

КЕЗИМАНА ПАРФЭ

**ПОЛИМОРФИЗМ ГЕНОВ СЕМЕЙСТВ *FAD*, *SAD* И *UGT* И ИХ РОЛЬ В
ОПРЕДЕЛЕНИИ ЖИРНОКИСЛОТНОГО СОСТАВА МАСЛА И СОДЕРЖАНИЯ
ЛИГНАНОВ В СЕМЕНАХ ЛЬНА (*Linum usitatissimum* L.)**

Специальности

03.02.07 – Генетика

06.01.05 – Селекция и семеноводство сельскохозяйственных растений

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Москва – 2021

Диссертационная работа выполнена в агробиотехнологическом департаменте Аграрно-технологического института ФГАОУ ВО «Российский университет дружбы народов» в 2016-2020 годах и в лаборатории постгеномных исследований ФГБУН «Института молекулярной биологии им. В.А. Энгельгардта» РАН.

Научные руководители:

кандидат сельскохозяйственных наук,
доцент агробиотехнологического департамента
Аграрно-технологического института
Российского университета
дружбы народов

Елена Валерьевна Романова

кандидат биологических наук
с.н.с. лаборатории постгеномных исследований
ФГБУН Института молекулярной биологии
им. В.А. Энгельгардта РАН

Наталья Владимировна Мельникова

Официальные оппоненты:

доктор биологических наук,
главный научный сотрудник лаборатории
селекции и семеноводства озимой пшеницы,
профессор; ФГБНУ "Белгородский федеральный
аграрный научный центр" РАН

Нецветаев Владимир Павлович

кандидат сельскохозяйственных наук,
ведущий научный сотрудник лаборатории селекции
и семеноводства овощных бобовых культур
ФГБНУ «ФНЦ овощеводства»

Пронина Екатерина Павловна

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский государственный аграрный университет -МСХА имени К. А. Тимирязева»

Защита диссертации состоится «23» декабря 2021 г. в 15:00 часов на заседании диссертационного Совета ПДС 0300.005 при ФГАОУ ВО «Российский университет дружбы народов» (РУДН) по адресу: 117198, ул. Миклухо-Маклая, д.8.

С диссертацией можно ознакомиться в библиотеке в УНИБЦ (Научной библиотеке) ФГАОУ ВО «Российский университет дружбы народов» по адресу: 117198, ул. Миклухо-Маклая, д.6 и на сайте <http://dissovet.rudn.ru>

Объявление о защите и автореферат диссертации размещены на сайтах <http://vak2.ed.gov.ru> и <http://dissovet.rudn.ru>

Автореферат разослан « » _____ 2021 г.

Ученый секретарь
диссертационного совета ПДС 0300.005
кандидат биологических наук, доцент

Гигани Ольга Олеговна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы и степень ее разработанности.

Мир сталкивается с огромными проблемами, такими, как необходимость повышения производительности для обеспечения потребностей растущего населения нашей планеты при одновременном снижении воздействия на окружающую среду и преодоления последствий изменения климата. В связи с этим, на сегодняшний день используются современные технологии генетики, геномики и геномной инженерии для создания новых сортов культур, удовлетворяющих требованиям меняющегося мира, в том числе, производство биологически активных веществ (БАВ), применяемых в различных индустриях. Лен (*Linum usitatissimum* L.) является одной из важнейших сельскохозяйственных культур, сорта которого представляют собой источник таких биологически активных веществ (БАВ), как растительные масла и лигнаны, применяемые в пищевой и фармацевтической промышленности [Bekhit и др., 2018; El-Beltagi, Salama, El-Hariri, 2007; Jhala, Hall, 2010]. Именно благодаря высокому содержанию в семенах льна таких БАВ, как α -линоленовая кислота (LIN) и лигнанов, обладающих широким спектром фармакологического действия [Morris, 2007a; Pilar и др., 2017; Touré, Xueming, 2010], вырос интерес к этому растению.

Современные сорта льна содержат до 50% масла, состоящего из пяти основных жирных кислот (ЖК): пальмитиновой (PAL) (C16:0; ~ 6%), стеариновой (STE) (C18:0; ~ 4%), олеиновой (OLE) (C18:1cis ^{Δ^9} ; ~ 24%), линолевой (LIO) (C18: 2cis ^{$\Delta^9,12$} ; ~ 15%) и линоленовой (LIN) (C18:3cis ^{$\Delta^9,12,15$} ; ~ 50%) [Muir, Westcott, 2003; Thambugala, Cloutier, 2014; You и др., 2014]. Высокое содержание LIN и умеренное содержание LIO в льняном масле способствует здоровой диете [Damude, Kinney, 2007; Thambugala, Cloutier, 2014]. Более того, LIN и LIO, также называемые полиненасыщенными ЖК (ПНЖК), являются важными компонентами клеточных мембран, участвующих в метаболизме растений в качестве источника запасной энергии в виде триацилглицеринов и в качестве предшественников сигнальных молекул, таких как жасмоновые кислоты [Ohlrogge, Browse, 1995]. Они также являются предшественниками других длинноцепочечных ПНЖК, таких как эйкозапентаеновая кислота (C20:5 ^{$\Delta^5,8,11,14,17$}) и докозагексановая кислота (C22: 6 ^{$\Delta^4,7,10,13,16,19$}) [Morris, 2007b; Pali, Mehta, 2014], чьи лечебно-профилактические свойства проявляются в снижении уровня холестерина в сыворотке крови и профилактике сердечно-сосудистых заболеваний, что доказано в ряде исследований [Kim, Pich, 2011; Wiesenfeld и др., 2003]. Сорта льна различаются по жирнокислотному составу масла (прежде всего, содержанию LIN) и, в зависимости от соотношения жирных кислот, могут применяться в фармацевтической, пищевой или лакокрасочной

промышленности, а также для производства биодизеля [Muir, Westcott, 2003; Брач, Пороховинова, Шеленга, 2016; Зубцов, Осипова, Антипова, 2007; Марков, Девянин, Трифонов, 2015], поэтому изучение генов, участвующих в биосинтезе ЖК, позволит использовать их в маркер-ориентированной селекции, тем самым повышая эффективность селекционного процесса.

Помимо ЖК, лен также является одним из самых богатых растительных источников лигнанов [Morris, 2007b], действующих как антиоксиданты [Hosseini и др., 2006]. Лигнаны являются классом полифенольных фитоэстрогенов, обладающих противораковыми и антидиабетическими свойствами, а также оказывающих профилактические эффекты против сердечно-сосудистых заболеваний [Adolphe и др., 2010; Imran и др., 2015; Muir, Westcott, 2003; Parikh, Netticadan, Pierce, 2018]. Лигнаны у льна обнаружены в гликозилированных формах, а гликозилирование влияет на биологическую активность и биодоступность метаболитов [Morris, 2007c; Parikh, Netticadan, Pierce, 2018; Struijs, 2008; Touré, Xueming, 2010], поэтому гены гликозилирования играют важную роль в биосинтезе лигнанов льна.

Таким образом, исходя из важности этих БАВ, изучен процесс их биосинтеза и определены гены, играющие роль в их формировании у льна. Показано, что гены десатураз играют важную роль в определении соотношения ЖК в льняном масле. Гены десатураз льна идентифицированы, и известно, что десатурацию STE в OLE, OLE в LIO, LIO и LIN последовательно выполняют *SAD* (stearoyl-ACP (acyl-carrier-protein) desaturase – стеароил - ацилпереносящий белок (АПБ)-десатуразы), *FAD2* (fatty acid desaturase 2 – десатуразы жирных кислот 2) и *FAD3* (fatty acid desaturase 3 – десатуразы жирных кислот 3) [Banik, Duguid, Cloutier, 2011; Fofana и др., 2006; You и др., 2014; Лось, 2001].

В биосинтезе лигнанов принимает участие ряд ферментов, включая пинорезинол-ларицирезинол редуктазы (PLRs – pinoresinol/lariciresinol reductases) и 1-уридиндифосфат гликозилтрансферазы (UGTs – uridine 5'-diphospho-glucuronosyltransferases) [Kim и др., 2009; Okazawa и др., 2014]. У растений существует большое семейство *UGT*, которые добавляют уридиндифосфат-активированные сахарные компоненты к акцепторным молекулам [Dhaubhadel, Farhangkhoe, Chapman, 2008], что обеспечивает их химическую стабильность и сниженную активность. У льна идентифицированы гены *PLR*, преобразующие пинорезинол в секоизоларицирезинол (SECO) [Hemmati и др., 2010; Hemmati, Schmidt, Fuss, 2007], и показано, что основная роль в превращении SECO в диглюкозид секоизоларицирезинола (SDG) играет ген *UGT74S1* [Fofana и др., 2017a; Ghose и др., 2014; Kezimana и др., 2018].

Несмотря на значение льна и наличие информации о генетическом контроле биосинтеза ЖК и лигнанов, исследования по оценке генетического разнообразия генов, вовлеченных в их синтез, ограничены. Более того, мало известно о взаимосвязи между полиморфизмами в этих генах и содержании ЖК и лигнанов. Получение таких данных способствует применению технологий генетики, геномики и генной инженерии для создания сортов льна специализированного назначения.

Цель и задачи исследования. Цель работы – идентифицировать полиморфизмы генов льна *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, *UGT74S1* и определить влияние полиморфизмов на содержание жирных кислот и лигнанов, а также разработать подходы для использования полученных данных в маркер-ориентированной селекции масличного льна.

Для реализации этой цели были поставлены **следующие задачи:**

1. Создание представительной коллекции ДНК сортов/линий льна, включающей различающиеся по жирнокислотному составу масла и содержанию лигнанов, генотипы.
2. Разработка подхода для определения нуклеотидных последовательностей генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, *UGT74S1* с использованием глубокого секвенирования.
3. Подготовка ДНК-библиотек и секвенирование генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1* для образцов из созданной коллекции.
4. Анализ данных секвенирования и выявление однонуклеотидных полиморфизмов (SNP, single nucleotide polymorphism), инсерций и делеций в генах *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1*.
5. Установление ассоциаций выявленных полиморфизмов генов с составом жирных кислот (PAL, STE, OLE, LIO, LIN) и содержанием лигнанов в семенах генотипов льна.
6. Разработка подхода для использования полученных данных о полиморфизмах изученных генов в маркер-ориентированной селекции масличного льна.

Научная новизна.

