

<https://doi.org/10.17816/ecogen16126>

## ВЫСОКОПРОИЗВОДИТЕЛЬНОЕ СЕКВЕНИРОВАНИЕ В ГЕНЕТИКЕ И СЕЛЕКЦИИ ЛЬНА

© А.О. Ахметшина, К.В. Стрыгина, Е.К. Хлесткина, Е.А. Пороховинова, Н.Б. Брач

ФГБНУ «ФИЦ Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова», Санкт-Петербург

Для цитирования: Ахметшина А.О., Стрыгина К.В., Хлесткина Е.К., и др. Высокопроизводительное секвенирование в генетике и селекции льна // Экологическая генетика. — 2020. — Т. 18. — № 1. — С. 103–124. <https://doi.org/10.17816/ecogen16126>.

Поступила: 18.09.2019

Одобрена: 17.02.2020

Принята: 19.03.2020

☼ Лен (*Linum usitatissimum* L.) — важная масличная и прядильная культура. Применение современных методов селекции льна позволяет ускорять введение некоторых желаемых вариантов генов в генотипы будущих сортов. На сегодняшний день важным условием для их создания является развитие исследований, основанных на секвенировании нового поколения (NGS, Next Generation Sequencing). В данной обзорной статье обобщены результаты, полученные на основе применения NGS-секвенирования в исследовании льна. К настоящему моменту для *L. usitatissimum* получена генетическая карта сцепления с высокой плотностью маркеров, которую уже используют для более эффективного поиска локусов количественных признаков. На основе сравнительных исследований транскриптомов и микроРНК льна в контрольных и стрессовых условиях уточнены молекулярно-генетические механизмы ответов на биотические и абиотические стрессы. Была построена отличающаяся высокой точностью модель геномной селекции льна на устойчивость к пасмо. Также, благодаря применению NGS-секвенирования, удалось уточнить некоторые особенности эволюции геномов представителей рода *Linum*. Систематизированные в обзоре знания могут быть полезны для исследователей, ведущих работу по селекции льна, и представлять фундаментальный интерес для понимания филогенетических взаимоотношений внутри рода *Linum*, особенностей онтогенеза и механизмов ответа растений льна на различные факторы стресса.

☼ **Ключевые слова:** *Linum usitatissimum*; секвенирование РНК; RNA-seq; геномная селекция; лен масличный; лен-долгунец; маркер-ориентированная селекция; полногеномный анализ ассоциаций; NGS-секвенирование; транскриптом.

## HIGH-THROUGHPUT SEQUENCING TECHNIQUES TO FLAX GENETICS AND BREEDING

© А.О. Akhmetshina, K.V. Strygina, E.K. Khlestkina, E.A. Porokhovinova, N.B. Brutch

Federal Research Center “N.I. Vavilov All-Russian Institute of Plant Genetic Resources”, St. Petersburg, Russia

Cite this article as: Akhmetshina AO, Strygina KV, Khlestkina E.K, et al.

High-throughput sequencing techniques to flax genetics and breeding, University. *Ecological genetics*. 2020;18(1):103-124. <https://doi.org/10.17816/ecogen16126>.

Received: 18.09.2019

Revised: 17.02.2020

Accepted: 19.03.2020

☼ Flax (*Linum usitatissimum* L.) is an important oil and fiber crop. Using modern methods for flax breeding allows accelerating the introduction of some desired genes into the genotypes of future varieties. Today, an important condition for their creation is the development of research, that is based on next-generation sequencing (NGS). This review summarizes the results obtained using NGS in flax research. To date, a linkage map with a high marker density has been obtained for *L. usitatissimum*, which is already being used for a more efficient search for quantitative traits loci. Comparative studies of transcriptomes and miRNomes of flax under stress and in control conditions elucidated molecular-genetic mechanisms of abiotic and biotic stress responses. The very accurate model for genomic selection of flax resistant to pasmo was constructed. Based on NGS-sequencing also some details of the genus *Linum* evolution were clarified. The knowledge systematized in the review can be useful for researchers working in flax breeding and whereas fundamental interest for understanding the phylogenetic relationships within the genus *Linum*, the ontogenesis, and the mechanisms of the response of flax plants to various stress factors.

☼ **Keywords:** flax; genome-wide association studies; genomic selection; linseed; *Linum usitatissimum*; marker-assisted selection; RNA-seq; transcriptome; NGS-sequencing.

## ВВЕДЕНИЕ

Лен (*Linum usitatissimum* L.) — ценная прядильная и масличная культура. В составе вида выделяют 4 разновидности: возделываемый как яровая прядильная и масличная культура лен-долгунец *L. usitatissimum* L. var. *usitatissimum* sensu Rechinger (1974), а также масличные культуры — лен-кудряш *L. usitatissimum* var. *humile* (Mill.) Pers., лен-межеумок *L. usitatissimum* L. var. *intermedium* (Czernom.) Kutuz. и лен карликовый курдюш *L. usitatissimum* var. *nanum* Kutuz. [1]. Лен возделывают на всех континентах, кроме Антарктиды. Лен-долгунец занимает четвертое место в мире по площади возделывания среди растений, выращиваемых на волокно, а типы льна курдюш и межеумок — десятое среди масличных культур [2].

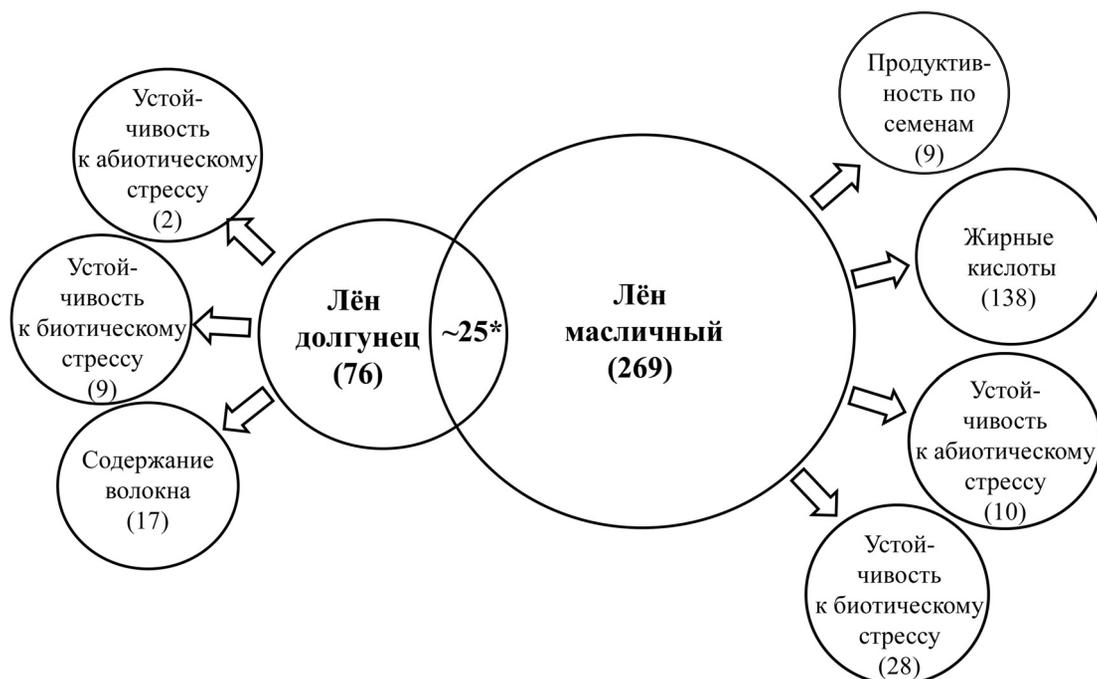
Задачи льноводства в России соответствуют мировым тенденциям. Однако особое внимание уделяется созданию сортов с большей экологической пластичностью и адаптивностью к неблагоприятным условиям, а также морозоустойчивых форм культуры с пониженной фотопериодической чувствительностью [3, 4]. Для льна, высеваемого

на севере России, актуальна селекция на устойчивость растений к токсичному действию ионов алюминия [5, 6].

По сравнению с другими культурами, в селекции льна использование современных методов (например, отбор при помощи ДНК-маркеров) пока затруднено ввиду недостаточной изученности селекционно значимых генов. Но поиск эффективных генетических маркеров для селекции ведется — по числу публикаций в этом направлении Россия входит в пятерку мировых лидеров, наряду с Канадой, США, Индией и Китаем (приложение 1).

При этом на масличном льне проводится примерно втрое больше подобного рода исследований, чем на льне-долгунце. Наибольшее внимание уделяют маркированию локусов, связанных с жирнокислотным составом масла, в меньшей степени — с устойчивостью, семенной продуктивностью (рис. 1).

Идентификация маркеров для диагностики аллельных вариантов генов является основой дальнейшей маркер-ориентированной селекции (marker-assisted selection, MAS), которую используют для отбора по признакам с моно- и олигоценным



**Рис. 1.** Основные направления исследований льна с использованием ДНК-маркеров. Число публикаций (цифры в скобках), связанных с применением ДНК-маркеров по основным направлениям селекции льна, на основе публикаций в базе данных Scopus ([www.scopus.com](http://www.scopus.com), дата обращения 13.09.2019) по пересечению ключевого слова «Marker» с «Fiber Flax» (лен, используемый для производства волокна) / «Linseed» (лен масличный) и с «Resistance» (устойчивость к биотическому стрессу) / «Tolerance» (устойчивость к абиотическому стрессу) / «Fiber content» (содержание волокна) / «Seed yield» (семенная продуктивность) / «Fatty acid» (жирные кислоты). \* Не во всех статьях возможно определить о каком именно типе льна идет речь

контролем, и способна дополнять традиционные подходы селекции, ускоряя отбор по некоторым признакам. Однако наиболее прогрессивным методом улучшения растения по таким признакам является геномное редактирование. Так, Sauer et al. (2016) использовали систему CRISPR/Cas для создания генотипа, устойчивого к гербициду глифосату [7].

Для отбора по признакам с полигенным контролем перспективным методом является геномная селекция. Это альтернативная стратегия, использующая статистические модели для предсказания, какие растения будут иметь оптимальные фенотипы на основе большого количества маркеров для индивидуального отбора [8]. У льна такой подход использован для геномной селекции на устойчивость к пасмо (*Septoria linicola* (Speg.) Gar.) [9].

Для выявления новых генетических маркеров для MAS или генов-мишеней для геномного редактирования, а также для геномной селекции все чаще привлекают методы, основанные на секвенировании нового поколения (next generation sequencing, NGS).

Целью настоящего обзора является анализ и систематизация данных, описывающих применение подходов, основанных на NGS, в селекционно-генетических исследованиях льна, а также оценка перспективы их дальнейшего использования в льноводстве.

#### СЕКВЕНИРОВАНИЕ ГЕНОМА *LINUM USITATISSIMUM*

Число хромосом у представителей рода *Linum L.* варьирует в широком диапазоне: от  $2n = 16$  до  $2n = 72$  [10]. Возделываемый вид *L. usitatissimum* имеет диплоидный геном  $2n = 30$ , который, по результатам сравнительного анализа геномов различных представителей рода *Linum*, проходил в ходе эволюции через аллотетраплоидизацию за счет гибридизации двух предковых диплоидных видов [11–13].

Первая полногеномная физическая карта льна была представлена в 2011 г. международным коллективом под руководством канадских ученых по программе Total Utilization Flax GENomics. Для сорта CDC Bethune создали библиотеку, состоящую из 43776 BAC (Bacterial artificial chromosome) клонов, и на их основе получили 87552 BES (BAC-end sequence) последовательностей [14]. Полученная

карта состояла из 416 контигов, охватывающих ~368 млн п. н., один контиг представлял хлоропластный геном [14]. Среди известных повторяющихся фракций, гены, кодирующие рибосомальную РНК, составляли ~13,8 %, а мобильные генетические элементы — 6,1 %. Функции 45,1 % последовательностей остались неизвестными [14].

