

<https://doi.org/10.17816/ecogen17153-63>

АНАЛИЗ СОБЫТИЙ АЛЬТЕРНАТИВНОГО СПЛАЙСИНГА В КОНЧИКАХ КОРНЕЙ И КЛУБЕНЬКАХ *PISUM SATIVUM* L.

© Е.А. Зорин¹, О.А. Кулаева¹, А.М. Афонин¹, В.А. Жуков¹, И.А. Тихонович^{1,2}¹ ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Санкт-Петербург;² ФГБОУ ВО «Санкт-Петербургский государственный университет», Санкт-Петербург

Для цитирования: Зорин Е.А., Кулаева О.А., Афонин А.М., и др. Анализ событий альтернативного сплайсинга в кончиках корней и клубеньках *Pisum sativum* L. // Экологическая генетика. — 2019. — Т. 17. — № 1. — С. 53–63. <https://doi.org/10.17816/ecogen17153-63>.

Поступила: 06.12.2018

Одобрена: 21.12.2018

Принята: 25.03.2019

☼ Горох посевной (*Pisum sativum* L.) способен вступать в симбиотические отношения с азотфиксирующими бактериями из группы ризобий (*Rhizobia*). В обмен на питательные вещества бактерии поставляют фиксированный азот, необходимый для поддержания роста растения. Фиксация азота происходит в специализированных органах — клубеньках. Формирование и функционирование клубенька — сложноорганизованный процесс, требующий тонкой регуляции работы генов. Один из ее механизмов — альтернативный сплайсинг (АС). На данный момент информация о роли АС в установлении и поддержании симбиотических отношений между азотфиксирующими бактериями и бобовыми растениями практически отсутствует, но известно, что это мощный способ регуляции разнообразия протеома клетки, а потому АС наблюдается в реакциях на присутствие микросимбионта. В рамках работы был проведен анализ событий АС в клубеньках и кончиках корней гороха посевного. Показано, что среди функциональных групп генов, демонстрирующих события АС, одной из наиболее представленных является группа, принимающая участие в стрессовых реакциях растения на патогены. Вероятно, транскрипты с удержанным интроном подвергаются воздействию NMD-системы либо меняют свою функцию вследствие изменения первичной и вторичной структуры. Также выявлена *in silico*, экспериментально подтверждена и оценена количественно с применением метода ПЦР в реальном времени представленность мРНК изоформ четырех генов гороха: *PsIGN1*, *PsSIP1*, *PsPR-10*, *PsWRKY40*.

☼ **Ключевые слова:** горох посевной; симбиоз; альтернативный сплайсинг; транскриптомика.

ANALYSIS OF ALTERNATIVE SPLICING EVENTS IN THE ROOT TIPS AND NODULES OF *PISUM SATIVUM* L.

© Е.А. Zorin¹, О.А. Kulaeva¹, А.М. Afonin¹, V.A. Zhukov¹, I.A. Tikhonovich^{1,2}¹ All-Russia Research Institute for Agricultural Microbiology, Saint Petersburg, Russia;² Saint Petersburg State University, Saint Petersburg, Russia

For citation: Zorin EA, Kulaeva OA, Afonin AM, et al. Analysis of alternative splicing events in the root tips and nodules of *Pisum sativum* L. *Ecological genetics*. 2019;17(1):53-63. <https://doi.org/10.17816/ecogen17153-63>.

Received: 06.12.2018

Revised: 21.12.2018

Accepted: 25.03.2019

☼ **Background.** Legumes establish symbioses with nitrogen-fixing bacteria from the Rhizobium group. In exchange for nutrients, bacteria provide fixed nitrogen needed to support plant growth. At the moment, information about the involvement of alternative splicing (AS) in the establishment and maintenance this symbiotic relationships is almost absent, but, as it is a powerful mechanism for the regulation of proteome diversity of the cell, it therefore may participate in cellular response to microsymbionts. **Materials and methods.** Alternative splicing was analyzed using the assembly of “supertranscripts” and alignment of the reads from nodules and root tips to this reference. Target genes expression levels was estimated in tips of non-inoculated roots, and in nodules (2, 4, and 6 weeks post inoculation) with use of RT-qPCR. **Results.** In this study, the analysis of AS events in the nodules and root tips of the pea was carried out. The presence of isoforms of four pea genes (*PsSIP1*, *PsIGN*, *PsWRKY40*, *PsPR-10*) was confirmed and their expression level was estimated. **Conclusion.** Pea nodules were shown to be more enriched with AS events compared to root tips. Among the functional groups of genes that demonstrate AS events, one of the most enriched functional groups is the pathogens stress response. Intron retention probably leads to degradation of the transcript via NMD-system or to change of the protein function, that modulates the activity of genes in nodules.

☼ **Keywords:** pea; symbiosis; alternative splicing; transcriptomics.

ВВЕДЕНИЕ

Бобовые растения, в том числе горох посевной (*Pisum sativum* L.), способны формировать симбиозы с чрез-

вычайно широким спектром полезных почвенных микроорганизмов, что служит примером мутуалистических растительно-микробных взаимодействий. Одной из наи-

более экологически важных и хорошо изученных систем является симбиоз бобовых растений с азотфиксирующими бактериями. Такой тип взаимодействий выгоден и растению, и бактерии, поскольку растение получает атмосферный азот, который способны фиксировать бактерии, а бактерии — питательные вещества и иные метаболиты. Фиксация азота происходит в специализированных органах — клубеньках [1].

На данный момент, благодаря исследованию мутантов модельных бобовых растений, наилучшим образом изучен начальный этап клубенькообразования, и использование феномена синтении геномов для клонирования гомологичных генов у гороха посевного привело к выявлению последовательностей генов гороха, ответственных за рецепцию сигнала и специфичность взаимодействия растения и штаммов клубеньковых бактерий [2]. Однако число генетических детерминант, участвующих в клубенькообразовании, все еще превышает число известных симбиотических генов бобовых растений; кроме того, чрезвычайно мало известно о роли различных регуляторных механизмов в процессе формирования и функционирования азотфиксирующих клубеньков.

Одним из механизмов регуляции белкового разнообразия эукариотической клетки является альтернативный сплайсинг (АС) [3]. Данный механизм задействован в реакциях растения на биотический и абиотический стресс, при этом происходит модуляция количественного соотношения транскриптов в клетке и регуляция экспрессии генов [4–6]. Альтернативный сплайсинг у растений, в отличие, например, от млекопитающих, изучен значительно хуже, не говоря уже об исследованиях на «немодельных» организмах. Однако показано, что АС играет ключевую роль в регуляции протеома клетки растения в различных условиях, в том числе в стрессовых [6–10].

