

Правительство Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(СПбГУ)

Индекс УДК 631.52 630*165.3

Рег. № НИОКТР АААА-А20-120073190047-4

УТВЕРЖДАЮ
Проректор по научной работе
СПбГУ

С.В. Микушев
«19» декабря 2022 г.



ОТЧЁТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

ИЗУЧЕНИЕ МЕХАНИЗМОВ НИТРАТ-ОПОСРЕДОВАННОЙ РЕГУЛЯЦИИ РАЗВИТИЯ
СИМБИОТИЧЕСКИХ КЛУБЕНЬКОВ У ГОРОХА ПОСЕВНОГО (*PISUM SATIVUM* L.)
И ЛЮЦЕРНЫ СЛАБОУСЕЧЕННОЙ (*MEDICAGO TRUNCATULA* GAERTN.) С
ИСПОЛЬЗОВАНИЕМ ГЕНОМНЫХ ТЕХНОЛОГИЙ
(заключительный)

Руководитель НИР,
ведущий научный сотрудник,
кандидат биологических наук

[Handwritten Signature]
подпись, дата 24.11.2022

В.А. Жуков

Санкт-Петербург 2022

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель
темы,

гл. н.с., к.б.н.



23.11.22.

(подпись, дата)

Жуков В.А.

(все разделы)

Исполнители:

м.н.с.



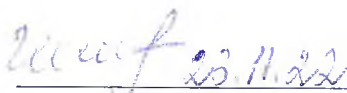
23.11.22

(подпись, дата)

Жернаков А.И.

(раздел 3.3.1. и 3.3.3.)

в.н.с., к.б.н.

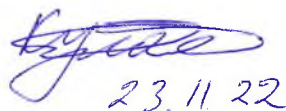


(подпись, дата)

Штарк О.Ю.

(раздел 3.3.1. и 3.3.2.)

с.н.с., к.б.н.



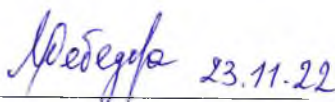
23.11.22

(подпись, дата)

Кулаева О.А.

(раздел 3.3.1. и 3.3.2.)

с.н.с., к.б.н.

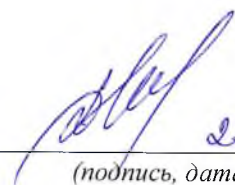


(подпись, дата)

Лебедева М.А.

(раздел 3.3.1., 3.3.2. и 3.3.4.)

доц., к.б.н.



23.11.22

(подпись, дата)

Додуева И.Е.

(раздел 3.3.4.)

М.Н.С.



Ганчева М.С.

23.11.22

(раздел 3.3.4.)

(подпись, дата)

ИНЖ.-ИССЛ.

Добычкина Д.А.



23.11.22

(раздел 3.3.4.)

(подпись, дата)

Нормоконтроль



Н.В. Квадришус

30.11.2022

РЕФЕРАТ

Отчет 69 с., 33 рис., 5 табл., 60 источников, 1 приложение.

Ключевые слова: бобово-ризобиальный симбиоз, нитрат, авторегуляция клубенькообразования, пептиды CLE, пептиды CEP, транскриптомика

Проект направлен на изучение механизмов системного контроля развития симбиотических клубеньков с участием нитрат-регулируемых пептидов CLE и CEP, а также изучение роли этих пептидов в контроле поглощения нитрата из почвы у бобовых растений - люцерны слабоусеченной и у гороха посевного. Объектом исследования являются ключевые компоненты генетических систем гороха и люцерны, ответственные за оценку азотного статуса бобовых растений. Цель работы - идентификация регулируемых нитратом пептидов CLE и CEP у люцерны слабоусеченной *Medicago truncatula* и у гороха посевного *Pisum sativum*, а также описание молекулярных механизмов их действия при клубенькообразовании и развитии растения. В работе использованы стандартные методы молекулярной биологии, такие как ПЦР, в том числе в реальном времени, секвенирование, клонирование фрагментов в векторах, а также разнообразные методы биоинформатики. На первом этапе работы проведено секвенирование геномов четырех линий гороха посевного. С использованием полученных данных идентифицированы гены, кодирующие пептиды CLE и CEP, описано их разнообразие и проведена оценка профиля их экспрессии. Кроме того, с помощью количественной ПЦР в реальном времени был определен уровень экспрессии генов CLE и CEP у люцерны и гороха в проведенных экспериментах, при клубенькообразовании и при обработке нитратом. Созданы конструкции для сверхэкспрессии генов CLE и CEP у гороха и люцерны. На втором этапе работы проанализировано разнообразие последовательностей четырех генов *MtCLE12*, *MtCLE13*, *MtCLE34*, *MtCLE35* у 262 линий *M. truncatula* и показано, что в гене *MtCLE13* присутствуют синонимичные замены, тогда как в остальных генах выявлены как синонимичные, так и несинонимичные замены. В отличие от люцерны, гомологичные гены гороха посевного *PsCLE12*, *PsCLE13* и *PsCLE34* у 99 генотипов гороха не являются полиморфными. При помощи подхода GWAS (Genome-wide association study) выявлена область генома люцерны (на хромосоме 2), расположенная перед кодирующей последовательностью гена *MtCLE34*, в которой SNP (однонуклеотидные полиморфизмы) демонстрируют ассоциацию с количеством клубеньков, образующихся на растениях. При помощи транскриптомного

анализа корней гороха после обработки нитратом выявлены более 100 дифференциально экспрессирующихся генов, связанных с ответом растения на нитрат. Также при помощи транскриптомного анализа показано, что в трансгенных корнях люцерны, сверхэкспрессирующих ген *MtCLE35*, повышена экспрессия генов, ответственных за метаболические процессы и белковый метаболизм, а также кодирующих компоненты клеточной стенки, а также снижена экспрессии ключевых генов-регуляторов формирования клубеньков (*NIN*, *NSP1*, *NF-YA1/HAP2* и *ERN1*), что приводит к подавлению клубенькообразования. Также оптимизированы методы трансформации гороха с помощью *Agrobacterium rhizogenes* и получен материал для транскриптомного анализа корней со сверхэкспрессией генов *PsCLE12*, *PsCLE13* и *PsCEP10*. На третьем этапе работы проанализирован полиморфизм генов *PsSym29* и *PsNRLK1*, кодирующих рецепторы CLE-пептидов, у гороха посевного. Не выявлено ассоциаций аллельного состояния данных генов и значений хозяйственно-значимых признаков у растений, выращенных в условиях инокуляции клубеньковыми бактериями. Секвенирован геном мутанта гороха по гену *fas*, предположительно кодирующему компонент системы CLAVATA. Секвенирован транскриптом корневой системы мутанта гороха по гену *fas* и исходного генотипа при инокуляции клубеньковыми бактериями и при воздействии нитрата, проанализирована роль гена *FAS* в клубеньковом симбиозе и в ответе растений гороха на нитрат в субстрате. С помощью метаболомного анализа проведена оценка уровня азотсодержащих соединений у растений люцерны со сверхэкспрессией гена *MtCLE35*. Показано, что ген *MtCLE35* задействован не только в контроле развития симбиотических клубеньков, но и в других нитрат-регулируемых процессах, в частности, процессах накопления азота в растениях. Результаты работы будут способствовать расширению понимания механизмов нитрат-опосредованной регуляции развития растений и станут основой для дальнейших исследований.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	8
ОСНОВНАЯ ЧАСТЬ	10
1 Обоснование направлений исследований	10
2 Методы исследований	12
3 Результаты исследований	20
ЗАКЛЮЧЕНИЕ	63
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	65

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ДНК – дезоксирибонуклеиновая кислота, носитель наследственной информации

РНК – рибонуклеиновая кислота, осуществляющая реализацию наследственной информации (например, путем экспрессии генов)

мРНК (матричная РНК) – РНК, кодирующая белок

кДНК (комплементарная ДНК) – ДНК, полученная путем обратной транскрипции мРНК

Секвенирование – процесс определения последовательности нуклеотидов в нуклеиновой кислоте

Транскриптом – совокупность транскриптов, т.е. последовательностей РНК, соответствующих генам, экспрессирующимся в данной ткани

ДЭГ – дифференциально экспрессирующиеся гены

ВВЕДЕНИЕ

Симбиотические взаимодействия бобовых растений с почвенными бактериями ризобиями приводят к развитию на корнях растений азотфиксирующих клубеньков, в которых происходит процесс биологической азотфиксации – восстановления молекулярного азота и его включения в органические соединения. Благодаря симбиотической азотфиксации происходит обогащение природных экосистем азотом, необходимым для роста и развития растений, что снижает потребность в применении азотсодержащих удобрений.

Развитие клубеньков находится под системным контролем со стороны растения. Клубеньки формируются при низком содержании азота, тогда как наличие в среде достаточного количества азота, в частности, нитрата, подавляет развитие клубеньков. Избыточное количество клубеньков для растений энергетически не выгодно, и поэтому количество клубеньков строго контролируется растениями на системном уровне – с помощью системы авторегуляции клубенькообразования. Важными участниками системного контроля клубенькообразования являются пептиды CLE – они были описаны у ряда бобовых растений как мобильные сигнальные молекулы, которые вырабатываются в корнях в ходе развития клубеньков и поступают по ксилеме в побег [1]. В побеге они связываются со своими рецепторами, CLV1-подобными киназами. Мутации в генах CLV1-подобных киназ у разных бобовых растений приводят к суперклубенькообразованию – образованию избыточного числа клубеньков [2,3]. Активация CLV1-подобной киназы в побеге при связывании с пептидами CLE приводит к выработке сигналов, поступающих в корни и подавляющих закладку новых клубеньков; кроме того, в побеге снижается уровень сигнальных молекул – индукторов – клубенькообразования, что в совокупности приводит к ограничению количества закладывающихся клубеньков [4].

У ряда бобовых были описаны пептиды CLE, синтезируемые в корнях не только в ответ на взаимодействие с ризобиями, но и при воздействии нитрата. Так, у *Lotus japonicus* экспрессия двух из трех генов CLE, индуцируемых при симбиозе с ризобиями - *LjCLE-RS2* и *LjCLE-RS3*, также индуцируется в ответ на обработку нитратом [1,5]. У сои *Glycine max* были выявлены 2 группы пептидов CLE, различающиеся по механизму активации экспрессии их генов. Наряду с индуцируемыми ризобиями пептидами группы RIC GmRIC1 и GmRIC2, которые задействованы в системном контроле клубенькообразования с участием работающей в побеге CLV1-подобной киназы GmNARK [6], у сои также были выявлены нитрат-индуцируемые CLE, относящиеся к группе NIC (от англ. NITRATE INDUCED CLE). Экспрессия генов *GmNIC1* и *GmNIC2* активируется в ответ на присутствие

нитрата в почве и приводит к ингибированию формирования клубеньков. Тогда как у люцерны пептиды *LjCLE-RS2* и *LjCLE-RS3* действуют системно [1,5], через рецептор, работающий в побеге, у сои нитрат-зависимый механизм подавления развития симбиотических клубеньков у сои действует локально с участием корневой CLV1-подобной киназы [4,7]. У люцерны и у гороха к настоящему времени не были охарактеризованы нитрат-регулируемые пептиды CLE, регулирующие развитие клубеньков, и их идентификации и изучению механизмов из действия посвящен настоящий проект.

Наряду с пептидами CLE, выступающих в роли ингибиторов клубенькообразования, регуляторные пептиды другой группы, CEP, выполняют противоположную функцию – они образуются при недостатке азота в среде и активируют закладку симбиотических клубеньков [8]. Как было показано недавно, у люцерны пептид *MtCEP7* активируется при клубенькообразовании и системно, действуя через рецептор, работающий в побеге, *MtCRA2*, стимулирует развитие инфекции [9]. Однако роль пептидов CEP в нитрат-зависимой регуляции развития клубеньков и механизм их регуляции остаются не изученными. В геноме гороха пептиды CEP к настоящему времени не были идентифицированы.

В задачи настоящего проекта входила идентификация регулируемых нитратом пептидов CLE и CEP у люцерны и гороха, а также расшифровка молекулярных механизмов их действия при клубенькообразовании и в регуляции азотного статуса растений. Полученные результаты расширяют представления о механизмах системного контроля клубенькообразования у бобовых растений и о роли нитрата в регуляции развития симбиотических клубеньков и развития растений в целом.

ОСНОВНАЯ ЧАСТЬ

1. Обоснование направлений исследований

Горох посевной (*Pisum sativum* L.) является древней сельскохозяйственной культурой, которая активно возделывается и в настоящее время. Семена гороха богаты белком, крахмалом, а также содержат широкий набор витаминов и минеральных веществ. Различные сорта используются для получения овощного, зернового и кормового гороха. Горох также считается первым объектом генетических исследований, благодаря которому были открыты законы наследования признаков. Поскольку горох является широко распространенным культурным растением [10], многие исследовательские центры имеют обширные коллекции генетических ресурсов (семян) гороха. К сожалению, работа с горохом осложнена его крупным геномом с большим количеством мобильных элементов [11], для которого лишь недавно была получена первая сборка [12].

Для проведения ряда исследований, касающихся таких важных компонентов, как малые РНК, регуляторные элементы, факторы эпигенетической регуляции, а также гены, кодирующие короткие пептиды, критически важны знания о геномных последовательностях. Секвенирование генома позволяет поднять качество исследований всех происходящих в растении процессов на более высокий уровень, к примеру, значительно повысить качество сборки и анализа транскриптомных данных, в частности, идентифицировать транскрипты с очень низким уровнем экспрессии. Точные сведения об архитектуре генома, включая и его некодирующую часть, помогут пролить свет на проблемы геномной регуляции транскрипции и эволюции бобовых растений. Таким образом, очевидна необходимость секвенирования генома различных генотипов гороха, не являющихся родственными друг другу.

