригентных белков (dirigent proteins, DIR), участвующих в процессе лигнификации. *Pdh* содействует укреплению стенки бобов, усиливая их скручивание при высыхании, что приводит к естественному растрескиванию и выбросу семян. Этот ген впервые описали в своем исследовании Н. Funatsuki с соавторами (Funatsuki et al., 2014). Использование метода секвенирования нового поколения помогло уточнить его роль и влияние на устойчивость к растрескиванию бобов у сои. Было доказано, что в генотипах с дефектной формой гена *Pdh1*, которая снижает вероятность растрескивания бобов, можно существенно уменьшить потери семян. В устойчивых к растрескиванию сортах сои обнаружили преждевременный стоп-кодон в гене *Pdh1*, который предотвращает его нормальное функционирование, что значительно снижает вероятность растрескивания. С использованием CRISPR/Cas9 ген *Pdh1* был нокаутирован, что привело к снижению растрескивания бобов и сохранению семян при сборе урожая. Технология создала мутации, вызывающие преждевременный стоп-кодон, что отключило функцию гена и предотвратило скручивание бобов. Это позволило уменьшить потери семян в условиях засухи и повысить стабильность урожайности сои.

GmJAGGED1 (*GmJAG*) – ген, который влияет на число семян на боб и форму листьев у сои. Замена нуклеотида С на G в гене приводит к потере функции *GmJAG1*, что повышает количество семян и урожайность. *GmJAG1* – гомолог гена *AtJAG1* у *Arabidopsis thaliana* – является ключевым транскрипционным фактором, участвующим в регуляции множества процессов развития. Соответствующее исследование показало, что *AtJAG1* кодирует белок с доменом «цинковые пальцы», участвующий в развитии тканей листьев и других органов растения, а нарушение его работы приводит к деформациям листьев. В исследовании М. Huang с соавторами (Huang M. et al., 2021) метод ChIP-seq эффективно помог исследовать участки связывания транскрипционного фактора *GmJAGGED1* в геноме сои, выявив новые мотивы и эпигенетические особенности. С помощью CRISPR/Cas9 был нокаутирован *GmJAG1* у сорта 'Huachun 6', что привело к увеличению урожайности.

GmKIX8-1 – регулятор размера органов у сои и негативный регулятор клеточной пролиферации, то есть разрастания ткани. *GmKIX8-1* кодирует белок с доменом KIX, участвующий в белок-белковых взаимодействиях, кото-

рые регулируют экспрессию генов. Эти домены распространены среди различных организмов, включая растения и животных. Ген *GmKIX8-1* идентифицировали как причинный ген для важного QTL, связанного с весом семян, – qSw17-1. Также в этом исследовании с помощью технологии CRISPR/Cas9 был нокаутирован *GmKIX8-1*, что привело к увеличению размеров листьев и веса семян, способствуя тем самым повышению продуктивности растений. G. T. Park с соавторами (Park et al., 2023) провели всесторонний анализ семейства генов с доменами KIX у сои, включая *GmKIX8-1*. Исследование выявило значительную экспрессию этих генов на ранних стадиях развития семян и их роль в регуляции органогенеза.

Гены-мишени, значимые для гибридной селекции

GmAMS1 и *GmAMS2* участвуют в развитии микроспора и клеток тапетума, которые необходимы для формирования пыльцы. Эти гены идентифицированы как ортологи гена *AMS* (*ABORTED MICROSPORES*) у *Arabidopsis thaliana*. *AMS* кодирует транскрипционный фактор класса MYC, который играет ключевую роль в развитии микроспора и клеток тапетума. Мутации в этом гене приводят к дефектам в развитии пыльцы, вызывая мужскую стерильность. В исследовании X. Chen с соавторами (Chen et al., 2021) технология CRISPR/Cas9 успешно доказала, что нокаут одного или обоих этих генов приводит к созданию мужских стерильных линий, которые используются для контроля опыления при гибридной селекции, особенно для повышения урожайности и качества культур.