В зависимости от механизма и конечной экзон-интронной структуры транскрипта выделяют несколько типов сплайсинга. Первые четыре описаны наиболее полно и часто встречаются в литературе: 1) пропуск экзона (ES) — пропускается, то есть не включается в зрелую мРНК, целый экзон, в сравнении с первичным, содержащим этот экзон, транскриптом; 2) удержание интрона (IR) — интрон не сплайсируется и остается частью зрелой мРНК; 3) альтернативный донорный сайт (AD) — донорный сайт, также известный как 5'-сайт сплайсинга, отличается от такового в зрелой мРНК; 4) альтернативный акцептор (AA) — в изоформе отличается 3'-сайт сплайсинга. На основе структуры первичного транскрипта выделяют три дополнительных типа событий АС: 1) альтернативные сплайс-сайты — оба, и акцепторный и донорный, сайта меняются; 2) новый интрон (NI) — сайт сплайсинга возникает в ранее известном экзоне; 3) удержанный экзон (RE) — новый экзон заменяет ранее аннотированный экзон в зрелой мРНК [11–13].

Взаимодействие растения с бактериями (хотя и непатогенными, в случае клубенькового симбиоза), возможно, имеет черты, общие со стрессовыми реакциями, поэтому весьма вероятно, что события АС могут протекать в тканях азотфиксирующих клубеньков. Так, например, сотрудниками нашей лаборатории в ходе анализа транскриптома клубеньков гороха посевного был обнаружен АС гена, гомологичного гену *IGN1* лядвенца японского (от англ. назв. мутантной линии *ign* — ineffective greenish nodules 1, неэффективные зеленоватые клубеньки) [14]. Данный ген кодирует белок, содержащий трансмембранные домены и анкириновые повторы. У лядвенца экспрессия гена *IGN1* не ограничивается симбиотическими органами (транскрипты обнаружены в клубеньках, корнях, листьях и цветках), однако у мутантов по гену *IGN1* нарушено лишь функционирование клубеньков — образуются бледно-зеленые клубеньки, в которых не происходит фиксации азота [15]. Клубеньки демонстрируют признаки преждевременного старения, начиная с этапа, когда в нормальных клубеньках начинается азотфиксация. Вероятная функция белка *IGN1* — закоривание на мембране клетки других белков, — возможно, связана с контролем защитных реакций растения; в случае отсутствия белка *IGN1* активность защитных систем растения приводит к гибели бактерий и деградации симбиотических структур [15].

Более длинная изоформа гена *IGN1* у гороха содержит в своем составе первый интрон (тип «удержание интрона») и, вследствие этого, может транслироваться в белок, не содержащий начальные 21 аминокислоты с N-конца [14]. Но, поскольку белок, вероятно, нарабатывается с альтернативного старт-кодона, он имеет другую сигнальную последовательность на N-конце и иную локализацию в клетке: возможно, транспортируется в компартменты, связанные с симбиозом.

Другой случай участия АС в клубенькообразовании был показан в работе Чао Вонга и др., которые продемонстрировали [9], что мРНК гена *LjSIP1* подвергается АС в клубеньках, таким образом давая две структурно различные изоформы: короткую (*SIP1S*) и длинную (*SIP1L*). SymRK-interacting protein 1 (*SIP1*) способен взаимодействовать с рецепторной киназой, SymRK, у *Lotus japonicus* и, вероятно, участвует в общем симбиотическом сигнальном пути. Более длинный вариант (*SIP1L*) содержит дополнительные 17 аминокислот, которые формируют полноценный heat shock protein 20 (*Hsp20*)-like домен. Примечательно, что у лядвенца конечный продукт данной изоформы не способен взаимодействовать с рецепторной киназой. Важно, что короткая (*SIP1S*) форма не была обнаружена у других высших растений, а длинная изоформа (*SIP1L*) даже способна взаимодействовать с ортологами рецепторной киназы SymRK. Поэтому, вероятнее всего, потеря способности *SIP1L* взаимодействовать с SymRK

у *L. japonicus* компенсируется повышенной экспрессией более короткого варианта (SIP1S), который связывает SymRK и играет роль в симбиотическом сигналинге [10].

Таким образом, в различных тканях и органах растений, в том числе в клубеньках различных бобовых, должны наблюдаться события АС. Некоторые гены, имеющие специфичный паттерн экспрессии в этих органах, могут принимать участие в установлении симбиотических отношений между растением и азотфиксирующими почвенными бактериями. Этот факт служит основанием для продолжения исследований в данной области. Однако ситуацию осложняет то, что для полноценного и максимально эффективного анализа событий АС необходим собранный геном организма-объекта исследования. Геном гороха еще не секвенирован, однако в нашем распоряжении имеются данные экспериментов по секвенированию транскриптома различных тканей этого растения. Стоит отметить, что для работы с такого рода данными в рамках поставленной цели требуются особые методы компьютерного анализа.

В данной работе описаны события АС в клубеньках и кончиках корней гороха посевного с применением метода сборки «супертранскриптов» из имеющихся транскриптомных данных клубеньков и кончиков корней с последующим анализом паттернов картирования прочтений на собранные «супертранскрипты». С помощью методов ПЦП с обратной транскрипцией и количественной ПЦП в реальном времени подтверждено наличие АС и количественно оценена представленность мРНК различных изоформ в кончиках корней и клубеньках гороха на трех временных точках после инокуляции для четырех генов: *PsSIP1*, *PsIGN1*, *PsWRKY40*, *PsPR-10*.

МАТЕРИАЛЫ И МЕТОДЫ

Биоинформатические методы

В качестве материалов для компьютерного анализа данных использовали результаты RNA-seq экспе-

риментов: сборка и сырые прочтения клубеньков линии SGE, сборка и сырые прочтения кончиков корней линии SGE [14], сборка транскриптома различных тканей гороха линии Sameog [16], сборка транскриптома различных тканей гороха линий J12822, Terese и J1992 (GEUU01.1) [17], а также линий Kaspera и Parafield [18].

Для достижения поставленной цели использовали алгоритм, рекомендуемый авторами инструмента Lase (доступен на <https://github.com/Oshlack>), с небольшими изменениями (рис. 1).

Качество сырых прочтений контролировали в FastQC [19]. Тримминг хвостов и удаление прочтений низкого качества осуществляли с использованием Trimmomatic [20]. Деконтаминацию проводили с применением пакета программ BBtools [21]. Прочтения, прошедшие контроль качества, на собранный транскриптом выравнивали при помощи программы STAR [22], поскольку данный алгоритм позволяет детектировать сплайс-сайты.