Последовательность генома льна у сорта CDC Bethune была секвенирована *de novo* в 2012 г. методом дробовика (whole-genome shotgun, WGS) [15]. Полученные прочтения были собраны в 116602 контига (302 млн п. н.), из которых были выделены 88384 скаффолда общим размером 318 млн п. н., что соответствует ~81 % генома льна. На полученную последовательность генома картировали более 96 % EST (Expressed Sequence Tags) льна, полученных из базы данных Национального центра биотехнологической информации США (National Center for Biotechnology Information, NCBI). Было показано, что пятая часть генома представлена транспозонами, а доля генов микроРНК составляет менее 1 %. В основной фракции авторами было выявлено 43484 кодирующих гена [15]. В процессе сборки выявлен 31 скаффолд, содержащий хлоропластную ДНК, однако соединить их в единую молекулу хлоропластного генома не удалось [15].

С помощью комплекса методов прямой и обратной генетики можно выявлять индивидуальную изменчивость организма и связывать ее с отличиями в проявлении хозяйственно ценных признаков. Одним из таких методов является оптическое картирование, основанное на непосредственной идентификации сайтов рестрикции в молекулах ДНК при помощи лазерной конфокальной микроскопии. Совокупный спектр полученных фрагментов ДНК служит уникальным «штрихкодом» для исходной последовательности [16]. Оптическая карта генома (BioNano genome (BNG) optical map) льна с использованием BioNano IRIS platform позволила улучшить качество сборки и пополнить последовательность генома сорта CDC Bethune, созданную на основе физической и генетической карты [13, 14, 15]. Было проведено выравнивание 211 контигов BNG (298,6 млн п. н., 94,2 %) и 622 скаффолдов (286,6 млн п. н., 94,9 %), полученных Z. Wang et al. [15]. На основе уточненных скаффолдов, полученных F.M. You et al. [13],

и подтвержденной физической карты “CDC Bethune”, 211 BNG контигов были объединены в 94 суперконтига, которые затем были соотнесены с картами сцепления пятнадцати хромосом *L. usitatissimum*. Размер псевдомолекул составил суммарно 316 млн п. н., что позволило охватить 97 % аннотированных генов льна [13].

Секвенирование, сборку и детальный анализ пластидного генома *L. usitatissimum* произвела группа A. de Santana Lopes et al. [17]. Пластидный геном представил собой кольцевую молекулу ДНК размером 156721 п. н. с типичной структурой, состоящей из четырех частей: два инвертированных повтора по 31990 п. н., между которыми располагаются большая (large single copy, LSC) и малая (small single copy, SSC) уникальные области, включающие 81767 п. н. и 10974 п. н. соответственно. В геноме идентифицировано 109 уникальных генов и 2 псевдогена, и показана утрата консервативных интронов гена *clpP* и полной последовательности гена *rps16* [17].

Таким образом, в результате проведенных многочисленными авторами исследований, полученный геном *L. usitatissimum* может быть использован как основа для быстрой и эффективной сборки геномов других представителей данного вида, а также близкородственных видов.

Сведения о сборке генома льна и сопутствующие данные представлены в приложении 2.

## ПРИМЕНЕНИЕ МЕТОДОВ NGS-СЕКВЕНИРОВАНИЯ ДЛЯ АНАЛИЗА ПОЛИМОРФИЗМА ДНК

До использования методов NGS-секвенирования для генотипирования существовало около двух десятков различных способов анализа однонуклеотидного полиморфизма (single nucleotide polymorphism, SNP), но с развитием технологий NGS у исследователей появилась возможность идентифицировать тысячи маркеров одновременно [21]. Выявленные SNP используют для оценки генетического разнообразия, картирования генов и локусов количественных признаков (quantitative trait loci, QTL) для маркер-ориентированной селекции [22–31].

Пиросеквенирование — это один из методов NGS-секвенирования, основанный на детекции освобождающихся пирофосфатов, используемый для выявления SNP [32–35]. В одной из пер-

вых таких работ изучали виды *L. usitatissimum* и *L. bienne*. С применением вычислительного программного конвейера для идентификации *de novo* аллельных вариантов в геноме льна было выявлено 1067 SNP на 713 контигах. Секвенирование по Сэнгеру подтвердило большинство установленных и определило дополнительные SNP, а также однонуклеотидные инсерции и делеции [36].

В 2015 г. L. Galindo-González et al. [24] показали эффективность ионного полупроводникового секвенирования (Ion Torrent sequencing) для выявления мутаций в определенных областях генома для больших популяций льна. Авторы провели скрининг 768 мутантов, полученных обработкой этилметансульфонатом, с использованием чипов. Из 29 потенциальных мутаций, выявленных после NGS-анализа, 16 были подтверждены с помощью секвенирования по Сэнгеру [24].

Развитие технологий NGS с использованием таких методов уменьшения сложности геномов, как секвенирование, связанное с рестрикционными сайтами (restriction site associated DNA sequencing, RAD-seq) [37, 38], секвенирование с уменьшенной репрезентативной библиотекой, основанное на применении эндонуклеаз рестрикции (restriction enzyme-based reduced representation library sequencing, RRL-seq) [23, 39], генотипирование путем секвенирования (Genotyping-By-Sequencing, GBS) [40] и секвенирование локус-специфичных амплифицированных фрагментов (specific locus amplified fragment sequencing, SLAF-seq) [29] позволяет обнаружить множество SNP за короткое время при низких затратах.

S. Kumar et al. [23] провели RRL-seq геномов восьми сортов льна на платформе Illumina (секвенирование синтезом за счет встраивания 3'-модифицированных нуклеотидов с присоединенными флуоресцентными метками) [23]. Было обнаружено 55465 SNP, почти четверть из которых расположена в генах. 4863 SNP, выявленные у сортов CDC Bethune и Macbeth, были подтверждены при GBS растений из популяции рекомбинантных инбредных линий, полученных от их скрещивания. Средняя частота встречаемости полиморфных локусов составила 0,17 на 1000 п. н. [23].

Для поиска новых SNP используют также метод SLAF-seq, который позволяет снизить затраты на проведение анализа и сохранить воз-

возможность эффективного секвенирования с глубоким покрытием [27, 41–46]. SLAF-seq — это оптимизированная версия RAD-seq, специально предназначенная для крупномасштабных экспериментов по генотипированию на основе серии предварительных экспериментов [47]. Этим методом в 2017 г. L. Yi et al. [27] обнаружили 4145 SNP со средним расстоянием между ними 0,64 сМ, которые использовали для построения высоко насыщенной маркерами генетической карты [27].

Важно, что через SLAF-маркеры удается привязать QTL к конкретным скаффолдам, что служит основой для предварительных выводов о близлежащих генах или генах с плейотропным эффектом, которые влияют на хозяйственно ценные признаки льна. Так J. Wu et al. [31] построили высоконасыщенную маркерами карту в сочетании с QTL селекционно значимых признаков. По результатам исследования F2 Diane x NY1 с использованием 2339 маркеров построена генетическая карта (1483,25 сМ), охватившая 15 групп сцепления со средним расстоянием между маркерами 0,63 сМ. Описаны 12 локусов, ассоциированных с хозяйственно ценными признаками льна (общая и техническая высота, семенная продуктивность, содержание волокна в стеблях и продуктивность по волокну) [31].

Быстрое внедрение методов высокопроизводительного генотипирования позволяет создавать несколько генетических карт для одного вида, которые объединяются в единую консенсусную карту, интегрирующую дополнительные маркеры в районы, где находится QTL [49]. Так, в 2013 г. P. Asgarinia et al. [49] при использовании 143 полиморфных маркеров SSR (simple sequence repeat) на основе анализа картирующей популяции NorMan/Linda составили карты всех 15 групп сцепления общей длиной 1241 сМ, провели QTL-анализ устойчивости к мучнистой росе и выявили локусы в группах сцепления 1, 7 и 9, ассоциированные с этими SSR-маркерами. Физическое месторасположение этих локусов в геноме было определено поиском по гомологии в базе данных полногеномного секвенирования, для чего использовали информацию о нуклеотидных последовательностях исходных для их идентификации EST- и BAC-клов [48, 54].

Таким образом, высокопроизводительное секвенирование может быть эффективно использовано для обнаружения большого количества маркеров, а также картирования локусов селекционно-значимых признаков льна с одновременной идентификацией не только их расположения на генетической карте, но и выявлением физического местоположения в геноме.

### ПОЛНОГЕНОМНЫЙ АНАЛИЗ АССОЦИАЦИЙ (GWAS)

Традиционно картирование QTL у растений выполняют с использованием картирующих популяций [50], что имеет ряд ограничений. Наиболее критическое из них — отсутствие аллельного разнообразия, так как популяция происходит от двух родителей, и генетическая изменчивость ограничена родительскими линиями. В последние годы для выявления генов-кандидатов, связанных с селекционно значимыми признаками растений, активно используют полногеномный анализ ассоциаций (genome-wide association studies, GWAS) [51–53]. Данный метод, требующий использования задомо большого количества маркеров и применяемый для анализа широкого спектра генотипов, открывает новые возможности для расшифровки молекулярно-генетических механизмов, контролирующих сложные полигенные признаки. Успех метода сильно зависит от генетической структуры анализируемой выборки, от количества и распределения молекулярных маркеров [54]. GWAS стал более доступным благодаря данным полногеномного секвенирования и созданию на его основе подходов к высокопроизводительному генотипированию.

В последние несколько лет GWAS стал неотъемлемым методом и для выявления новых хозяйственно ценных генов льна [55–58]. Эта работа начата сравнительно недавно, и большинство исследований были в первую очередь направлены на идентификацию локусов, определяющих качество получаемого растительного сырья.

GWAS на основе SLAF-seq является хорошо разработанной технологией для идентификации аллелей, что доказали D. Xie et al. [56, 57], воспользовавшись сочетанием этих методов при генотипировании 224 сортов льна. Секвенирование позволило идентифицировать 584 987 SNP, со средней частотой встречаемости 1 на 1,2 тыс. п. н.

С использованием статистических моделей EMMA (efficient mixed-model association eXpedited), GLM (generalized linear model) и MLM (multilevel modeling) получены данные и исследованы районы локализации значимых маркеров и SNP-локусов, ассоциированных с масличностью, высотой растения, числом ветвей и коробочек, массой 1000 семян, что позволило выявить гены-кандидаты, контролирующие эти признаки [56, 57] (приложение 3).

Важно, что GWAS может быть дополнен поиском в геноме следов положительного отбора (например, GW3S (genome-wide selective sweep scan)), что выражается в снижении частоты нуклеотидных вариаций в участках, граничащих с новой полезной мутацией [60].

F. You et al. [59] описали выявленные с помощью GWAS, дополненного GW3S, участки генов, связанные с продуктивностью и качественным составом масла льна. В работе были изучены и проанализированы с использованием GBS 260 линий, принадлежавших к трем разным двуродительским картирующим популяциям. Было идентифицировано более 500 тыс. SNP. Основываясь на результатах проведенного анализа, авторы выявили 33 QTL для важных хозяйственно ценных признаков льна. Из них 7 связали со сроками созревания, 5 — с высотой растения, а остальные — с жирнокислотным составом масла [59].

В 2018 г. B. Soto-Cerda et al. [55] выявили локусы, влияющие на количество слизи и семенной кожуры семян льна сортов канадской коллекции. Уменьшение этих показателей является желательным для повышения кормовой ценности семян для животноводства. Были идентифицированы 7 локусов, ассоциированных с содержанием слизи в семенах. Определены четыре локуса, связанных с низкой долей фракции семенных оболочек [55] (приложение 3).