Гены-мишени, значимые для селекции сои как масличной культуры.

Гены из семейства *FAD2* (*FAD2-1A*, *FAD2-1B*, *FAD2-2A*) кодируют десатуразу омегу-6 жирных кислот – фермент, преобразующий олеиновую кислоту в линолевою, – что играет важную роль в биосинтезе полиненасыщенных жирных кислот. Нарушение функции этих генов приводит к снижению уровня линолевой кислоты, что повышает качество масла и позволяет создавать улучшенные сорта масличного направления. История открытия у сои первого из этих генов связана с периодом, когда подходы с использованием NGS еще не были внедрены, однако идентификация разных представителей данного семейства, их характеристика и сама возможность редактирования – результат применения методов, тесно связанных с применением NGS.

FAD впервые идентифицировали у *Arabidopsis thaliana* с помощью мутантного скрининга, который показал, что мутации в этом гене повышают содержание олеиновой кислоты. E. P. Heppard с соавторами (Heppard et al., 1996) исследовали экспрессию микросомальных генов десатуразы у сои в зависимости от стадии развития и температуры роста, что показало важность этих генов для адаптации растений и регулирования липидного состава. Нокаут гена *FAD2* привел к снижению активности фермента, что повысило содержание олеиновой кислоты в семенах сои, улучшив стабильность масла и его устойчивость к окислению.

В исследовании M. A. Berestovoy с соавторами (Berestovoy et al., 2020) рассмотрена роль десатураз в жизнедеятельности растений, при этом особенно подчеркнута их важность для синтеза ненасыщенных жирных кислот,

необходимых для сохранения текучести клеточных мембран и адаптации к стрессовым условиям. Биотехнологический потенциал десатураз, таких как *FAD2*, заключается в улучшении состава масла сельскохозяйственных культур, что делает его более стабильным и полезным.

В исследовании J. Zhou с соавторами (Zhou et al., 2023) рассматривались различные стратегии генного редактирования системы CRISPR/Cas9 для изменения метаболизма насыщенных и ненасыщенных жирных кислот у сои. Редактирование методом направленной инактивации гена *FAD2-1A* приводит к повышению содержания мононенасыщенной жирной кислоты – олеиновой – и снижению уровня полиненасыщенной – линолевой. Олеиновая кислота более стабильна и окисляется медленнее, делая растительные масла устойчивее к прогорканию и термическим воздействиям, что важно для пищевой и косметической промышленности.

Ген *GmFAD2-1B* впервые идентифицировали с использованием методов прямой генетики, таких как сравнительный анализ по гомологии и молекулярное клонирование. В исследовании G. Tang с соавторами (Tang et al., 2005) изучена регуляция ферментов десатураз у сои, которые преобразуют олеиновую кислоту в линолевою. Было показано, что активность этих ферментов контролируется через дифференциальную стабильность и фосфорилирование, что важно для управления уровнем ненасыщенных жирных кислот в клетках и улучшении состава масла. Систему CRISPR/Cas9 с двумя направляющими РНК (dual gRNA) использовали для одновременного редактирования генов *GmFAD2-1A* и *GmFAD2-1B* у сои. Это привело к значительному увеличению содержания олеиновой кислоты и снижению уровней линолевой и альфа-линоленовой кислот, что сделало масло сои более стабильным и полезным благодаря высокому содержанию мононенасыщенных жирных кислот.