Геном люцерны слабоусеченной был секвенирован около 10 лет назад [13], однако многие его особенности, например, характеристики генов, кодирующих короткие пептиды, не исследованы должным образом до настоящего времени. В частности, значительный интерес представляют гены пептидных гормонов, продукция которых регулируется уровнем нитрата в среде: CEP (C-TERMINALLY ENCODED PEPTIDES) и CLE (CLAVATA3/EMBRYO SURROUNDING REGION-related) (см. обзор Oh et al., 2018 [14]). Пептиды CLE и CEP относятся к группе посттрансляционно модифицируемых пептидов: они синтезируются из более протяженного пептида-предшественника, из которого в результате протеолиза и посттрансляционных модификаций образуется активный регуляторный пептид длиной около 12-15 аминокислот (см. обзор Oh et al., 2018 [14]). Пептиды CEP и CLE, как стало известно за последние годы, опосредуют системный ответ

на содержание азота в среде: они синтезируются в тканях корня и способны перемещаться в наземную часть растения, где во флоэме листа они узнаются специфическими рецепторами (см. обзор Notaguchi and Okamoto, 2015 [15]). Таким образом, сигналы, поступающие из подземных органов в побеги, запускают ответные реакции, регулирующие процессы развития с учетом потребностей растения в целом.

Однако, у люцерны к настоящему времени не были идентифицированы нитрат-регулируемые гены CLE, их идентификация и изучение их возможной регуляции с участием NIN и NLP представляет значительный интерес для понимания механизмов нитрат-опосредованной регуляции развития клубеньков. У гороха нитрат-индуцируемых пептидов CLE ранее также не было выявлено.

В отличие от CLE, пептиды CEP синтезируются в корнях растений в ответ на недостаток азота. Известно, что пептиды CEP ингибируют рост корня, подавляя пролиферацию клеток меристемы, а также закладку боковых корней [16]. Роль пептидов CEP в клубенькообразовании также была изучена: у модельного бобового растения *Medicago truncatula* пептид MtCEP1 стимулирует формирование клубеньков и при этом подавляет закладку боковых корней [8]. Однако механизмы действия нитрат-регулируемых пептидов CEP, их мишени в развитии симбиотических клубеньков, остаются малоизученными.

В настоящей работе осуществлён детальный анализ функций пептидов CLE и CEP с помощью подходов геномики и транскриптомики. Так, получены трансгенные растения люцерны слабоусечённой с изменённым уровнем экспрессии генов, кодирующих пептиды CEP и CLE, проведен анализ транскриптомов этих растений, а также изучено влияние данных генов на процессы накопления азотсодержащих соединений в тканях растений. С помощью анализа последовательностей геномов различных линий гороха и люцерны был проанализирован полиморфизм генов, кодирующих пептиды CLE и CEP, а также генов, кодирующих рецепторные молекулы, связывающие эти пептиды. Кроме этого, был проведён анализ ассоциаций, в ходе которого была оценена связь аллельных состояний генов с такими признаками, как число образующихся симбиотических клубеньков и содержание азота в тканях растений. В результате работы было расширено понимание механизмов участия нитрата в системном контроле развития растений на модели клубенькообразования у бобовых растений, что должно способствовать формированию новых принципов земледелия, позволяющих оптимизировать количество вносимых азотных удобрений при культивировании различных сельскохозяйственно-значимых культур.

2. Методы исследований

Растительный материал

В работе использовали линию гороха посевного *Pisum sativum* Frisson (из коллекции ФГБНУ ВНИИСХМ) и линии люцерны слабоусечённой *Medicago truncatula* Gaertn A17 и R108 (Университет г. Вагенингена, Нидерланды).

Семена стерилизовали путем обработки концентрированной серной кислотой (95-97%) в течение 10 минут, после чего промывали стерильной водой 8-10 раз. Затем высевали на 1% водный агар и оставляли при температуре +4 на 24 часа, а затем проращивали в темноте при комнатной температуре в течение 48 часов. Полученные проростки переносили на питательную среду Farhaeus. Выращивание растений осуществляли в фитотроне (KBW F240, Binder, Германия) с режимом день/ночь 16/8 часов, 21°C при освещенности 30 тыс. люкс.

Для получения симбиотических клубеньков растения выращивали в горшках, наполненных вермикулитом. Для инокуляции гороха использовали штамм *Rhizobium leguminosarum* RCAM1026 (из коллекции ФГБНУ ВНИИСХМ), растения люцерны инокулировали штаммом *Sinorhizobium meliloti* 2011 (Университет г. Вагенингена, Нидерланды). Бактерий культивировали в жидкой питательной среде YEM (0,5 г/л H_2PO_4 – 0.5 г, $MgSO_4 \cdot 7H_2O$ - 0,2 г, NaCl - 0,1 г, маннитол - 10 г, дрожжевой экстракт - 1 г на один литр среды) в термостате при температуре $+28 \pm 1^\circ C$ до достижения оптической плотности $A_{600}=0,7-1$. Инокуляцию производили из расчета 1 мл бактериальной суспензии на растение.

Обработку проростков растений нитратом проводили в гидропонной системе, содержащей среду Farhaeus с добавлением 10 mM KNO_3 в течение 24 часов, после чего собирали корни для последующего анализа экспрессии генов.

Трансформация растений с помощью *Agrobacterium rhizogenes*

Растения *M. truncatula* трансформировали штаммом *A. rhizogenes* *Arqua-1*, содержащим конструкцию для сверхэкспрессии нитрат-регулируемого гена *MtCLE35* люцерны (35S:MtCLE35), а также конструкцию для сверхэкспрессии репортерного гена бета-глюкуронидазы, используемую в качестве контроля (35S:GUS). Такие конструкции были получены ранее на основе вектора pB7WG2D (VIB, Гент, Бельгия), который дополнительно содержит ген *GFP* под контролем промотора 35S для возможности отбора трансгенных корней. Трансформацию проростков люцерны осуществляли согласно

протоколу [17] следующим образом: проростки *M. truncatula* обрезали в области гипокотыля и на срез наносили суспензию агробактерий. Культивирование таких эксплантов проводили в течение 5 дней на твердой питательной среде Fahraeus [18]. Затем растения переносили на среду с добавлением антибиотика (цефотаксим, 300 мкг/мл) и культивировали до регенерации корней, после чего пересаживали в горшки, заполненные вермикулитом и инокулировали жидкой культурой бактерий *Sinorhizobium meliloti*, штамм 2011, для индукции клубенькообразования (1 мл культуры бактерий при оптической плотности OD600 на одно растение). Через 12 дней после инокуляции отбирали трансгенные корни по флуоресценции GFP с использованием флуоресцентного стереомикроскопа Leica M205FA (Leica, Германия) в ресурсном центре «Развитие молекулярных и клеточных технологий».

Растения гороха трансформировали штаммами *A. rhizogenes Arqua-1*, содержащими конструкции для сверхэкспрессии генов *PsCLE13*, *PsCLE12*, *PsCEP10*, а также конструкцию для сверхэкспрессии репортерного гена бета-глюкуронидазы, используемую в качестве контроля (35S:GUS). Агробактерии выращивали на твердой агаризированной среде YEP с добавлением селективного антибиотика спектиномицина в течение 2 дней. Трёхдневные проростки гороха помещали на чашки Петри, содержащие среду Farhaeus, по четыре проростка на чашку. С помощью иглы от шприца с нанесенной суспензией агробактерий, осуществляли укол в верхней части корня, и дополнительно наносили суспензию агробактерий в область укола. Через 7 дней растения переносили в горшки с вермикулитом, увлажненным жидкой средой Farhaeus. Из каллусов в последствие формировались трансгенные корни, которые отбирали по флуоресценции репортерного белка GFP.

Анализ экспрессии генов

Выделение РНК проводили с использованием реагента TriZol (BioRad Laboratories) согласно протоколу производителя. Очистку полученных проб от примеси ДНК осуществляли с помощью набора Rapid Out DNA Removal Kit (Thermo Fisher Scientific, США), согласно протоколу производителя. Концентрацию РНК в пробе определяли с помощью спектрофотометра UV-Vis Nano Drop 2000c (Thermo Scientific, США).

Синтез кДНК проводили с использованием 1000 нг тотальной РНК с помощью обратной транскриптазы Revert Aid (Thermo Fisher Scientific, США) согласно протоколу производителя. Полученные пробы кДНК разводили до конечного объема 100 мкл.

ПЦР с детекцией в реальном времени проводили с помощью амплификатора CFX 96 (Bio-Rad Laboratories, США) с использованием набора с интеркалирующим красителем Eva Green (Синтол, Россия), а также по 10 пМ праймеров к исследуемым генам. Анализ активности генов интереса проводили путем оценки относительных уровней экспрессии по методу Livak [19]. В качестве референсных генов использовали гены убикуитина (*PsUbi*) и фактора инициации трансляции (*PsTIF*). Последовательности праймеров, используемых для анализа экспрессии генов, приведены в Таблице 1.

Таблица 1 — Список праймеров, используемых для анализа экспрессии генов

Название гена	Прямой праймер (5' -> 3')	Обратный праймер (5' -> 3')
<i>PsUbi</i>	ATGCAGATYTTTGTGAAGAC	ACCACCACGRAGACGGAG
<i>PsTIF</i>	TCCCCTTTCTATCTCCTCTCATC AA	CAATCGCCTTCACACCTTCC
<i>PsCLE12</i>	CTTGAAGGTGGTGGGAAGACATC A	TGGATACAATAAGGTTGGACTAGAAAA TC
<i>PsCLE13</i>	GGAAAATCACCACCAAGCAAG ТААС	CACACAAAGTAATAGAAACGATAATCA СС
<i>PsCEP10</i>	AACTAGTCCTGGTGTGGGCA	CGGACTGTGACTAGGGTTGGT

Транскриптомный анализ

Для транскриптомного анализа тотальную РНК была выделяли из корней гороха и люцерны с помощью реагента TriZol (BioRad Laboratories) согласно протоколу производителя. Очистку полученных проб от примеси ДНК осуществляли с помощью набора Rapid Out DNA Removal Kit (Thermo Fisher Scientific, США), согласно протоколу производителя. Концентрацию РНК в пробе определяли с помощью спектрофотометра UV-Vis Nano Drop 2000c (Thermo Scientific, США). Приготовление библиотек для транскриптомного анализа методом 3'MACE были приготовлены по протоколу производителя GenXPro GmbH (Франкфурт-на-Майне, Германия). Полученные библиотеки были секвенировали в компании Macrogen (Сеул, Южная Корея).

Приготовление библиотек для транскриптомного анализа корней гороха после нитратной обработки проводили в Ресурсном центре СПбГУ “Центр Биобанк” с использованием технологии Illumina. Для приготовления библиотек использовали набор NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Секвенирование осуществляли на приборе HiSeq4000 с режимом секвенирования PE151. Для анализа качества “сырых” данных секвенирования использовалась программа FastQC. Обрезку последовательностей проводили с помощью Trimmomatic [20].

Выравнивание на референсный геном (сборка MedtrA17_4.0 [21] в случае люцерны, а также сборка генома *Pisum_sativum_v1a* <https://urgi.versailles.inra.fr/download/pea/> и сборка генома гороха линии Frisson, полученный в лаборатории генетики растительно-микробных взаимодействий ФГБНУ ВНИИСХМ) проводили с помощью HISAT2 [22], после чего осуществляли перевод файлов формата “sam” в “bam” с помощью Samtools [23] и подсчёт прочтений с помощью Stringtie [24]. Пакет R DESeq2 [25] использовались для анализа дифференциальной экспрессии генов. Функциональную аннотацию выявленных дифференциально экспрессирующихся генов люцерны проводили с помощью ресурса Medicago Classification Super Viewer http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_superviewer_medicago.cgi с использованием параметров по умолчанию.

Поиск варибельных позиций в последовательностях генов *CLE* у линий *M. truncatula* осуществлялся с помощью базы данных HarMap (<http://www.medicagoharmap.org/>) [26]. Для поиска ассоциаций полиморфных сайтов с количеством образующихся клубеньков у люцерны использовали алгоритмы Multidimensional Scaling (MDS), General Linear Model (GLM) и Mixed Linear Model (MLM) (Principal Component Analysis (PCA) + K) в программе Tassel (<https://www.maizegenetics.net/tassel/>) [27].

Анализ полиморфизма генов гороха

Анализ полиморфизма генов гороха проведён на материале 99 генотипов из коллекции Всероссийского института генетических ресурсов растений им. Н.И. Вавилова (ВИР). Данный набор генотипов, отобранный на основании максимального разнообразия их происхождения (из более 25 стран), содержит образцы, относящиеся к следующим подвидам: *P. sativum* subsp. *sativum* (85), *asiaticum* (8), *abyssinicum* (3) and *transcaucasicum* (3) [28]. Ранее была проведена оценка серии параметров (PW – plant shoot weight, PL – plant shoot length, PN – pod number, SN – seed number per plant, SW – seed weight per plant, TSW – thousand-seed weight, NtS – seed nitrogen content, PhS – seed phosphorus content), определённых для растений, выращенных в условиях инокуляции клубеньковыми бактериями и грибами арбускулярной микоризы (АМ). ДНК для генотипирования растений была выделена из 7-недельных проростков при помощи СТАВ-метода.

Фрагменты генов гороха *PsSym29* и *PsNRLK1* были амплифицированы при помощи специфичных праймеров (Таблица А.7 Приложения) и секвенированы. Полученные

последовательности были выровнены в программе Mega 5.0 [29]; анализ полиморфизма был проведён в программе DNAsp 5.0 [30]. Критерии D Таджимы и H Фэя и Ву были рассчитаны в скользящем окне размером 100 п.н. с шагом 25 п.н. в программе DNAsp 5.0 [30]. Анализ ассоциаций наблюдаемых однонуклеотидных вариантов с исследуемыми признаками был проведён при помощи двухфакторного дисперсионного анализа с коррекцией FDR, а также при помощи регрессионных моделей, подробно описанных в [28]. Статистический анализ ассоциаций проведён в среде R Statistical Software (version 3.5.3). Анализировали варианты, присутствующие не менее чем в пяти образцах из коллекции гороха.