С помощью метода глубокого секвенирования впервые на представительной выборке образцов льна (279 сортов/линий, не менее 50 растений для каждого генотипа) идентифицированы полиморфизмы в генах *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, играющих ключевую роль в синтезе жирных кислот, и *UGT74S1*, влияющего на синтез лигнанов. Установлена неоднородность (гетерогенность) сортов/линий льна по изученным генам. Найдены ассоциации между содержанием PAL, STE, OLE, LIO и LIN в льняном масле и полиморфизмами генов семейств *SAD* и *FAD*. Показано, что наиболее информативными и

перспективными для использования в маркер-ориентированной селекции масличного льна являются три полиморфизма в генах *FAD3A* и *FAD3B*, главным образом определяющие содержание линолевой и линоленовой кислот в льняном масле.

Теоретическая и практическая значимость работы.

Жирнокислотный состав масла и содержание лигнанов определяют, как правило, на поздних этапах селекционного процесса, что не позволяет вести ранний и эффективный отбор на данный признак, однако разработка ДНК-маркеров для идентификации аллелей генов, играющих ключевую роль в детерминации содержания жирных кислот и лигнанов, значительно повысит эффективность селекции. Полученные нами данные по полиморфизмам генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1* и их ассоциации с жирнокислотным составом масла и содержанием лигнанов, позволили провести ДНК-паспортизацию сортов/линий льна, оценить сортовую гетерогенность и идентифицировать ключевые полиморфизмы генов *FAD3A* и *FAD3B*, перспективные для использования в маркер-ориентированной селекции льна для повышения эффективности селекционного процесса и ускорения создания сортов льна пищевого, фармацевтического и промышленного назначения. Также собран и аннотирован геном льна сорта Атлант, что является основой для дальнейших молекулярно-генетических исследований этой культуры.

Методология и методы исследования.

Семена льна, полученные из Института льна, выращивали в чашках Петри для получения проростков, из которых выделяли ДНК ЦТАБ-протоколом. Подготовку ДНК-библиотеки для секвенирования проводили с использованием двухступенчатой ПЦР и на основе модифицированного протокола Illumina [Illumina, 2013].

Праймеры подбирали с использованием программ MEGA и NCBI Primer-BLAST. Качество ДНК-библиотек оценивали на системе капиллярного фореза Bioanalyzer 2100 (Agilent Technologies, США), концентрацию – на флуориметре Qubit 2.0 (Life Technologies, США). Секвенирование проводили на платформе MiSeq (Illumina, США). Данные секвенирования были обработаны и выровнены на референсный геном льна (assembly GCA_000224295.2/ASM22429v2). Для определения аллельных вариантов использовали приложения VarScan и freeBayes. Для корреляционного анализа между идентифицированными полиморфизмами и содержанием ЖК и лигнанов использовались коэффициенты ранговой корреляции Спирмана и Кендалла и р-значение. Кластерный анализ проводился методом Ward D2, а визуализация дендрограмм и тепловых карт выполнена с использованием пакетов pheatmap, d3heatmap, ggtree и gheatmap.

ДНК для полногеномного секвенирования на платформах Oxford Nanopore Technologies и Illumina выделяли набором ДНК-ЭКСТРАН-3 (Синтол, Россия) с последующей очисткой набором Blood & Cell Culture DNA Mini Kit (Qiagen, США). Качество и концентрацию ДНК оценивали на флуориметре Qubit 2.0 (Life Technologies, США) и спектрофотометре NanoDrop 2000C (Thermo Fisher Scientific, США). Секвенирование проводили на приборах MinION (Oxford Nanopore Technologies) и HiSeq 2500 (Illumina). Для сборки генома льна использовали четыре ассемблера: Canu 2.0, Flye 2.7, Shasta 0.5.0 и wtdbg2. 2.5. Оценку полученных сборок генома проводили с помощью QUAST 5.0.2.

Положения, выносимые на защиту:

1. Идентифицированы полиморфизмы генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, играющих ключевую роль в синтезе жирных кислот, и *UGT74S1*, участвующего в синтезе лигнанов, для выборки из 279 сортов/линий льна, и показано наибольшее генетическое разнообразие для генов *FAD3A* и *FAD3B*.
2. Сорта/линии льна характеризуются значительной внутренней гетерогенностью по генам семейств *SAD* и *FAD*.
3. Разнообразие льна по гену *UGT74S1* низкое, и полиморфизмы этого гена не оказывают существенного влияния на содержание лигнанов в семенах изученной выборки льна.
4. Данные о полиморфизмах генов *SAD* и *FAD* представляют ценность для генетической паспортизации сортов/линий льна, подбора пар для скрещиваний и контроля сортовой чистоты.
5. Полиморфизмы, приводящие к замене триптофана на стоп-кодон в *FAD3A*, гистидина на тирозин в *FAD3B* и аргинина на стоп-кодон в *FAD3A*, являются ключевыми и определяют содержание линоленовой и линолевой кислот в льняном масле.
6. Ключевые полиморфизмы генов *FAD3A* и *FAD3B* перспективны для использования в маркер-ориентированной селекции льна для отбора генотипов с определенными аллелями этих генов в гомозиготном состоянии, что повышает эффективность создания сортов льна специального назначения.

Апробация результатов работы.

Результаты работы представлены на международных и всероссийских съездах, симпозиумах, конгрессах, конференциях, в том числе:

- The 43rd FEBS Congress Biochemistry Forever, 7-12 July 2018, Prague, Czech Republic;
- The 12th Congress of the International Plant Molecular Biology, 5-10 August 2018, Montpellier, France;

- The Eleventh International Conference on Bioinformatics of Genome Regulation and Structure\Systems Biology (BGRS\SB-2018), 20-25 August 2018, Novosibirsk, Russia; Biodiversity: Genomics and Evolution, BioGenEvo-2018 Symposium, 21-24 August 2018, Novosibirsk, Russia;
- The 1st Cologne Conference on Food for Future, 5-7 September 2018, Köln, Germany;
- Plant Science Center Summer School: Responsible Research and Innovation in Plant Sciences, 10-14 September 2018, Einsiedeln, Switzerland;
- Международная научно-практическая конференция «Гармонизация подходов к фармацевтической разработке», 28 ноября 2018, Москва, Россия;
- IV Международная научно-методическая конференция «Гаммермановские чтения», 30-31 января 2019, Санкт-Петербург, Россия;
- International CEPLAS Summer School Transatlantic Summer School - Frontiers in Plant Sciences, 27-31 May 2019, Wermelskirchen, Germany;
- The Fifth International Scientific Conference Plant Genetics, Genomics, Bioinformatics, and Biotechnology (PlantGen2019), 24-29 June 2019, Novosibirsk, Russia;
- The 44th FEBS Congress From Molecules to Living Systems, 6-11 July 2019, Krakow, Poland;
- Plant Genomes in a Changing Environment, 16-18 October 2019, Wellcome Genome Campus, Hinxton, Cambridge, UK.

Публикации

Общее число работ по теме диссертации, включая сборники трудов конференций, составляет 13, в том числе 7 в изданиях, рекомендованных ВАК, 6 из них индексируются в Scopus и Web of Science.

Личный вклад автора заключается в выполнении основного объема теоретических и экспериментальных исследований по теме данной работы, включая анализ известных в литературе данных, планирование и проведение экспериментов. Диссертационная работа является результатом исследований, проведенных автором в агробиотехнологическом департаменте АТИ РУДН в 2016–2020 гг. и ФГБУН Институте молекулярной биологии им. В.А. Энгельгардта Российской академии наук в 2017-2020 гг.

Структура и объем диссертации.

Диссертация изложена на 150 страницах печатного текста, содержит 12 таблиц и 23 рисунка. Состоит из введения, обзора литературы, описания материалов и методов,

результатов исследования и их обсуждения, заключения, выводов, приложения и списка цитируемой литературы, содержащего 257 источников.

ОБЗОР ЛИТЕРАТУРЫ

В разделе «Обзор литературы» диссертационной работы представлены направления использования льняного масла и лигнанов, обобщены результаты молекулярно-генетических исследований по льну, и описан генетический контроль биосинтеза жирных кислот и лигнанов в семенах льна.

МАТЕРИАЛЫ И МЕТОДЫ

Растительный материал. Для определения полиморфизма генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, играющих ключевую роль в синтезе жирных кислот, и *UGT74S1*, играющего важную роль в синтезе лигнанов, использованы 279 сортов/линий льна посевного (*L. usitatissimum* L.), полученных из Института льна (г. Торжок). Жироокислотный состав известен для 84 линий/сортов, а содержание лигнанов – для 57 линий/сортов (табл. 1). Семена обрабатывали 1,5 % р-ром гипохлорана, а затем проращивали на фильтровальной бумаге в чашках Петри. На 5-7-й день проводили сбор растительного материала с 50 или более проростков для каждого сорта/линии, которые затем использовали для выделения ДНК.

Выделение ДНК. Выделение ДНК проводили из проростков льна с использованием модифицированного ЦТАБ-метода [Clarke, 2009].

Подготовка библиотек ампликонов. Для подготовки к секвенированию генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1* применен модифицированный протокол Illumina, заключающийся в использовании двустадийной полимеразной цепной реакции (ПЦР): 1^{ая} стадия – амплификация выбранных участков генома и добавление на ампликоны универсальной последовательности, и 2^{ая} стадия – добавление последовательностей, необходимых для высокопроизводительного секвенирования, и двойных индексов для идентификации образцов. Для амплификации подобрали 12 пар праймеров для генов *SAD*, 10 для генов *FAD2*, 18 для генов *FAD3* и 8 для гена *UGT74S1*. При разработке праймеров использовали собранный нам геном сорта Атлант, который депонирован в базе данных NCBI под регистрационным номером BioProject PRJNA648016. Концентрацию и качество ДНК-библиотек оценивали с помощью флуориметра Qubit 2.0 (Life Technologies) и системы капиллярного фореза Bioanalyzer 2100 (Agilent Technologies, США). Средняя длина библиотек ДНК составила около 550–600 п.н. и соответствовала ожидаемой.