GmFAD2-2A идентифицировали с использованием методов геномного анализа, сравнительной геномики и молекулярного клонирования, аналогичных подходам к другим членам семейства *FAD2*. В исследовании N. Lakhssassi с соавторами (Lakhssassi et al., 2017) проведена характеристика семейства генов *FAD2* у сои, которые играют ключевую роль в биосинтезе полиненасыщенных жирных кислот, включая преобразование олеиновой кислоты в линолевою. Особое внимание уделялось мутациям в генах *FAD2-1A* и *FAD2-1B*, способным повышать содержание олеиновой кислоты в масле сои, что важно для улучшения его качества для пищевого и технического использования. Исследование показало ограничения метода TILLING при работе с генами с большим количеством копий, как в случае семейства *FAD2*, из-за сложности идентификации мутантов. С использованием технологии CRISPR/Cas9 созданы мутанты сои, в которых редактировались гены *GmFAD2-1A* и *GmFAD2-2A*. Нокаут этих генов приводит к увеличению белка и уменьшению содержания линолевой кислоты в семенах сои, что положительно сказывается на их питательных свойствах и повышает качество масла.

Гены *GmFATB1a* и *GmFATB1b* кодируют ферменты семейства ацил-АСР-тиоэстераз, которые участвуют в синтезе насыщенных жирных кислот, таких как пальмитиновая и стеариновая. Эти ферменты катализируют разрыв тиоэфирной связи, завершая синтез жирных кислот и высвобождая свободные жирные кислоты. Гены *FATB* впервые описаны у *Arabidopsis thaliana* с использованием методов мутантного скрининга, молекулярного клонирования и технологий высокопроизводительного секвенирования.

нирования (NGS). Изучили субстратное предпочтение тиюэстераз FatA и FatB и выявили ключевую роль N-концевого домена в специфичности ферментов. Мутация в *FATB* у арабидопсиса приводит к снижению содержания насыщенных жирных кислот, замедлению роста растений и нарушению мембранной структуры, что подчеркивает важность насыщенных жирных кислот для нормального роста.

Редактирование генов *GmFAT* может улучшить профиль жирных кислот в семенах за счет повышения уровня олеиновой кислоты. В исследовании J. Ma с соавторами (Ma J. et al., 2021) с помощью CRISPR/Cas9 был нокаутирован ген *GmFATB1*, что привело к заметному снижению содержания пальмитиновой и стеариновой кислот – основных насыщенных жирных кислот в семенах сои, при этом уровень ненасыщенных жирных кислот, таких как олеиновая, оставался стабильным. Такое соотношение способствует повышению качества масла, делая его более полезным для здоровья благодаря увеличению доли полезных ненасыщенных жиров и улучшению термостабильности, что особенно важно для кулинарного применения и продления срока хранения.

Таким образом, благодаря широкому спектру методов, основанных на применении NGS, поиск генов-мишеней и их редактирование стали важным инструментом в создании улучшенных линий сои и открыли перспективу для селекции следующего поколения (NGB).

Заключение

С развитием подходов NGS удалось провести секвенирование различных сортов и видов сои, тем самым сформировать базу для последующих геномных исследований. Большое количество исследуемых сортов позволяет оценить разнообразие генофонда культуры, а также дает возможность выявлять локусы, сцепленные с селекционными значимыми признаками. В настоящее время результаты секвенирования в совокупности с технологиями высокопроизводительного генотипирования можно использовать для эффективного направленного отбора нужных генотипов среди селекционных линий, что существенно ускорит создание новых сортов сои с заданными характеристиками.

References / Литература

Barabaschi D., Tondelli A., Desiderio F., Volante A., Vaccino P., Valè G. et al. Next generation breeding. *Plant Science*. 2016;242:3-13. DOI: 10.1016/j.plantsci.2015.07.010

Berestovoy M.A., Pavlenko O.S., Goldenkova-Pavlova I.V. Plant fatty acid desaturases: role in the life of plants and biotechnological potential. *Biology Bulletin Reviews*. 2020;10(2):127-139. DOI: 10.1134/S2079086420020024

Bhat J.A., Ali S., Salgotra R.K., Mir Z.A., Dutta S., Jadon V. et al. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Frontiers in Genetics*. 2016;7:221. DOI: 10.3389/fgene.2016.00221

Bhat J.A., Yu D. High-throughput NGS-based genotyping and phenotyping: Role in genomics-assisted breeding for soybean improvement. *Legume Science*. 2021;3(3):e81. DOI: 10.1002/leg3.81