Эксперимент по нитратному ингибированию клубенькообразования у гороха

Для изучения влияния мутации в гене *fas* на экспрессию генов были секвенированы образцы мРНК корней с клубеньками сорта гороха посевного Штамбовый (несущий мутантную копию гена *fas*) и исходный сорт Немчиновский. Для этого был поставлен эксперимент по выращиванию растений в условиях фитотрона. Оба генотипа были обработаны 15мМ KNO₃. Корни с клубеньками были собраны на сроке 4 недели после инокуляции. Далее корни с клубеньками были растёрты в жидком азоте и хранились в морозильной камере при температуре -80 С.

РНК из полученного материала выделяли с помощью TRIzol (Thermo Fisher Scientific) по методике производителя с небольшими изменениями. Количество РНК оценивалось на приборе NanoDrop OneC (Thermo Fisher Scientific). 3'МАСЕ Библиотеки изготавливали с использованием 3'МАСЕ набора, поставляемого производителем и секвенировали на приборе Illumina NovaSeq 6000 по договору с компанией Evrogen (Москва, Россия).

Качество сырых прочтений оценивали в программе FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Удаление коротких и низкокачественных (Q<20) прочтений выполняли в программе BBDuk (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>), дедупликацию осуществляли по UMI с помощью программы UMI-tools [31] и samtools [32]. Прочтения высокого качества картировали на геном гороха посевного линии Frisson с помощью STAR (Dobin et al., 2013), а квантификацию генов проводили в программе featureCounts [33].

Анализ дифференциальной экспрессии выполняли в пакете DESeq2 [25] в среде программирования R. Диаграммы Венна построены с помощью пакета Eulerr (<https://github.com/jolars/eulerr>), анализ обогащения в терминах Gene Ontology проведён в пакете TopGO [34]. Вся визуализация выполнена с помощью пакета ggplot2 [35] и ggrepel [36].

Метаболомный анализ

В работе использовали линию люцерны диплоидной *Medicago truncatula* Gaertn дикого типа R108, а также растения люцерны диплоидной со стабильной сверхэкспрессией гена *MtCLE35* двух семей MtCLE35-OE-2 и MtCLE35-OE-3, полученные нами ранее. Семена стерилизовали путем обработки концентрированной серной кислотой (95-97%) в течение 10 минут, после чего промывали стерильной водой 8-10 раз. Семена высевали на 1% водный агар и оставляли при температуре +4 на 24 часа, а затем проращивали в темноте при комнатной температуре в течение 48 часов. Полученные проростки переносили на питательную среду Farhaeus (CaCl₂ – 0,1 г, MgSO₄ • 7H₂O – 0,12 г, KН₂PO₄ – 0,1 г, Na₂HPO₄ – 0,15 г, Fe-цитрат – 0,005 г, Mn, Cu, Zn, B, Mo - среды, на один литр среды, pH = 6.5) [18] и выращивали в фитотроне в течение 10 дней. Далее проростки переносили в аэропонную систему, содержащую 8 л раствора для аэропоники (см. Таблицу А.1 Приложения) и выращивали в фитокомнате с режимом день/ночь 16/8 часов, 21°C, при освещенности 30 тыс. люкс.

Растительный материал растирали с жидким азотом в фарфоровой ступке и добавляли 1 мл тризола. Инкубировали в течение 5 минут, а затем центрифугировали образцы при скорости 12000об/мин в течение 10 минут при температуре +4°C. После центрифугирования осуществляли перенос верхней фракции в чистый эппендорф. Далее добавляли 200 мкл. хлороформа, переворачивали в течение 15 секунд и инкубировали при комнатной температуре в течение 3 минут. Затем центрифугировали образцы при скорости 12000 об/мин в течение 15 минут при температуре +4°C. После осуществляли перенос верхней фракции в чистый эппендорф и добавляли 500 мкл. изопропанола. Далее инкубировали пробы в течение 20 минут при -20°C и центрифугировали при скорости 12000 об/мин в течение 15 минут при температуре +4°C. Полученный осадок промывали два раза 400 мкл 70% этанола. После этого осадки растворяли в 25 мкл воды.

Обработку ДНКазой проводили в объеме 30 мкл, включающем 25 мкл РНК, 3 мкл 10-кратного буфера и 2 мкл ДНКазы (2 U). Образец инкубировали в течение 30 минут при температуре 37°C. После осуществляли очистку от ДНКазы. В образец добавляли 3 мкл реагента DRR (DNA Removal Reagent, Thermo Fisher Scientific, США) и инкубировали 2 минуты, периодически переворачивая. Далее центрифугировали образцы при скорости 800 об/мин в течение минуты. После центрифугирования осуществляли перенос верхней фракции в чистый эппендорф. Концентрацию РНК измеряли с помощью спектрофотометра NanoDrop™ 2000.

Для синтеза кДНК брали по 1000 нг РНК во всех исследуемых пробах. К РНК добавляли реакционную смесь, содержащую 1 мкл обратной транскриптазы Revert Aid (200 ед./мкл Thermo Fisher Scientific, США), 1 мкл блокатора РНКаз RNase-lock (40 ед./мкл.) (Thermo Fisher Scientific, США), 1 мкл олигонуклеотида oligodT (100мМ), 2 мкл dNTPs (10мМ), 4 мкл 5x RT буфера, воду до объема 20 мкл. Пробы инкубировали в течение 1 часа при +42°C с последующим прогреванием в течение 10 минут при +70°C.

Оценку уровней экспрессии гена *MtCLE35* проводили с помощью ПЦР в реальном времени с использованием амплификатора с системой детекции CFX 96 (BioRad, США). В качестве референсного гена детектировали уровень экспрессии убиквитина в пробах. В состав 20 мкл реакционной смеси, помимо кДНК (1мкл), входили следующие компоненты: 8 мкл 2,5-кратной готовой смеси для ПЦР (Синтол, Россия), содержащей все необходимые компоненты для ПЦР, а также флуоресцентный краситель Eva Green, по 1 мкл прямого и обратного праймера к гену *MtCLE35* (концентрация 10 пм/мкл) (*MtCLE35_CDS_FOR*; ATGGCAAACACACAAATAАСТАТАТТТ, *MtCLE35_CDS_REV*; СТАСТТГТТТТГТГГАССТГСА) и 9 мкл воды. Программа амплификации: +95°C - 30 сек, +56°C - 30 сек и +72°C - 40 сек (45 циклов). Детекцию флуоресценции интеркалирующего красителя Eva Green осуществляли на стадии элонгации (+72°C) в конце каждого цикла. Анализ активности генов интереса проводили путем оценки относительных уровней экспрессии по методу Delta-Delta Ct, описанному Livak и Schmittgen [19]. В качестве референсного гена использовали убиквитин.

Анализ содержания метаболитов проводили с помощью метода газовой хромато-масс-спектрометрии (ГХ-МС). Для этого 200 мг растительного материала растирали в ступке с помощью пестика в жидком азоте, заливали 1 мл метанола. Далее проводили экстракцию на термошейкере (TS-100С, Biosan) при скорости 900 об/мин в течение 45 минут при температуре +10°C. Затем гомогенат центрифугировали при скорости 15000 об/мин в течение 15 минут при температуре +4°C. После центрифугирования осуществляли перенос верхней фракции в чистый эппендорф. Осадок промывали 0,5 мл метанола и проводили экстракцию на термошейкере при скорости 900 об/мин в течение 30 минут при температуре +10°C. Затем гомогенат центрифугировали при скорости 15000 об/мин в течение 15 минут. Получившуюся верхнюю фракцию объединяли с первичным экстрактом. Промывку повторяли несколько раз. Затем суммарный экстракт выпаривали в центрифужном вакуумном испарителе SpeedVac (Centrivar, Labconco) досуха и хранили при -80°C. Высушенный материал растворяли в пиридине с трикозаном (нормальный углеводород) в качестве внутреннего стандарта. Затем добавляли силилирующий агент BSTFA:TMCS 99:1 (Sigma-Aldrich) и дериватизировали при 90 °С в течение 20 мин. **Анализ**

ГХ-МС проводили на хроматографе Agilent 5860, оснащенном капиллярной колонкой DB-5 MS и сопряженным с квадрупольным масс-селективным детектором Agilent 5975 (Agilent Technologies, Санта-Клара, Калифорния, США). Температура на входе составляла 250 °С в режиме без разделения. Температура колонки начиналась с 70 °С и повышалась до 320 °С со скоростью 6°С в минуту. Скорость потока гелия составляла 1 мл/мин. Ионизацию электронным ударом проводили при 70 В. **Анализ данных ГХ-МС** был проведен с использованием программного обеспечения PARADISE (Department of Food Science Faculty of Science, University of Copenhagen, Denmark, Johnsen et al., 2017) в сочетании с NIST MS (National Institute of Standards and Technology (NIST), USA). Кроме того, для деконволюции и идентификации метаболитов использовали AMDIS (Automated Mass Spectral Deconvolution and Identification System, NIST, USA). Соединения идентифицировали по полученным масс-спектрам и индексам удерживания с использованием библиотек: NIST2010, Golm Metabolome Database (GMD) и собственной библиотеки.

Статистический анализ метаболомных данных проводили в языковой среде R 4.1.1 («Funny looking kid», R Core Team, 2020). Данные были нормализованы относительно выборочной медианы. Данные были логарифмированы и стандартизированы. Если соединение не было обнаружено в образце, но присутствовало в других повторностях, это считалось технической ошибкой и проводилась импутация методом KNN (к-ближайших соседей) с помощью пакета `impute` [37]. Тепловая карта была сделана с помощью пакета `ComplexHeatmap` [38]. OPLS-DA (Orthogonal Partial Least Squares – Discriminant Analysis) выполнялись с помощью `gorpls` [39]. Для анализа обогащения набора метаболитов (MSEA) использовали алгоритм `fgsea` [40]. Метаболомный анализ проведен совместно с сотрудниками СПбГУ и Ботанического института им. Комарова Емельяновым В. В. и Пузанским Р. К.

3. Результаты исследований

3.1. Результаты, полученные на первом этапе выполнения проекта

3.1.1. Секвенирование геномов четырех линий гороха посевного.

В ходе выполнения первой задачи проекта было проведено секвенирование геномов четырех линий гороха посевного (*Pisum sativum* L.), использованных в работе. Сырые данные секвенирования были получены путем прочтения библиотек на одной дорожке прибора Illumina NovaSeq6000 в режиме 2x150, paired end. Секвенирование проведено на коммерческой основе в компании Macrogen Inc. (Сеул, Южная Корея). Числовые характеристики результатов секвенирования представлены в таблице 2.

Таблица 2 — Результаты секвенирования геномов четырех линий гороха посевного

Линия	Количество прочтений млн п.н.	Теоретическое покрытие
Frisson	438,3	43x
Finale	385,6	37x
Sparkle	325,5	33x
SGE	304,9	27x

3.1.2. Идентификация генов, кодирующих пептиды CEP, в геноме гороха и оценка профиля их экспрессии в открытых базах данных

В референсном геноме гороха сорта Frisson была проведена идентификация генов, кодирующих пептиды CEP, с использованием программы SPADA (Small Peptide Alignment Discovery Application) [41]. В качестве основы для создания модели для поиска были использованы последовательности, кодирующие пептиды CEP у *M.truncatula*, взятые из базы данных малых пептидов люцерны [42].

В результате проведенной работы было идентифицировано 15 последовательностей гороха, кодирующих CEP-пептиды. Также были идентифицированы соответствующие аминокислотные последовательности (рис. 1). Анализ данных генов у нескольких линий

гороха позволил установить низкий уровень полиморфизма между выявленными последовательностями.

PsCEP1	MARFYAMQVVLLIIIFGSFCSIEGRQLHIRSEKLNKKVKNKPSSTDSLFLASLPKGTVPSSAPSKRGNSEVVEDEKIVARHLSNEPEVRILLRSVPSGAGH	
PsCEP3	MAQNKSLISLIFLALLVFNQGFKSIEGRFLKSNVFNQSLKHNETNDDIIHVSVSNTSTVTNMSPNVVYNGATGEPSPPPAHDITDFRPTAPGHSPIGHSVHN	
PsCEP9	MAKFSQMLQHFTLFFLALIVFNYSLQSHARLLKPLDQDNVPASTSEKKIDSSMNSHINEDSSENHTYAFQPTTPGNSPGVGHRIFTEEKSKTVVQRPDDNISVTEDTKNDFKKTDPGHSPGVGHAYQNKIGN	
PsCEP10	MGLFEVTTKHLVVFVLYNIAVYNNFLATQARPIKPFNQSSLSIENSATKETGFRPTTPGTSPOGVGHRNFVVEDKNMKTMMVVHSPNIEAFATQRSDDGFKPTNPSSHSPGVGHAYHTKIGNLN	
PsCEP12	MGELAMQKYFVIFLVLTCHIFLPTQARKIKPTIPFTKTDVNVLSPTLESKVDSSLMPNHVVASLGDSTGDTDAFRPTTPGNSPGVGHRKFGREEDKAIKVQSPDVVKVIVITTEGLKNVFKPLDMFAKTQMGN	
PsCEP4	MAHLARIFLYVLLLSHELLTTTEGRILRQSIHPNTESSKIMNTATSTTAFVAKPSHHKRFLEGDACAFRPTNPGHSPGVGHSINS	
PsCEP14	MVLSLVFLMCCQIEISATSRIRSKENSCEISSGIASHVEVSVMDYSDENGSPPPPDDYDFYRKHGDIPSPGAGH	
PsCEP6	MGNTRKLIQITLFFMFFILNQQLFVEGRNLRSLRCGCKPHKSSIIHHERRKSSRAEYEVDDFRPTSPGHSPGVGHSINN	
PsCEP5	MTHFTLIFLFLVSSDILPIEGRSIRRESIGSFESSIERVTRSFVLSRKLRSRRNLEENVEAFRPTTPGHSPGVGHSVKN	
PsCEP7	MPEKTMILTFLIIIMQHNFSGMEASRLNINHPPTIPQSPQAPSLDYYWYSINDDNKGGDDAFRPTSPGHSPGVGHQTPPP	
PsCEP2	MARSSTLLMFLVLFASFNCLEGRNLVVGGEKQWKMMNMKPPSRDSLFLSALPKGTVPSSSPTKGAHVYVDEKLIERHLSMDRRLIIVSAPSGIGH	
PsCEP13	MAISKLVFTISSILFFNLCLSTLGRPIKHTINKLEYASTYKNIVNWRNMLENDTMTATLDSPLNLIHVNAEAKFVDDFRPTDPGHSPGAGHSTPAIPRDANVLRPP	
PsCEP8	MAEKTIMLFLVCLILVQNFGLMEVIGKKVQAAPIPRVLIIRSPQAPSIQFYTKNNDKDGQDAFRPTSPGHSPGVGHDAAPPNRP	
PsCEP11	MKTMVVVHSPNIEAFATQRSDDGFKPTNPSSHSPGVGHAYHTKIGNLN	
PsCEP15	MFMVLSLVFLMCCQIEISATSRIRSKENSCEISSGIASHVEVSVMDYSDENGSPPPPDDYDFYRKHGDIPSPGAGH	

Рисунок 1 – Последовательности СЕР-пептидов гороха посевного.