Таблица 1. Линии и сорта льна, используемые в исследовании, с известным составом жирных кислот и содержанием лигнанов

№	Название	PAL, %	STE, %	OLE, %	LIO, %	LIN, %	SECO, мг/кг ⁻¹	№	Название	PAL, %	STE, %	OLE, %	LIO, %	LIN, %	SECO, мг/кг ⁻¹
A6	Lola	7	2,7	12,9	68	9,4	2900	A53	AGT 427/10	6,4	2,7	13,1	13,4	64,3	5125
A7	Recital	6,2	3,9	18,1	14,7	57,1	3125	A54	IDG 4102	6,4	4	14,6	12,1	62,9	3250
A8	Oural	6,2	4,5	19,2	15	55,1	2500	A55	IDG 4101	6,7	6	19,8	13,6	53,8	3375
A9	Baikal	5,4	4,6	20,4	17,1	52,5	2050	A56	IDG4105	5,9	3,5	12,9	12,5	65,3	2000
A10	Jantar	6,2	3,5	15,3	66,8	8,3	3750	A57	AGT 1535/07	7,6	4,7	18,8	64	4,9	2125
A11	Bethune	5,7	3,9	19,2	16,2	55	2600	A58	Leane	6,4	3,5	14,6	17,8	57,7	2550
A13	Bachmalskii 1056	6	4,1	21	13,6	55,5	3625	A60	McDuff	5,5	3,9	16,5	17,3	56,8	2875
A14	Atalante	6,4	4,3	16,4	14,7	58,2	4650	A61	Lf-L37	6,6	4,5	16	13	59,9	4000
A19	AGT 422/10	7,2	6,4	18,3	35,5	32,6	3625	A62	Jaeger	6,8	4,3	16,5	13,2	59,2	3100
A20	AGT 439/ 10	6,6	4,2	17,8	15,9	55,4	3500	A63	Natural	7,1	5,2	18,4	14	55,2	2750
A21	AGT461/10	6,7	3,9	18,9	13,6	56,8	2875	A64	AGT 987/02	6,8	4,2	17,4	53,1	18,5	3725
A22	AGT 464/10	6	3,8	16,4	24,9	48,9	4400	A65	AGT 583/05	6,6	3,5	17	15,7	57,2	3750
A23	AGT 494/10	6,8	4,5	21,8	16	50,9	3450	A66	LS-L68	6,4	4,4	18,3	13,4	57,5	3050
A24	AGT 545/10	6,4	4,4	14,2	14,3	60,6	4125	A68	Kinel'skij 2000	6,4	2,8	16,2	17	57,6	-
A25	AGT 316/02	6,8	3,1	14,9	72,4	2,7	4300	A71	l. 1-1 Walaga	7,2	3,6	18	66,8	4,5	-
A26	Benvenuto Real	6,4	4,8	19,7	14,2	55	3625	A73	l. 1-3-2 Macovi M.A.G.	5,6	3,9	24	16,1	50,5	-
A27	Entre-Rios	6,4	4,4	20,1	16,5	52,5	3900	A74	l.1-2 Bolley Golden	6,4	4	16,2	13,5	59,9	-
A28	Vitagold	5,1	3,7	18,3	14,8	58,1	3000	A75	l.1-1 Mermilloid	5,9	4,2	13,8	14,5	61,5	-
A29	Astral	6,4	3,5	16,4	12,5	61,9	2375	A76	l.1-1-2 Rabat 12	6,3	4,6	22	13,9	53,2	-
A30	Kaolin	6,5	4	17,1	15	57,4	3100	A77	l.1-1 Ottawa 2152	5,8	4,3	18,7	17,6	53,6	-
A31	Alaska	6,3	4	16,6	15,3	57,8	4250	A78	l. 1-1-1 Liral Dominion	5	3,6	16,9	18,2	56,3	-
A32	Baladin	6,2	5,4	17,5	14	56,9	2750	A87	l.1-1-1 Bionda	5,1	2,7	16,7	18,3	57,2	-
A33	Omega	5,2	3,7	20,5	16,8	53,8	3000	A90	l. 1-1-2 Eyre	7	4,1	19,6	64,8	4,5	-
A34	Amon	6,7	3,5	14,8	71,8	3,1	4250	A93	Cian	5,7	4,7	20,3	14,9	54,3	-
A35	Bilton	6,6	3,8	13,4	15,5	60,7	2750	A98	l.1 K-1352	6,2	4,3	18,1	11,9	59,5	-
A36	Prairie Blue	6,8	4	13,1	15,6	60,5	2600	A103	Culbert	5,1	3,6	20,1	16,8	54,3	-
A37	Raciol	6,3	4	15,5	39,2	35	3900	A104	No 3857	6,1	3,6	20,6	13,7	56	-
A38	AGT 1538/07	7,2	2,9	15,2	71,5	3,2	3125	A105	Eurodor	5,6	4,8	16,9	18,8	53,9	-
A39	AGT 1568/07	6,6	4,2	17,1	28,6	43,5	3150	A106	LM 96	5,4	3,5	23,7	15,3	52,1	-
A40	AGT 467/08	6,3	4,2	16,1	26,7	46,8	3875	A107	LM 92	5,9	3,7	22,3	14,1	54,1	-
A41	AGT 470/08	7,3	3,9	18,6	14,3	55,9	2625	A109	Clark	5,9	3,7	21,9	16	52,6	-
A42	AGT 510/08	5,9	4	17,4	13,8	59	3625	A110	No 3814	6,1	3,8	19,3	14,7	56	-
A43	AGT 302/10	6	4,4	18,3	15,4	56	3400	A111	No 3872	5,7	3,4	21,2	16,5	53,2	-
A44	AGT 368/10	7	3,4	13,9	72,3	3,5	3125	A112	LM 91	5,9	3,8	23,8	19,4	47	-
A45	AGT 390/10	7,1	4,9	14,1	14,9	59	2250	A113	KL 24	5,7	3,3	23,2	14,9	52,9	-
A46	AGT 399/10	7	4,9	15,1	14,4	58,6	4875	A114	Norlin	5,9	3,1	22,2	14,5	54,3	-
A47	AGT 409/ 10	7	4,4	14,7	15,4	58,6	1700	A115	M2965	5,6	2,7	17,5	16	58,3	-
A48	AGT 306/110	5,7	3,7	15,7	15,8	59,2	3000	A127	Ilim	5,4	3,8	19,3	16,2	55,2	-
A49	AGT 305/10	6,3	4,2	14,2	15	60,3	3200	A148	N 3871	5,9	3,9	21,4	16,3	52,4	-
A50	AGT 307/10	5,7	3,8	17,8	16,6	56,1	2500	A317	MC Beth	5,9	3,8	15,3	15,2	59,8	2600
A51	AGT 308/10	5,5	4,2	18,4	16,8	55,1	4250	A318	Eole	6,4	4,5	20,5	13,2	55,4	-
A52	AGT 393/10	7,1	4,2	15,6	16,8	56,3	3375	A321	Flanders	5,9	4,4	20,2	15,1	54,4	3000

Секвенирование библиотек ДНК проведено на секвенаторе Illumina MiSeq с использованием набора MiSeq Reagent Kit v3 600-cycles, согласно протоколу производителя.

Биоинформатическая и статистическая обработки данных секвенирования

Проведена оценка качества полученных чтений и с помощью программы Trimmomatic [Bolger, Lohse, Usadel, 2014], выполнены обрезка и фильтрацию чтений, содержащих адаптерную последовательность или имеющих низкое качество прочтения ($Q < 20$). Затем чтения картировали на референсный геном *L. usitatissimum* (сборка GCA_000224295.2/ASM22429v2) с помощью алгоритма BWA-MEM. Картированные чтения группировали с помощью инструментов Picard, и выполняли поиск полиморфизмов инструментами VarScan [Koboldt и др., 2012] и freeBayes [Garrison, Marth, 2012] с фильтрацией вариантов с VAF (Variant Allele Frequency) $< 20\%$ во всех образцах.

Для корреляционного анализ содержания PAL, STE, OLE, LIO, LIN и полиморфизмов в генах *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, а также содержания лигнанов и полиморфизмов гена *UGT74S1*, рассчитывали коэффициенты ранговой корреляции Спирмана и Кендалла и р-значения. Особое внимание уделяли значению относительного стандартного отклонения (rStD) значений VAF, так как, чем больше rStD, тем надежнее корреляция.

Секвенирование генома сорта Атлант проводилось на платформе Oxford Nanopore Technologies (прибор MinION) с проточной ячейкой R9.4.1 и на платформе Illumina (приборе HiSeq 2500) с длиной чтения 250 + 250 п.н. Сборка генома на основе чтений Nanopore выполнялась с использованием четырех ассемблеров: Canu 2.0 [Koren и др., 2017], Flye 2.7 [Kolmogorov и др., 2019], Shasta 0.5.0 [Shafin и др., 2020] и wtdbg2. 2.5 [Ruan, Li, 2020]. Статистический анализ качества сборок генома проводился с помощью QUAST 5.0.2 [Gurevich и др., 2013]. Сборки полировали для повышения точности контигов с помощью RacoN 1.4.3 [Vaser и др., 2017] и Medaka 1.0.3, а также POLCA. Полнота сборок оценивалась с использованием BUSCO v4 [Seppey, Manni, Zdobnov, 2019].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Полиморфизмы изучаемых генов, определяющих жирнокислотный состав, и содержание лигнанов в семенах льна

Разработан подход для точного определения последовательностей генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1* с использованием высокопроизводительного секвенирования. Проведено таргетное секвенирование генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1* на платформе Illumina для выборки из 279 сортов/линий льна, 84 из которых имели известный состав жирных кислот и 57 – известное содержание лигнанов. В среднем получено 400-кратное покрытие исследуемых генов.

1.1 Полиморфизмы генов десатураз жирных кислот

1.1.1 Полиморфизм генов *SAD*, *FAD2*, и *FAD3*

С использованием приложения VarScan идентифицированы 232 полиморфизма генов, участвующих в синтезе ЖК: *SAD1* - 14, *SAD2* - 14, *FAD2A* - 21, *FAD2B* - 11, *FAD3A* - 101, *FAD3B* - 71. Таким образом, наибольшим генетическим разнообразием характеризовались гены *FAD3A* и *FAD3B*. Используемый нами подход, основанный на секвенировании ДНК, выделенной из пулов растений (около 50 проростков для каждого образца), позволил оценить гетерогенность каждого исследованного сорта/линии. Большинство исследованных образцов были гетерогенными, по крайней мере, по одному из секвенированных генов. Данные о частотах полиморфизмов генов семейств *FAD* и *SAD* представлены на рис. 1.