Cai Y., Chen L., Hou W. Genome editing technologies accelerate innovation in soybean breeding. *Agronomy*. 2023;13(8):2045. DOI: 10.3390/agronomy13082045

Carter Jr.T.E., Hymowitz T., Nelson R.L. Biogeography, local adaptation, Vavilov, and genetic diversity in soybean. In:

D. Werner (ed.). *Biological Resources and Migration. Conference Proceedings*. Berlin; Heidelberg: Springer-Verlag; 2004. p.47-59. DOI: 10.1007/978-3-662-06083-4_5

Carter Jr.T.E., Nelson R.L., Sneller C.H., Cui Z. Genetic diversity in soybean. In: R.M. Shibles, J.E. Harper, R.F. Wilson, R.C. Shoemaker (eds). *Agronomy Monographs. Vol. 16. Soybeans: Improvement, Production, and Uses*. 3rd ed. Madison, WI: ASA-CSSA-SSSA; 2004. p.303-416. DOI: 10.2134/agronmonogr16.3ed.c8

Che Z., Zhang S., Pu Y., Yang Y., Liu H., Yang H. et al. A novel soybean malectin-like receptor kinase-encoding gene, *GmMLRKL1*, provides resistance to soybean mosaic virus. *Journal of Experimental Botany*. 2023;74(8):2692-2706. DOI: 10.1093/jxb/erad046

Chen X., Yang S., Zhang Y., Zhu X., Yang X., Zhang C. et al. Generation of male-sterile soybean lines with the CRISPR/Cas9 system. *The Crop Journal*. 2021;9(6):1270-1277. DOI: 10.1016/j.cj.2021.05.003

Chu J.C.C., Peng B., Tang K., Yi X., Zhou H., Wang H. et al. Eight soybean reference genome resources from varying latitudes and agronomic traits. *Scientific Data*. 2021;8(1):164. DOI: 10.1038/s41597-021-00947-2

Eathington S.R., Crosbie T.M., Edwards M.D., Reiter R.S., Bull J.K. Molecular markers in a commercial breeding program. *Crop Science*. 2007;47(53):154-163. DOI: 10.2135/crosci2007.04.0015IPBS

Egorov O.S., Borisova N.Yu., Borisova E.Ya., Rezhabbaev M.L., Afanas'eva E.Yu., Arzamastsev E.V. Structure and biological action of analogs and derivatives of biogenic polyamines. *Fine Chemical Technologies*. 2021;16(4):287-306. [in Russian] (Егоров О.С., Борисова Н.Ю., Борисова Е.Я., Режаббаев М.Л., Афанасьева Е.Ю., Арзамасцев Е.В. Структура и биологическое действие аналогов и производных биогенных полиаминов. *Тонкие химические технологии*. 2021;16(4):287-306). DOI: 10.32362/2410-6593-2021-16-4-287-306

Excoffier L., Lischer H. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010;10(3):564-567. DOI: 10.1111/j.1755-0998.2010.02847.x

FAOSTAT. Food and Agriculture Organization of the United Nations. Food and Agriculture Data: [website]. Available from: <https://www.fao.org/faostat/en/#data/QCL> [accessed Mar. 30, 2024].

Federal State Statistics Service: [website]. [in Russian] (Федеральная служба государственной статистики: [сайт]). URL: <https://rosstat.gov.ru/compendium/document/13277> [дата обращения: 30.03.2024].