GWAS также актуален для анализа устойчивости льна к патогенам и факторам абиотического стресса, имеющим сложный полигенный контроль. Так, L. He et al. [58] провели GWAS для 370 образцов льна канадской коллекции по результатам пятилетних полевых исследований устойчивости к пасмо. В результате GBS было выявлено 258873 SNP, распределенных по всем 15 хромосомам льна. На основании ортологии с генами *A. thaliana*

выявлено два гена-кандидата устойчивости льна к этому патогену [58].

Используя GWAS и данные пятилетних полевых наблюдений, L. He et al. [9] провели поиск локусов, ассоциированных с устойчивостью льна к *S. linicola*. На коллекции 370 образцов в качестве обучающей и тестовой популяции ими была создана высокоэффективная модель предсказания генетической устойчивости льна к данному патогену с точностью прогнозирования 0,92, что является наиболее точным среди всех моделей геномных предсказаний для устойчивости к болезням растений [9].

Таким образом, работы по GWAS на льне указывают на необходимость использования нескольких статистических моделей при идентификации локусов, что способствует поиску генетических маркеров, для дальнейшего использования в селекционных программах.

## ПУТИ К ГЕНОМНОЙ СЕЛЕКЦИИ

Геномная селекция — это метод, позволяющий преодолеть ограничения MAS для количественных признаков. Цель GS состоит в определении генетического потенциала изучаемого объекта вместо определения конкретного QTL [61].

У льна было предпринято несколько попыток использования геномной селекции. Так, в 2016 г. F.M. You et al. [62] с использованием GS провели отбор образцов исходя из их селекционной значимости, предсказанной статистической моделью, основанной на взаимосвязи между фенотипами и генетическими маркерами. Для семенной продуктивности, содержания масла в семенах, иодного числа и содержания линолевой и линоленовой кислот в каждой из трех изученных родительских популяций была проведена оценка точности прогноза и эффективности геномной селекции относительно фенотипического отбора на основе трех прогностических моделей: RR-BLUP (random regression best linear unbiased prediction), BL (Bayesian LASSO) и BRR (Bayesian ridge regression) [62]. Несмотря на то что эмпирические результаты показали, что геномная селекция может повысить урожайность льна в процессе селекции, необходимы дальнейшие исследования для создания оптимальных обучающих популяций и наборов маркеров для применения GS в льноводстве [62].

Таким образом, массовое использование GS еще затруднено, поскольку суммарные затраты на анализ на порядок выше, чем на MAS. Однако с развитием технологий и расширением геномных ресурсов ожидается более широкое применение геномной селекции.

### ИСПОЛЬЗОВАНИЕ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ ДЛЯ РЕШЕНИЯ ВОПРОСОВ ФИЛОГЕНИИ ПРЕДСТАВИТЕЛЕЙ РОДА *LINUM*

Технологии высокопроизводительного секвенирования стали перспективным подходом для анализа немодельных объектов с обнаружением филогенетических взаимосвязей [63]. Среди основных методов выделяют RAD-seq, GBS, реже применяют полногеномное секвенирование [40, 64–66]. Иными словами, филогенетическая реконструкция оперирует данными о SNP, которые теперь технически возможно получить из любых геномов любых видов за короткое время при низких затратах [67, 68].

В исследовании Y.-B. Fu et al. [26] геномная ДНК из 18 образцов льна, принадлежащих к 16 видам четырех секций, была секвенирована с использованием платформы Illumina. В исследовании были выявлены 6143 SNP в хлоропластной, 2673 в митохондриальной и 19562 в ядерной ДНК. Филогенетические деревья, построенные на основании анализа ядерных, пластидных и митохондриальных маркеров, отражающие дивергенцию четырех секций рода *Linum*, совпадают по топологии. Это демонстрирует наличие конгруэнтных филогенетических связей четырех секций внутри рода *Linum*. Также были представлены три главные ветви, выявляющие две основные эволюционные стадии, приведшие в итоге к появлению возделываемого льна (*L. usitatissimum*). Этот вид и его предок (*L. angustifolium*) сформировались как обособленная ветвь, генетически более тесно связанная с *L. decumbens* и *L. grandiflorum* [26].

Высокопроизводительное секвенирование семейств генов мультикопийной рРНК позволило внести дополнения в филогению голубоцветкового льна, а также выявить внутри- и межвидовую дивергенцию последовательностей генов рРНК. На основе данных секвенирования N.L. Bolshva et al. [12] разделили представителей секции *Linum* на 4 ветви: первая — диплоиды

*L. decumbens* и *L. grandiflorum*; вторая — тетраплоиды *L. narbonense*; третья — тетраплоид *L. usitatissimum* и его дикорастущий предок *L. angustifolium*; четвертая — полиплоидный вид *L. marginale*. Предполагают, что *L. usitatissimum* и *L. angustifolium* могли возникнуть в результате гибридизации двух диплоидных видов ( $2n = 16$ ), возможных прародителей современных видов *L. grandiflorum* и *L. decumbens*, или гибридизации предков *L. narbonense* с  $2n = 14$  и диплоидного вида с  $2n = 16$  [12].

Изменчивость по морфологии, размеру и количеству хромосом в кариотипах определяется набором повторяющихся последовательностей, сравнительное исследование которых в геноме льна провели N.L. Bolshva et al. [69]. На основе результатов полногеномного секвенирования с низким покрытием для 12 видов льна, принадлежащих к шести секциям, был проведен биоинформатический анализ, который установил, что геномы отличались количеством повторяющихся фракций, однако их наборы были схожи у всех видов [69]. Ретротранспозоны семейства *Ty3/Gypsy* представляли собой большую часть повторяющихся последовательностей у всех проанализированных видов. Образование форм голубоцветкового льна в эволюции сопровождалось волнами увеличения количества tandemных повторов (сателлитной ДНК) и ретротранспозонов с длинными концевыми повторами (long terminal repeats, LTR). Наименьшее число диспергированных повторов и наибольшее содержание сателлитных ДНК в геномах льна из секции *Linum*, вероятно, являются результатом аллотетраплоидного происхождения растений из этой группы [69].

Таким образом, методами высокопроизводительного секвенирования выявлены новые детали эволюции представителей рода *Linum*, которые можно использовать в исследовании геномов и филогенетических взаимоотношений внутри этой группы растений.

### ТРАНСКРИПТОМНЫЙ АНАЛИЗ

Доступ к обновленной и дополненной последовательности генома льна [13] позволит сделать селекцию более эффективной. Но одного лишь изучения генома льна недостаточно. Требуются обширные данные транскриптомики, протеомики

и метаболомики, поскольку только комбинация всех омиксных данных позволит понять важные физиологические и молекулярные механизмы, уникальные для данного растения. Метод высокопроизводительного секвенирования РНК (RNA-seq) эффективно используют для анализа транскриптомов и обнаружения генов, участвующих в формировании селекционно-значимых признаков, особое внимание уделяют устойчивости к факторам биотического и абиотического стресса [70–72]. Идентификация дифференциально экспрессирующихся генов (ДЭГ) в настоящее время позволяет лучше понять механизмы, лежащие в основе формирования изучаемых признаков, однако технологии секвенирования транскриптома еще не в полной мере используют для изучения льна [73].

Основной идеей транскриптомного эксперимента на растениях является сравнение экспрессии генов в двух контрастных условиях, например в оптимальных и при неблагоприятном воздействии среды. Особо эффективно сравнение в разных условиях набора генотипов — устойчивых и неустойчивых к изучаемому виду стресса. Идентификация среди ДЭГ новых генов устойчивости позволит разработать новые диагностические маркеры для ускорения отбора в процессе селекции.

Среди абиотических стрессов неблагоприятный состав почвы является одним из основных факторов, ограничивающих площади возделывания льна. Неудивительно, что одним из первых исследований, проведенных в этом направлении на льне, считается изучение транскриптома растения в условиях защелачивания и засоления почвы методом RNA-seq, где были проанализированы паттерны экспрессии пяти категорий генов, включая транскрипционные факторы, сигнальные белки трансдукции, фитогормоны, ферменты антиоксидантной защиты и молекулярные транспортеры. Среди ДЭГ были выявлены гены, относящиеся к ключевым регуляторным семействам, вовлеченным в ответ на абиотический стресс, таким как *WRKY*, *MAPKKK* (mitogen-activated protein kinase kinase), *ABA* (abscisic acid), *PrxR* (NADH-peroxiredoxin reductase) и гены, кодирующие белки, входящие в состав ионных каналов [74].

В поисках генов, связанных с осмотическим стрессом у льна, J. Wu et al. [73] провели секвенирование транскриптома проростков, выращен-

ных при нормальных условиях и при стрессовом воздействии, вызванном полиэтиленгликолем 6000. Авторами были аннотированы 2533 гена, из которых 239 определены как ДЭГ. Значительное число генов было отнесено к семействам, кодирующим транскрипционные факторы, такие как NAC (nascent polypeptide-associated complex), LEA (late embryogenesis—abundant), WRKY, ERF (ethylene responsive factors) и bZIP (basic-leucine zipper) [73].

В 2019 г. J. Wu et al. [75] применили высокопроизводительное секвенирование РНК для изучения транскриптома и обнаружения генов солеустойчивости льна. Из 2582 коэкспрессирующихся генов 2482 были аннотированы. Показано, что основные механизмы солеустойчивости льна ассоциированы с гормональной системой, белками светособирающих комплексов и биосинтезом аминокислот [75].

Геном льна высокопластичен, поэтому у некоторых образцов при дисбалансе питательных веществ в почве происходят наследуемые в нескольких поколениях генетические изменения. Такие генотипы называют генотрофами [76]. А.А. Dmitriev et al. [6] при помощи транскриптомного анализа выявили ДЭГ льна при выращивании в условиях несбалансированности элементов питания по сравнению с оптимальными условиями. На основе данных высокопроизводительного секвенирования авторы выбрали 17 генов для дальнейшего анализа методом количественной ПЦР на расширенной выборке образцов. В условиях избыточного питания выявлено изменение экспрессии генов, кодирующих белки семейства WRKY. В условиях недостатка фосфатов изменялась экспрессия генов, кодирующих белки семейства JAZ (jasmonate ZIM-domain), HARBI1 (harbinger transposase-derived nuclease) и ING1 (inhibitor of growth 1) [6].

В 2016 г. А.А. Dmitriev et al. [77] для изучения устойчивости льна к токсичным свойствам алюминия секвенировали 16 транскриптомов четырех образцов льна (двух устойчивых и двух чувствительных к действию алюминия). На основании полученных данных высокопроизводительного секвенирования обнаружены гены с дифференциальной экспрессией, большинство продуктов которых имели транспортную функцию. Значительное повышение экспрессии генов глутатион-S-трансферазы и UDP-гликозилтрансферазы позволило авторам выдвинуть предположение, что именно эти гены

участвуют в защитных механизмах у растений, связанных с устойчивостью к токсичному действию алюминия посредством утилизации активных форм кислорода и модификации клеточной стенки [77] (приложение 3). Позже, A.V. Zyablitsin et al. [78] установили, что в корнях проростков льна экспрессируются три транскрипт-варианта гена *СХЗ*. Данные количественной ПЦР хорошо согласуются с данными, полученными в результате высокопроизводительного секвенирования тех же образцов растений. Таким образом *СХЗ*- $\text{Ca}^{2+}/\text{H}^{+}$ -антипортёр может участвовать в реакции льна на высокую кислотность почвы и высокую концентрацию алюминия в почве посредством  $\text{Ca}^{2+}$ -опосредованной внутриклеточной регуляции [78] (приложение 3).