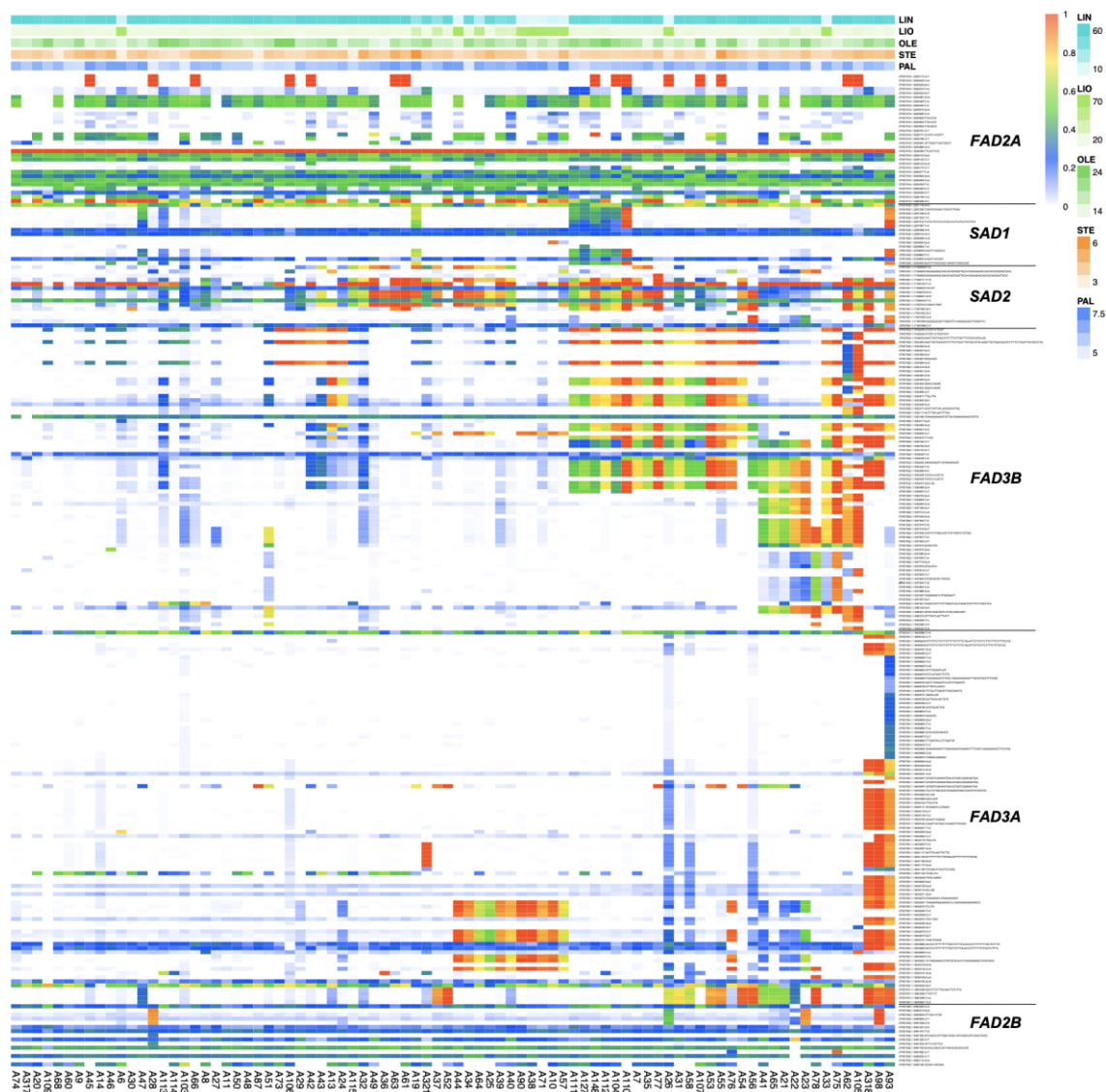


Рис. 1. Полиморфизмы генов семейств *SAD* и *FAD* (белый цвет – преобладание референсного аллеля; красный цвет - преобладание альтернативного аллеля, другие цвета - промежуточные варианты соотношения референсного и альтернативного аллелей).

На основе полученных данных о полиморфизмах генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A* и *FAD3B*, разработаны генетические паспорта для сортов/линий льна, которые являются основой для проверки соответствия генотипов сорту/линии, контроля сортовой чистоты и подбора родительских пар для скрещиваний с целью получения сортов с заданным жирнокислотным составом масла.

В изученной коллекции *L. usitatissimum* вариабельность содержания LIN и LIO была самой высокой (24- и 6-кратное различие соответственно) по сравнению с содержанием OLE, STE и PAL (от 1,5 до 2,4-кратного различия), что может быть причиной выявления наибольшего количества полиморфизма для генов *FAD3A* и *FAD3B*. Полученный результат согласуется с данными других исследований [Rajwade и др., 2016; Thambugala и др., 2013].

Необходимо отметить, что подвыборка из 84 сортов, контрастных по жирнокислотному составу масла, максимально охватывала разнообразие генов семейств *FAD* и *SAD*, для остальных проанализированных сортов и линий льна не удалось выявить полиморфизмы, которые отсутствовали бы в этой подвыборке. Полиморфизм генов *SAD* и *FAD* у льна-долгунца был значительно ниже по сравнению с масличным льном.

1.1.2 Полиморфизмы, ассоциированные с жирнокислотным составом масла

Корреляционный анализ между жирнокислотным составом льняного масла и частотой полиморфизмов в генах *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A* и *FAD3B*, выявил ассоциации между конкретными полиморфизмами и содержанием PAL, STE, OLE, LIO и LIN. В таблице 2 представлены данные об ассоциированных с ЖК-составом полиморфизмах с коэффициентами корреляции Спирмена более 0,25 или менее -0,25 (р-значение <0,05): обнаружен 31 такой полиморфизм, в том числе, 29 SNP и 2 вставки/делеции. Для *SAD1* 5 SNP коррелировали с ЖК составом, для *SAD2* – 2 SNP и 1 вставка, для *FAD2A* – 4 SNP, для *FAD2B* – 0, для *FAD3A* - 8 SNP и 1 вставка, для *FAD3B* - 11 SNP.

Выявленные корреляционные связи были преимущественно между полиморфизмами генов десатураз и содержанием тех жирных кислот, в биосинтез которых они вовлечены: 5 полиморфизмов в *SAD1* были связаны с содержанием OLE, 2 в *SAD2* - с содержанием STE и OLE, 9 в *FAD3A* и 10 в *FAD3B* - с содержанием LIO и LIN. Однако, также были выделены полиморфизмы, ассоциированные с ЖК, в биосинтезе которых они не участвуют. Большинство полиморфизмов, ассоциированных с жирнокислотным составом, идентифицировали в генах *FAD3A* и *FAD3B*. Сходные результаты получены в исследованиях [Thambugala и др., 2013] и [You и др., 2018b].

Таблица 2. Корреляции полиморфизмов генов *SAD* и *FAD*, и жирнокислотного состава масла льна (содержания *PAL*, *STE*, *OLE*, *LIO*, *LIN*).

Сайт	Расположение	Реф.	Альт.	ЖК*	VarScan	
					r_s	p -значене
<i>SAD1</i> (Chr. CP027620.1)						
2257330	Перед геном	A	G	OLE	0.37	4x10 ⁻⁴
2257353	Перед геном	T	C	OLE	0.35	0.001
2258938	Интрон	C	G	OLE	0.27	0.01
2258980	Интрон	T	C	OLE	0.34	0.002
2260202	После гена	T	G	OLE	0.35	0.001
<i>SAD2</i> (Chr. CP027621.1)						
17364707	Перед геном	A	AG	OLE	0.31	0.004
17365164	Экзон	T	C	STE	0.31	0.009
17367383	После гена	G	T	LIO	-0.28	0.01
<i>FAD2A</i> (Chr. CP027619.1)						
5296364	Экзон	G	A	PAL	-0.28	0.01
5296364	Экзон	G	A	STE	-0.34	0.002
5296658	Экзон	C	T	STE	0.38	4x10 ⁻⁴
5296705	После гена	C	A	PAL	0.35	0.006
<i>FAD3A</i> (Chr. CP027631.1)						
16092241	Интрон	T	C	LIN	-0.32	0.003
16092241	Интрон	T	C	PAL	0.29	0.008
16092273	Интрон	C	A	LIN	-0.32	0.003
16092294	Экзон	T	C	LIN	-0.33	0.002
16092294	Экзон	T	C	LIO	0.32	0.003
16092294	Экзон	T	C	PAL	0.36	7x10 ⁻⁴
16092348	Экзон	C	T	LIN	-0.29	0.008
16092348	Экзон	C	T	PAL	0.33	0.002
16092575	Экзон	C	T	LIN	-0.36	8x10 ⁻⁴
16092674	Перед геном	G	T	LIN	-0.32	0.003
16092741	Перед геном	T	TAA	LIN	-0.37	5x10 ⁻⁴
16093029	Перед геном	T	C	LIN	-0.36	8x10 ⁻⁴
16093029	Перед геном	T	C	LIO	0.27	0.01
16093029	Перед геном	T	C	PAL	0.39	3x10 ⁻⁴
16093040	Перед геном	C	A	LIN	-0.43	7x10 ⁻⁶
16093040	Перед геном	C	A	LIO	0.33	0.003
16093040	Перед геном	C	A	PAL	0.35	0.001
<i>FAD3B</i> (Chr. CP027622.1)						
1034358	Перед геном	G	A	STE	0.28	0.01
1034389	Перед геном	A	G	LIO	-0.31	0.005
1034526	Перед геном	C	A	LIO	-0.26	0.02
1034873	Перед геном	G	A	LIO	-0.32	0.004
1034904	Перед геном	G	C	LIO	-0.30	0.007
1035028	Перед геном	A	G	LIO	-0.28	0.01
1035480	Экзон	A	G	LIO	-0.29	0.007
1035655	Экзон	C	T	LIO	0.28	0.009
1035655	Экзон	C	T	LIN	-0.37	6x10 ⁻⁴
1035655	Экзон	C	T	PAL	0.32	0.003
1035674	Экзон	T	G	LIO	-0.26	0.02
1036195	Интрон	G	T	LIO	-0.26	0.02
1037964	Интрон	G	C	LIN	-0.26	0.02

* - жирная кислота, содержание которой ассоциировано с полиморфизмом.

VAF для всех пяти полиморфизмов *SAD1*, ассоциированных с ЖК, имела близкие соотношения в каждом отдельном генотипе, в то время как для *SAD2*, такая закономерность не наблюдалась. Для полиморфизмов гена *FAD2A*, ассоциированных с ЖК, VAF варьировала от 0,2 до 0,5 в большинстве образцов и расходилась в отдельных образцах. Для гена *FAD3A*, VAF всех 9 полиморфизмов, ассоциированных с содержанием LIN, была близка в большинстве образцов, однако, в некоторых образцах для сайтов 16092294, 16092348, 16093029 и 16093040 были выявлены значительные различия в VAF по сравнению с другими сайтами: в образцах A58, A26, A56, A318, A98 и A93 - для всех четырех сайтов, и в A76, A22 и A23 - только для сайта 16092348. Эти сайты также были связаны с содержанием ЛЮ. Большая вариация VAF, ассоциированных с жирнокислотным составом полиморфизмов, выявлена в отдельных сортах/линиях для гена *FAD3B*.

Следует отметить, что по всем полиморфизмам, ассоциированным с ЖК составом масла, по крайней мере, один или несколько сортов/линий льна, были неоднородными, что свидетельствует о сортовой гетерогенности, и может приводить к различиям в составе жирных кислот для образцов одного сорта/линии.