Филогенетический анализ соответствующих аминокислотных последовательностей позволил выявить гомологичные последовательности у гороха и люцерны (рис. 2).

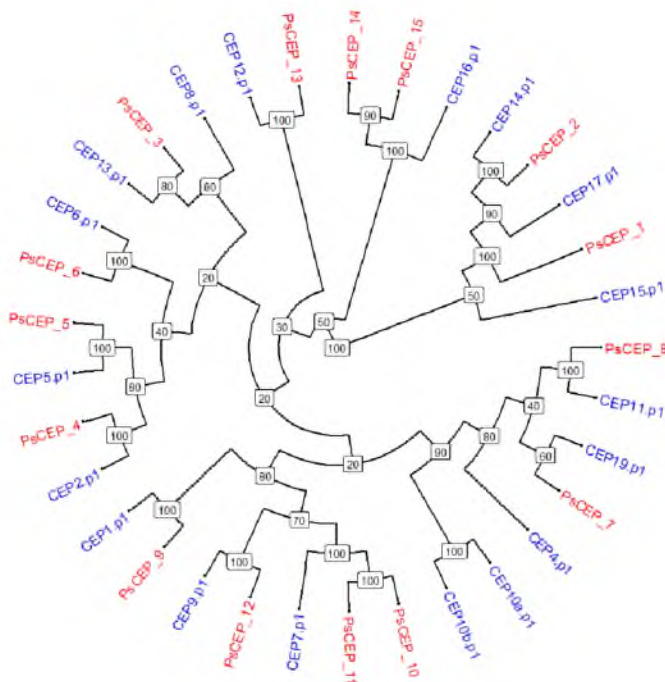


Рисунок 2 – Филогенетическое дерево, отражающее родство аминокислотных последовательностей белков СЕР у гороха (красный цвет) и люцерны (синий цвет).

Как видно из рис. 2, для большинства последовательностей была идентифицирована гомологичная пара. В том числе, были выявлены гомологи СЕР-пептидов люцерны, принимающих участие в регуляции клубенькообразования и развития корня.

С использованием информации, представленной в открытых базах данных (NCBI, EMBL) и экспрессионном атласе гороха [43] был проведен анализ экспрессии выявленных генов в корнях, клубеньках, побегах, листьях и бобах гороха (рис. 3).

ID CEP гороха	Roots_M_A_LN	Roots_M_A_LN	Root_B_LN	Root_F_LN	Nodule_A_LN	Nodule_B_LN	Nodule_G_LN	Shoot_A_LN	Shoot_A_LN	Leaf_B_LN	Lowerleaf_C_LN	Upperleaf_C_LN	Tendril_BC_LN	Stem_BC_LN	Peduncle_C_LN	ApicalNodule_B_LN	Flower_B_LN	Pod_C_LN	Seeds_13 dep	Seed_3 dbi
PsCEP_2	0,00543	0,00741	0,00287	0,00428	0,00658	0,00428	0,00797	0,00171	4,70E-04	0,00449	8,70E-04	0,00224	0,0053	0,00359	0,00538	0	0,00282	1,80E-04	0	0,0022
PsCEP_10	0,00257	0,01219	0,02735	0,01461	0,0128	0,00807	0,0194	0	0	0	0	0	6,60E-04	0,00417	0,00103	0	2,20E-04	0	0	2,30E-04
PsCEP_11	0,00543	0,00741	0,00287	0,00428	0,00658	0,00428	0,00797	0,00171	4,70E-04	0,00449	8,70E-04	0,00224	0,0053	0,00359	0,00538	0	0,00282	1,80E-04	0	0,0022
PsCEP_14	0,01237	0,01892	0,02539	0,0055	4,00E-04	4,10E-04	0,0199	0,00475	0,00424	0,00191	0,00136	5,20E-04	0,00721	0,03211	0,01203	8,90E-04	5,30E-04	0,001362	0,0083	5,40E-04
PsCEP_15	0,01237	0,01892	0,02539	0,0055	4,00E-04	4,10E-04	0,0199	0,00475	0,00424	0,00191	0,00136	5,20E-04	0,00721	0,03211	0,01203	8,90E-04	5,30E-04	0,001362	0,0083	5,40E-04
PsCEP_1	0,0043	0,01175	0,00457	0,00222	0,015	0,00384	0,0126	0,00115	0,00136	0,00432	0,00309	0,01738	0,00498	0,00379	0,00351	0	2,20E-04	2,93E-04	2,20E-04	0,00299
PsCEP_3	0,00161	0,0031	0,00757	0,0203	0	0,0011	0,0011	0	0	0	0	0	0	0	0	4,40E-04	0,00218	8,15E-04	8,02356	8,888E-04

Рисунок 3 — Тепловая карта, отражающая экспрессию генов, кодирующих CEP пептиды, в разных органах гороха посевного.

Как видно из рис.3, гены *PsCEP10*, *PsCEP14* и *PsCEP15* максимально экспрессируются на 10 день после инокуляции (проба «Nodule_G_LN»). Для этих же генов показана высокая экспрессия в корне, особенно у растений, находящихся в начале цветения (проба «Root_B_LN»). В побеге гороха детектируется повышенный уровень экспрессии *PsCEP14* и *PsCEP15* (проба «Stem_BC_LN»). При выращивании в среде, содержащей пониженное и повышенное количество нитрата, значительной разницы в экспрессии генов, кодирующих CEP пептиды, выявлено не было.

3.1.3. Идентификация генов, кодирующих пептиды *CLE* в геноме гороха, и оценка профиля их экспрессии в открытых базах данных

В открытых базах данных транскриптома гороха нами были идентифицированы 48 транскриптов *CLE*, демонстрирующие сходство с последовательностями генов *CLE* у люцерны (см. Рисунки А.1 и А.2 Приложения). Филогенетический анализ соответствующих аминокислотных последовательностей позволил выявить у гороха ближайших гомологов *CLE* других бобовых растений, для которых ранее было показано участие в регуляции развития симбиотических клубеньков. Так, последовательности *PsCam040153*, *PsCam040702*, *PsCam040984* образуют единую ветвь с последовательностями *MtCLE12*, *MtCLE13*, *GmRIC1*, *GmRIC2* для которых ранее было показано участие в системном контроле развития симбиотических клубеньков: сверхэкспрессия соответствующих генов у сои и люцерны системно подавляет клубенькообразование [6,7,44]. Эта ветвь также включает *MtCLE35*, который согласно нашим данным, экспрессируется в клубеньках и опосредует нитрат-зависимое подавление клубенькообразования [45]. Последовательность *PsCam041632* оказалась близкой последовательности генов нитрат-индуцируемых *CLE* у сои, *GmRIC1*, *GmRIC2*, а также гену *MtCLE34*, который, как мы показали, также активируется при добавлении нитрата (см. Рисунок 4, рисунок А.1 Приложения).

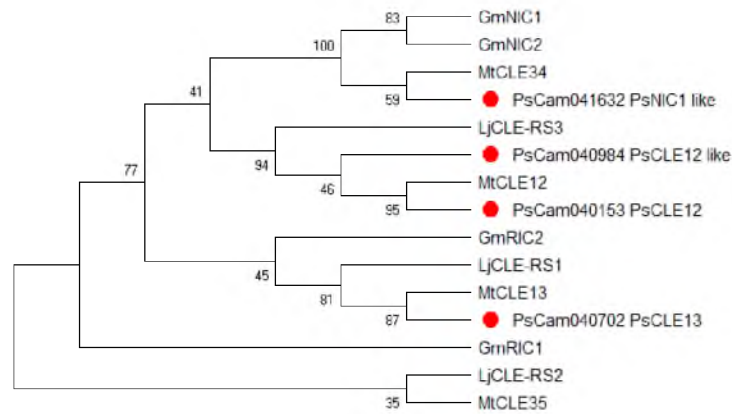


Рисунок 4 — Филогенетическое дерево, отражающее родство аминокислотных последовательностей белков CLE, регулирующих развитие симбиотических клубеньков у бобовых растений, и их ближайших гомологов у гороха (отмечены красными кружками).

Последовательность домена CLE пептида PsCam040702 (названного нами PsCLE13 на основании близкого сходства с MtCLE13) оказалась идентичной последовательности пептидов CLE LjCLE-RS1 и LjCLE-RS1 *Lotus japonicus*, для которых ранее было показано участие в системном подавлении клубенькообразования у лядвенца [1]. Домен CLE PsCam040984 в целом демонстрирует сходство с пептидом LjCLE-RS3 лядвенца: в его начале располагаются аминокислоты триптофан и изолейцин, а не аргинин и лейцин, как у всех других известных пептидов CLE других бобовых растений (см. рис. 5). Домен CLE пептида PsCam041632 (названный нами PsNIC-like на основании сходства с последовательностями GmNIC) оказался идентичен домену CLE у MtCLE34. Домен CLE у PsCam040153 (названный PsCLE12 на основании сходства с MtCLE12) имеет наибольшее сходство с пептидом MtCLE12 (см. Рисунок 2).

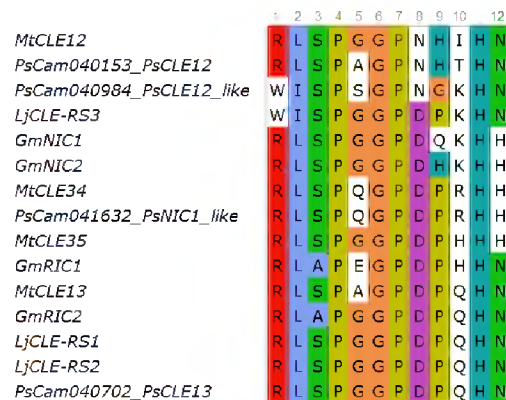


Рисунок 5 — Выравнивание аминокислотных последовательностей домена CLE у гороха, регулирующих развитие симбиотических клубеньков у бобовых растений

На следующем этапе был оценен уровень экспрессии выявленных гомологов *CLE* в разных органах растений гороха (корень, клубенек, побег, усик, плод, цветонос, семя) по транскриптомным данным [43]. Среди всех генов *CLE* у гороха выделяется кластер генов, демонстрирующих специфичную экспрессию в клубеньках (выделен зеленым прямоугольником на Рисунке 3). Этот кластер включает гены *PsCam040702* (*PsCLE13*), *PsCam040984* (*PsCLE12-like*), *PsCam041632* (*PsNIC-like*), которые как мы показали, являются ближайшими гомологами генов *CLE* – регуляторов клубенькообразования у других бобовых. Ген *PsCam040153* (*PsCLE12*), помимо экспрессии в клубеньках, также демонстрирует высокий уровень экспрессии и в корнях, в особенности при высоком содержании азота (проба RootSys_A_HN на рис. 6).

Кроме того, в кластер генов со специфичной экспрессией в клубеньках также входят гены *PsCam0141636* и *PsCam044761*. Ближайшими гомологами этих генов у люцерны являются *MiCLE29* и *MiCLE42*, соответственно. Эти гены у люцерны также экспрессируются в клубеньках [46].