Также А.А. Dmitriev et al. [79] с помощью транскриптомного анализа изучили реакцию сортов льна на пониженную кислотность и недостаток цинка. У изучаемых сортов льна, выращенных при pH 5,5, выявлены более значительное ингибирование роста и изменения экспрессии генов по сравнению с контрольной группой и группой, выращенной в условиях дефицита цинка. Выявленные авторами ДЭГ участвуют в различных процессах, включая ионный транспорт, биогенез клеточной стенки, активность оксидоредуктаз и фотосинтез, и могут играть важную роль в реакции льна на изученные стрессы [79].

Немаловажным является изучение устойчивости растений льна к биотическим стрессам. В 2016 г. L. Galindo-González et al. [80] одними из первых применили транскриптомный анализ для изучения механизмов устойчивости к *Fusarium oxysporum*. Умеренно устойчивый к фузариозному увяданию сорт CDC Bethune использовали для полного исследования транскриптома методом RNA-seq. Результаты во многом подтвердили известные механизмы реакции льна на патоген и дополнили существующие данные, что позволило авторам построить обновленную и наиболее полную модель устойчивости льна к *F. oxysporum* [80] (приложение 3).

Реакцию на заражение устойчивых к фузариозному увяданию гибридов льна изучали А.А. Dmitriev et al. [81]. Для идентификации генов-кандидатов устойчивости к фузариозу у культуры поиск вели по генам с повышенной или пониженной экспрессией в устойчивых (по сравнению с восприимчивыми) сортах и популяциях при инфицировании. Среди ДЭГ особое внимание было уделено

генам, кодирующим белки SRG (salt response protein), UDP-гликозилтрансферазу, АТФазы, связанные с разнообразными клеточными активностями (AAA-ATPase), глюкан-эндо-1,3-бета-глюкозидазу, факторы транскрипции MYB, дегидрины и ауксин-чувствительный белок SAUR. Авторы предположили, что идентифицированные гены со специфически индуцированной экспрессией в ответ на заражение *F. oxysporum* у устойчивых сортов и популяций являются наиболее перспективными генами-кандидатами устойчивости [81].

Еще одно направление в области транскриптомики растений связано с изучением молекулярно-генетических механизмов онтогенеза, для чего проводят сравнение транскрипции генов в разных частях растений и на разных стадиях дифференцировки. Такие исследования помогают понять известные и выявлять новые молекулярно-генетические механизмы эволюционных и онтогенетических изменений, а также идентифицировать потенциальные гены-мишени для направленного мутагенеза с целью изменения типа роста и развития, ускорения цветения, созревания и т. д.

RNA-seq использовали для сравнения транскриптов в двух сегментах вегетативного побега 14-дневного растения льна, у которого были удалены все листья [82]. Выявленные различия в экспрессии между апикальной областью побега и большей частью стебля будут дополнять другие подходы (например, скрининг мутантов), направленные на изучение первичной дифференцировки волокон флоэмы [82].

В 2017 г. N. Mokshina et al. [83] исследовали экспрессию генов комплекса целлюлозосинтазы у льна (*LusCESA*) на разных стадиях развития и в процессе гравитропической реакции растения с помощью RNA-seq и количественной ПЦР. Гравитимуляция растений временно увеличивала количество транскриптов целлюлозосинтазного комплекса. Таким образом, волокна внутренней (третичной) клеточной стенки имеют специфический механизм биосинтеза целлюлозы и особенности его регуляции [83].

Используя данные транскриптомного анализа, S.M. Shivaraj et al. [84] изучили гены интегральных мембранных белков, в большинстве представленных аквапоринами. Анализ данных RNA-seq и микрочипов льна показал высокую экспрессию

белков плазматической мембраны во множественных тканях, низкую экспрессию интегральных нодулин-26-подобных белков и специфическую для семян экспрессию 17 интегральных белков тонопласта. Столь масштабный анализ этих белков позволил лучше понять их физиологические функции и роль в развитии льна [84].

Р.К. Dash et al. [85] осуществили сборку транскриптома умеренно устойчивого к засухе индийского сорта льна T-397. Полученные данные помогут идентифицировать локусы и создать маркеры SSR для селекции на засухоустойчивость [85].

Для интрузивно растущих волокон, полученных методом лазерной микродиссекции, Т. Gorshkova et al. [86] провели RNA-seq первичной флоэмы льна. Так как изменения в уровнях мРНК отдельных генов напрямую не связаны с синтезом белка и, особенно, с активностью фермента, использовано профилирование транскриптома в качестве высокоинформативного подхода для выявления метаболических путей и белков, играющих ключевую роль при интрузивном росте клеток. Выявлено, что удлинение клеток сопряжено с активацией фотосинтеза и интенсивной экспрессией локализованных в хлоропластах шаперонинов и тиоредоксинов. Также установлено, что высокая доля специфически активированных генов кодирует экспансины, ферменты для модификации пектинов и несколько белков, локализованных в клеточной стенке [86].

Имеющиеся данные об исследовании транскриптома льна представлены в базах данных NCBI Sequence Read Archive [88], NCBI Gene Expression Omnibus [73, 82, 85, 86, 88].

Следует отметить, что в области транскриптомики льна больше работ по изучению его устойчивости к абиотическим факторам, чем к биотическим. Вероятно, это связано с селекционными программами и целевыми характеристиками растений для конкретного региона. Но все это оставляет возможность для дальнейших перспективных исследований ДЭГ при разных видах стресса, а также в процессе онтогенеза. В конечном счете, совокупность данных транскриптомного анализа позволит предоставить справочную карту для обнаружения дополнительных генов устойчивости, полезных для использования в программах по MAS льна или для его улучшения при помощи генетического редактирования.

## ИСПОЛЬЗОВАНИЕ МЕТОДОВ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ В ИССЛЕДОВАНИИ МИКРОРНК

МикроРНК (miRNA) — это класс малых некодирующих РНК, обнаруженных у различных организмов, включая растения. Они состоят примерно из 20–24 нуклеотидов и могут контролировать многочисленные биологические процессы посредством негативной регуляции экспрессии генов путем специфического связывания и ингибирования мРНК-мишеней [89]. Уровень микроРНК изменяется в условиях стресса, вызванного такими факторами, как засуха, гипоксия, холод, засоленные почвы, наличие тяжелых металлов и высокий или низкий уровень питательных веществ [90–95].

Для изучения микроРНК в растениях часто используют высокопроизводительное секвенирование. N.V. Melnikova et al. [93] идентифицировали 96 консервативных гомологов микроРНК из 21 семейства, 12 из которых были впервые обнаружены. Самое распространенное семейство miR165/166 у льна, которое, по предположению авторов, активно участвует в регуляции экспрессии генов, играет роль в координации метаболизма растения во время стресса, а также в специализации растительных тканей. Экспрессия 7 микроРНК (miR168, miR169, miR395, miR398, miR399, miR408 и lus-miR-N1) в условиях дефицита питательных веществ на основании данных высокопроизводительного секвенирования была оценена на расширенном числе образцов с использованием ПЦР в реальном времени. В условиях дефицита фосфата отмечено изменение экспрессии lus-miR-N1 и miR399. Выявили отрицательную корреляцию экспрессии lus-miR-N1 и его прогнозируемой мишени, гена убиквитин-активирующего фермента E1, а также miR399 и его прогнозируемой мишени — гена убиквитин-конъюгирующего фермента E2 [93, 96].

В ответ на стресс, вызванный токсическим действием ионов алюминия у льна, А.А. Dmitriev et al. [95] отметили изменение экспрессии miR319, miR390 и miR393. Возможно, они играют важную роль в стрессовых реакциях растения посредством регуляции процессов их роста [95].

Y. Yu et al. в 2016 г. [94] исследовали профиль микроРНК льна в условиях повышенного содержания солей и щелочей в почве с использованием высо-

копроизводительного секвенирования. Были идентифицированы 124 известных микроРНК, принадлежащих к 23 консервативным семействам, и 394 новых микроРНК. После оценки профилей дифференциальной экспрессии были отобраны 17 известных *lus-miRNA* и 36 новых *lus-miRNA*, которые использовали для прогнозирования генов-мишеней. Параллельный анализ концевых прочтений РНК и профилирование транскриптома показали изменения в экспрессии 29 пар микроРНК при изучаемых стрессовых условиях. Выдвинуто предположение, что ген-мишень *miR398*, кодирующий супероксиддисмутазу, и *miR530*, мишенью которой являются транскрипционные факторы семейства WRK, могут играть значительную роль в стрессоустойчивости льнов [94].

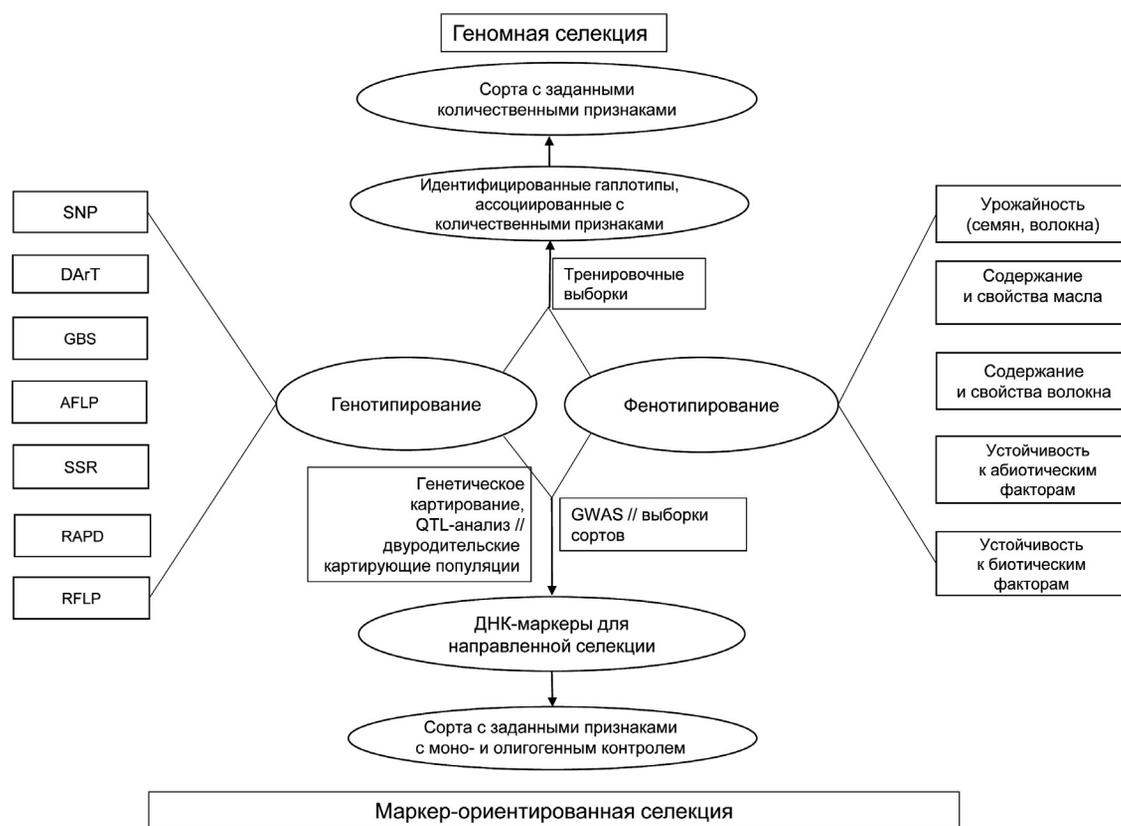
Данные по микроРНК льна депонируются в базе данных miRBase. Ресурс содержит репозиторий, в котором представлены в том числе и последо-

вательности 124 микроРНК *L. usitatissimum* [97]. Для них приведены первичная и вторичная структуры, локализация в геноме льна сборки v1.0.