Для более исчерпывающего и полного анализа АС конструирование «супертранскриптов» должно осуществляться из транскриптомов разных тканей. Для этой задачи сборки клубеньков и кончиков корней были объединены. После объединения двух сборок с целью создания более полноценного по экзон-интронному составу набора контигов сложность сборки была понижена с помощью программы CD-HIT [23], результат работы которой сводится к кластеризации похожих транскриптов и удалению идентичных. Таким образом, в объединенном наборе данных из 95 459 контигов осталось 63 395.

Необходимым этапом построения супертранскриптов является кластеризация исходных контигов между собой. Первый этап кластеризации — выравнивание прочтений на собранный транскриптом. Для более точного выравнивания и поиска сплайс-сайтов использовали программу STAR в режиме 2-pass. Полученные

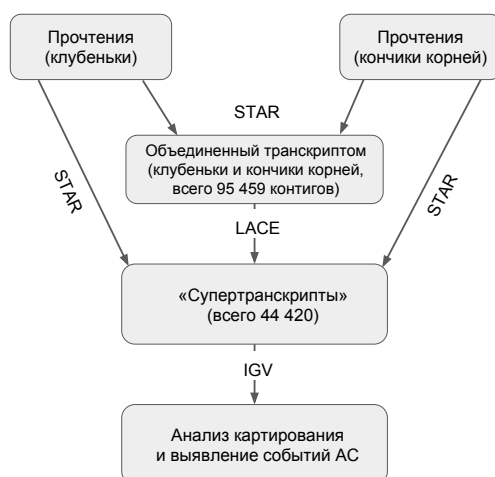


Рис. 1. Схема сборки «супертранскриптов» с последующим анализом событий альтернативного сплайсинга

выравнивания были обработаны утилитой Corset [24], результат работы которой — сгруппированные по паттерну выравнивания контиги. В программе Lase [25] из собранных из сырых прочтений контигов и файла с данными о том, как контиги кластеризуются между собой, была сконструирована «супертранскриптомная» сборка, содержащая «супертранскрипты», комбинирующие в себе данные из клубеньковой и корневой тканей.

Визуализацию выравнивания и анализ последовательностей выполняли в геномном браузере IGV (Integrative Genomics Viewer) [26]. Дополнительную валидацию присутствия конкретных изоформ проводили путем локального выравнивания последовательностей транскриптов с помощью Blast на транскриптомные сборки линий гороха Kasper, Parafield, Cameog, JI2822, Terese и JI992.

Дополнительный функциональный анализ последовательностей, демонстрирующих события АС, включал в себя анализ рамок считывания и эффекта, оказываемого АС на организацию кодирующей последовательности. Предсказание рамок считывания и трансляцию в аминокислотную последовательность выполняли инструментами Augustus [29] и ExPASy [30]. Белковые последовательности выравнивали с помощью алгоритма Смита — Ватермана, реализованного в BLAST, на базу UniProt [31] и при помощи алгоритма скрытых марковских моделей (HMM) на базу Pfam [32] с целью предсказания потенциальной функции белка по гомологии. Функциональный анализ последовательностей, демонстрирующих АС, проводили в программе REVIGO [27], аннотацию супертранскриптов в категориях Gene Ontology — через базу EggNog [28].

Биологические методы

В качестве объекта использовали линию гороха посевного (*Pisum sativum* L.) SGE [34]. Для получения биологического материала кончиков корней семена стерилизовали в концентрированной серной кислоте в течение 20 мин с последующей промывкой дистиллированной водой пять раз. Растения выращивали в фи-

тотроне Heraeus Voetsch (Германия) при следующих условиях: день/ночь — 16/8 ч, температура — 21 °С, относительная влажность воздуха — 75 %.

Для выделения РНК ткани кончиков корней (с каждого бокового корня отрезали по 5 мм) были гомогенизированы с применением жидкого азота. РНК выделяли по протоколу NucleoSpin miRNA (Macherey-Nagel, Германия). Качество РНК оценивали с использованием гель-электрофореза в 1,5 % агарозном геле. Концентрацию оценивали на спектрофотометре Shimadzu UV mini-1240 (Shimadzu, Япония). Для реакции обратной транскрипции применяли РНК в количестве 1,5 мкг. кДНК синтезировали с помощью ThermoScientific RevertAid First Strand cDNA Syntesis kit (Thermo Fisher Scientific, США) по протоколу производителя в амплификаторе T 100Thermal Cycler (Bio-Rad Laboratories, США). Пробы на присутствие ДНК оценивали с помощью ПЦР с праймерами к промотору фактора инициации трансляции 4E (номер в базе NCBI — DQ641471).

Процедуры стерилизации семян, инокуляции растений, условия выращивания, выделение РНК и синтез кДНК клубеньков проводили по методике, описанной ранее [35]. Материалы для анализа были собраны на 2, 4 и 6-й неделе после инокуляции штаммом *Rhizobium leguminosarum* bv. *viciae* Rlv 3841. Эксперимент был выполнен в трех биологических повторностях.

Оценку представленности мРНК обнаруженных генов осуществляли с применением количественной ПЦР в реальном времени. Праймеры для процедуры ПЦР в реальном времени подбирали таким образом, чтобы последовательность прямого праймера приходилась на участок экзона выше 5'-сайта сплайсинга примерно на 100 пар оснований, а последовательность обратного праймера — на стык двух экзонов. Такой способ подбора позволил оценить представленность транскрипта, подвергшегося конститутивному сплайсингу, то есть не демонстрирующего паттерн АС, так как ожидалось, что в продукте нет интрона (рис. 2). Для оценки количества длинной изоформы (с интроном в составе) использовали праймеры, разработанные также в соответствии со схемой (см. рис. 2):

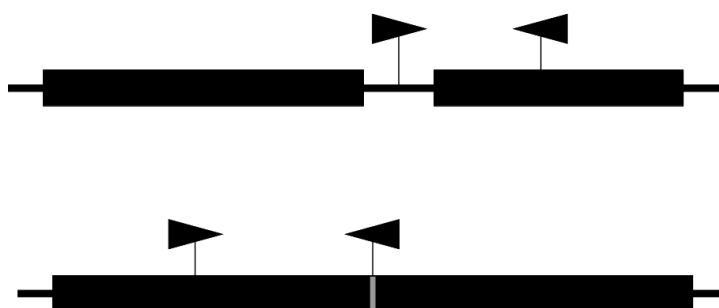


Рис. 2. Схема расположения праймеров для количественной оценки изоформ. Верхний рисунок — схема изоформы с интроном (широкие блоки — экзоны, узкий — интрон). Нижний рисунок — схема изоформы, не содержащей интрон (черные блоки — экзоны, серый — граница экзонов)

прямой праймер приходится на середину интрона, а обратный — на участок экзона, расположенного ниже 3'-сайта сплайсинга.