1.1.3 Аминокислотные замены, вызванные ассоциированными с ЖК составом полиморфизмами

Идентифицированные нами полиморфизмы, ассоциированные с ЖК составом, локализованы в разных областях генов - промоторах, экзонах и интронах. Поиск аминокислотных замен, вызванных этими полиморфизмами в генах *SAD* и *FAD*, выявил миссенс-мутации в генах *SAD2*, *FAD3A* и *FAD3B* и нонсенс-мутацию в *FAD3A* (табл. 3).

В *SAD1* SNP идентифицированы в интроне (2), перед геном (2) и после гена (1). В *SAD2* полиморфизмы локализованы в экзоне (замена GGT на AGT в CP027621.1: 17365164 приводит к замене глицина на серин), интроне и после гена. В *FAD2A* в экзоне обнаружены два SNP, не приводящих к замене аминокислот, и один - после гена. В *FAD3A*, два SNP приводят к замене аминокислот: аланина на треонин (CP027631.1: 16092575) и триптофана на стоп-кодон (CP027631.1: 16092348). В *FAD3B*, замена CAT на TAT приводит к замене гистидина на тирозин (CP027622.1: 1035655), а замена ATC на AGC приводит к замене серина на изолейцин (CP027622.1: 1035674). Другие идентифицированные в *FAD3A* и *FAD3B* генах полиморфизмы не приводят к аминокислотным заменам.

Таблица 3. Аминокислотные замены, вызванные полиморфизмами генов *SAD* и *FAD*, ассоциированными с ЖК составом

Сайт	Реф.	Альт.	Реф. аминокислота	Альт. аминокислота
<i>SAD2 (Chr. CP027621.1)</i>				
17365164	T	C	Glycine	Serine
<i>FAD2A (Chr. CP027619.1)</i>				
5296364	G	A	Lysine	Lysine
5296658	C	T	Asparagine	Asparagine
<i>FAD3A (Chr. CP027631.1)</i>				
16092294	T	C	Glycine	Glycine
16092348	C	T	Tryptophan	Stop
16092575	C	T	Alanine	Threonine
<i>FAD3B (Chr. CP027622.1)</i>				
1035480	A	G	Alanine	Alanine
1035655	C	T	Histidine	Tyrosine
1035674	T	G	Serine	Isoleucine

Полиморфизмы, приводящие к аминокислотным заменам в генах *SAD*, *FAD2* и *FAD3*, а также появлению стоп-кодонов в генах *FAD3*, выявлены ранее в ряде исследований [Banik, Duguid, Cloutier, 2011; Porokhovinova и др., 2019; Rajwade и др., 2014; Thambugala и др., 2013; Vrinten и др., 2005], и наши результаты согласуются с ними.

1.1.4 Кластеризация изучаемых сортов и линий на основе полиморфизмов генов *SAD* и *FAD*

Кластеризацию сортов/линий льна проводили на основе полиморфизмов генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A* и *FAD3B* (Рис. 2). На данной дендрограмме выявлен кластер, включающий 9 из 10 сортов/линий с низким содержанием LIN и высоким содержанием LIO (A57, A10, A71, A38, A90, A34, A44, A25 и A64). В данном кластере не находился только один сорт с низким содержанием LIN (A6). В других кластерах не выявлены четкие ассоциации с ЖК-составом, что могло быть связано с преобладанием полиморфизмов в генах *FAD3A* и *FAD3B*, участвующих в десатурации LIO в LIN.

Анализ полиморфизмов в генах *FAD3A* и *FAD3B* выполнен для образца A6 с низким содержанием LIN, не входящего в кластер других сортов/линий с низким содержанием LIN. Мы обнаружили один полиморфизм в гене *FAD3A* (сайт CP027631.1: 16090340), приводящий к замене аргинина на стоп-кодон, который может объяснить низкое содержание LIN. Данный полиморфизм также идентифицирован в образцах A22, A42 и A64, однако с низкой VAF, около 5–10%, в то время как в A6, около 80% растений имели данный SNP.

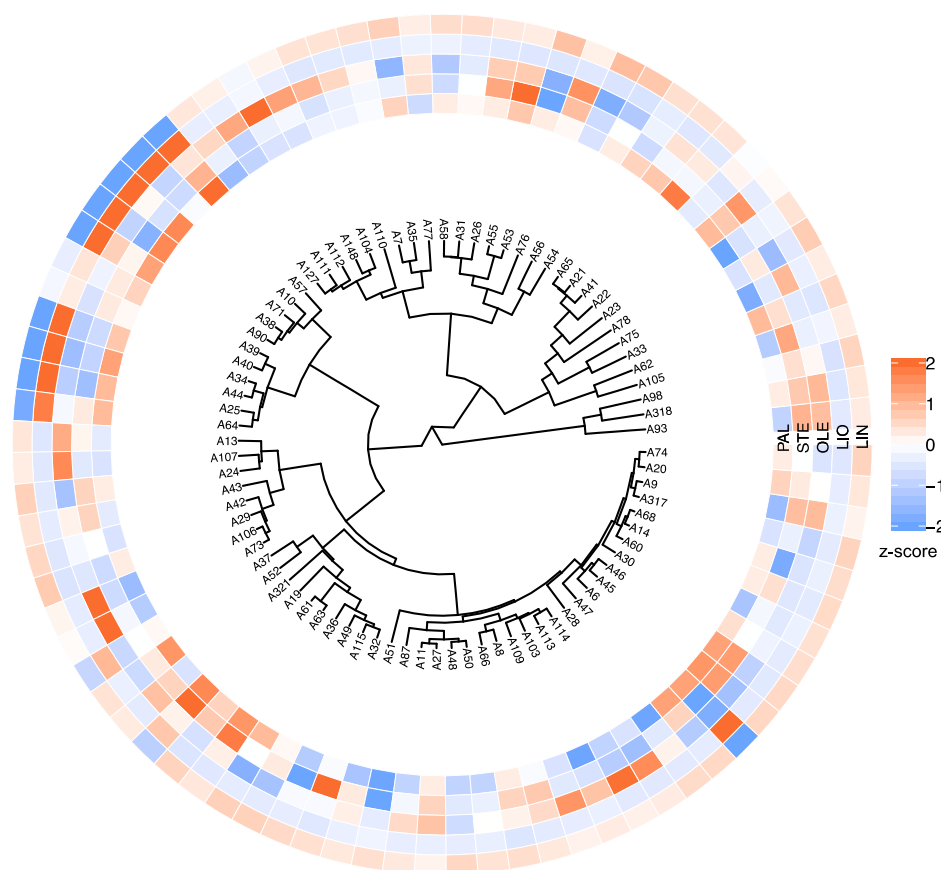


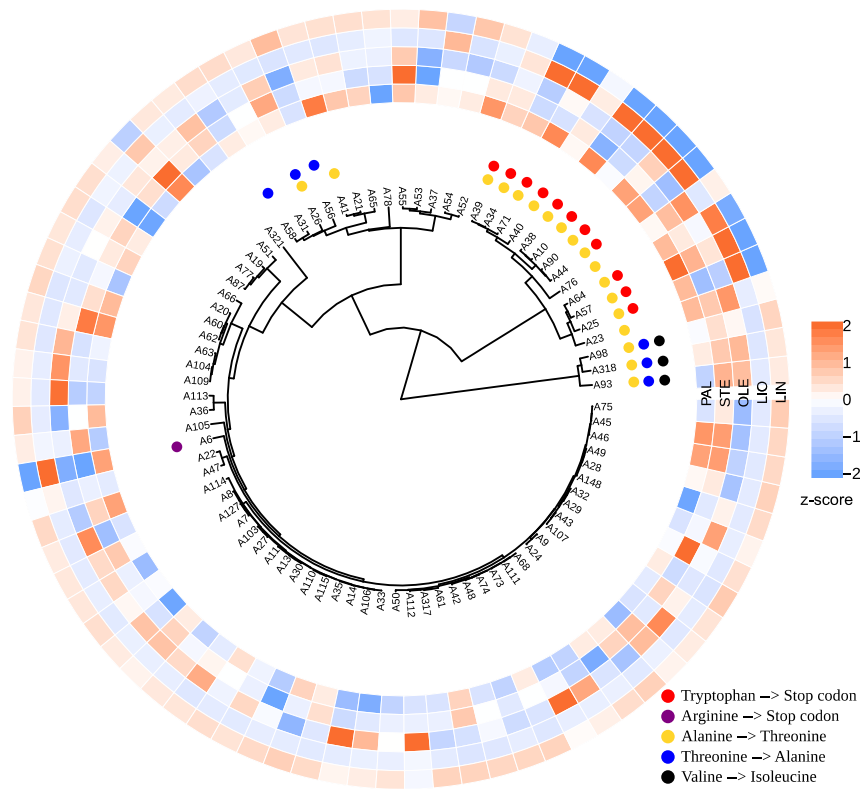
Рис. 2. Кластеризация 84 сортов/линий на основе полиморфизмов генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, и *FAD3B* (*VarScan*)

Анализ полиморфизмов гена *FAD3A*, ассоциированных с ЖК-составом, показал, что образцы с низким и средним содержаниями LIN, имеющие стоп-кодон из-за замены в сайте CP027631.1: 16092348, также имели остальные 8, ассоциированных с жирнокислотным составом, полиморфизмов *FAD3A*. Однако, два образца, A76 и A23, несущие все полиморфизмы, ассоциированные с жирнокислотным составом в *FAD3A*, кроме замены в сайте CP027631.1: 16092348, имели высокое содержание LIN, что указывает на ключевую роль полиморфизма, приводящего к замене аргинина на стоп-кодон, в определении низкого содержания LIN и высокого содержания LIO у сортов/линий льна.