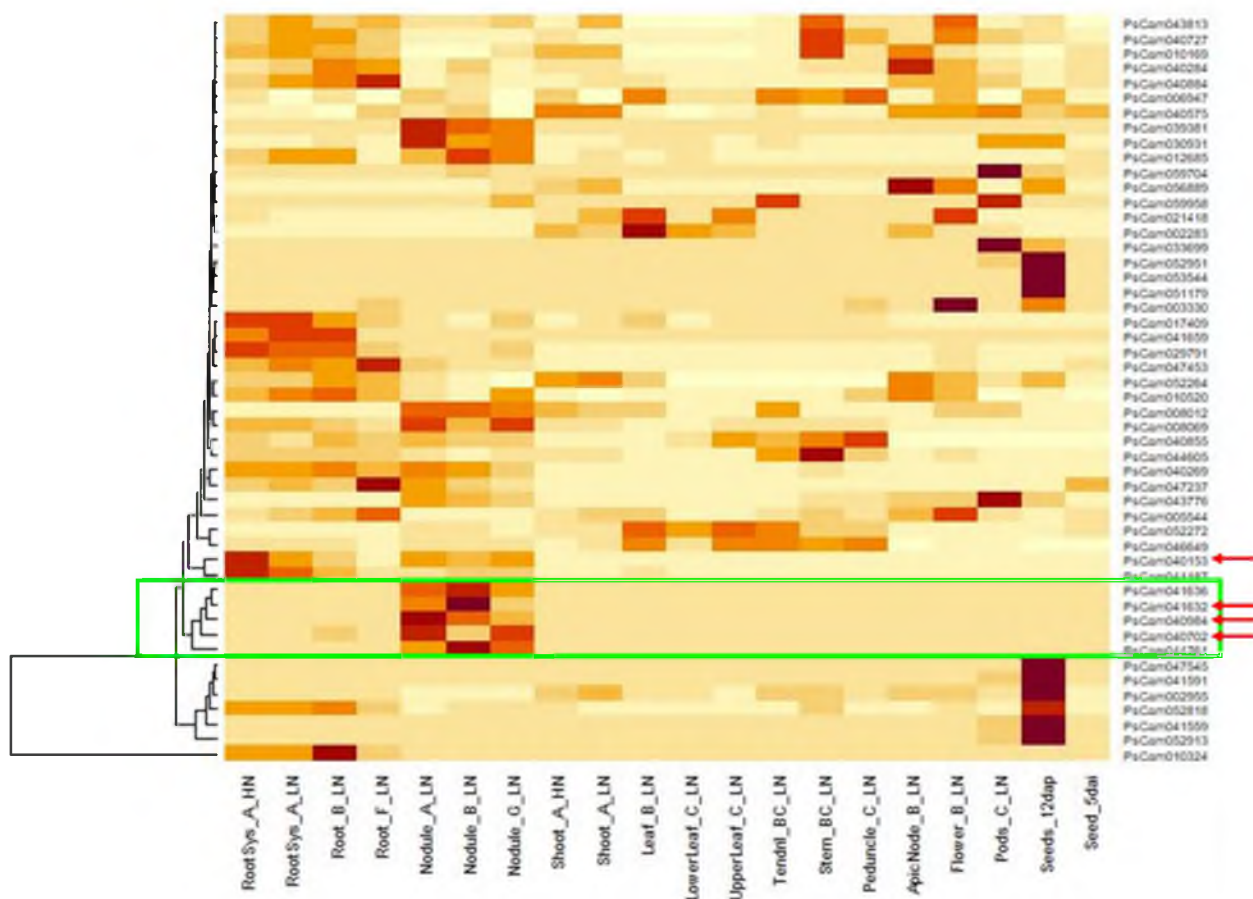


Рисунок 6 — Экспрессия генов *PsCLE* в различных органах гороха. Тепловая карта построена по данным транскриптомного анализа Alves-Carvalho et al., 2015, интенсивность окрашивание ячеек соответствует уровням экспрессии генов (темно-коричневый – более высокие уровни экспрессии, светло-коричневый – более низкие уровни экспрессии).

Для пептида MtCLE42 было изучено влияние на развитие симбиотических клубеньков, и в отличие от пептидов MtCLE12 и MtCLE13 этот пептид не был способен подавлять клубенькообразование [47]. Роль пептидов MtCLE29 и MtCLE42 в развитии симбиотических клубеньков остается не известной. Еще один ген *CLE* гороха, экспрессия которого специфично усиливается в клубеньках – *PsCam039381*, ближайшим гомологом которого является ген *MtCLE44* у люцерны. По данным De Vang и соавт. при развитии клубеньков и в ответ на обработку Nod-факторами усиления экспрессии *MtCLE44* не происходит [46], а согласно данным транскриптомного анализа, доступным в базе <https://iant.toulouse.inra.fr/symbimics/> [48], ген *MtCLE44* демонстрирует слабый уровень экспрессии в клубеньках, тогда как в корне этот ген не экспрессируется. Также интересно отметить, что гены кодирующие пептиды CLE, по аминокислотной последовательности близкие к MtCLE44, были обнаружены в геноме гриба арбускулярной микоризы *Rhizophagus irregularis* [49].

Таким образом, среди идентифицированных генов *CLE* у гороха нами были выбраны четыре гена (*PsCam040702 (PsCLE13)*, *PsCam041632 (PsCLE12)*, *PsCam040984 (PsCLE12-like)* и *PsCam041632 (PsNIC-like)*), которые, как мы предполагаем на основании их сходства с последовательностями *CLE* других бобовых и относительно высоких уровнях экспрессии в клубеньках гороха, могут быть вовлечены в регуляцию клубенькообразования.

3.1.4. Оценка уровней экспрессии генов *PsCLE* и у гороха при клубенькообразовании и при обработке нитратом с помощью количественной ПЦР в реальном времени

На следующем этапе мы оценили уровни экспрессии генов *PsCam040702 (PsCLE13)*, *PsCam041632 (PsCLE12)*, *PsCam040984 (PsCLE12-like)* и *PsCam041632 (PsNIC-like)* в развивающихся примордиях клубеньков у гороха линии Frisson на 7 и 14 день после инокуляции (дпи) ризобиями с помощью количественной ПЦР в реальном времени. Для всех проанализированных генов были выявлены относительно высокие уровни экспрессии в формирующихся клубеньках, по сравнению с неинокулированными корнями (рис. 7).

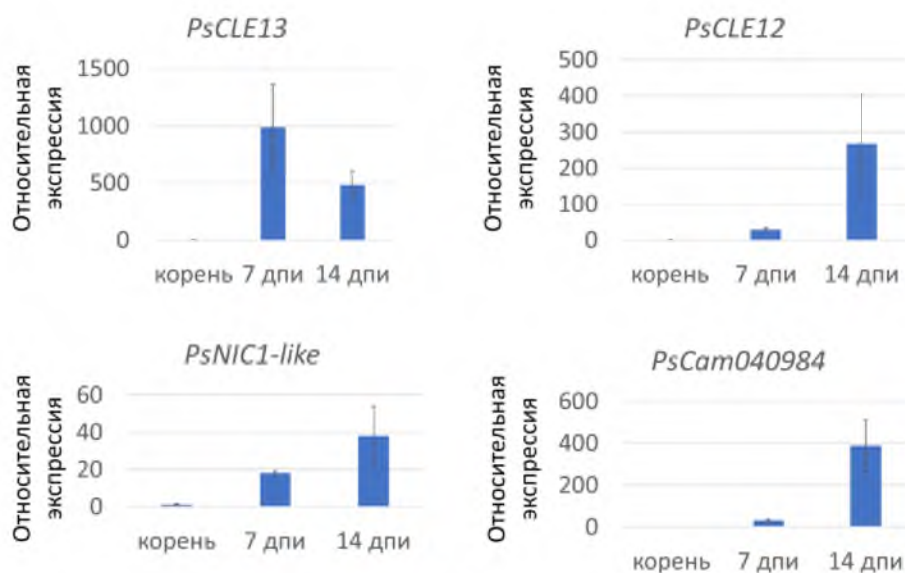


Рисунок 7 — Относительные уровни экспрессии генов *PsCam040702* (*PsCLE13*), *PsCam041632* (*PsCLE12*), *PsCam040984* (*PsCLE12-like*) и *PsCam041632* (*PsNIC-like*) в корне (контроль) и в формирующихся клубеньках на 7 и 14 дни после инокуляции (дни).

Для всех проанализированных генов были выявлены относительно высокие уровни экспрессии в формирующихся клубеньках, по сравнению с неинокулированными корнями (рис. 7). При этом для гена *PsCLE13* были выявлены наиболее высокие уровни экспрессии на 7 дни, тогда как для генов *PsCam041632* (*PsCLE12*), *PsCam040984* (*PsCLE12-like*) и *PsCam041632* (*PsNIC-like*) экспрессия в формирующихся клубеньках была выше на 14 дни. Результаты анализа экспрессии с помощью количественной ПЦР подтвердили данные транскриптомного анализа, полученные Alves-Carvalho с соавторами [43], свидетельствующие об экспрессии изучаемых генов *CLE* в формирующихся клубеньках гороха.

Далее мы проанализировали экспрессию генов *CLE* в ответ на обработку нитратом (10 mM KNO₃) в течение 24 часов. Экспрессия генов *PsCam041632* (*PsCLE12*) и *PsCam041632* (*PsNIC-like*) усиливалась при обработке нитратом, тогда как значительных изменений в уровнях экспрессии гена *PsCam040984* при нитратной обработке по сравнению с контролем обнаружено не было (рис. 8). Согласно нашим данным, экспрессия генов *PsCam041632* (*PsCLE12*) и *PsCam041632* (*PsNIC-like*) усиливается при добавлении нитрата в 5,5, 2,5 и 5 раз, соответственно.

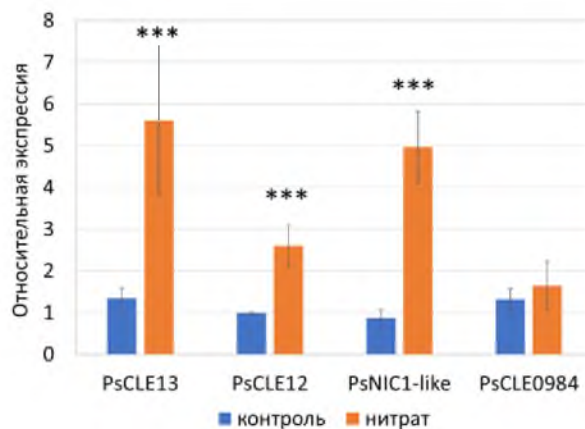


Рисунок 8 — Относительные уровни экспрессии генов *PsCam040702* (*PsCLE13*), *PsCam041632* (*PsCLE12*), *PsCam040984* (*PsCLE12-like*) и *PsCam041632* (*PsNIC-like*) при обработке нитратом (10 mM KNO₃) в течение 24 часов и в контроле. Приведены средние значения для трех биологических повторностей, планки погрешностей обозначают стандартные отклонения (***) $p < 0,01$).

По данным транскриптомного анализа, полученным Alves-Carvalho с соавторами [43], ген *PsCam040153* (*PsCLE12*) экспрессируется на более высоком уровне в корнях при высоком содержании азота (проба RootSys_A_HN на рис. 6), по сравнению с корнями при низком содержании азота (RootSys_A_LN на рис. 6), что также указывает на нитрат-зависимую регуляцию экспрессии этого гена. В дальнейшем мы планируем провести дополнительные эксперименты с использованием различных сроков обработки и различных концентраций KNO₃ и других источников азота для выявления особенностей экспрессии генов CLE в ответ на нитрат и другие источники азота.

Таким образом, согласно нашим предварительным данным, гены *PsCam040702* (*PsCLE13*), *PsCam041632* (*PsCLE12*) и *PsCam041632* (*PsNIC-like*) являются нитрат-регулируемыми. Их экспрессия активируется как при клубенькообразовании, так и при обработке нитратом, как это было показано ранее для генов *LjCLE-RS2*, *LjCLE-RS3*, а также для генов *MiCLE34* и *MiCLE35* (см. раздел 3.5, Lebedeva et al., 2020). Дальнейшие эксперименты по изучению влияния сверхэкспрессии генов *PsCLE* на клубенькообразование у гороха позволят выявить их функцию в регуляции развития клубеньков.

3.1.5. Оценка уровней экспрессии генов *MtCLE* у люцерны при клубенькообразовании и при обработке нитратом с помощью количественной ПЦР в реальном времени

Среди генов люцерны для более детального изучения характера экспрессии и функции мы выбрали два гена, *MtCLE34* и *MtCLE35*, которые являются близкими гомологами других генов бобовых растений, вовлеченных в регуляцию клубенькообразования. Согласно данным количественной ПЦР в реальном времени оба эти гена экспрессируются в развивающихся клубеньках (рис. 9).

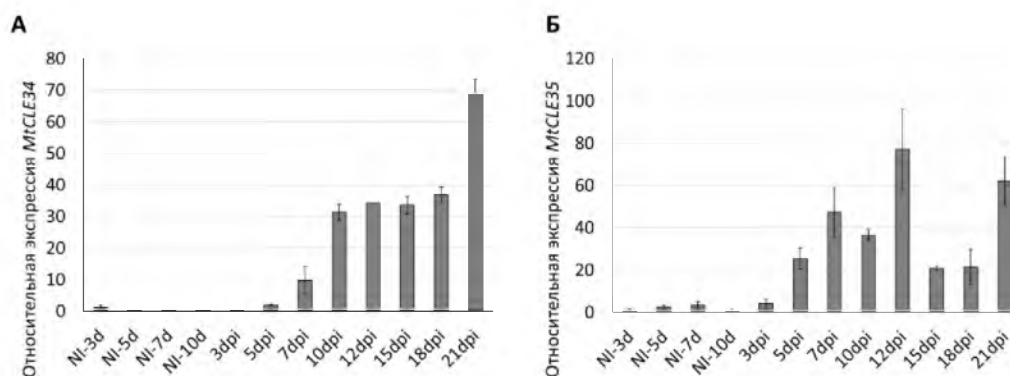


Рисунок 9 — Относительные уровни экспрессии генов *MtCLE34* (А) и *MtCLE35* (Б) на разных сроках после инокуляции ризобиями (dpi, days post inoculation) и в не инокулированных корнях (NI, контроль). Планки погрешностей обозначают стандартные отклонения для средних значений из трех технических повторностей.

Экспрессия гена *MtCLE34* возрастала на 7 дни, более высокий уровень экспрессии сохранялся на 10-18 дни, тогда как на 21 дни наблюдалось еще большее увеличение уровня экспрессии *MtCLE34*. Увеличение экспрессии гена *MtCLE35* наблюдалось на 5 дни, максимум его экспрессии наблюдался на 12 дни, после чего экспрессия *MtCLE35* снижалась на 15 и 18 дни и вновь возрастала на 21 дни (рис. 9).

Далее мы оценили влияние нитрата на экспрессию генов *MtCLE34* и *MtCLE35*, а также двух других генов *CLE* люцерны, *MtCLE12* и *MtCLE13*, которые, как было показано ранее, подавляют клубенькообразование у люцерны (Mortier et al., 2010). При обработке нитратом (10 mM KNO_3) в течение 24 часов мы наблюдали значительное увеличение экспрессии генов *MtCLE34* и *MtCLE35* (в среднем в 57 и 25 раз, соответственно), по сравнению с контролем. При этом экспрессия генов *MtCLE12* и *MtCLE13* значительно не изменялась при обработке нитратом (рис. 10).