Funatsuki H., Suzuki M., Hirose A., Inaba H., Yamada T., Hajika M. et al. Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(50):17797-17802. DOI: 10.1073/pnas.1417282111

Gaunt T.R., Rodriguez S., Zapata C., Day I.N.M. MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics*. 2006;7:227. DOI: 10.1186/1471-2105-7-227

Gorissen S.H.M., Crombag J.J.R., Senden J.M.G., Waterval W.A.H., Bierau J., Verdijk L.B. et al. Protein content and amino acid composition of commercially available plant-based protein isolates. *Amino Acids*. 2018;50(12):1685-1695. DOI: 10.1007/s00726-018-2640-5

Guit J., Tavares M.B.L., Hul J., Ye C., Loos K., Jager J. et al. Photopolymer resins with biobased methacrylates based on soybean oil for stereolithography. *ACS Applied*

- Polymer Materials*. 2020;2(2):949-957. DOI: 10.1021/acsapm.9b01143
- He J., Zhao X., Laroche A., Lu Z.H., Liu H., Li Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*. 2014;5:484. DOI: 10.3389/fpls.2014.00484
- Heppard E.P., Kinney A.J., Stecca K.L., Miao G.H. Developmental and growth temperature regulation of two different microsomal [ω]-6 desaturase genes in soybeans. *Plant Physiology*. 1996;110(1):311-319. DOI: org/10.1104/pp.110.1.311
- Huang M., Zhang L., Zhou L., Wang M., Yung W.S., Wang Z. et al. An expedient survey and characterization of the soybean JAGGED 1 (GmJAG1) transcription factor binding preference in the soybean genome by modified ChIPmentation on soybean protoplasts. *Genomics*. 2021;113(1 Pt 1):344-355. DOI: 10.1016/j.ygeno.2020.12.026
- Huang Y., Koo D.H., Mao Y., Herman E.M., Zhang J., Schmidt M.A. A complete reference genome for the soybean cv. Jack. *Plant Communications*. 2024;5(2):100765. DOI: 10.1016/j.xplc.2023.100765
- Jeong S.C., Moon J.K., Park S.K., Kim M.S., Lee K., Lee S.R. et al. Genetic diversity patterns and domestication origin of soybean. *Theoretical and Applied Genetics*. 2019;132(4):1179-1193. DOI: 10.1007/s00122-018-3271-7
- Kim J.M., Kim K.H., Jung J., Kang B.K., Lee J., Ha B.K. et al. Validation of marker-assisted selection in soybean breeding program for pod shattering resistance. *Euphytica*. 2020;216(11):166. DOI: 10.1007/s10681-020-02703-w
- Kim M.S., Lee T., Baek J., Kim J.H., Kim C., Jeong S.C. Genome assembly of the popular Korean soybean cultivar Hwangkeum. *G3 Genes/Genomes/Genetics*. 2021;11(10):jkab272. DOI: 10.1093/g3journal/jkab272
- Kim M.Y., Lee S., Van K., Kim T.H., Jeong S.C., Choi I.Y. et al. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(51):22032-22037. DOI: 10.1073/pnas.1009526107
- Lakhssassi N., Zhou Z., Liu S., Colantonio V., AbuGhazaleh A., Meksem K. Characterization of the *FAD2* gene family in soybean reveals the limitations of gel-based TILLING in genes with high copy number. *Frontiers in Plant Science*. 2017;8:324. DOI: 10.3389/fpls.2017.00324
- Lavrent'yeva S.I., Bondarenko O.N., Blinova A.A., Penzin A.A., Fokina E.M., Ivachenko L.E. Study of the morphological and agriculturally important traits of the wild forms and cultivated varieties of soybean from the All-Russia Soybean Research Institute and soybean identification using microsatellites. *Russian Agricultural Sciences*. 2023;49(4):341-347. DOI: 10.3103/S1068367423040080
- Li Z., Liu Z.B., Xing A., Moon B.P., Koellhoffer J.P., Huang L. et al. Cas9-Guide RNA directed genome editing in soybean. *Plant Physiology*. 2015;169(2):960-970. DOI: 10.1104/pp.15.00783
- Liu J., Xie H., Lin T., Tie C., Luo H., Yang B. et al. Putative variants, genetic diversity and population structure among soybean cultivars bred at different ages in Huang-Huai-Hai region. *Scientific Reports*. 2022;12(1):2372. DOI: 10.1038/s41598-022-06447-6