Статистическая обработка данных

Для сравнения экспрессии генов между образцами использовали однофакторный дисперсионный анализ с последующим применением критерия Тьюки. Это дало возможность выявить, какие именно группы значительно различаются между собой и внести поправку на множественные сравнения [33]. Анализ и визуализацию данных проводили в среде программирования R (версия 3.4.3).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В общей сложности супертранскриптомная сборка содержала 44 420 «супертранскриптов» (см. рис. 1). Таким образом, каждый «супертранскрипт» представляет собой совокупность экзонов и интронов из всех транскриптов, синтезированных с одного предполагаемого гена из различных тканей одного организма. Среди 44 420 «супертранскриптов» утилита Augustus предсказала 40 884 кодирующих последовательностей (CDS).

Повторное выравнивание корневых и клубеньковых прочтений на «супертранскриптом» позволило детектировать и визуализировать события АС. В общей сложности для клубеньковой ткани нами было об-

наружено 311477 сайтов сплайсинга. 278265 сайтов были представлены GT/AG, 4355 — GC/AG и 1285 — AT/AC. 27572 сайтов были представлены неканоническими динуклеотидными последовательностями (рис. 3). Благодаря анализу сборки кончиков корней удалось детектировать 47917 сайтов сплайсинга, среди которых GT/AG — 40583, GC/AG — 912, AT/AC — 216 и 6206 неканонических.

Анализ экзон-интронной структуры транскриптов клубеньковой сборки с помощью Augustus показал, что как минимум 10233 транскрипта содержат один и более интронов в своем составе. Общее количество событий удержания интрона — 16119. Общая длина всех интронов — 2198563, а средняя его длина — 136 пар оснований. Общая длина в парах нуклеотидов всех интронов в сборке кончиков корней составила всего лишь 928087, то есть в 2,3 раза меньше, что, скорее всего, связано не только с меньшим разнообразием событий АС, но и с невысокой глубиной покрытия. Средняя длина интрона составила 113, общее количество интронов — 8224, а число транскриптов, содержащих хотя бы один интрон, — 5977 (табл. 1).

Анализ выравнивания позволил выявить 5486 транскриптов, демонстрирующих события АС. Аннотацию обнаруженных транскриптов проводили через базу EggNog. Из 5486 уникальных транскриптов 812 было аннотировано.

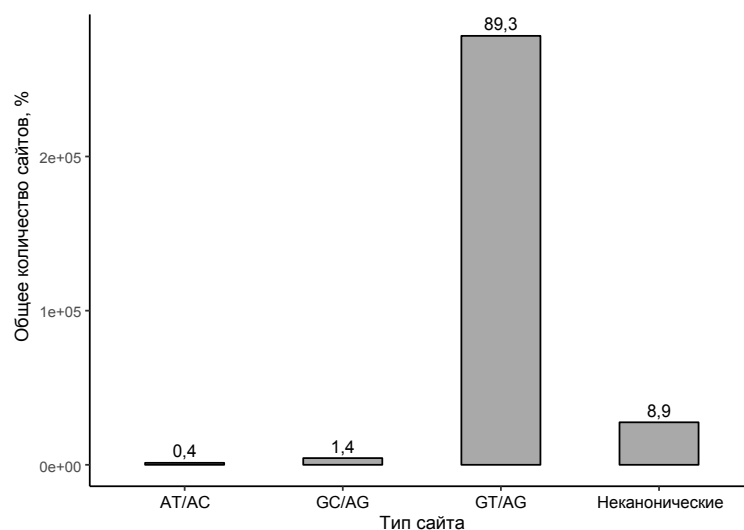


Рис. 3. Распределение различных типов сайтов сплайсинга в транскриптом клубеньков гороха

Таблица 1

Количественное описание событий сплайсинга в сборках клубеньков и кончиков корней

Орган растения	Количество транскриптов с >1 интроном (процент от общего числа контигов)	Общее количество невырезанных интронов	Общая длина интронов (в парах нуклеотидов)	Средняя длина интрона (в парах нуклеотидов)
Клубеньки	10233 (17,5 %)	16119	2198563	136
Кончики корней	5977 (16 %)	8224	928087	113

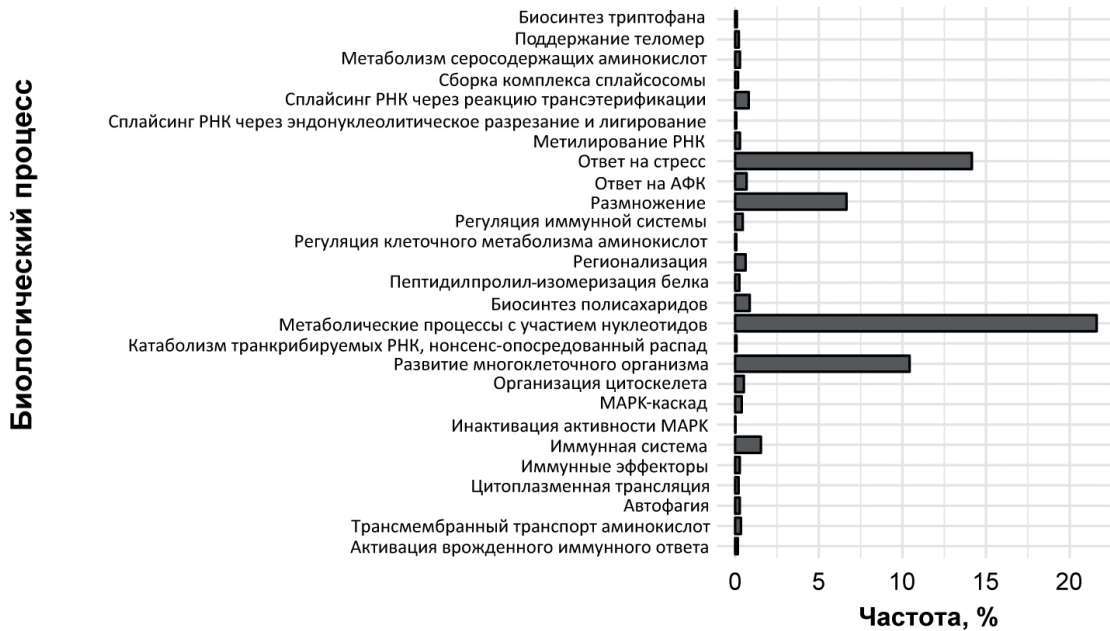


Рис. 4. Обогащение различных функциональных групп событиями альтернативного сплайсинга. Частота представляет собой процент транскриптов от общего количества, демонстрирующих альтернативный сплайсинг, попадающих в данную функциональную группу

Функциональный анализ последовательностей (в терминах Gene Ontology), демонстрирующих АС, показал, что событиями АС обогащена группа генов, связанных с нуклеотидным обменом. Альтернативный сплайсинг в рамках этой группы, скорее всего, оказывает влияние на процессы репликации и транскрипции, то есть имеет общий эффект в отношении функционирования клубенька (рис. 4). Вторая по обогащенности событиями АС группа включает в себя гены, отвечающие за реакцию на стресс. Дальнейший анализ событий АС был проведен в рамках данного кластера.