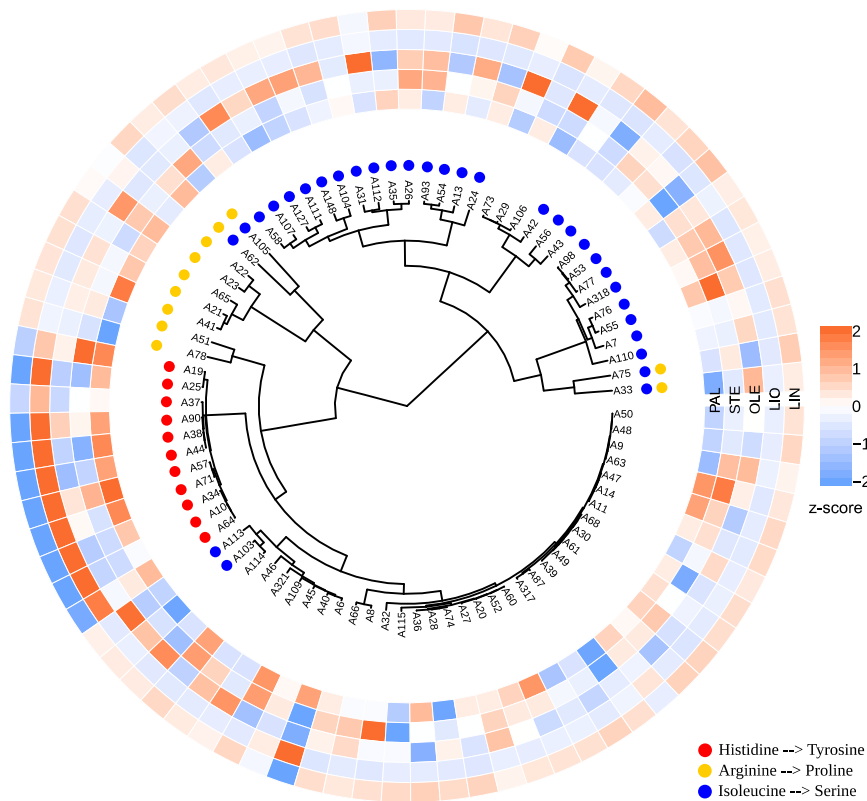
Большинство полиморфизмов гена *FAD3B* имели отрицательную корреляцию с содержанием LIO. Только один, ассоциированный с жирнокислотным составом полиморфизм в гене *FAD3B* (сайт CP027622.1: 1035655), имел отрицательную корреляцию с содержанием LIN и приводил к замене гистидина на тирозин. Вероятно, данный SNP, представленный в образцах с низким и средним содержаниями LIN, при отсутствии в них остальных 10 ассоциированных с жирнокислотным составом полиморфизмов гена *FAD3B*, играет ключевую роль в определении низкого содержания линоленовой кислоты в льняном масле.

Мы также провели кластеризацию образцов льна на основе полиморфизмов в отдельных генах: *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A* и *FAD3B*. Полиморфизмы, идентифицированные методами VarScan и freeBayes, давали сходные дендрограммы для изученных генов. Кластеризации для *FAD3A* и *FAD3B* представлены на рис. 3. Большинство образцов с низким содержанием LIN и высоким содержанием LIO, имели мутацию в CP027631.1: 16092348, приводящую к появлению стоп-кодона в *FAD3A*, и мутацию в CP027622.1: 1035655, приводящую к замене гистидина на тирозин в *FAD3B*, тогда как все сорта и линии с высоким LIN не имели этой мутации. Образцы A39 и A40 с содержанием LIN – 43,5% и 46,8% соответственно, имели только мутацию, приводящую к стоп-кодону в *FAD3A*, в то время, как A19 и A37 с содержанием LIN – 32,6% и 35% имели только мутацию, приводящую к замене гистидина на тирозин в *FAD3B*. Таким образом, образцы, только с одной из этих двух мутаций, имели среднее содержание LIN. Образец A6, низколиноленовый сорт (9,4% LIN), не имел ни одной из вышеуказанных мутаций в генах *FAD3A* и *FAD3B*, но, как уже сказано, имел другую мутацию на сайте CP027631.1: 16090340 в *FAD3A*, в результате которой формируется стоп-кодон вместо аргинина, и мы предполагаем, что для этого сорта данная мутация является вероятной причиной низкого содержания LIN, несмотря на то, что полиморфизмы в *FAD3B* не выявлены.

Несмотря , что значительное число полиморфизмов ассоциировано с ЖК составом в генах *FAD3A* и *FAD3B*, именно три полиморфизма играли ключевую роль в определении соотношения линолевой и линоленовой кислот в нашей выборке сортов/линий: полиморфизм в первом экзоне гена *FAD3A* (замена G на A в сайте CP027631.1:16092348, красные точки на Рис. 3. А), приводящий к замене триптофана на стоп-кодон (результат замены триплета TGG на TGA); полиморфизм во втором экзоне гена *FAD3B* (замена C на T в сайте CP027622.1:1035655, красные точки на Рис. 3.В), приводящий к замене гистидина на тирозин (результат замены триплета CAT на TAT); и полиморфизм в пятом экзоне гена *FAD3A* (замена C на T в сайте CP027631.1:16090340, фиолетовые точки на рис. 3. А), приводящий к замене аргинина на стоп-кодон (результат замены триплета CGA на TGA). Данные полиморфизмы представляют наибольший интерес для использования в маркер-ориентированной селекции масличного льна специального назначения с определенным содержанием LIN и LIO: с использованием ДНК-маркеров можно выбрать растения, которые являются гомозиготными по генам *FAD3A* и *FAD3B*, и имеют варианты аллелей, ассоциированные с высоким, низким или средним содержанием линоленовой кислоты уже на стадии F₂ или F₃ гибридов, что значительно повысит эффективность селекции.



А. Кластеризация 84 сортов/линий на основе полиморфизмов гена *FAD3A* (VarScan)



В. Кластеризация 84 сортов/линий на основе полиморфизмов гена *FAD3B* (VarScan)

Рис. 3. Кластеризация 84 сортов/линий льна на основе полиморфизмов генов *FAD3A* и *FAD3B*

1.1.5 Селекция льна по содержанию линоленовой кислоты в семенах льна на основе полиморфизмов в генах *FAD3*

Наши результаты по оценке полиморфизма генов десатураз ЖК выявили, что сорта/линии льна с низким содержанием LIN и высоким содержанием ЛЮ, имели мутацию в CP027631.1: 16092348, приводящую к появлению стоп-кодона в *FAD3A* и мутацию в CP027622.1: 1035655, приводящую к замене гистидина на тирозин в *FAD3B*, а сорта/линии со средним содержанием LIN имели только одну из этих двух мутаций. Была поставлена задача использовать эти результаты для создания улучшенных сортов льна пищевого назначения с низким или средним содержанием линоленовой кислоты. Для выполнения данной задачи были выбраны родительские сорта/линии ЛМ98, Raciol, AGT 427/10.

ЛМ98 - один из немногих российских сортов льна пищевого назначения с низким содержанием LIN (имеет два ключевых полиморфизма, приводящих к появлению стоп-кодона в *FAD3A* и замене гистидина на тирозин в *FAD3B*), однако, он неустойчив к патогену *Fusarium oxysporum* f.sp. lini, поэтому его целесообразно использовать для создания других сортов пищевого использования с низким или средним содержанием линоленовой кислоты. Сорт Raciol отличается редким для льна средним содержанием линоленовой кислоты, определяющимся полиморфизмом, приводящим к замене гистидина на тирозин в *FAD3B*, также перспективен для селекционной работы. Линия AGT 427/10 представляет интерес по комплексу хозяйственно ценных признаков, но является высоколиноленовой, поэтому ее нужно скрещивать с низко- или среднелиноленовыми сортами/линиями в селекционных работах для создания сортов пищевого назначения с низким или средним содержанием LIN.

В Институте льна проведены скрещивания сортов льна Raciol x AGT 427/10 и ЛМ 98 x AGT 427/10. После скрещивания и получения поколения F₃, нами собран материал с 20 растений от каждой семьи, и секвенированы участки генов *FAD3A* и *FAD3B*, в которых локализованы два ключевых полиморфизма, определяющих содержание LIN и ЛЮ. По данным секвенирования определены семьи, гомозиготные по ключевым полиморфизмам генов *FAD3A* и *FAD3B* (рис.4, 5), и несущие нужные аллельные варианты, приводящие к среднему или низкому содержанию LIN, что повысит эффективность селекции сортов льна с заданным ЖК составом. В настоящее время в Институте льна из этих семей ведут отбор наиболее перспективных генотипов, обладающих комплексом хозяйственно ценных признаков, которые в дальнейшем могут стать сортами льна пищевого назначения.

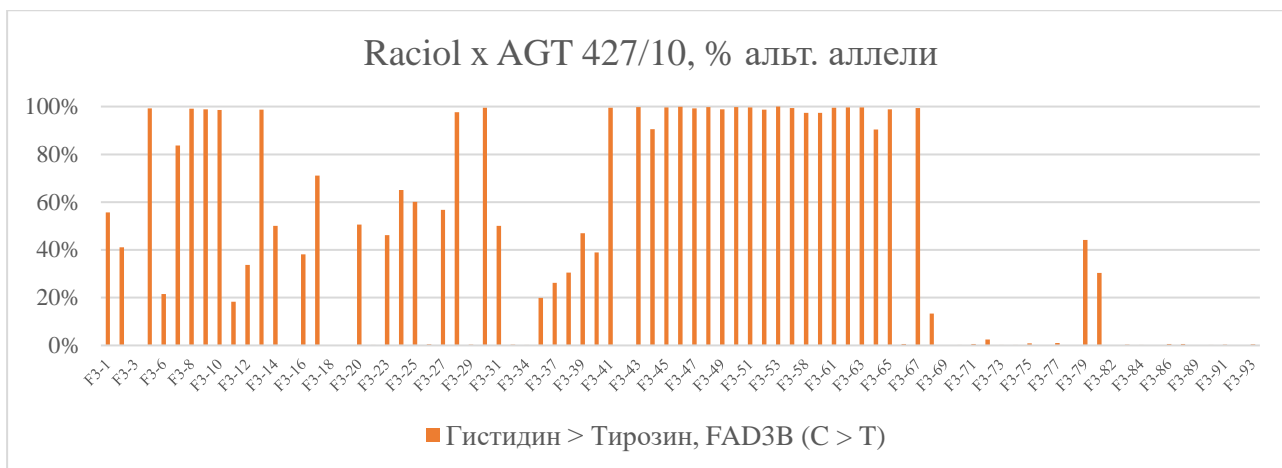


Рис. 4. Полиморфизм гена *FAD3B* при скрещивании *Raciol* x *AGT 427/10*

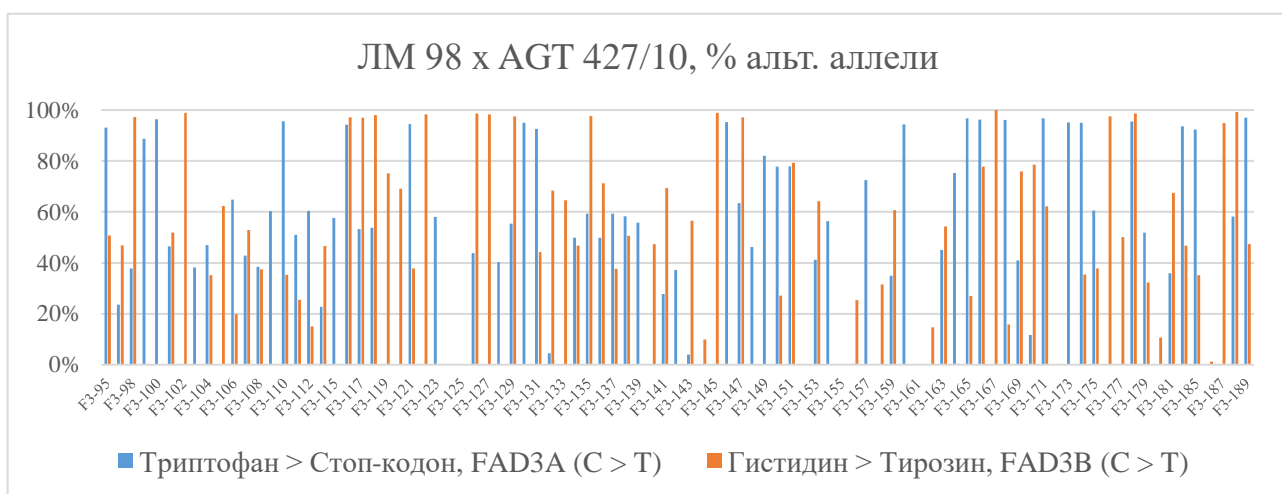


Рис. 5. Полиморфизмы генов *FAD3A* и *FAD3B* при скрещивании *ЛМ 98* x *AGT 427/10*