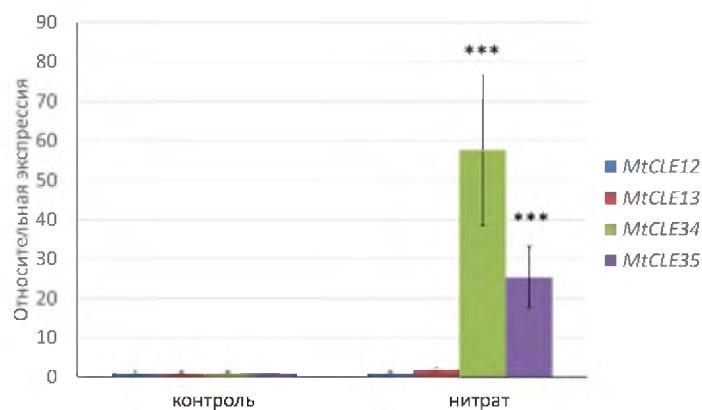


Рисунок 10 — Относительные уровни экспрессии генов *MtCLE34* и *MtCLE35* при обработке нитратом (10 mM KNO₃) в течение 24 часов и в контроле. Приведены средние значения для трех биологических повторов, планки погрешностей обозначают стандартные отклонения (***) $p < 0,01$).

Таким образом, гены *MtCLE34* и *MtCLE35* экспрессируются как при развитии симбиотических клубеньков, так и обработке нитратом. Для изучения функции этих генов в развитии симбиотических клубеньков нами были созданы генетические конструкции для их сверхэкспрессии (см. Раздел 3.6).

3.1.6. Оценка уровней экспрессии генов *MtCLE* у люцерны при клубенькообразовании и при обработке нитратом

Известно, что у арабидопсиса экспрессия генов *CEP* активируется при недостатке нитрата. В работе De Bang и соавт. были идентифицированы 17 генов, кодирующих пептиды CEP. Мы провели количественный анализ экспрессии некоторых генов *MtCEP*, (*MtCEP1*, *MtCEP2*, *MtCEP5*, *MtCEP11*) при обработке нитратом. Обработка нитратом (10 mM KNO₃, 24 часа) подавляла экспрессию всех проанализированных генов *MtCEP* (рис. 11).

Поскольку ранее было показано, что *MtCEP1* стимулирует клубенькообразование у люцерны [8], можно предположить, что подавление активности *MtCEP* при действии нитрата может представлять собой один из путей ингибирующего действия нитрата на развитие симбиотических клубеньков.

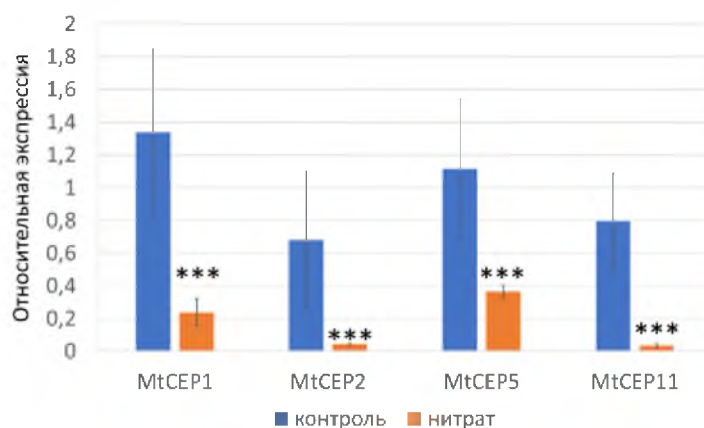


Рисунок 11 — Относительные уровни экспрессии генов *MtCEP1*, *MtCEP2*, *MtCEP5*, *MtCEP11* при обработке нитратом (10 mM KNO₃) в течение 24 часов и в контроле. Приведены средние значения для трех биологических повторов, планки погрешностей обозначают стандартные отклонения (***) $p < 0,01$).

Мы также оценили уровни экспрессии всех идентифицированных у люцерны генов *MtCEP* в базе <https://iant.toulouse.inra.fr/symbimics/>, содержащей данные транскриптомного анализа, полученные Roux и соавт. [48]. Среди всех генов *MtCEP* экспрессия двух генов, *MtCEP16* и *MtCEP17*, была значительно увеличена в клубеньках, по сравнению с корнем (рис. 12).

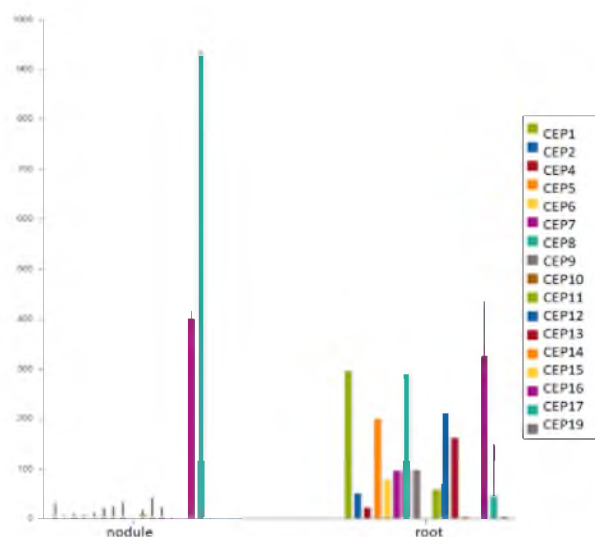


Рисунок 12 — Относительные уровни экспрессии генов *MtCEP* в корне (root) и в клубеньках (nodule) люцерны по данным транскриптомного анализа Roux и соавт. (<https://iant.toulouse.inra.fr/symbimics/>).

Роль генов *MtCEP16* и *MtCEP17* в клубенькообразовании остается неизученной. На следующем этапе работы мы планируем изучить функцию этих генов с использованием конструкций для их сверхэкспрессии.

3.1.7. Создание конструкций для сверхэкспрессии генов *CLE* и *CEP* у гороха и люцерны

Для исследования роли генов *CLE* и *CEP* в развитии симбиотических клубеньков у люцерны и гороха нами были созданы конструкции для их сверхэкспрессии. Кодирующие области генов люцерны *MtCLE34*, *MtCLE35*, *MtCEP17* *MtCEP16*, а также генов гороха *PsCam040702* (*PsCLE13*), *PsCam041632* (*PsCLE12*), *PsCam040984* (*PsCLE12-like*) и *PsCam041632* (*PsNIC-like*) были клонированы в векторах ввода pDONR221 или pDONR207, секвенированы, и затем помещены с помощью клонирования Gateway в вектор назначения pB7WG2D под контроль конститутивного промотора 35S (см. рис. 13, А). Вектор pB7WG2D содержит кассету для экспрессии флуоресцентного маркера GFP, что позволяет проводить отбор трансгенных корней по флуоресценции GFP после трансформации с помощью *A. rhizogenes*. Полученные конструкции проверяли путем рестрикции с помощью *Bam*HI (см. рис. 13, Б), а также секвенирования с использованием праймера к 3'-области промотора 35S.

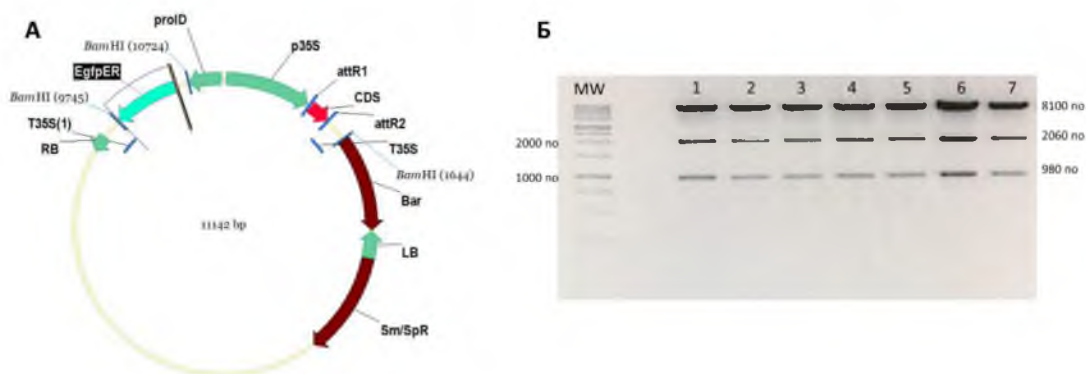


Рисунок 13 — Схема вектора pB7WG2D, используемого для сверхэкспрессии изучаемых генов (А). Пример результатов рестриционного анализа полученных конструкций с использованием рестриктазы *Bam*HI (Б). Приведена электрофореграмма фрагментов рестрикции. MW — маркер молекулярного веса, 1-7 — плазмидная ДНК, обработанная рестриктазой *Bam*HI.

В дальнейшем с помощью полученных конструкций будут трансформированы растения люцерны и гороха для изучения функции генов *CLE* и *CEP* в регуляции развития симбиотических клубеньков и в определении азотного статуса растения.

3.2. Результаты, полученные на втором этапе выполнения проекта

3.2.1. Анализ полиморфизма генов *CLE* у люцерны

Ранее было показано, что гены *MtCLE12* и *MtCLE13*, оказывают ингибирующий эффект на развитие симбиотических клубеньков у люцерны, и их экспрессия активируется с участием компонентов сигнального пути, активируемого ризобиями [44]. Продукт гена *MtCLE35*, как нами было показано ранее, также вовлечен в системный контроль клубенькообразования, и экспрессия этого гена активируется как в ответ на инокуляцию ризобиями, так и при обработке нитратом [45,50]. Наряду с геном *MtCLE35*, близкий ему по последовательности ген *MtCLE34* также активировался при обработке нитратом и в ходе развития симбиотических клубеньков, и таким образом, также является потенциальным регулятором развития симбиотических клубеньков в ответ на присутствие нитрата в почве.

Для данных четырех генов (*MtCLE12*, *MtCLE13*, *MtCLE34*, *MtCLE35*) мы провели поиск однонуклеотидных полиморфизмов (SNP) в последовательностях генов люцерны у 262 линий *M. truncatula*, последовательности которых доступны в базе данных НарМар <http://www.medicagohapmap.org/>. Выявленные SNP приведены в таблице 3.

Таблица 3 — SNP в генах *MtCLE12*, *MtCLE13*, *MtCLE34* и *MtCLE35*, выявленные в базе данных НарМар

Ген	SNP	Codon Change	Effect	Частота альтернативной аллели
<i>MtCLE12</i>	30800355	(A/C) tcA/tcC	Синонимичная	0.0035
	30800406	(T/C) tcT/tcC	Синонимичная	0.064
	30800452	(A/C) Aat/Cat	Несинонимичная (N37H)	0.142
	30800472	(C/T) ctC/ctT	Синонимичная	0.0035
<i>MtCLE13</i>	30793865	(A/G) tcA/tcG	Синонимичная	0.169
	30793889	(C/T) cgC/cgT	Синонимичная	0.01
<i>MtCLE34</i>	39237030	(T/A) ATG/AAG	Несинонимичная	0.201
	39237043		(M30N)	0.0071
	39237092		Синонимичная	0.983

		(T/C) AAT/ AAC (T/C) TAG /CAG	Stop-кодон	
MtCLE35	39246840	(C/T) Cta/Tta	Синонимичная	0.042
	39246899	(C/T) caC/caT	Синонимичная	0.023
	39246937	(A/G) cAt/cGt	Несинонимичная (H43R)	0.01
	39246939	(G/A) Gtg/Atg	Несинонимичная (V44R)	0.092
	39246968	(A/G) ctA/ctG	Синонимичная	0.08

В гене *MtCLE13* были обнаружены только синонимичные замены, тогда как в генах *MtCLE12*, *MtCLE34* и *MtCLE35* выявлены как синонимичные, так и несинонимичные замены, что может указывать на консервативную роль гена *MtCLE13* и значимость его функции для жизнедеятельности растения. Кроме того, у четырех линий (линии НМ101 (линия А17), НМ256, НМ038 и НМ058 из базы НарМар) был обнаружен стоп-кодон в гене *MtCLE34*, наличие которого должно приводить к синтезу укороченного нефункционального продукта без функционального домена CLE. Таким образом, данные линии, которые включают в том числе и линию НМ101 (А17), наиболее широко используемую в лабораторных условиях линию для изучения симбиотических реакций, являются природными мутантами по гену *MtCLE34*. У другой лабораторной линии *M truncatula*, линии R108, согласно последовательностям, представленным в базе данных НарМар, стоп-кодон отсутствует, и таким образом, у линии R108 ген *CLE34* может кодировать функциональный продукт.

3.2.2. Поиск ассоциаций полиморфизма генов *MtCLE* с количеством образующихся клубеньков

В рамках данной работы мы также предприняли попытку оценить, существуют ли возможные ассоциации SNP в геномных участках, охватывающих гены *MtCLE12*, *MtCLE13*, *MtCLE34* и *MtCLE35*, с количеством образующихся клубеньков. Поскольку известно, что продукты генов *MtCLE12*, *MtCLE13*, *MtCLE35* оказывают ингибирующий эффект на развитие симбиотических клубеньков [44,45,50], а ген *MtCLE34* также является потенциальным регулятором развития симбиотических клубеньков, мы предположили, что SNP, приводящие к изменению уровня активности данных генов, могут оказывать влияние на количество образующихся симбиотических клубеньков. Данные о количестве

образующихся клубеньков у разных линий *M. truncatula* были взяты из работы Stanton-Geddes и соавторов [51], которые отдельно учитывали количество клубеньков в верхней части корневой системы (признак *nodule_above*), что отражает более ранние события инокуляции, и количество клубеньков в нижней части корневой системы (*nodule_below*), отражающее более поздние события инокуляции.

Прежде всего, нами была построена точечная диаграмма (XYScatter), отражающая распределение линий *M. truncatula* в соответствии с количеством у них клубеньков в верхней и нижней части корневой системы (рисунок 14). На этой диаграмме мы отметили синим цветом линии, у которых имеется стоп-кодон в гене *MtCLE34* (SNP 39237092) (это линии НМ038, НМ058, НМ101 (линия А17), НМ256), чтобы первоначально оценить, насколько в целом отличается количество клубеньков у данных конкретных линий от количества клубеньков у других линий.