Для экспериментального подтверждения и количественной оценки представленности мРНК были выбраны четыре гена: *PsSIP1*, *PsPR-10*, *PsWRKY40*, *PsIGN1*.

В результате анализа методом количественной ПЦР в реальном времени было установлено, что на 2-й и 4-й неделях после инокуляции происходит небольшое повышение экспрессии обеих изоформ гена *SIP1*, что может свидетельствовать о важности как длинной, так и короткой изоформы, а также позволяет предположить, что у гороха, как и у лядвенца, короткая изоформа является участником симбиотического сигналинга и компенсирует неспособность длинного варианта взаимодействовать с SymRK (рис. 5), но эти различия статистически не значимы. Однако, поскольку экспрессия длинной изоформы также индуцируется в клубеньках на 2, 4 и 6-й неделях после инокуляции, можно предположить, что она способна выполнять какую-либо другую функцию, важную для функционирования клубенька. Паттерн экспрессии гена гороха *SIP1*

согласуется с результатами, полученными Чао Вонг и др. в исследовании, проведенном на лядвенце [9]. Для более детального описания функции гена *SIP1* необходимы дальнейшие исследования.

Анализ транскриптомных данных, выравнивание последовательностей мРНК гена *PsIGN1* на базу Transcriptome Shotgun Assembly (через NCBI) позволили предположить, что пре-мРНК этого гена действительно подвергается АС. Как уже было сказано выше, данный ген кодирует белок, содержащий трансмембранные домены и анкириновые повторы. Ген имеет две изоформы: более длинную, содержащую в своем составе первый невырезанный интрон с преждевременным стоп-кодоном, и короткую, без интрона в этой позиции. По всей видимости, длинная форма (с интроном) присутствует в клубеньках, но отсутствует в других тканях, поскольку об этом свидетельствовал анализ транскриптомных данных линии SGE. По данным экспериментальной проверки обе изоформы представлены и в кончиках корней и в клубеньках, так как в результате ПЦР с обратной транскрипцией и анализа продуктов амплификации с помощью гель-электрофореза наблюдались продукты ожидаемого размера, амплифицированные как с праймеров, подобранных к интрону, так и с праймеров, положение которых исключало амплификацию интрона. Причем количественный анализ представленности мРНК обеих изоформ методом ПЦР в реальном времени показал, что их соотношение в клубеньках разного возраста находится на одном уровне и значимо не изменяется, что говорит о важно-