1.2 Полиморфизмы гена *UGT74S1*, вовлеченного в синтез лигнанов

Полученные нами методом глубокого секвенирования последовательности гена *UGT74S1*, участвующего в биосинтезе SDG, картировали на геном *L. usitatissimum* (GCA_000224295.2/ASM22429v2), что позволило нам уменьшить вероятность ошибочной идентификации полиморфизмов в изучаемом гене из-за неправильного праймирования и секвенирования псевдогенов с высокой гомологией на уровне ДНК.

Для гена *UGT74S1* идентифицированы 30 полиморфизмов (метод VarScan), из которых 9 находились в экзонах, 14 - в интронах, 6 - до гена и 1 - после гена. Однако, как видно из рис.6А, у большинства образцов отсутствовали полиморфизмы с высоким значением VAF. Более того, корреляционный анализ между выявленными полиморфизмами гена *UGT74S1* и содержанием лигнанов в семенах льна, не выявил значимых ассоциаций, и кластеризация сортов/линий льна на основе полиморфизмов гена *UGT74S1* не позволила выделить кластеры

с различным содержанием лигнанов (рис. 6.В). Таким образом, в исследованной выборке сортов/линий льна, полиморфизмы *UGT74S1* не оказывали существенного влияния на содержание лигнанов в семенах льна, и, вероятно, полиморфизмы других генов, вовлеченных в синтез лигнанов, имели большее влияние на данный признак.

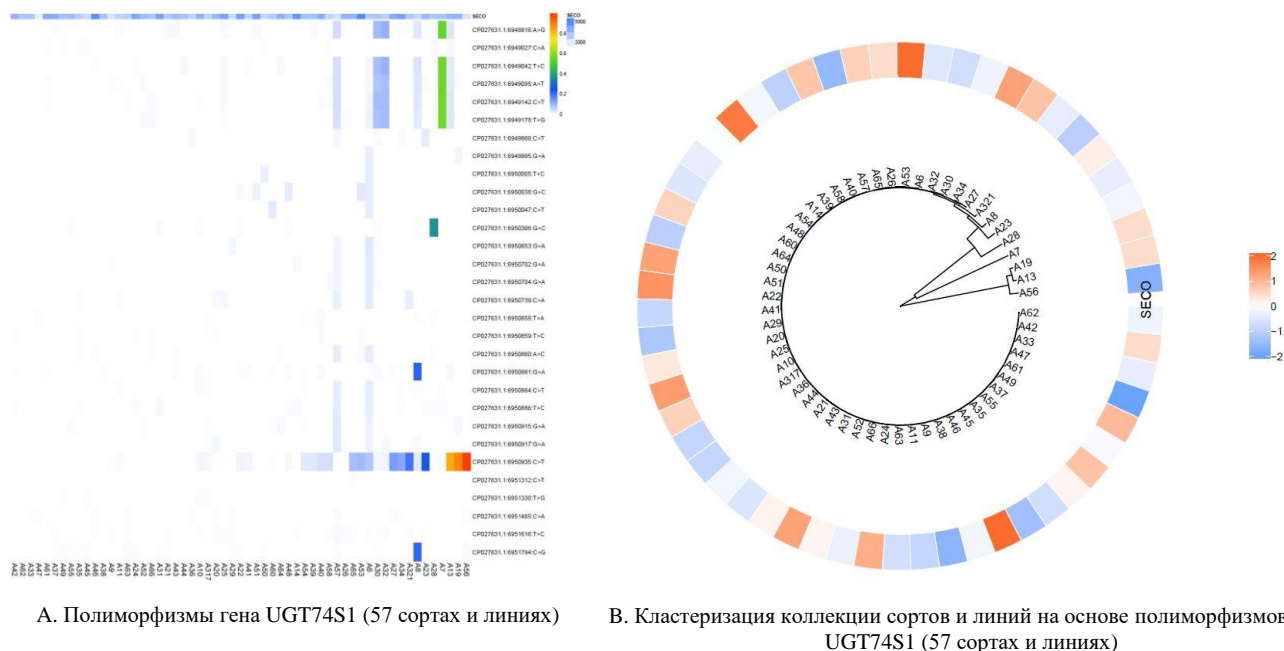


Рис. 6. Полиморфизмы гена *UGT74S1*

2. Гетерогенность сортов льна

Секвенирование пулов растений (около 50 для каждого сорта и линии) позволило оценить гетерогенность сортов/линий. Методами freeBayes и VarScan мы обнаружили, что большинство образцов показали гетерогенность, по крайней мере, для одного из изучаемых генов.

1.3 Внутрисортная гетерогенность по генам десатураз жирных кислот

Данные секвенирования генов десатураз ЖК выявили внутрисортную гетерогенность по этим генам, как показано на тепловой карте (рис. 1). Ген *FAD3B* является наиболее показательным примером этой гетерогенности, так как в некоторых образцах преобладали референсные аллели, в других образцах преобладали альтернативные аллели, а также были обнаружены образцы с референсными и альтернативными аллелями с близкими соотношениями.

Генетическая гетерогенность сортов/линий по генам десатураз ЖК указывает на возможность использования данных секвенирования генов *SAD* и *FAD* в селекции и семеноводстве льна. Полученные результаты перспективны для разработки дальнейших

стратегий отбора селекционного материала. Более того, наши данные показывают необходимость разработки маркерной системы для сертификации сортов/линий и контроля сортовой чистоты. Даже по ключевым полиморфизмам генов *FAD3A* и *FAD3B*, определяющим содержание LIN и LIO, некоторые сорта/линии оказались неоднородны (табл. 4), поэтому следует на основе ДНК-маркеров проводить очистку или разделение на биотипы сортов/линий льна для таких образцов, как, например, А57, в котором только 57% растений имели полиморфизм, приводящий к замене триптофана на стоп-кодон, что свидетельствует о различиях по ЖК-составу между растениями этого образца.

Таблица 4. Гетерогенность сортов по ключевым полиморфизмам, определяющим содержание линолевой и линоленовой кислот льнового масла

№	LIO, %	LIN, %	Аргинин > Стоп-кон, <i>FAD3A</i> , %	Триптофан > Стоп-кодон, <i>FAD3A</i> , %	Гистидин > Тирозин, <i>FAD3B</i> , %
А6	68,0	9,4	77	0	0
А19	35,5	32,6	0	0	71
А21	13,6	56,8	0	14	19
А24	14,3	60,6	0	16	0
А25	72,4	2,7	0	57	72
А39	28,6	43,5	2	85	0
А40	26,7	46,8	0	83	0
А57	64,0	4,9	0	69	83
А64	53,1	18,5	5	67	92
А71	66,8	4,5	0	82	81

3. Геном сорта льна Атлант

В результате секвенирования генома льна-долгунца сорта Атлант получено 23-кратное покрытие (8,4 млрд. п.н., N50 = 12 тыс. н.) на платформе Oxford Nanopore Technologies и 30-кратное покрытие (22,6 млн прочтений длиной 250+250 н.) на платформе Illumina. Наилучшая сборка генома получена с использованием Canu (длина 361,7 млн п.н., N50=350 т.п.н.). После полировки по схеме Rason+Medaka+POLCA полнота сборки по оценке BUSCO составила 97.40%. Полученный геном использовался нами при подборе праймеров для амплификации генов семейств *SAD*, *FAD*, *UGT*, а также при анализе данных секвенирования этих генов. Кроме того, полученный геном сорта Атлант представляет ценность для оценки полиморфизма *L. usitatissimum* на уровне генома, идентификации последовательностей, специфичных для льна-долгунца, и использования в качестве референса в исследованиях льна-долгунца, анализа метилирования ДНК льна на полногеномном уровне, а также для разработки методов геномной селекции и редактирования генома льна.

ЗАКЛЮЧЕНИЕ

Нами впервые проведена оценка полиморфизма генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, играющих ключевую роль в синтезе жирных кислот, и *UGT74S1*, вовлеченного в синтез лигнанов, на выборке из 279 сортов/линий льна методом высокопроизводительного секвенирования. В процессе выполнения данной работы подобраны праймеры для подготовки ДНК-библиотек и оптимизированы условия проведения ПЦР. В результате получены ДНК-библиотеки генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1* высокого качества для 279 сортов/линий льна, которые затем секвенировали на MiSeq (Illumina). Высокопроизводительное секвенирование позволило с высокой точностью не только идентифицировать полиморфизмы исследуемых генов, но и изучить гетерогенность сортов/линий льна.

Корреляционный анализ между содержанием жирных кислот и полиморфизмами генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, а также содержанием лигнанов и полиморфизмами *UGT74S1*, позволил идентифицировать полиморфизмы генов *FAD3A* и *FAD3B*, играющие ключевую роль в определении содержания линоленовой и линолевой кислот. Наши результаты позволили выбрать из селекционного материала F₃ (*Raciol* x AGT 427/10 и ЛМ 98 x AGT 427/10) семьи, гомозиготные по *FAD3A* и *FAD3B* генам с аллельными вариантами, определяющими низкое или среднее содержание линоленовой кислоты. Полученные данные также перспективны для разработки молекулярно-генетических маркеров для идентификации определенных аллелей генов *FAD3A* и *FAD3B*, представляющих селекционную ценность.