Исходя из опубликованных данных, наибольший интерес представляет изучение устойчивости к абиотическим факторам. Очевидно, что исследование микроРНК и их функций у льна все еще находится на начальных этапах, и ряд ключевых вопросов остается без ответа. Накопление знаний о механизмах микроРНК-регуляции у растений позволит разработать искусственные микроРНК для использования в качестве эффективных инструментов для контроля экспрессии генов.

### ЗАКЛЮЧЕНИЕ

Использование современных подходов, дополняющих методы традиционной селекции, способствует более эффективному созданию сортов льна с улучшенными свойствами (рис. 2.). Разработка новых



**Рис. 2.** Схематическое представление современных селекционных подходов, основанных на использовании комбинирования подходов для маркер-ориентированной и геномной селекции. SNP — single nucleotide polymorphism (однонуклеотидный полиморфизм), DArT — diversity arrays technology (ДНК-чип технология для изучения разнообразия), GBS — Genotyping by Sequencing (генотипирование путем секвенирования), AFLP — amplified fragment length polymorphism (полиморфизм длины амплифицированных фрагментов), SSR — simple sequence repeat (однонуклеотидный полиморфизм), RAPD — random amplification of polymorphic DNA (случайно амплифицированная полиморфная ДНК), RFLP — restriction fragment length polymorphism (полиморфизм длины рестрикционных фрагментов), QTL — quantitative trait loci (локус количественных признаков), GWAS — genome-wide association studies (полногеномный анализ ассоциаций)

генетических маркеров для селекции и выявления новых генов-мишеней для генетического редактирования, несомненно, будет содействовать расширению работ, основанных на применении методов высокопроизводительного секвенирования в количественной генетике, транскриптомике льна. Получаемые в этом направлении результаты также представляют фундаментальный интерес, ведущий к пониманию особенностей филогении рода *Linum*, эволюции геномов представителей этого рода, механизмов, лежащих

в основе онтогенетического развития растения и фенотипической изменчивости, а также разнообразия реакций на различные стимулы окружающей среды.

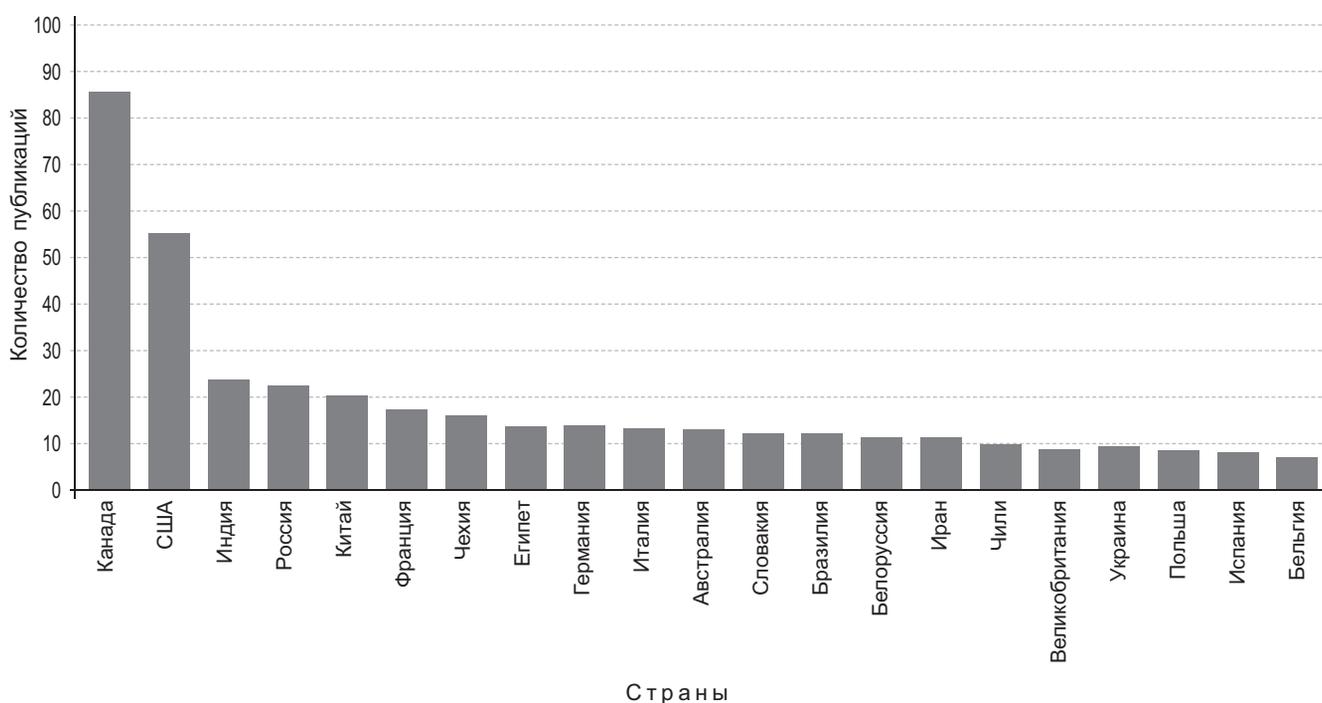
### Благодарности

Данный обзор был проведен в рамках государственного задания по теме № 0481-2019-0001.

*Конфликт интересов.* Авторы заявляют об отсутствии конфликта интересов.

### Приложение 1

Использование ДНК-маркеров льна в исследованиях в разных странах. Составлено на основе поиска публикаций в базе данных Scopus ([www.scopus.com](http://www.scopus.com), дата обращения 01.02.2019) по пересечению ключевых слов «Marker» и «Linum»



### Приложение 2

#### Электронные ресурсы, содержащие данные полногеномного секвенирования льна

Результаты секвенирования и сборки генома, а также последние данные о сборке псевдомолекул *L. usitatissimum*, доступны в базе нуклеотидных последовательностей NCBI GenBank под идентификатором GenomeProject #68161 [18]. Номера доступа от CP027619 до CP027633 для каждой из 15 хромосом. Секвенированные последовательности, полученные на основе фосмидных и ВАС-библиотек, были представлены в GenBank под номерами HQ902252, JN133299-JN133301 и JX174444-JX174449.

Последовательность аннотированного генома и информация о кодирующих генах также доступны в базе данных Phytozome [13, 15, 19]. Характеристики сборки: общая длина собранных последовательностей ~318,3 млн п. н., количество скаффолдов — 88420,

количество контигов — 110390. Аннотированный геном содержит 43471 гена и 43484 транскрипта. Кроме того, в базе представлены EST из NCBI GenBank, картированные на геном при помощи PASA (the program to assembly spliced alignments) [15].

Некоторые данные о геномах видов льна можно найти в Европейском архиве нуклеотидных последовательностей (ENA) [20]. Например, здесь представлены необработанные парные прочтения, полученные в результате секвенирования методом shotgun таких диких видов льна, как *L. leonii* (SRR1592650), *L. lewissii* (SRR1592654), *L. perenne* (SRR1592548), *L. narbonne* (SRR1592545), *L. grandiflorum* (SRR1592647), *L. decumbens* (SRR1592610) и *L. angustifolium* (SRR1592607).

**Гены-кандидаты, ассоциированные с хозяйственно ценными признаками льна**

Признак	Ген-кандидат, ассоциированный с признаком	Кодируемый продукт	Интерпретация ассоциации	Использованный метод	Литературная ссылка
Содержание пальмитиновой кислоты	<i>PIP5K (Lus10022606)</i> ( <i>PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE</i> )	Фосфатидилинозитол-4-фосфат-5-киназа	PIP5K участвует в метаболизме инозитолфосфата, в результате гидролиза которого образуются предшественники пальмитиновой кислоты	GWAS	[56]
Высота растения	<i>ABC (Lus10016125)</i> ( <i>ATP-BINDING CASSETTE</i> )	АТФ-связывающие кассетные транспортеры	Обеспечивают транспорт липидов, сахаров, аминокислот и др., таким образом играют важную роль в развитии и росте растений.	GWAS	[56]
	<i>UGT (UDP-GLYCOSYLTRANSFERASE)</i>	UDP-гликозилтрансфераза	Избыточная экспрессия <i>UGT84B1</i> и <i>UGT74E2</i> у арабидопсиса приводит к образованию короткого стебля с большим количеством ветвей. Влияет на развитие растения, участвуя в метаболизме ауксина.	GWAS	[57]
	<i>PL (PECTATE LYASE)</i>	Пектатлиаза	У риса и арабидопсиса ген <i>PL</i> тесно связан с ростом и развитием растений за счет регуляции скорости деления клеток и участия модификации клеточной стенки	GWAS	[57]
Содержание волокна	<i>Lus10016354</i>	Ксантоксиндегидрогеназа	Участвует в синтезе абсцизовой кислоты (АБК). Ранее сообщалось о влиянии АБК на снижение синтеза гемицеллюлозы и целлюлозы.	GWAS	[56]
Содержание стеариновой кислоты	<i>Lus10021171</i>	Протеинфосфатаза	У льна с пониженной экспрессией гена наблюдается повышенное содержание стеариновой кислоты	GWAS, RNA-seq	[56]
Масса 1000 семян	<i>PHO1 (PHOSPHATE PERMEASE)</i>	Фосфатпермеаза	Участвует в передаче фосфора от корней к надземным частям растений. Увеличение потребляемого фосфора увеличивает размер семян	GWAS	[56]
Ветвистость стебля растения	<i>GRAS (GIBBERELLIC ACID INSENSITIVE + REPRESSOR OF GA1-3 + SCARECROW)</i>	Транскрипционные факторы семейства GRAS (GIBBERELLIC ACID INSENSITIVE + REPRESSOR OF GA1-3 + SCARECROW)	Играют ключевую роль в развитии и передаче сигналов. LS и MOC1 — члены семейства белков GRAS. Отсутствие экспрессии гена <i>LS</i> ингибирует образование пазушной меристемы и уменьшает количество пазушных почек. Отсутствие экспрессии гена <i>MOC1</i> приводит к почти полному отсутствию кущения у риса ( <i>O. sativa</i> ), так как продукт гена участвует регуляции клеточного цикла.	GWAS	[57]
	<i>XTH (XYLOGLUCAN ENDOTRANSGLUCOSYLASE / HYDROLASE)</i>	Ксилоглюкан эндотрансглюкозилаза/ гидролаза	У <i>A. thaliana</i> ген <i>XTH9</i> экспрессируется в апикальной меристеме цветочных почек и цветоножек и связан с удлинением этих частей растения; уменьшение экспрессии приводит к короткой длине междоузлия	GWAS	[57]