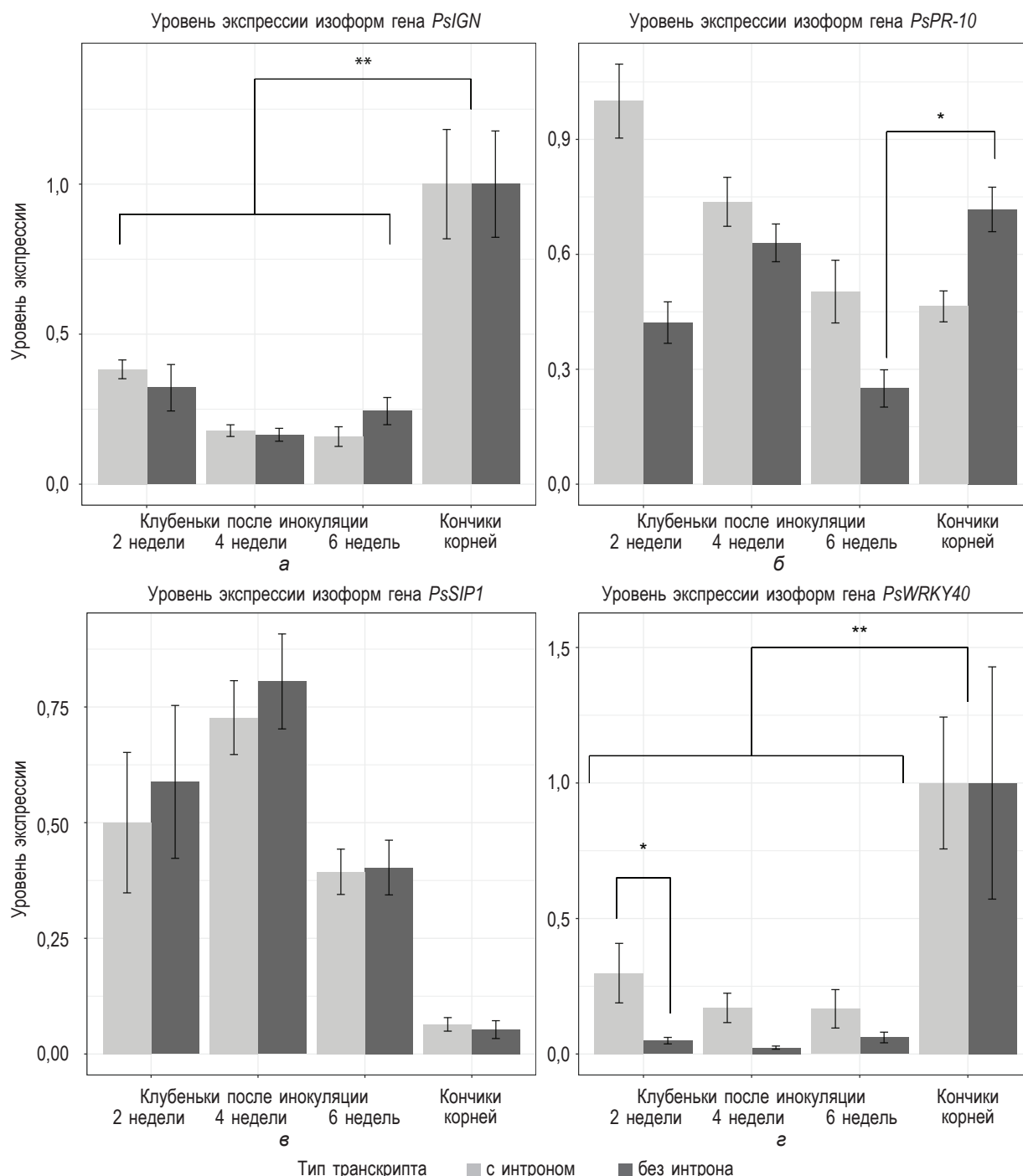


Рис. 5. Сравнение экспрессии изоформ генов гороха *PsPR-10*, *PsWRKY40*, *PsIGN*, *PsSIP1*. На рис. а линиями соединены группы сравнения, демонстрирующие статистически значимые различия (корни относительно клубеньков на сроках 2, 4, 6 недель), на рис. б значимые различия показаны для изоформ в корнях относительно клубеньков на сроке 2, 4, 6 недель и отдельно изоформы с интроном и без в клубеньках на сроке 2 недели (* $p < 0,05$, ** $p < 0,0001$)

сти обеих изоформ (см. рис. 5). Вероятно, удержание интрона в случае с длинной изоформой приводит к тому, что белок нарабатывается с альтернативного старт-кодона, находящегося выше удержанного интрона, вследствие чего конечный белковый продукт приобретает другую по своей структуре сигнальную последователь-

ность на N-конце и, как следствие, иную локализацию в клетке. Примечательно, что изоформа с удержанным интроном, по-видимому, характерна только для гороха, поскольку выравнивание данной изоформы на транскриптом *M. truncatula*, *L. japonicus* и *Glycine max* показало ее отсутствие, в то время как в транскриптомных

сборках гороха линий SGE, Kaspa и Parafield данная изоформа присутствует.

Среди группы генов, демонстрирующих события AC, также был выявлен ген из семейства WRKY. WRKY — большое семейство транскрипционных факторов с WRKY-доменом. Промоторы многих генов защитных систем обогащены W-box-элементами, которые служат сайтами-мишенями транскрипционных факторов семейства WRKY [36]. Результаты выравнивания с помощью blast показали, что ближайшими гомологами этого гена являются *MtWRKY40* (GenBank accession XP_003604245) у люцерны и *TpWRKY56* (GenBank accession PNY02664.1) у клевера, поскольку процент идентичности аминокислотных последовательностей продукта гена гороха и этих гомологов составляет 87 и 88 % соответственно. Известно, что *WRKY40* позитивно регулирует защитные реакции растений посредством индуцирования экспрессии ряда защитных генов, однако лишь при коэкспрессии с двумя другими членами семейства — *WRKY18*, *WRKY60*, возможно, посредством взаимодействия или образования гетеродимеров [38].

В транскриптом гороха посевного обнаружены две изоформы гена *PsWRKY40*: длинная, с невырезанным интроном, находящемся в позиции с 667-й по 825-ю и содержащем в своей последовательности преждевременный стоп-кодон, и короткая, из состава которой интрон вырезан в ходе сплайсинга. Как следствие, с короткой изоформы транслируется полноразмерный белок с функциональным WRKY-доменом, сохраняющим способность связывать сайты-мишени и индуцировать экспрессию защитных генов.

Количественный анализ показал, что экспрессия обеих изоформ гена *PsWRKY40* в кончиках корней находится на одном уровне, в то время как в клубеньках наблюдается резкое понижение количества обеих изоформ на двух-, четырех- и шестинедельном сроке. При этом количество изоформы с удержанным интроном повышается относительно изоформы без интрона на статистически значимом уровне (p -value — 0,0166) на сроке 2 недели (см. рис. 5). Эти данные могут свидетельствовать о том, что для полноценного функционирования клубеньков растению необходимо подавить экспрессию транскрипционных факторов, которые позитивно регулируют работу защитных генов. Помимо понижения общего уровня экспрессии гена, одним из способов подавления служит повышение количества изоформы с удержанным интроном с преждевременным стоп-кодоном в своем составе посредством AC, что мы и наблюдаем согласно результатам ПЦР в реальном времени. Транскрипты с преждевременным стоп-кодомом могут становиться мишенями для системы Nonsense-Mediated Decay (NMD). Система идентифицирует aberrантные мРНК, несущие преждевременный стоп-кодон в интроне, и избирает их для дальнейшей

деградации, тем самым препятствуя накоплению потенциально вредных и опасных усеченных белков [37, 39]. Такой механизм также является более простым и быстрым (в эволюционном смысле) способом регуляции экспрессии гена, активность которого необходимо подавить, чем, например, изменение последовательности промотора.

Еще один ген, подвергающийся AC, — это *PR-10*. Белки PR-10 (Pathogenesis-related proteins) — это семейство маленьких мономерных белков, для которых характерна конформация с внутренним гидрофобным карманом, сформированным семью антипараллельными бета-слоями и С-терминальной альфа-спиралью. Не так давно установлено, что некоторые белки из семейства PR-10 принимают участие в процессах клубенькообразования [40]. Помимо прочего, белки этого семейства обладают сродством к широкому диапазону лигандов: цитокинины, гиббереллины, жирные кислоты и флавоноиды. Таким образом, белки PR-10 могут участвовать в гормон-опосредованной устойчивости к патогенам или регуляции развития клубеньков, но поскольку конкретный PR-10 белок может нести в своей структуре сайты связывания разных лигандов одновременно, то его функции могут различаться в зависимости от ткани и условий, в которых он экспрессируется [41].

В транскриптом клубеньков гороха посевного были обнаружены две изоформы гена *PsPR-10* гомолога *MtPhBP* люцерны (GenBank accession number XP_003600177.1): длинная, содержащая дополнительный интрон в позиции с 241-й по 541-ю с преждевременным стоп-кодомом в своем составе, и короткая, подвергающаяся конститутивному сплайсингу и, следовательно, не содержащая интрона. Таким образом, активной является форма без интрона, в то время как транскрипт с интроном транслируется в усеченный белок либо подвергается деградации посредством NMD-системы еще до этапа трансляции.