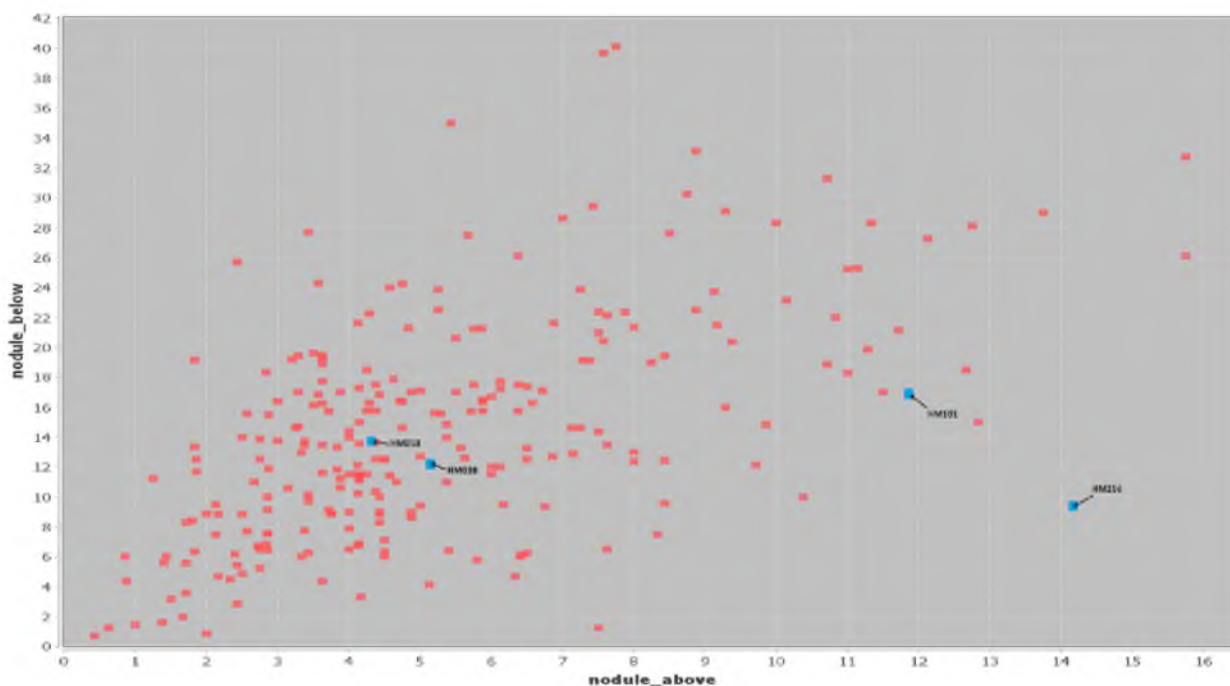


Рисунок 14 — Точечная диаграмма (XYScatter), отражающая распределение линий *M. truncatula* в соответствии с количеством клубеньков в верхней (*nodule_above*) и нижней части (*nodule_below*) корневой системы. Синим цветом отмечены линии, имеющие SNP 39237092 – стоп-кодон в гене *MtCLE34*.

Как видно из данного распределения, у двух линий, НМ038 и НМ058, количество клубеньков в верхней и нижней части корня находится в пределах значений признака, характерных для большинства других линий, тогда как точки, соответствующие двум другим линиям, НМ101 (линия А17) и НМ256, отклоняются от общего распределения – для этих двух линий характерны более высокие значения признака «*nodule_above*». Таким

образом, на основании данного распределения, невозможно однозначно определить, связана ли мутация в гене *MtCLE34*, приводящая к потере функции *MtCLE34* вследствие наличия преждевременного стоп-кодона, к изменению количества клубеньков.

Для выявления генов, ответственных за встречающиеся в природе фенотипические вариации, часто используют полногеномный поиск ассоциаций (Genome-wide association study, GWAS). Этот подход позволяет выявить участки генома, полиморфизм в которых демонстрирует ассоциации с проявлением изучаемого признака. В нашей работе с помощью GWAS в программе Tassel мы провели поиск ассоциаций наличия SNP в участках генома, включающих кластеры генов *CLE* (для *MtCLE12* и *MtCLE13* - в хромосоме 4, *MtCLE34* и *MtCLE35* в хромосоме 2), с количеством образующихся клубеньков на корнях разных линий *M. truncatula*. Данные по SNP были взяты из базы данных HarMap Medicago (<http://www.medicagoharmap.org/>). Нами были использованы модели для поиска ассоциаций, доступные в программе Tassel: Multidimensional Scaling (MDS), General Linear Model (GLM) и Mixed Linear Model (MLM) (Principal Component Analysis (PCA) + K).

В результате проведенного анализа были построены графики Manhattan, отражающие вероятностное распределение участков кластеров генов *CLE* с SNP на хромосомах 2 и 4, и их связь с фенотипическими характеристиками – количество клубеньков в верхней части корневой системы (*nodule_above*) и в нижней части корневой системы (*nodule_below*). График Manhattan представляет собой точечный график, который суммирует результаты GWAS (рисунок А.3 и А.4 Приложения), где по оси X указана геномная позиция SNP, а по оси Y – отрицательный логарифм значения *P* (вероятность наличия ассоциации SNP в данной позиции с изучаемым признаком). Таким образом, более высокие пики на графике, указывают на то, что SNP в данном конкретном положении имеет более высокую ассоциацию с изучаемым признаком (в нашем случае, количество образующихся клубеньков в нижней или верхней части корневой системы). В частности, в позиции 39236000 и соседних точках на хромосоме 2 можно отметить наличие более высоких пиков, что позволяет предположить, что SNP в данном участке генома имеют более сильную связь с количеством образующихся клубеньков в верхней части растения (рисунок А.4 Приложения). Эта область генома расположена перед кодирующей последовательностью гена *MtCLE34*, и возможно, включает в себя регуляторные элементы, влияющие на его уровень экспрессии.

Гены *MtCLE12* и *MtCLE13* на графиках Manhattan представлены позициями 30793500 и 30799500-30800500, и в этих участках не было выявлено SNP,

демонстрирующих высокие показатели степени ассоциации с количеством образующихся клубеньков (рисунок А.5 и А.6 Приложения).

Таким образом, с помощью подхода GWAS нами был выявлен участок, соответствующий позиции 39236000 и соседним точкам в хромосоме 2, который расположен перед кодирующей последовательностью гена *MtCLE34* и в пределах которого наблюдаются SNP, демонстрирующие более высокую степень ассоциации с количеством симбиотических клубеньков, по сравнению с соседними участками генома. Для SNP, расположенных в области других проанализированных генов *MtCLE*, таких ассоциаций выявлено не было.

3.2.3. Анализ полиморфизма генов *CLE* и *CEP* у гороха посевного

На предыдущем этапе работы в ходе анализа генома гороха были определены последовательности серии генов семейств *CLE* и *CEP* гороха посевного. На данном этапе работы была предпринята попытка проведения анализа ассоциаций аллельного состояния некоторых генов из этих семейств с проявлением хозяйственно значимых признаков (масса растений, масса семян, количество семян, содержание в семенах азота и фосфора). Эти признаки были оценены в проведенном ранее вегетационном эксперименте на 99 генотипах культурного гороха [52]. Для анализа были выбраны три гена, гомологичные генам люцерны, для которых нами был проведен подобный анализ, - гены *PsCLE12*, *PsCLE13* и *PsCLE34*. В результате секвенирования кодирующих последовательностей данных генов на выборке из 99 генотипов гороха различий между последовательностями этих генов у разных генотипов выявлено не было. Вероятно, это связано с сильным давлением отбора, действовавшего на гены *PsCLE12*, *PsCLE13* и *PsCLE34* в недавней эволюционной истории вида горох посевной, а также с относительно небольшим размером данных генов. Соответственно, проведение ассоциативного анализа оказалось невозможным ввиду отсутствия полиморфных сайтов в исследованных частях генов. Вероятно, анализ полиморфизма промоторных областей этих и других генов на указанной выборке из 99 генотипов гороха может оказаться полезным для выявления ассоциаций с хозяйственно ценными признаками.

3.2.4. Анализ транскриптома растений гороха и люцерны после обработки нитратом

Транскриптомный анализ с использованием РНК, выделенной из корней гороха после нитратной обработки, проводили в Ресурсном центре СПбГУ «Центр Биобанк», с помощью прибора HiSeq4000 по технологии Illumina в режиме парных прочтений. После обработки “сырых” прочтений (их обрезки и объединения парных прочтений) для каждого образца было получено от 14,8 до 29,73 млн прочтений (см. таблицу 4), при этом их длина составила от 75 до 118 нуклеотидов.

Таблица 4 — Количество прочтений для проанализированных образцов, полученных после очистки и объединения парных прочтений

Название образца	Количество прочтений на образец после фильтрации, млн
Control 1	22,9
Control 2	14,8
Control 3	27,5
Nitrate 1	21,64
Nitrate 2	15,74
Nitrate 3	29,73

Выравнивание прочтений на референсный геном проводили с помощью программы HISAT2. В качестве референса использовали две разные сборки генома гороха: геном *Pisum sativum*_v1a <https://urgi.versailles.inra.fr/download/pea/>, а также геном гороха линии Frisson, полученный в лаборатории генетики растительно-микробных взаимодействий ФГБНУ ВНИИСХМ.

Для оценки вариации между различными образцами использовали метод главных компоненты (PCA), его результаты приведены на рисунке 15.

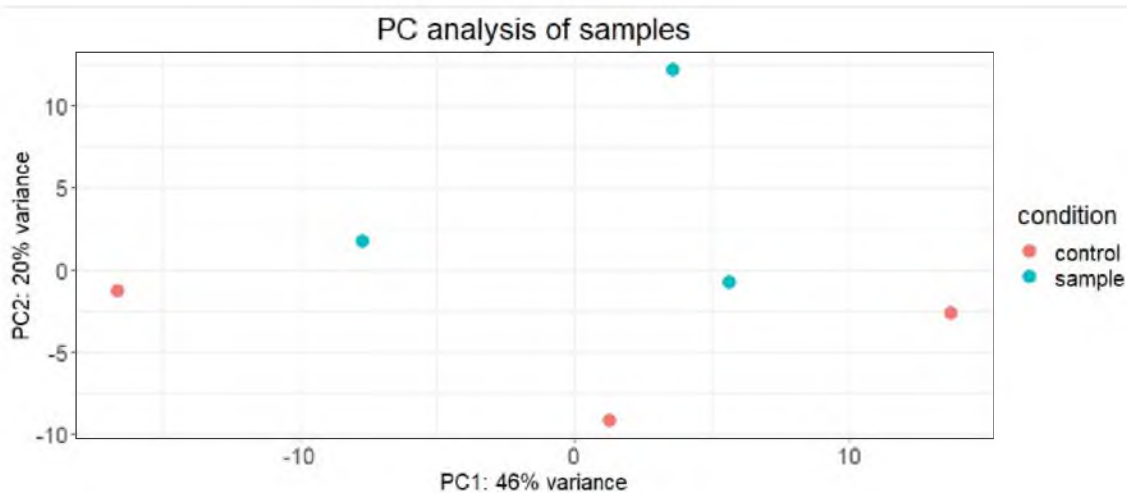


Рисунок 15 — Результаты анализа образцов с помощью метода главных компонент (PCA).

Красным цветом отмечены контрольные образцы без обработки нитратом (control), голубым цветом – образцы после нитратной обработки (sample) – 10 мМ KNO_3 в течение 24 часов.

Согласно результатам PCA, контрольные образцы и образцы после нитратной обработки не четко различаются между собой. Это может свидетельствовать о том, что обработка нитратом в данном эксперименте не вызвала масштабных изменений в спектре экспрессирующихся генов, которые обусловили бы выявляемые с помощью PCA различия между образцами. В связи с этим в дальнейшем следует оптимизировать результаты эксперимента, увеличив продолжительность нитратной обработки и/или концентрацию нитрата, используемую в эксперименте.

Тем не менее, для данных образцов была произведена оценка дифференциально экспрессирующихся генов с помощью программы DESeq2 с пороговыми значениями p -value (adjusted p -value) и \log_2 fold change (кратность изменения уровней экспрессии), равным 0,01 и 1,0, соответственно. При использовании в качестве референса сборки генома гороха линии Frisson (ФГБНУ ВНИИСХМ) было выявлено 24 дифференциально экспрессирующихся гена, из них 22 гена с повышенной экспрессией в образцах с нитратной обработкой, и только три гена – с пониженной экспрессией (см. таблицу А.2. Приложения). При использовании в качестве референса генома *Pisum sativum_v1a* французской сборки с помощью DESeq2 также были выявлены дифференциально экспрессирующиеся гены. В результате был выявлен 101 дифференциально экспрессирующийся ген, в том числе 45

генов с повышенной экспрессией, и 56 генов с пониженной экспрессией в образцах после нитратной обработки (см. таблицы А.3 и А.4 Приложения).