Продолжение приложения 3

Признак	Ген-кандидат, ассоциированный с признаком	Кодируемый продукт	Интерпретация ассоциации	Использованный метод	Литературная ссылка
Содержание слизи в семени	<i>GATL5 (Lus10009311)</i> ( <i>GALACTUROSYL TRANSFERASE-LIKE 5</i> )	Белок, подобный галактозилтрансферазе 5	Участвует в синтезе остова рамно-галактуронана I.	GWAS	[55]
	<i>MUM4 (Lus10009288)</i> ( <i>MUCILAGE-MODIFIED 4</i> )	UDP-L-рамноз-синтаза	Необходим для производства рамнозы — ключевого субстрата для биосинтеза слизи.	GWAS	[55]
	<i>PME36 (Lus10009287)</i> ( <i>PECTIN METHYLESTERASE 36</i> )	Метилэстераза пектина 36	Участвует в синтезе пектина и модификациях клеточной стенки.	GWAS	[55]
	<i>SBT1.7 (Lus10009313, Lus10007083)</i> ( <i>SUBTILISIN-LIKE SERINE PROTEASE</i> )	Субтилизинподобная сериновая протеаза	Запускает активацию ферментов, модифицирующих клеточную стенку для выделения слизи.	GWAS	[55]
	<i>TT8 (Lus10007101)</i> ( <i>TRANSPARENT TESTA 8</i> )	Семейство транскрипционных факторов TT	Наряду с транскрипционным фактором <i>GLABRA2 (GL2)</i> регулирует экспрессию гена <i>MUM4</i>	GWAS	[55]
Содержание семенной кожуры	<i>AGL62 (Lus10035456)</i> ( <i>AGAMOUS-LIKE MADS-BOX PROTEIN</i> )	Семейство транскрипционных факторов AGL	Мутанты по <i>AGL62</i> могут инициировать образование зародыша и эндосперма, но не в состоянии сформировать семенную оболочку.	GWAS	[55]
	<i>GH17 (Lus10018306)</i> ( <i>GLYCOSYL HYDROLASE FAMILY 17</i> )	Гликозилгидролаза 17	Экспрессируется совместно с <i>TT12</i> , <i>ANA10</i> и <i>BAN</i> , которые могут модифицировать гликозильированные флаван-3-ол мономеры, что приводит к накоплению проантоцианидинов в оболочке семян.	GWAS	[55]
	<i>UGT79B1 (Lus10026926)</i> ( <i>UDP-GLUCOSE FLAVONOL 3-O-GLUCOSYLTRANSFERASE</i> )	UDP-глюкоза: флавоноид 3-О-глюкозилтрансфераза	Ключевой фермент, катализирует конечную стадию биосинтеза антоцианинов	GWAS	[55]
Устойчивость к высокой кислотности почвы	<i>CAX3 (CA<sup>2+</sup>/H<sup>+</sup>-EXCHANGER)</i>	Ca <sup>2+</sup> /H <sup>+</sup> -антипортер	Участвует в регуляции роста растений и усвоения питательных веществ, а также в регуляции транспорта фосфатов. <i>CAX3</i> находится в тонопласте. <i>CAX3</i> мутанты <i>A. thaliana</i> чувствительны к стрессу, вызванному повышенной соленостью и кислотностью. <i>CAX1</i> и <i>CAX3</i> участвуют в регуляции pH в апопласте	RNA-seq	[78]

Продолжение приложения 3

Признак	Ген-кандидат, ассоциированный с признаком	Кодируемый продукт	Интерпретация ассоциации	Использованный метод	Литературная ссылка
Устойчивость к токсичному действию алюминия	<i>GST (GLUTATHIONE S-TRANSFERASE)</i>	Глутатион-S-трансфераза	Участвует в реакции на стресс, включая окислительный, и может действовать как глутатион-зависимая пероксидаза. Воздействие алюминия на растения в условиях пониженного pH приводит к образованию активных форм кислорода и перекисному окислению липидов. В ответ на повышенное содержание алюминия увеличивается экспрессия гена <i>GST</i> как у устойчивых, так и у чувствительных к токсичному действию этого вещества у кукурузы, арабидопсиса и др. Наблюдалось повышение экспрессии гена <i>GST</i> у льна при Al-стрессе, особенно у устойчивых сортов.	RNA-seq	[77]
	<i>CAX3 (CA<sup>2+</sup>/H<sup>+</sup>-EXC)</i>	Ca <sup>2+</sup> /H <sup>+</sup> -антипортер	Выявлена повышенная регуляция гена <i>CAX3</i> у льна, толерантного к алюминию. Активация гена <i>CAX3</i> способствует компарментализации Ca <sup>2+</sup> в вакуоли. Это может быть механизмом, обеспечивающим устойчивость льна к стрессу, вызванному повышенными концентрациями алюминия.	RNA-seq	[78]
	<i>UGT (UDP-GLYCOSYLTRANSFERASE)</i>	UDP-гликозилтрансфераза	Участвуют в биосинтезе вторичных метаболитов, гормональном гомеостазе, детоксикации ксенобиотиков и в реакции растений на стресс. Изменения в экспрессии <i>UGT</i> при воздействии повышенных концентраций алюминия наблюдались на рисе, кукурузе, льне и других культурах. Участвуют в биосинтезе лигнанов и детоксикации клеток от вторичных метаболитов реактивных форм кислорода, что может быть механизмом устойчивости льна к Al-стрессу	RNA-seq	[77]
Устойчивость к <i>Fusarium oxysporum</i>	<i>WRKY3, WRKY70, WRKY75</i>	Транскрипционные факторы семейства WRKY	Сверхэкспрессия этих генов у <i>A. thaliana</i> наблюдается при биотическом стрессе. Транскрипция ортологов <i>WRKY75</i> , усиливается у <i>Brassica napus</i> после заражения <i>Sclerotinia sclerotiorum</i> и <i>Alternaria brassicae</i> .	RNA-seq	[80]
	<i>MYB113, MYB108</i>	Транскрипционные факторы семейства MYB	Ранее показано, что у <i>A. thaliana</i> <i>MYB113</i> индуцируется при инокуляции <i>F.oxysporum</i> и имеет решающее значение для производства антоцианов, которые включают специфические стадии метаболизма фенилпропанондов	RNA-seq	[80]

Окончание приложения 3

Признак	Ген-кандидат, ассоциированный с признаком	Кодируемый продукт	Интерпретация ассоциации	Использованный метод	Литературная ссылка
Устойчивость к <i>Fusarium oxysporum</i>	<i>ERF1</i> , <i>ERF14</i> ( <i>ETHYLENE RESPONSE FACTOR</i> )	Семейство транскрипционных факторов, отвечающих на этилен	Наблюдается повышенная экспрессия генов из семейства <i>ERF</i> у зараженных растений. При инокуляции растений <i>Fusarium</i> наблюдалась активация генов <i>ERF</i> . Ген <i>ERF1</i> , вероятно, является одним из наиболее важных генов, участвующих в защите растений от грибковых патогенов, и связан с устойчивостью арабидопсиса к <i>F. oxysporum</i> sp. <i>conglutinans</i> и <i>F. oxysporum</i> sp. <i>lycopersici</i> .	RNA-seq	[80]
	<i>CYP79B2</i> , <i>CYP79B3</i> ( <i>CYTOCHROME P450</i> )	Цитохром P450	CYP450 способствует превращению триптофана в индол-3-ацетальдоксим (предшественник индол-3-уксусной кислоты — ауксина). Активация пула ауксина имеет решающее значение для прорастания и роста растений, что в свою очередь так же может сказать на защите от патогенов.	RNA-seq	[80]
	<i>PRX52</i> ( <i>PEROXIDASE 52</i> )	Пероксидаза	Участвует в образовании лигнина. Наблюдали усиление экспрессии гена более чем в 40 раз при инфицировании арабидопсиса <i>Verticillium longisporum</i> .	RNA-seq	[80]
	<i>CHS</i> ( <i>CHALCONE SYNTHASE</i> )	Халконсинтаза	Ключевой фермент биосинтеза флавоноидов, которые обладают высокой антиоксидантной способностью. Это используется для создания форм с повышенной устойчивости к <i>F. oxysporum</i> и <i>F. culmorum</i> посредством конструирования трансгенного льна с мультиконструкцией, включающей халконсинтазу ( <i>CHS</i> ), халконизомеразу ( <i>CHI</i> ) и дигидрофлавонол редуктазу ( <i>DFR</i> ) петунии.	RNA-seq	[80]
	<i>DFR</i> ( <i>DIHYDROFLAVONOL REDUCTASE</i> )	Дигидрофлаванол редуктаза	Один из ключевых ферментов синтеза антоцианов. Используется, как и <i>CHS</i> .	RNA-seq	[80]
<i>RIPK</i> ( <i>RPMI-INDUCED PROTEIN KINASE</i> )	Протеинкиназа индуцируемая RPM-1	Ген является ортологом гена <i>A. thaliana</i> <i>ATG05940</i> , который кодирует протеинкиназу, активирующую эффекторный иммунитет	RNA-seq	[80]	

## ЛИТЕРАТУРА

1. Кутузова С.Н., Чухина И.Г. Уточнение внутривидовой классификации культурного льна (*Linum usitatissimum* L.) методами классической генетики //

Труды по прикладной ботанике, генетике и селекции. — 2017. — Т. 178. — № 3. — С. 97-109. [Kutuzova SN, Chukhina IG. Updating the intraspecific classification of cultivated flax (*Linum usitatissimum* L.) using

- methods of classical genetics. *Trudy po prikladnoy botanike, genetike i seleksii*. 2017;178(3):97-109. (In Russ.]. <https://doi.org/10.30901/2227-8834-2017-3-97-109>.
2. FAOSTAT [Internet]. Food and Agriculture Organization of the United Nation; 2017. [cited 2019 May 14] Available from: <http://www.fao.org/faostat/ru/#data/QC>.
  3. Лукомец В.М., Пивень В.Т., Тишков Н.М. Интегрированный подход к защите посевов льна масличного от вредных организмов // Защита и карантин растений. — 2010. — № 5. — С. 52-56. [Lukomets VM, Piven' VT, Tishkov NM. Integrirovannyy podkhod k zashchite posevov l'na maslichnogo ot vrednykh organizmov. *Zashchita i karantin rasteniy*. 2010;(5): 52-56. (In Russ.)]
  4. Рябенко Л.Г., Мошненко Е.В., Зеленцов В.С., и др. Выделение морозоустойчивых форм льна с пониженной фотопериодической чувствительностью // Междунар. научно-практ. конф. «Инновационные исследования и разработки для научного обеспечения производства и хранения экологически безопасной сельскохозяйственной и пищевой продукции». — Краснодар, 2015. — 463 с. [Ryabenko LG, Moshnenko YeV, Zelentsov VS, et al. Vydeleniye morozoustoychivyykh form l'na s ponizhennoy fotoperiodicheskoy chuvstvitel'nost'yu. (Conference proceedings) Mezhdunar. nauchno-prakt. konf. "Innovatsionnyye issledovaniya i razrabotki dlya nauchnogo obespecheniya proizvodstva i khraneniya ekologicheskii bezopasnoy sel'sko-khozyaystvennoy i pishchevoy produktsii". Krasnodar; 2015. 463 p. (In Russ.)]
  5. Кишлян Н.В., Рожмина Т.А., Кудрявцева Л.П., Киселева Т.С. Изучение сортообразцов в коллекции масличного льна на устойчивость к почвенной кислотности // Масличные культуры. — 2010. — № 2. — С. 107–112. [Kishlyan NV, Rozhmina TA, Kudryavtseva LP, Kisilyova TS. Studying of oil flax varieties in collection on resistance to soil acidity. *Maslichnyye kul'tury*. 2010;(2):107-122. (In Russ.)]
  6. Dmitriev AA, Kudryavtseva AV, Krasnov GS, et al. Gene expression profiling of flax (*Linum usitatissimum* L.) under edaphic stress. *BMC Plant Biol*. 2016;16(S3):237. <https://doi.org/10.1186/s12870-016-0927-9>.
  7. Sauer NJ, Narváez-Vásquez J, Mozoruk J, et al. Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiol*. 2016;170(4):1917-1928. <https://doi.org/10.1104/pp.15.01696>.
  8. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics*. 2010;9(2):166-177. <https://doi.org/10.1093/bfpg/elq001>.
  9. He L, Xiao J, Rashid K, et al. Evaluation of genomic prediction for pasmo resistance in flax. *Int J Mol Sci*. 2019;20(2):359. <https://doi.org/10.3390/ijms20020359>.
  10. Rogers CM. The systematics of *Linum* sect. *Linopsis* (*Linaceae*). *Plant Syst Evol*. 1982;140(2-3):225-234. <https://doi.org/10.1007/BF02407299>.
  11. Muravenko OV, Yurkevich OY, Bolsheva NL, et al. Comparison of genomes of eight species of sections *Linum* and *Adenolinum* from the genus *Linum* based on chromosome banding, molecular markers and RAPD analysis. *Genetica*. 2009;135(2):245-255. <https://doi.org/10.1007/s10709-008-9273-7>.
  12. Bolsheva NL, Melnikova NV, Kirov IV, et al. Evolution of blue-flowered species of genus *Linum* based on high-throughput sequencing of ribosomal RNA genes. *BMC Evol Biol*. 2017;17(S2):253. <https://doi.org/10.1186/s12862-017-1105-x>.
  13. You FM, Xiao J, Li P, et al. Chromosome-scale pseudomolecules refined by optical, physical and genetic maps in flax. *Plant J*. 2018;95(2):371-384. <https://doi.org/10.1111/tpj.13944>.
  14. Ragupathy R, Rathinavelu R, Cloutier S. Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. *BMC Genom*. 2011;12(1):217. <https://doi.org/10.1186/1471-2164-12-217>.
  15. Wang Z, Hobson N, Galindo L, et al. The genome of flax (*Linum usitatissimum* L.) assembled *de novo* from short shotgun sequence reads. *Plant J*. 2012;72(3):461-473. <https://doi.org/10.1111/j.1365-313X.2012.05093.x>.
  16. Aston C, Mishra B, Schwartz DC. Optical mapping and its potential for large-scale sequencing projects. *Trends Biotechnol*. 1999;17(7):297-302. [https://doi.org/10.1016/S0167-7799\(99\)01326-8](https://doi.org/10.1016/S0167-7799(99)01326-8).
  17. de Santana Lopes A, Pacheco TG, Santos KG, et al. The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep*. 2018;37(2):307-328. <https://doi.org/10.1007/s00299-017-2231-z>.