- Liu N., Niu Y., Zhang G., Feng Z., Bo Y., Lian J. et al. Genome sequencing and population resequencing provide insights into the genetic basis of domestication and diversity of vegetable soybean. *Horticulture Research*. 2022;9:uhab052. DOI: 10.1093/hr/uhab052
- Liu Q., Chang S., Hartman G.L., Domier L.L. Assembly and annotation of a draft genome sequence for *Glycine latifolia*, a perennial wild relative of soybean. *The Plant Journal*. 2018;95(1):71-85. DOI: 10.1111/tpl.13931
- Liu Y., Du H., Li P., Shen Y., Peng H., Liu S. et al. Pan-genome of wild and cultivated soybeans. *Cell*. 2020;182(1):162-176. DOI: 10.1016/j.cell.2020.05.023
- Liu Y., Tian Z. From one linear genome to a graph-based pan-genome: a new era for genomics. *Science China. Life Sciences*. 2020;63(12):1938-1941. DOI: 10.1007/s11427-020-1808-0
- Liu Z., Li H., Wen Z., Fan X., Li Y., Guan R. et al. Comparison of genetic diversity between Chinese and American soybean (*Glycine max* (L.)) Accessions revealed by high-density SNPs. *Frontiers in Plant Science*. 2017;8:2014. DOI: 10.3389/fpls.2017.02014
- Liu Z., Yang Q., Liu B., Li C., Shi X., Wei Y. et al. *De novo* genome assembly of a high-protein soybean variety HJ117. *BMC Genomic Data*. 2024;25(1):25. DOI: 10.1186/s12863-024-01213-1
- Luo J., Zhou Y., Gao Q., Li J., Yan N. From wastes to functions: a new soybean meal and bark-based adhesive. *ACS Sustainable Chemistry and Engineering*. 2020;8(29):10767-10773. DOI: 10.1021/acssuschemeng.0c02413
- Lyu J. Pan-genome upgrade. *Nature Plants*. 2020;6:732. DOI: 10.1038/s41477-020-0731-2
- Ma J., Sun S., Whelan J., Shou H. CRISPR/Cas9-mediated knockout of *GmFATB1* significantly reduced the amount of saturated fatty acids in soybean seeds. *International Journal of Molecular Sciences*. 2021;22(8):3877. DOI: 10.3390/ijms22083877
- Ma X., Fan L., Zhang Z., Yang X., Liu Y., Ma Y. et al. Global dissection of the recombination landscape in soybean using a high-density 600K SoySNP array. *Plant Biotechnology Journal*. 2023;21(3):606-620. DOI: 10.1111/pbi.13975
- Padgett S.R., Kolacz K.H., Delannay X., Re D.B., LaVallee B.J., Tinius C.N. et al. Development, identification, and characterization of a glyphosate-tolerant soybean line. *Crop Science*. 1995;35(5):1451-1461. DOI: 10.2135/cropsci1995.0011183X003500050032x
- Park G.T., Moon J.K., Park S., Park S.K., Baek J., Seo M.S. Genome-wide analysis of *KIX* gene family for organ size regulation in soybean (*Glycine max* L.). *Frontiers in Plant Science*. 2023;14:1252016. DOI: 10.3389/fpls.2023.1252016
- Pritchard J.K., Stephens M., Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945-959. DOI: 10.1093/genetics/155.2.945
- Pritchard J.K., Stephens M., Rosenberg N.A., Donnelly P. Association mapping in structured populations. *American Journal of Human Genetics*. 2000;67(1):170-181. DOI: 10.1086/302959
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 2007;81(3):559-575. DOI: 10.1086/519795
- Reyes V.P., Kitony J.K., Nishiuchi S., Makihara D., Doi K. Utilization of genotyping-by-sequencing (GBS) for rice pre-breeding and improvement: a review. *Life (Basel)*. 2022;12(11):1752. DOI: 10.3390/life12111752
- Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W. et al. Genome sequence of the palaeopolyploid soybean. *Nature*. 2010;463(7278):178-183. DOI: 10.1038/nature08670
- Seferova I.V., Vishnyakova M.A. Role of agricultural institutions of Chinese-Eastern Railway in the formation of the soybean collection at the Vavilov Institute of Plant Industry and its breeding in the USSR. *Vavilov Journal of Genetics and Breed-*