Таким образом, идентификация определенных аллельных вариантов генов *FAD3A* и *FAD3B* может быть использована для эффективной и быстрой селекции новых сортов льна с желаемым содержанием линоленовой и линолевой кислот для пищевой, фармацевтической или технической индустрии. Более того, данные о полиморфизмах генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* могут быть применены для разработки подходов, основанных на CRISPR, направленных на изменение жирнокислотного состава льняного масла.

Впервые получен и аннотирован геном льна сорта Атлант с использованием платформ Oxford Nanopore Technologies и Illumina, что важно для оценки полиморфизма *L. usitatissimum* на уровне генома, а также для разработки подходов для геномной и маркер-ориентированной селекции и редактирования генома льна.

ВЫВОДЫ

1. Методом глубокого секвенирования определены последовательности генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1*, вовлеченных в синтез жирных кислот и лигнанов, для представительной коллекции из 279 сотов/линий льна.
2. Идентифицированы ДНК-полиморфизмы генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1*, и проведена оценка генетического разнообразия на основе данных глубокого секвенирования. Показано, что наиболее полиморфными являются гены *FAD3A* и *FAD3B* (101 и 71 полиморфизм, соответственно).
3. Показано низкое разнообразие льна по гену *UGT74S1* и отсутствие существенного влияния полиморфизмов этого гена на содержание лигнанов в семенах льна у изученной выборки.
4. Установлена гетерогенность сортов/линий льна по генам, вовлеченным в синтез жирных кислот.
5. Выявлены полиморфизмы, играющие ключевую роль в определении содержания линолевой и линоленовой кислот в масле льна: замена С на Т (CP027631.1: 16092348), приводящая к замене триптофана на стоп-кодон, и G на A (CP027631.1: 16090340), приводящая к замене аргинина на стоп-кодон, для гена *FAD3A*, и замена С на Т (CP027622.1: 1035655), приводящая к замене гистидина на тирозин, для гена *FAD3B*.
6. Созданы ДНК-паспорта сортов/линий льна, содержащие информацию об аллельных вариантах генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1*.
7. Предложен подход для использования данных о полиморфизмах генов *FAD3A* и *FAD3B* в маркер-ориентированной селекции льна для идентификации высоко-, средне- и низколиноленовых генотипов, оценки гетерогенности сортов/линий, отбора генотипов с нужными аллельными вариантами генов *FAD3A* и *FAD3B* в гомозиготном состоянии, а также подбора пар для скрещиваний.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Dmitriev, A.A., **Kezimana, P.**, Rozhmina, T.A. Zhuchenko A.A, Povkhova L.V., Pushkova E.N., Novakovskiy R.O., Pavelek M., Vladimirov G.N., Nikolaev E.N., Kovaleva O.A., Kostyukevich Y.I., Chagovets V.V., Romanova E.V., Snezhkina A.V., Kudryavtseva A.V., Krasnov G.S., Melnikova N.V. (2020). Genetic diversity of *SAD* and *FAD* genes responsible for the fatty acid composition in flax cultivars and lines. BMC Plant Biology 20, 301.
2. **Kezimana, P.**, Dmitriev, A. A., Kudryavtseva, A. V., Romanova, E. V., Melnikova, N. V. (2018). Secoisolariciresinol diglucoside of flaxseed and its metabolites: Biosynthesis and potential for nutraceuticals. Frontiers in Genetics, 9, 641.
3. Dmitriev, A. A., Pushkova, E. N., Novakovskiy, R. O., Beniaminov, A. D., Rozhmina, T. A., Zhuchenko, A. A., Bolsheva, N. L., Muravenko, O. V., Povkhova, L. V., Dvorianinova, E. M.,

- Kezimana, P.,** Snezhkina, A. V., Kudryavtseva, A. V., Krasnov, G. S., Melnikova, N. V. (2021). Genome Sequencing of Fiber Flax Cultivar Atlant Using Oxford Nanopore and Illumina Platforms. *Frontiers in Genetics*, 11: 590282.
4. **Кезимана, П.,** Дмитриев, А. А., Мельникова, Н. В., Романова, Е. В. (2021). Молекулярно-генетическая характеристика льна (*Linum usitatissimum* L.). Теоретические и прикладные проблемы агропромышленного комплекса, 47(1), 39-42
 5. **Kezimana, P.,** Dmitriev, A., Rozhmina, T., Novakovskiy, R., Krasnov, G., Romanova, E., Melnikova, N. (2018). Polymorphism of glycosyltransferase *UGT74S1* gene in flax (*Linum usitatissimum* L.). *FEBS OPEN BIO* (Vol. 8, pp. 469-469).
 6. **Kezimana, P.,** Dmitriev, A. A., Rozhmina, T. A., Novakovskiy, R. O., Romanova, E. V., Kudryavtseva, A. V., Melnikova, N. V. (2018). Assessment of genetic variation for *SAD* and *FAD* genes in flax by high-throughput sequencing. *FEBS OPEN BIO* (Vol. 8, pp. 466-466).
 7. **Kezimana, P.,** Rozhmina, T. A., Novakovskiy, R. O., Pushkova, E. N., Povkhova, L. V., Romanova, E. V., Melnikova, N. V. (2019). Genetic control of lipid biosynthesis in a collection of flax (*Linum usitatissimum* L.) cultivars and lines. In *FEBS OPEN BIO* (Vol. 9, pp. 101-101).
 8. **Кезимана, П.,** Рожмина, Т.А., Краснов, Г.С., Повхова, Л.В., Новаковский, Р.О., Пушкова, Е. Н., Жученко, А. А., Vjelková, M., Pavelek, M., Дмитриев, А. А., Мельникова, Н. В. (2021) Оценка полиморфизма генов семейств *SAD* и *FAD* у сортов и линий льна (*Linum usitatissimum* L.) методом глубокого секвенирования. Теория и практика адаптивной селекции растений (Жученковские чтения VI): сб. научн. трудов по материалам Межд. научн.-практ. конф. С. 49-52.
 9. **Kezimana, P.,** Dmitriev, A. A., Melnikova, N. V., Romanova, E. V. (2018). α -linoleic acid: biosynthesis of a fatty acid with nutraceutical and pharmacological effects in flaxseed (*Linum usitatissimum*). In *Гармонизация подходов к фармацевтической разработке* (pp. 17-19).
 10. **Kezimana, P.,** Dmitriev, A., Rozhmina, T., Novakovskiy, R., Romanova, E., Kudryavtseva, A., Melnikova, N. (2018). Polymorphism in genes related to fatty acid composition in *Linum usitatissimum*. *Biodiversity: genomics and evolution (BioGenEvo-2018)*, 22.
 11. **Kezimana, P.,** Rozhmina, T. A., Krasnov, G. S., Novakovskiy, R. O., Povkhova, L. V., Pushkova, E. N., Melnikova, N. V. (2019). Genetic diversity of genes involved in fatty acid biosynthesis in a collection of flax cultivars. In *Plant Genetics, Genomics, Bioinformatics, and Biotechnology (PlantGen2019)* (pp. 99-99).
 12. **Kezimana, P.,** Rozhmina, T. A., Krasnov, G. S., Novakovskiy, R. O., Povkhova, L. V., Pushkova, E. N., Melnikova, N. V. (2019). Preparation of DNA libraries for sequencing of *SAD* and *FAD* genes in flax. *Current Challenges in Plant Genetics, Genomics, Bioinformatics, and Biotechnology*, 24, 153.
 13. **Kezimana, P.,** The pharmacological properties of flaxseed lignans. Сборник научных трудов IV Гаммермановские чтения: 30–31 января 2019 г., Москва: РУСАЙНС, 2019, 144-149

Список сокращений

FAD – fatty acid desaturase (десатуразы жирных кислот)	SDG – secoisolariciresinol diglucoside (секоизоларицирезинола диглюкозид)
LIN – linolenic acid (Линоленовая кислота)	SECO – secoisolariciresinol (секоизоларицирезинол)

LIO – linoleic acid (линолевая кислота)
OLE – oleic acid (олеиновая кислота)
PAL – palmitic acid (пальмитиновая кислота)
SAD – stearyl-ACP (acyl-carrier-protein) desaturase (стеарилацилпереносящий белок (АПБ)-десатуразы)

STE – stearic acid (стеариновая кислота)
UGTs – uridine 5'-diphosphoglucuronosyltransferase (уридин-5-дифосфат глюкурозилтрансферазы)
VAF – variant allele frequency (частота вариантных аллелей)

РЕЗЮМЕ

кандидатской диссертации Кезимана Парфэ «Полиморфизм генов семейств *FAD*, *SAD* и *UGT* и их роль в определении жирнокислотного состава масла и содержания лигнанов в семенах льна (*Linum usitatissimum* L.)»

Диссертационная работа посвящена оценке полиморфизмов генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1*, ассоциированных с определенным содержанием жирных кислот и лигнанов в коллекции из 279 сортов / линий льна. В результате оценены гетерогенность сортов / линий по всем изученным генам и полиморфизмы этих генов. Наиболее полиморфными были гены *FAD3A* и *FAD3B* – 91 и 62 полиморфизма, соответственно.

Также определены ключевые полиморфизмы в определении содержания LIO и LIN в масле льна для генов *FAD3A* и *FAD3B*. В ходе исследования также была использована полученная информация об аллельных вариантах генов *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B* и *UGT74S1*, для создания ДНК-паспортов сортов/линий льна. Полученные результаты позволяют предложить подход применения ключевых полиморфизмов генов *FAD3A* и *FAD3B* в маркер-ориентированной селекции льна по масличности и оценке гетерогенности сортов/линий.

SUMMARY

Kezimana Parfait. Polymorphisms of genes of the *FAD*, *SAD*, and *UGT* families and their role in determining oil's fatty acid composition and lignan content in flax seeds (*Linum usitatissimum* L.)

The present research aims to assess the polymorphisms of the *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, and *UGT74S1* genes associated with the content of fatty acids and lignans in a collection of 279 flax cultivars/lines. The heterogeneity of cultivars/lines for all studied genes and the polymorphisms of these genes were identified. The most polymorphic genes were *FAD3A* and *FAD3B* - 91 and 62 polymorphisms, respectively.

In addition, we identified the critical polymorphisms associated with the content of LIO and LIN in flax oil for the *FAD3A* and *FAD3B* genes. During the study, we also used obtained information on the allelic variants of the *SAD1*, *SAD2*, *FAD2A*, *FAD2B*, *FAD3A*, *FAD3B*, and *UGT74S1* genes to create a DNA passport for the flax cultivars/lines. The obtained results allow us to propose an approach to using the critical polymorphisms of the *FAD3A* and *FAD3B* genes in marker-oriented flax selection for oil content and to assess the heterogeneity of cultivars/lines.