18. BioProject [Internet]. Bethesda (MD): National Library of Medicine (US). *Linum usitatissimum* CDC Bethune genome sequencing project [cited 2019 Jun 12]. Available from: <https://www.ncbi.nlm.nih.gov/bioproject/68161>.
19. Phytozome [Internet]. The Regents of the University of California; 1997-2017 [cited 2019 Apr 23]. Available from: [https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Lusitatissimum](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Lusitatissimum).
20. EMBL-EBI [Internet]. Hinxton: European Molecular Biology Laboratory(UK) [cited 2019 Jun 18]. Available from: <https://www.ebi.ac.uk/>.
21. Khlestkina EK, Salina EA. [SNP markers: methods of analysis, ways of development, and comparison on an example of common wheat (In Russ.)]. *Genetika*. 2006;42(6):725-736. <https://doi.org/10.1134/s1022795406060019>.
22. Cortés AJ, Chavarro MC, Blair MW. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet*. 2011;123(5):827-845. <https://doi.org/10.1007/s00122-011-1630-8>.
23. Kumar S, You FM, Cloutier S. Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. *BMC Genom*. 2012;13(1):684. <https://doi.org/10.1186/1471-2164-13-684>.
24. Galindo-González L, Pinzón-Latorre D, Bergen EA, et al. Ion torrent sequencing as a tool for mutation discovery in the flax (*Linum usitatissimum* L.) genome. *Plant Methods*. 2015;11(1):19. <https://doi.org/10.1186/s13007-015-0062-x>.
25. Peterson G, Dong Y, Horbach C, et al. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity*. 2014;6(4):665-680. <https://doi.org/10.3390/d6040665>.
26. Fu Y-B, Dong Y, Yang MH. Multiplexed shotgun sequencing reveals congruent three-genome phylogenetic signals for four botanical sections of the flax genus *Linum*. *Mol Phylogenetics Evol*. 2016;101:122-132. <https://doi.org/10.1016/j.ympev.2016.05.010>.
27. Yi L, Gao F, Siqin B, et al. Construction of an SNP-based high-density linkage map for flax (*Linum usitatissimum* L.) using specific length amplified fragment sequencing (SLAF-seq) technology. *PLoS One*. 2017;12(12):e0189785. <https://doi.org/10.1371/journal.pone.0189785>.
28. Yang H, Li C, Lam HM, et al. Sequencing consolidates molecular markers with plant breeding practice. *Theor Appl Genet*. 2015;128(5):779-795. <https://doi.org/10.1007/s00122-015-2499-8>.
29. Sun X, Liu D, Zhang X, et al. SLAF-seq: an efficient method of large-scale *de novo* snp discovery and genotyping using high-throughput sequencing. *PLoS One*. 2013;8(3):e58700. <https://doi.org/10.1371/journal.pone.0058700>.
30. Zhang J, Long Y, Wang L, et al. Consensus genetic linkage map construction and QTL mapping for plant height-related traits in linseed flax (*Linum usitatissimum* L.). *BMC Plant Biol*. 2018;18(1):160. <https://doi.org/10.1186/s12870-018-1366-6>.
31. Wu J, Zhao Q, Zhang L, et al. QTL mapping of fiber-related traits based on a high-density genetic map in flax (*Linum usitatissimum* L.). *Front Plant Sci*. 2018;9:1-10. <https://doi.org/10.3389/fpls.2018.00885>.
32. Barbazuk WB, Emrich SJ, Chen HD, et al. SNP discovery via 454 transcriptome sequencing. *Plant J*. 2007;51(5):910-918. <https://doi.org/10.1111/j.1365-313X.2007.03193.x>.
33. Wiedmann RT, Smith TP, Nonneman DJ. SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genet*. 2008;9(1):81. <https://doi.org/10.1186/1471-2156-9-81>.
34. Maughan PJ, Yourstone SM, Jellen EN, et al. SNP discovery via genomic reduction, barcoding, and 454-pyrosequencing in amaranth. *Plant Genome*. 2009;2(3):260. <https://doi.org/10.3835/plantgenome2009.08.0022>.
35. Hyten DL, Song Q, Fickus EW, et al. High-throughput SNP discovery and assay development in common bean. *BMC Genom*. 2010;11:475. <https://doi.org/10.1186/1471-2164-11-475>.
36. Fu YB, Peterson GW. Developing genomic resources in two *Linum* species via 454 pyrosequencing and genomic reduction. *Mol Ecol Resour*. 2012;12(3):492-500. <https://doi.org/10.1111/j.1755-0998.2011.03100.x>.
37. Miller MR, Dunham JP, Amores A, et al. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*. 2007;17(2):240-248. <https://doi.org/10.1101/gr.5681207>.
38. Baird NA, Etter PD, Atwood TS, et al. Rapid SNP discovery and genetic mapping using sequenced rad markers. *PLoS One*. 2008;3(10):e3376. <https://doi.org/10.1371/journal.pone.0003376>.

39. Trebbi D, Maccaferri M, de Heer P, et al. High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). *Theor Appl Genet.* 2011;123(4):555-569. <https://doi.org/10.1007/s00122-011-1607-7>.
40. Elshire RJ, Glaubitz JC, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 2011;6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379>.
41. Zhang Y, Wang L, Xin H, et al. Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 2013;13(1):141. <https://doi.org/10.1186/1471-2229-13-141>.
42. Li B, Tian L, Zhang J, et al. Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max.* *BMC Genom.* 2014;15(1):1086. <https://doi.org/10.1186/1471-2164-15-1086>.
43. Liu T, Guo L, Pan Y, et al. Construction of the first high-density genetic linkage map of *Salvia miltiorrhiza* using specific length amplified fragment (SLAF) sequencing. *Sci Rep.* 2016;6(1):24070. <https://doi.org/10.1038/srep24070>.
44. Wei Q, Wang Y, Qin X, et al. An SNP-based saturated genetic map and QTL analysis of fruit-related traits in cucumber using specific-length amplified fragment (SLAF) sequencing. *BMC Genom.* 2014;15(1):1158. <https://doi.org/10.1186/1471-2164-15-1158>.
45. Jiang B, Liu W, Xie D, et al. High-density genetic map construction and gene mapping of pericarp color in wax gourd using specific-locus amplified fragment (SLAF) sequencing. *BMC Genom.* 2015;16(1):1035. <https://doi.org/10.1186/s12864-015-2220-y>.
46. Ma JQ, Huang L, Ma CL, et al. Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). *PLoS One.* 2015;10(6):e0128798. <https://doi.org/10.1371/journal.pone.0128798>.
47. Bohra A, Saxena RK, Gnanesh BN, et al. An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. *Theor Appl Genet.* 2012;125(6):1325-1338. <https://doi.org/10.1007/s00122-012-1916-5>.
48. Cloutier S, Ragupathy R, Miranda E, et al. Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). *Theor Appl Genet.* 2012;125(8):1783-1795. <https://doi.org/10.1007/s00122-012-1953-0>.
49. Asgarinia P, Cloutier S, Duguid SR, et al. Mapping quantitative trait loci for powdery mildew resistance in flax. *Crop Sci.* 2013;53(6):2462. <https://doi.org/10.2135/cropsci2013.05.0298>.
50. Würschum T. Mapping QTL for agronomic traits in breeding populations. *Theor Appl Genet.* 2012;125(2):201-210. <https://doi.org/10.1007/s00122-012-1887-6>.
51. Pauli D, Muehlbauer GJ, Smith KP, et al. Association mapping of agronomic QTLs in U.S. Spring barley breeding germplasm. *Plant Genome.* 2014;7(3):1-15. <https://doi.org/10.3835/plantgenome2013.11.0037>.
52. Atwell S, Huang YS, Vilhjálmsson BJ, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature.* 2010;465(7298):627-631. <https://doi.org/10.1038/nature08800>.
53. Tian F, Bradbury PJ, Brown PJ, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet.* 2011;43(2):159-162. <https://doi.org/10.1038/ng.746>.
54. Gondro C, Julius HJ, van der Werf J, Hayes BJ. Genome-wide association studies and genomic prediction. Humana Press; 2013. <https://doi.org/10.1007/978-1-62703-447-0>.
55. Soto-Cerda B, Cloutier S, Quian R, et al. Genome-wide association analysis of mucilage and hull content in flax (*Linum usitatissimum* L.) seeds. *Int J Mol Sci.* 2018;19(10):2870. <https://doi.org/10.3390/ijms19102870>.
56. Xie D, Dai Z, Yang Z, et al. Genomic variations and association study of agronomic traits in flax. *BMC Genom.* 2018;19(1):512. <https://doi.org/10.1186/s12864-018-4899-z>.
57. Xie D, Dai Z, Yang Z, et al. Genome-wide association study identifying candidate genes influencing important agronomic traits of flax (*Linum usitatissimum* L.) using SLAF-seq. *Front Plant Sci.* 2018;8:1-13. <https://doi.org/10.3389/fpls.2017.02232>.
58. He L, Xiao J, Rashid KY, et al. Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.). *Front Plant Sci.* 2019;9:1-15. <https://doi.org/10.3389/fpls.2018.01982>.
59. You F, Xiao J, Li P, et al. Genome-wide association study and selection signatures detect genomic regions