Количественный анализ экспрессии гена *PsPR-10* показал, что на сроке 2 и 6 недель наблюдается понижение экспрессии функционального варианта без интрона в клубеньках, а на сроке 2 и 4 недели — повышение экспрессии нефункциональной изоформы с интроном в клубеньках относительно кончиков корней (однако не на уровне статистической значимости) (см. рис. 5). Вероятно, это свидетельствует о том, что функциональный продукт данного гена может принимать участие в стрессовых реакциях растения в ответ на инфицирование, а потому подлежит негативной регуляции. Возможно, что экспрессия этой изоформы будет находиться на высоком уровне в первые 3–4 дня после инокуляции, а на более поздних сроках снижаться, как это происходит и у *M. truncatula* (см. экспрессионный атлас MtGEA [42]). Повышенный уровень экспрессии нефункциональной изоформы с удержанным интроном, однако не на уровне статистической значимости, может

свидетельствовать о том, что транскрипт не подвергается деградации посредством NMD-системы, а приобретает новую функцию (например, в сигналинге) по причине изменения вторичной и третичной структур.

Как и ожидалось, количество изоформы с вырезанным интроном (функциональная форма) генов *PsPR-10* и *PsWRKY40* оказалось понижено, поскольку растению, вероятно, необходимо подавлять защитные реакции в клубеньках при взаимодействии с азотфиксирующими бактериями, в то время как изоформа с удержанным интроном может быть широко представлена либо потому, что не способна служить транскрипционным фактором (в случае с *PsWRKY40*) и подлежит деградации посредством NMD-системы, либо потому, что приобретает новую функцию вследствие изменения структуры белка (в случае с *PsPR-10*).

ЗАКЛЮЧЕНИЕ

В работе впервые был применен метод сборки «супертранскриптов» [25] для анализа АС в клубеньках гороха посевного. Полученные результаты позволяют заключить, что в азотфиксирующих клубеньках наблюдаются специфичные соотношения изоформ некоторых генов (*PsPR-10*, *PsWRKY40*) и в целом более интенсивно осуществляется АС в сравнении с кончиками корней. Установлено также, что в клубеньках гороха посевного, как и в клубеньках лядвенца японского, присутствует две изоформы гена *PsSIP1*: с удержанным интроном и без, хотя ранее считалось [9], что присутствие короткой изоформы в клубеньках — явление, специфичное для *L. japonicus*. Для более детального и достоверного изучения АС у гороха требуется выполнение комплексного анализа на основе данных геномного и транскриптомного секвенирования, над которым в настоящее время и работают авторы.

БЛАГОДАРНОСТИ

Работа осуществлена при поддержке гранта РФФИ 17-76-30016. Авторы выражают благодарность Татьяне Александровне Серовой (ФГБНУ ВНИИСХМ, Санкт-Петербург) за участие в обсуждении результатов работы.

ЛИТЕРАТУРА

1. Тихонович И.А., Андронов Е.Е., Борисов А.Ю., и др. Принцип дополненности геномов в расширении адаптационного потенциала растений // Генетика. — 2015. — Т. 51. — № 9. — С. 973–990. [Tikhonovich IA, Andronov EE, Borisov AY. The principle of genome complementarity in the enhancement of plant adaptive capacities. *Genetika*. 2015;51(9): 973-990. (In Russ.)]. <https://doi.org/10.7868/S001667581509012X>.
2. Shtark OY, Sulima AS, Zhernakov AI, et al. Arbuscular mycorrhiza development in pea (*Pisum sativum* L.) mutants impaired in five early nodulation genes including putative orthologs of NSP1 and NSP2. *Symbiosis*. 2016;68(1-3):129-144. <https://doi.org/10.1007/s13199-016-0382-2>.
3. Reddy AS, Marquez Y, Kalyna M, Barta A. Complexity of the alternative splicing landscape in plants. *Plant Cell*. 2013;25(10):3657-3683. <https://doi.org/10.1105/tpc.113.117523>.
4. Gracz J. Alternative splicing in plant stress response. *Biotechnologia*. 2016;97(1):9-17. <https://doi.org/10.5114/bta.2016.57719>.
5. Laloum T, Martin G, Duque P. Alternative Splicing Control of Abiotic Stress Responses. *Trends Plant Sci*. 2018;23(2):140-150. <https://doi.org/10.1016/j.tplants.2017.09.019>.
6. Shang X, Cao Y, Ma L. Alternative Splicing in Plant Genes: A Means of Regulating the Environmental Fitness of Plants. *Int J Mol Sci*. 2017;18(2). <https://doi.org/10.3390/ijms18020432>.
7. Zhiguo E, Wang L, Zhou J. Splicing and alternative splicing in rice and humans. *BMB Rep*. 2013;46(9):439-447. <https://doi.org/10.5483/BMBRep.2013.46.9.161>.
8. Zenoni S, Ferrarini A, Giacomelli E, et al. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol*. 2010;152(4):1787-1795. <https://doi.org/10.1104/pp.109.149716>.
9. Wang C, Zhu H, Jin L, et al. Splice variants of the SIP1 transcripts play a role in nodule organogenesis in *Lotus japonicus*. *Plant Mol Biol*. 2013;82(1-2):97-111. <https://doi.org/10.1007/s11103-013-0042-3>.
10. Walters B, Lum G, Sablok G, Min XJ. Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. *DNA Res*. 2013;20(2):163-171. <https://doi.org/10.1093/dnares/dss041>.
11. Iniguez LP, Ramirez M, Barbazuk WB, Hernandez G. Identification and analysis of alternative splicing events in *Phaseolus vulgaris* and *Glycine max*. *BMC Genomics*. 2017;18(1):650. <https://doi.org/10.1186/s12864-017-4054-2>.
12. Kelemen O, Convertini P, Zhang Z, et al. Function of alternative splicing. *Gene*. 2013;514(1):1-30. <https://doi.org/10.1016/j.gene.2012.07.083>.
13. Wang Y, Liu J, Huang BO, et al. Mechanism of alternative splicing and its regulation. *Biomed Rep*. 2015;3(2):152-158. <https://doi.org/10.3892/br.2014.407>.
14. Zhukov VA, Zhernakov AI, Kulaeva OA, et al. De Novo Assembly of the Pea (*Pisum sativum* L.) Nodule Transcriptome. *Int J Genomics*. 2015;2015:695947. <https://doi.org/10.1155/2015/695947>.
15. Kumagai H, Hakoyama T, Umehara Y, et al. A Novel Ankyrin-Repeat Membrane Protein, IG1, Is Required for Persistence of Nitrogen-Fixing Symbiosis