- ing. 2014;18(3):572-577. [in Russian] (Сеферова И.В., Вишнякова М.А. Вклад сельскохозяйственных опытных учреждений китайско-восточной железной дороги в формирование коллекции сои ВИР и в развитие ее селекции в СССР. *Вавиловский журнал генетики и селекции*. 2014;18(3):572-577).
- Shen Y., Liu J., Geng H., Zhang J., Liu Y., Zhang H. et al. *De novo* assembly of a Chinese soybean genome. *Science China. Life Sciences*. 2018;61(8):871-884. DOI: 10.1007/s11427-018-9360-0
- Song Q., Hyten D.L., Jia G., Quigley C.V., Fickus E.W., Nelson R.L. et al. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One*. 2013;8(1):54985. DOI: 10.1371/journal.pone.0054985
- Song Q., Hyten D.L., Jia G., Quigley C.V., Fickus E.W., Nelson R.L. et al. Fingerprinting soybean germplasm and its utility in genomic research. *G3 Genes/Genomes/Genetics*. 2015;5(10):1999-2006. DOI: 10.1534/g3.115.019000
- Suyama Y., Matsuki Y. MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. *Scientific Reports*. 2015;5:16963. DOI: /10.1038/srep16963
- Tang G.Q., Novitzky W.P., Griffin H.C., Huber S.C., Dewey R.E. Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. *The Plant Journal*. 2005;44(3):433-446. DOI: 10.1111/j.1365-3113X.2005.02535.x
- Telfer E., Graham N., Macdonald L., Li Y., Klápště J., Resende M, et al. A high-density exome capture genotype-by-sequencing panel for forestry breeding in *Pinus radiata*. *PLoS One*. 2019;14(9):0222640. DOI: 10.1371/journal.pone.0222640
- Torkamaneh D., Lemay M., Belzile F. The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. *Plant Biotechnology Journal*. 2021;19(9):1852-1862. DOI: 10.1111/pbi.13600
- Ukhatova Y.V., Erastenkova M.V., Korshikova E.S., Krylova E.A., Mikhailova A.S., Semilet T.V. et al. Improvement of crops using the CRISPR/Cas system: new target genes. *Molecular Biology*. 2023;57(3):387-410. [in Russian] (Ухатова Ю.В., Ерастенкова М.В., Коршикова Е.С., Крылова Е.А., Михайлова А.С., Семилет Т.В. и др. Улучшение культурных растений при помощи системы CRISPR/Cas: новые гены-мишени. *Молекулярная биология*. 2023;57(3):387-410). DOI: 10.31857/S0026898423030151
- Valliyodan B., Brown A.V., Wang J., Patil G., Liu Y., Otyama P.I. et al. Genetic variation among 481 diverse soybean accessions, inferred from genomic re-sequencing. *Scientific Data*. 2021;8(1):50. DOI: 10.1038/s41597-021-00834-w
- Wang L., Lin F., Li L., Li W., Yan Z., Luan W. et al. Genetic diversity center of cultivated soybean (*Glycine max*) in China – New insight and evidence for the diversity center of Chinese cultivated soybean. *Journal of Integrative Agriculture*. 2016;15(11):2481-2487. DOI: 10.1016/S2095-3119(15)61289-8
- Zhou J., Li Z., Li Y., Zhao Q., Luan X., Wang L. et al. Effects of different gene editing modes of CRISPR/Cas9 on soybean fatty acid anabolic metabolism based on *GmFAD2* family. *International Journal of Molecular Sciences*. 2023;24(5):4769. DOI: 10.3390/ijms24054769
- Zhou Z., Jiang Y., Wang Z., Gou Z., Lyu J., Li W. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*. 2015;33(4):408-414. DOI: 10.1038/nbt.3096
- Zou M., Xia Z. Hyper-seq: A novel, effective, and flexible marker-assisted selection and genotyping approach. *Innovation*. 2022;3(4):100254. DOI: 10.1016/j.xinn.2022.100254