- associated with seed yield and oil quality in flax. *Int J Mol Sci.* 2018;19(10):2303. <https://doi.org/10.3390/ijms19082303>.
60. Xie W, Wang G, Yuan M, et al. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc Natl Acad Sci USA.* 2015;112(39):e5411-5419. <https://doi.org/10.1073/pnas.1515919112>.
  61. Robertsen C, Hjortshøj R, Janss L. Genomic selection in cereal breeding. *Agronomy.* 2019;9(2):95. <https://doi.org/10.3390/agronomy9020095>.
  62. You FM, Booker HM, Duguid SD, et al. Accuracy of genomic selection in biparental populations of flax (*Linum usitatissimum* L.). *Crop J.* 2016;4(4):290-303. <https://doi.org/10.1016/j.cj.2016.03.001>.
  63. Soltis DE, Gitzendanner MA, Stull G, et al. The potential of genomics in plant systematics. *Taxon.* 2013;62(5):886-898. <https://doi.org/10.12705/625.13>.
  64. Emerson KJ, Merz CR, Catchen JM, et al. Resolving postglacial phylogeography using high-throughput sequencing. *Proc Natl Acad Sci USA.* 2010;107(37):16196-16200. <https://doi.org/10.1073/pnas.1006538107>.
  65. Andolfatto P, Davison D, Erezyilmaz D, et al. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 2011;21(4):610-617. <https://doi.org/10.1101/gr.115402.110>.
  66. Ruhsam M, Rai HS, Mathews S, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour.* 2015;15(5):1067-1078. <https://doi.org/10.1111/1755-0998.12375>.
  67. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* (Edinb). 2011;107(1):1-15. <https://doi.org/10.1038/hdy.2010.152>.
  68. Straub SC, Parks M, Weitemier K, et al. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am J Bot.* 2012;99(2):349-364. <https://doi.org/10.3732/ajb.1100335>.
  69. Bolsheva NL, Melnikova NV, Kirov IV, et al. Characterization of repeated DNA sequences in genomes of blue-flowered flax. *BMC Evol Biol.* 2019;19(S1):49. <https://doi.org/10.1186/s12862-019-1375-6>.
  70. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57-63. <https://doi.org/10.1038/nrg2484>.
  71. Haas BJ, Zody MC. Advancing RNA-Seq analysis. *Nat Biotechnol.* 2010;28(5):421-423. <https://doi.org/10.1038/nbt0510-421>.
  72. Wang L, Li P, Brutnell TP. Exploring plant transcriptomes using ultra high-throughput sequencing. *Brief Funct Genom.* 2010;9(2):118-128. <https://doi.org/10.1093/bfgp/elp057>.
  73. Wu J, Zhao Q, Sun D, et al. Transcriptome analysis of flax (*Linum usitatissimum* L.) undergoing osmotic stress. *Ind Crops Prod.* 2018;116:215-223. <https://doi.org/10.1016/j.indcrop.2018.02.035>.
  74. Yu Y, Huang W, Chen H, et al. Identification of differentially expressed genes in flax (*Linum usitatissimum* L.) under saline-alkaline stress by digital gene expression. *Gene.* 2014;549(1):113-122. <https://doi.org/10.1016/j.gene.2014.07.053>.
  75. Wu J, Zhao Q, Wu G, et al. Comprehensive analysis of differentially expressed unigenes under nacl stress in flax (*Linum usitatissimum* L.) using RNA-Seq. *Int J Mol Sci.* 2019;20(2):369. <https://doi.org/10.3390/ijms20020369>.
  76. Durrant A. The environmental induction of heritable change in *Linum*. *Heredity.* 1962;17:27-61. <https://doi.org/10.1038/hdy.1962.2>.
  77. Dmitriev AA, Krasnov GS, Rozhmina TA, et al. Glutathione S-transferases and UDP-glycosyltransferases are involved in response to aluminum stress in flax. *Front Plant Sci.* 2016;7:1-10. <https://doi.org/10.3389/fpls.2016.01920>.
  78. Zyablitsin AV, Dmitriev AA, Krasnov GS, et al. [CAX3 gene is involved in flax response to high soil acidity and aluminum exposure. (In Russ.)]. *Mol Biol.* 2018;52(4):514-519. <https://doi.org/10.1134/S0026893318040192>.
  79. Dmitriev AA, Krasnov GS, Rozhmina TA, et al. Flax (*Linum usitatissimum* L.) response to non-optimal soil acidity and zinc deficiency. *BMC Plant Biol.* 2019;19(S1):54. <https://doi.org/10.1186/s12870-019-1641-1>.
  80. Galindo-González L, Deyholos MK. RNA-Seq transcriptome response of flax (*Linum usitatissimum* L.) to the pathogenic fungus *Fusarium oxysporum* f. sp. *lini*. *Front Plant Sci.* 2016;7:1-22. <https://doi.org/10.3389/fpls.2016.01766>.
  81. Dmitriev AA, Krasnov GS, Rozhmina TA, et al. Differential gene expression in response to *Fusarium oxysporum* infection in resistant and susceptible genotypes of flax (*Linum usitatissimum* L.). *BMC Plant*

- Biol.* 2017;17(S2):253. <https://doi.org/10.1186/s12870-017-1192-2>.
82. Zhang N, Deyholos MK. RNA-Seq analysis of the shoot apex of flax (*Linum usitatissimum* L.) to identify phloem fiber specification genes. *Front Plant Sci.* 2016;7:1-5. <https://doi.org/10.3389/fpls.2016.00950>.
83. Mokshina N, Gorshkov O, Ibragimova N, et al. Cellulosic fibres of flax recruit both primary and secondary cell wall cellulose synthases during deposition of thick tertiary cell walls and in the course of graviresponse. *Funct Plant Biol.* 2017;44(8):820. <https://doi.org/10.1071/FP17105>.
84. Shivaraj SM, Deshmukh RK, Rai R, et al. Genome-wide identification, characterization and expression profile of aquaporin gene family in flax (*Linum usitatissimum*). *Sci Rep.* 2017;7(1):46137. <https://doi.org/10.1038/srep46137>.
85. Dash PK, Rai R, Mahato AK, et al. Transcriptome landscape at different developmental stages of a drought tolerant cultivar of flax (*Linum usitatissimum*). *Front Chem.* 2017;5:1-4. <https://doi.org/10.3389/fchem.2017.00082>.
86. Gorshkova T, Chernova T, Mokshina N, et al. Transcriptome analysis of intrusively growing flax fibers isolated by laser microdissection. *Sci Rep.* 2018;8(1):14570. <https://doi.org/10.1038/s41598-018-32869-2>.
87. Sequence Read Archive [Internet]. Bethesda (MD): National Library of Medicine (US) [cited 2019 Jun 12]. Available from: <https://www.ncbi.nlm.nih.gov/sra>.
88. Gene Expression Omnibus [Internet]. Bethesda (MD): National Library of Medicine (US) [cited 2019 Jun 12]. Available from: <https://www.ncbi.nlm.nih.gov/geo/>.
89. Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol.* 2006;57(1):19-53. <https://doi.org/10.1146/annurev.arplant.57.032905.105218>.
90. Sunkar R, Chinnusamy V, Zhu J, et al. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci.* 2007;12(7):301-309. <https://doi.org/10.1016/j.tplants.2007.05.001>.
91. Guleria P, Mahajan M, Bhardwaj J, et al. Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. *Genomics Proteomics Bioinformatics.* 2011;9(6):183-199. [https://doi.org/10.1016/S1672-0229\(11\)60022-3](https://doi.org/10.1016/S1672-0229(11)60022-3).
92. Khraiweh B, Zhu JK, Zhu J. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim Biophys Acta.* 2012;1819(2):137-148. <https://doi.org/10.1016/j.bbagr.2011.05.001>.
93. Melnikova NV, Dmitriev AA, Belenikin MS, et al. Identification, expression analysis, and target prediction of flax genotroph micrnas under normal and nutrient stress conditions. *Front Plant Sci.* 2016;7:399. <https://doi.org/10.3389/fpls.2016.00399>.
94. Yu Y, Wu G, Yuan H, et al. Identification and characterization of miRNAs and targets in flax (*Linum usitatissimum*) under saline, alkaline, and saline-alkaline stresses. *BMC Plant Biol.* 2016;16(1):124. <https://doi.org/10.1186/s12870-016-0808-2>.
95. Dmitriev AA, Kudryavtseva AV, Bolsheva NL, et al. miR319, miR390, and miR393 are involved in aluminum response in flax (*Linum usitatissimum* L.). *BioMed Res Int.* 2017;2017:1-6. <https://doi.org/10.1155/2017/4975146>.
96. Melnikova NV, Dmitriev AA, Belenikin MS, et al. Excess fertilizer responsive miRNAs revealed in *Linum usitatissimum* L. *Biochimie.* 2015;109:36-41. <https://doi.org/10.1016/j.biochi.2014.11.017>.
97. miRBase [Internet]. Manchester: University of Manchester (UK) [cited 2019 Jul 19]. Available from: <http://www.mirbase.org/textsearch.shtml?q=Linum>.

## ✿ Информация об авторе

**Алёна Олеговна Ахметшина** — младший научный сотрудник, лаборатория постгеномных исследований, ФГБНУ «ФИЦ Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова», Санкт-Петербург. SPIN: 3947-2772. E-mail: [akhmetshinaalena@gmail.com](mailto:akhmetshinaalena@gmail.com).

## ✿ Author and affiliations

**Alena O. Akhmetshina** — Junior Researcher, Postgenomic Research Department. Federal Research Center "N.I. Vavilov All-Russian Institute of Plant Genetic Resources", Saint Petersburg, Russia. SPIN: 3947-2772. E-mail: [akhmetshina-alena@gmail.com](mailto:akhmetshina-alena@gmail.com).

## ✿ Информация об авторе

**Ксения Владимировна Стрыгина** — канд. биол. наук, старший научный сотрудник, лаборатория постгеномных исследований. ФГБНУ «ФИЦ Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова», Санкт-Петербург. E-mail: k.strygina@nw.vir.ru.

**Елена Константиновна Хлесткина** — профессор, д-р биол. наук, врио директора ВИР, заведующая лабораторией постгеномных исследований. ФГБНУ «ФИЦ Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова», Санкт-Петербург. SPIN: 3061-1429. E-mail: khlest@bionet.nsc.ru.

**Елизавета Александровна Пороховинова** — канд. биол. наук, старший научный сотрудник, отдел ГР масличных и прядильных культур. ФГБНУ «ФИЦ Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова», Санкт-Петербург. SPIN: 5033-3263. E-mail: e.porohovinova@mail.ru.

**Нина Борисовна Брutch** — д-р биол. наук, ведущий научный сотрудник, отдел ГР масличных и прядильных культур. ФГБНУ «ФИЦ Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова», Санкт-Петербург. SPIN: 1753-4382. E-mail: n.brutch@vir.nw.ru.

## ✿ Author and affiliations

**Ksenia V. Strygina** — PhD, Senior Researcher, Postgenomic research department. Federal Research Center “N.I. Vavilov All-Russian Institute of Plant Genetic Resources”, Saint Petersburg, Russia. E-mail: k.strygina@nw.vir.ru.

**Elena K. Khlestkina** — Prof., Doctor of Science, Director, Head of the Postgenomic Research Department. Federal Research Center “N.I. Vavilov All-Russian Institute of Plant Genetic Resources”, Saint Petersburg, Russia. SPIN: 3061-1429. E-mail: khlest@bionet.nsc.ru.

**Elizaveta A. Porokhovinova** — PhD, Senior Researcher, Oil and Fibre Crops Department. Federal Research Center “N.I. Vavilov All-Russian Institute of Plant Genetic Resources”, Saint Petersburg, Russia. SPIN: 5033-3263. E-mail: e.porohovinova@mail.ru.

**Nina B. Brutch** — Doctor of Science, Main Researcher, Oil and Fibre Crops Department. Federal Research Center “N.I. Vavilov All-Russian Institute of Plant Genetic Resources”, Saint Petersburg, Russia. SPIN: 1753-4382. E-mail: n.brutch@vir.nw.ru.