- in Root Nodules of *Lotus japonicus*. *Plant Physiol.* 2007;143(3):1293-1305. <https://doi.org/10.1104/pp.106.095356>.
16. Alves-Carvalho S, Aubert G, Carrere S, et al. Full-length *de novo* assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 2015;84(1):1-19. <https://doi.org/10.1111/tpj.12967>.
 17. Jiao K, Li X, Guo W, et al. High-Throughput RNA-Seq Data Analysis of the Single Nucleotide Polymorphisms (SNPs) and Zygomorphic Flower Development in Pea (*Pisum sativum* L.). *Int J Mol Sci.* 2017;18(12). <https://doi.org/10.3390/ijms18122710>.
 18. Sudheesh S, Sawbridge TI, Cogan NO, et al. *De novo* assembly and characterisation of the field pea transcriptome using RNA-Seq. *BMC Genomics.* 2015;16:611. <https://doi.org/10.1186/s12864-015-1815-7>.
 19. Bioinformatics.babraham.ac.uk [Internet]. FASTQC [cited 2018 Nov 2]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 20. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 21. jgi.doe.gov/data-and-tools [Internet]. bbtools; [cited 2018 Nov 2]. Available from: <https://jgi.doe.gov/data-and-tools/bbtools/>.
 22. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. <https://doi.org/10.1093/bioinformatics/bts635>.
 23. Fu L, Niu B, Zhu Z, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28(23):3150-3152. <https://doi.org/10.1093/bioinformatics/bts565>.
 24. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. *Genome Biol.* 2014;15(7):410. <https://doi.org/10.1186/s13059-014-0410-6>.
 25. Hawkins AD, Oshlack A, Davidson NM. SuperTranscript: a data driven reference for analysis and visualisation of transcriptomes. *BioRxiv.* 2016:077750. <https://doi.org/10.1101/077750>.
 26. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178-192. <https://doi.org/10.1093/bib/bbs017>.
 27. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6(7):e21800. <https://doi.org/10.1371/journal.pone.0021800>.
 28. Huerta-Cepas J, Forslund K, Coelho LP, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol.* 2017;34(8):2115-2122. <https://doi.org/10.1093/molbev/msx148>.
 29. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005;33(Web Server issue): W465-467. <https://doi.org/10.1093/nar/gki458>.
 30. Gasteiger E, Gattiker A, Hoogland C, et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003;31(13):3784-3788. <https://doi.org/10.1093/nar/gkg563>.
 31. UniProt C. The universal protein resource (UniProt). *Nucleic Acids Res.* 2008;36(Database issue): D190-195. <https://doi.org/10.1093/nar/gkm895>.
 32. Finn RD, Bateman A, Clements J, et al. Pfam: The protein families database. *Nucleic Acids Res.* 2014;42(D1):D222-D230. <https://doi.org/10.1093/nar/gkt1223>.
 33. spss-tutorials.com [Internet]. ANOVA – Simple Introduction [cited 2018 Nov 2]. Available from: <https://www.spss-tutorials.com/anova-what-is-it/>.
 34. Kosterin OE, Rozov SM. Mapping of the new mutation blb and the problem of integrity of linkage group I. *Pisum Genet.* 1993;25:27-31.
 35. Serova TA, Tsyganova AV, Tsyganov VE. Early nodule senescence is activated in symbiotic mutants of pea (*Pisum sativum* L.) forming ineffective nodules blocked at different nodule developmental stages. *Protoplasma.* 2018;255(5):1443-1459. <https://doi.org/10.1007/s00709-018-1246-9>.
 36. Chakraborty J, Ghosh P, Sen S, Das S. Epigenetic and transcriptional control of chickpea WRKY40 promoter activity under *Fusarium* stress and its heterologous expression in *Arabidopsis* leads to enhanced resistance against bacterial pathogen. *Plant Sci.* 2018;276:250-267. <https://doi.org/10.1016/j.plantsci.2018.07.014>.
 37. Birkenbihl RP, Kracher B, Ross A, et al. Principles and characteristics of the *Arabidopsis* WRKY regulatory network during early MAMP-triggered immunity. *Plant J.* 2018;96(3):487-502. <https://doi.org/10.1111/tpj.14043>.
 38. Xu X, Chen C, Fan B, Chen Z. Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell.* 2006;18(5):1310-1326. <https://doi.org/10.1105/tpc.105.037523>.
 39. Rayson S, Arciga-Reyes L, Wootton L, et al. A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in *Arabidopsis thaliana* NMD mutants. *PLoS One.* 2012;7(2):e31917. <https://doi.org/10.1371/journal.pone.0031917>.
 40. Ruskowski M, Szpotkowski K, Sikorski M, Jaskolski M. The landscape of cytokinin binding by a plant nodulin. *Acta Crystallogr D Biol Crystallogr.*

2013;69(Pt 12):2365-2380. <https://doi.org/10.1107/S0907444913021975>.

41. Liu JJ, Ekramoddoullah AKM. The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic

stresses. *Physiol Mol Plant Pathol.* 2006;68(1-3):3-13. <https://doi.org/10.1016/j.pmpp.2006.06.004>.

42. mtgea.noble.org/v3/ [Internet]. The *Medicago truncatula* Gene Expression Atlas (MtGEA) Project [cited 2018 Nov 3]. Available from: <https://mtgea.noble.org/v3/>.

✉ Информация об авторах

Евгений Андреевич Зорин — техник 1-й категории, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: kjokkjok8@gmail.com.

Ольга Алексеевна Кулаева — канд. биол. наук, старший научный сотрудник, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: okulaeva@arriam.ru.

Алексей Михайлович Афонин — инженер-исследователь, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: afoninalexeym@gmail.com.

Владимир Александрович Жуков — канд. биол. наук, заведующий лабораторией, лаборатория генетики растительно-микробных взаимодействий. ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург. E-mail: vzhukov@arriam.ru.

Игорь Анатольевич Тихонович — д-р биол. наук, научный руководитель института, академик РАН, ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии», Пушкин, Санкт-Петербург; декан, биологический факультет, ФГБУ ВПО «Санкт-Петербургский государственный университет», Санкт-Петербург. E-mail: arriam2008@yandex.ru.

✉ Information about the authors

Evgeny A. Zorin — Technician, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: kjokkjok8@gmail.com.

Olga A. Kulaeva — PhD, Senior Scientist, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: okulaeva@arriam.ru.

Alexey M. Afonin — Researcher, Laboratory of Genetics of Plant-microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: afoninalexeym@gmail.com.

Vladimir A. Zhukov — PhD, Head of the Lab, Laboratory of Genetics of Plant-Microbe Interactions. All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia. E-mail: vzhukov@arriam.ru.

Igor A. Tikhonovich — Sc.D., Professor PI, Academician of RAS, All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia; Dean of the Faculty, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia. E-mail: arriam2008@yandex.ru.