Информация об авторах

Михаил Тимофеевич Меньков, младший научный сотрудник, Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, 354340 Россия, Краснодарский край, федеральная территория «Сириус», пгт. Сириус, Олимпийский пр., 1, Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова, 190000 Россия, Санкт-Петербург, ул. Б. Морская, 42, 44, menkov.mt@talantiuspeh.ru, <https://orcid.org/0009-0005-7915-441X>

Ирина Вениаминовна Розанова, кандидат биологических наук, старший научный сотрудник, Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, 354340 Россия, Краснодарский край, федеральная территория «Сириус», пгт. Сириус, Олимпийский пр., 1, i.rozanova@vir.nw.ru, <https://orcid.org/0000-0003-4341-0766>

Анастасия Ярославовна Евлаш, младший научный сотрудник, Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, 354340 Россия, Краснодарский край, федеральная территория «Сириус», пгт. Сириус, Олимпийский пр., 1, evlash-nastja@rambler.ru, <https://orcid.org/0000-0003-0246-8929>

Елена Константиновна Хлесткина, доктор биологических наук, профессор РАН, директор, Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова, 190000 Россия, Санкт-Петербург, ул. Б. Морская, 42, 44, руководитель направления «Биология и биотехнология растений», Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, 354340 Россия, Краснодарский край, федеральная территория «Сириус», пгт. Сириус, Олимпийский пр., 1, director@vir.nw.ru, <https://orcid.org/0000-0002-8470-8254>

Information about the authors

Mikhail T. Menkov, Associate Researcher, Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, 1 Olimpiyskiy Ave., Sirius Settlement, Sirius Federal Territory, Krasnodar Territory 354340, Russia, N.I. Vavilov All-Russian Institute of Plant Genetic Resources, 42, 44 Bolshaya Morskaya Street, St. Petersburg 190000, Russia, menkov.mt@talantiuspeh.ru, <https://orcid.org/0009-0005-7915-441X>

Irina V. Rozanova, Cand. Sci. (Biology), Senior Researcher, Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, 1 Olimpiyskiy Ave., Sirius Settlement, Sirius Federal Territory, Krasnodar Territory 354340, Russia, i.rozanova@vir.nw.ru, <https://orcid.org/0000-0003-4341-0766>

Anastasia Ya. Evlash, Associate Researcher, Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, 1 Olimpiyskiy Ave., Sirius Settle., Sirius Federal Territory, Krasnodar Territory 354340, Russia, evlash-nastja@rambler.ru, <https://orcid.org/0000-0003-0246-8929>

Elena K. Khlestkina, Dr. Sci. (Biology), Professor of the RAS, Director, N.I. Vavilov All-Russian Institute of Plant Genetic Resources, 42, 44 Bolshaya Morskaya Street, St. Petersburg 190000, Russia, Plant Biology and Biotechnology Research Manager, Sirius University of Science and Technology, Research Center of Genetics and Life Sciences, 1 Olimpiyskiy Ave., Sirius Settle., Sirius Federal Territory, Krasnodar Territory 354340, Russia, director@vir.nw.ru, <https://orcid.org/0000-0002-8470-8254>

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Contribution of the authors: the authors contributed equally to this article.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflict of interests: the authors declare no conflicts of interests.

Статья поступила в редакцию 02.11.2024; одобрена после рецензирования 02.12.2024; принята к публикации 03.12.2024.
The article was submitted on 02.11.2024; approved after reviewing on 02.12.2024; accepted for publication on 03.12.2024.