3.2.4. Анализ транскриптома корней люцерны со сверхэкспрессией нитрат-регулируемого гена *MtCLE35*

В ходе предыдущего этапа выполнения работы нами было показано, что сверхэкспрессия активируемого нитратом гена *MtCLE35* под контролем конститутивного промотора 35S приводит к подавлению развития симбиотических клубеньков на корнях люцерны [45]. Более того, сверхэкспрессия этого гена приводила к системному подавлению развития клубеньков – закладка клубеньков значительно подавлялась как на трансгенных корнях, сверхэкспрессирующих *MtCLE35*, так и на нетрансгенных корнях, которые образовывались наряду с трансгенными корням у растений, трансформированных с помощью *A. rhizogenes*. Таким образом, продукт гена *MtCLE35*, регуляторный пептид семейства CLE, как мы предполагаем, способен осуществлять опосредуемое нитратом ингибирование развития клубеньков. Для выяснения механизмов действия пептида MtCLE35 и поиска мишеней пути, им активированного, нами был проведен транскриптомный анализ трансгенных корней со сверхэкспрессией гена *MtCLE35*. В качестве материала для анализа мы использовали трансгенные корни со сверхэкспрессией гена MtCLE35 (*MtCLE35_oe*), в качестве контроля – корни растений со сверхэкспрессией гена *GUS* (*GUS_oe*), кодирующего репортерный ген бета-глюкуронидазы. РНК выделяли из трансгенных корней через 12 дней после инокуляции ризобиями, для каждого варианта эксперимент проводили в трех повторностях. Транскриптомный анализ проводили с помощью технологии MACE (Massive Analysis of cDNA Ends, см. раздел «Материалы и методы»). В таблице 5 приведены названия образцов и количество прочтений для каждого из них.

Таблица 5 — Количество прочтений для анализированных образцов, полученных после этапов очистки и дедупликации

Название образца	Количество прочтений на образец после этапов очистки и дедупликации, млн
GUS-oe-1	1,42
GUS-oe-2	4,59
GUS-oe-3	3,21
MtCLE35-oe-1	2,64
MtCLE35-oe-2	2,07
MtCLE35-oe-3	5,32

Полученные риды выравнивали на референсный геном люцерны *M. truncatula* версии 4.0 (сборка MedtrA17_4.0 [21]) с помощью программы HISAT2 2.0.5 (<https://daehwankimlab.github.io/hisat2/>), количественную оценку транскриптов проводили с помощью программы Stringtie 1.2.4 8 (<http://www.ccb.jhu.edu/software/stringtie/>). Дифференциально экспрессирующиеся гены (ДЭГ) выявляли с помощью пакета DESeq2 (<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>) как гены, демонстрирующие двухкратное изменение уровней экспрессией ($\log_2 \text{fold change} > |1|$) при уровне значимости откорректированного значения $p\text{-value} < 0.05$. Анализ и визуализацию данных проводили в программе R-studio (см. Материалы и методы).

На рисунке 16 приведены результаты анализа проб с помощью метода главных компонент, который выявил различия между контрольными образцами со сверхэкспрессией GUS и образцами со сверхэкспрессией *MtCLE35*. Образцы, соответствующие различным биологическим повторностям в пределах двух вариантов, демонстрируют высокое сходство между собой, за исключением одного из образцов в варианте GUS-oe, который по минорной компоненте PC2, определяющей лишь 4 % вариации, отличается от двух других биологических повторностей в варианте со сверхэкспрессией GUS.

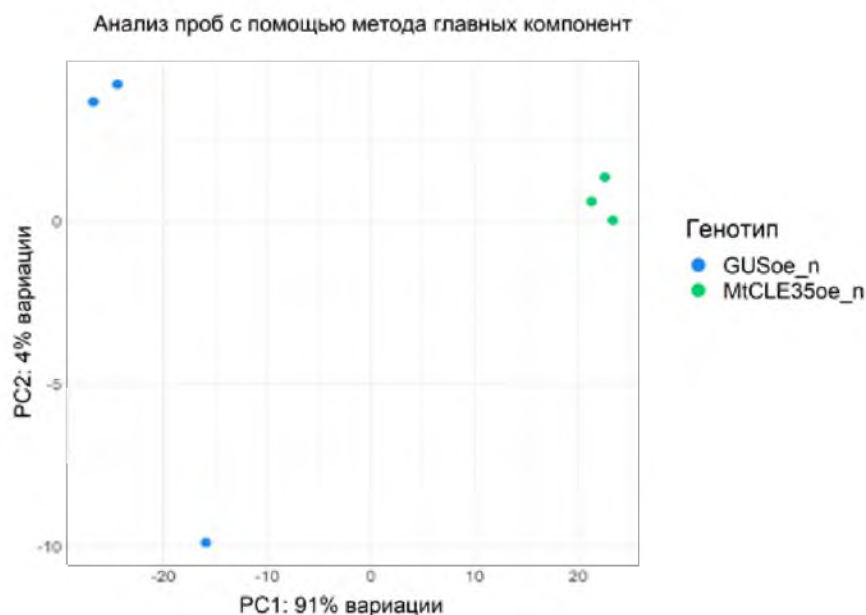


Рисунок 16 — Результаты анализа образцов с помощью метода главных компонент (PCA). Синим цветом отмечены образцы контрольные образцы со сверхэкспрессией GUS (GUS-oe), зеленым цветом – образцы со сверхэкспрессией MtCLE35 (MtCLE35-oe).

Всего при сравнении образцов GUS-oe и MtCLE35-oe было выявлено 889 дифференциально экспрессирующихся генов, из них 714 генов со сниженной экспрессией и 175 генов – с повышенной экспрессией при в варианте со сверхэкспрессией *MtCLE35* (рисунок 17).

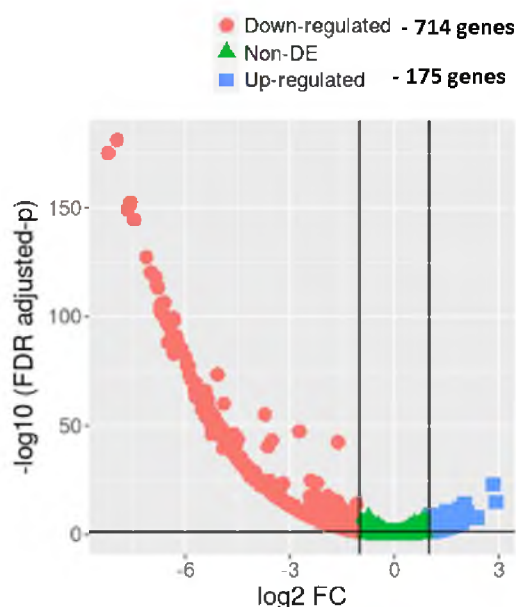


Рисунок 17 — График Volcano plot, иллюстрирующий дифференциально экспрессирующиеся гены при сверхэкспрессии *MtCLE35* (MtCLE35-oe). Гены с повышенным уровнем экспрессии (175 генов) обозначены синим цветом, гены с повышенном уровнем экспрессии (714 генов) обозначены красным цветом.

В образцах MtCLE35-oe максимальное увеличение ($\log_2\text{FoldChange}$ 4,34) при наименьшем значении вероятности ошибки ($p_{\text{adj}} = 5,65\text{E-}14$) было выявлено прежде всего для самого гена *MtCLE35* (Medtr2g091125) (таблица А.6 Приложения), что подтверждает эффективность работы конструкции для сверхэкспрессии этого гена в трансгенных корнях. Был произведен анализ обогащенных групп Gene Ontology (Biological process, Molecular function, Cellular component) с помощью классификатора Medicago Classification Super Viewer (http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_superviewer_medicago.cgi). Среди биологических процессов (Biological process) значительное обогащение продуктами генов с повышенной экспрессией в образцах MtCLE35-oe было обнаружено для метаболических процессов и белкового метаболизма (см. рисунок 18). Среди клеточных компонентов (Cellular component) обогащенными продуктами генов с повышенной экспрессией в значительной степени оказались компоненты клеточной стенки (Cell wall). Таким образом, конститутивная активация гена *MtCLE35*, продукт которого, помимо клубенькообразования, вероятно, опосредует и другие нитрат-зависимые реакции у растений, приводит к значительным метаболическим изменениям. В дальнейшем, мы планируем проанализировать экспрессию выявленных дифференциально экспрессирующихся генов с помощью количественной ПЦР, а также провести оценку метаболомный анализ и оценку содержания азота в тканях трансгенных растений со сверхэкспрессией гена *MtCLE35*.

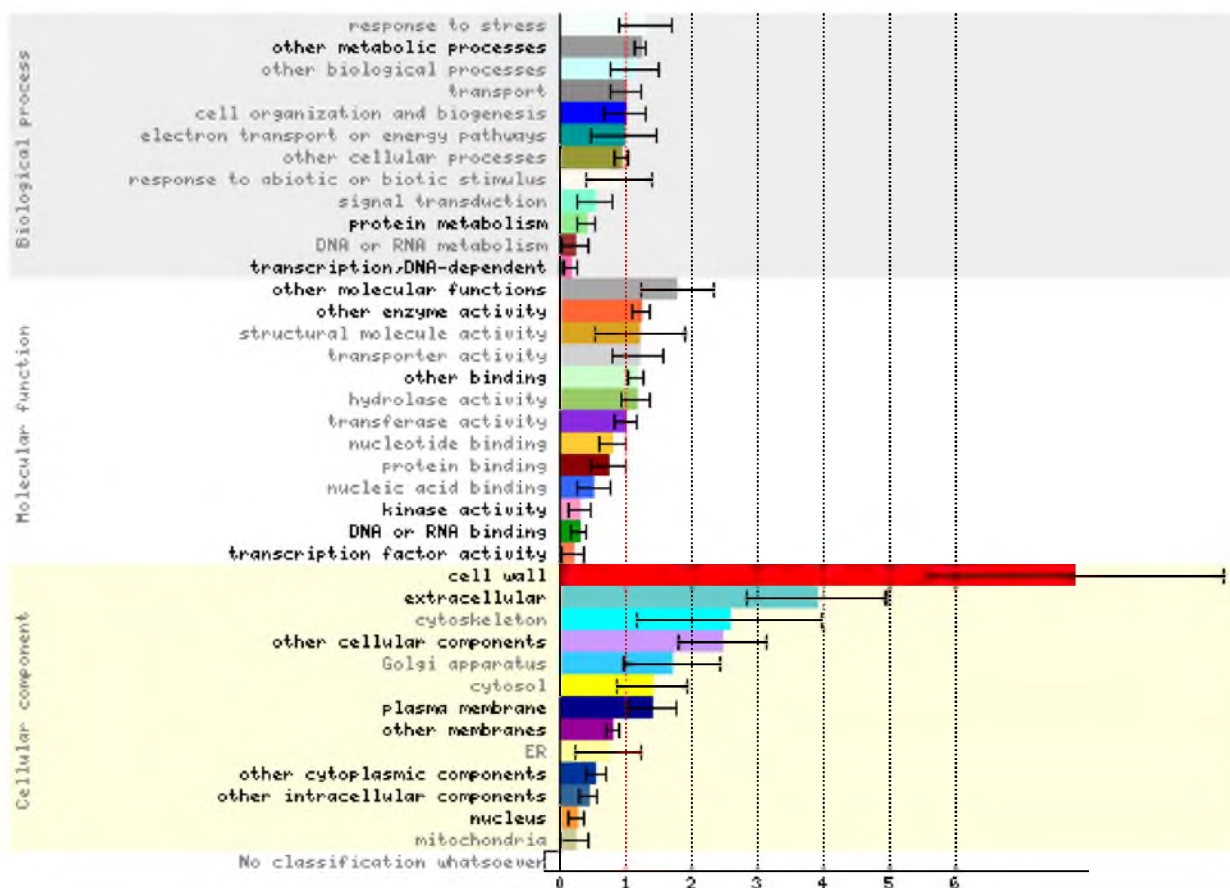


Рисунок 18 — Гистограмма, иллюстрирующая функциональную аннотацию (Gene Ontology, UniProt) продуктов генов, экспрессия которых увеличивается в корнях со сверхэкспрессией *MtCLE35*. Жирным шрифтом выделены процессы, для которых отмечено статистически значимое обогащение продуктами генов с повышенной экспрессией в образцах *MtCLE35_oe* ($p < 0.05$)

Сверхэкспрессия гена *MtCLE35* в корнях приводит к практически полному подавлению развития симбиотических клубеньков [45]. Выявленное нами значительное число генов (714 генов, около 80% от всех ДЭГ), экспрессия которых подавляется в образцах в образцах *MtCLE35-oe*, согласуется с ролью продукта этого гена как ингибитора клубенькообразования. Среди дифференциально экспрессирующихся генов, экспрессия которых значительно снижена в образцах – ключевые гены-регуляторы симбиотических клубеньков, кодирующие транскрипционные факторы NIN (NODULE INCEPTION), NSP1 (NODULATION SIGNALING PATHWAY 1), NF-YA1/HAP2, а также ERN1 (ETHYLENE-RESPONSIVE TRANSCRIPTION FACTOR 1) (см. рисунок 19, рисунок А.7 Приложения, таблица А.6 Приложения). Гены, кодирующие данные транскрипционные факторы, активируются с участием компонентов сигнального каскада, индуцируемого ризобиями. При этом следует отметить, что статистически значимого снижения экспрессии генов, кодирующих рецепторы Nod-факторов (NPF и LYK3), а также генов, кодирующих

компоненты сигнального каскада, кальций/кальмодулин-зависимую киназу DMI3 (DOESN'T MAKE INFECTIONS 3) не наблюдалось (см. рисунок 19).

Gene name	GID	log2FoldChange	padj
NFP	Medtr5g019040		0,783071
LYK3	Medtr5g086130	-2	0,408136
DMI1	Medtr2g005870		0,952514
DMI2	Medtr5g030920		0,003938
DMI3	Medtr8g043970		0,77634
IPD3	Medtr5g026850		9,24E-05
NSP1	Medtr8g020840	-1	0,001456
NSP2	Medtr3a072710	-1	0,485258
NF-YA1/HAP2	Medtr1g056530	-11	3,39E-19
ERN1	Medtr7g085810	-2	0,029442
NIN	Medtr5g099060	-4	3,38E-17
MtRR11	Medtr8g038620		0,667667
MtRR9	Medtr3g015490		0,973856

Рисунок 19 — Изменение уровней экспрессии (log2FoldChange) генов-регуляторов симбиоза в образцах MtCLE35-oe. Черный цвет указывает на отсутствие изменений, зеленый цвет обозначает значимое снижение экспрессии генов в образцах MtCLE35-oe. Красными прямоугольниками выделены гены, экспрессия которых статистически значимо снижается в образцах MtCLE35-oe ($padj < 0.05$).

Таким образом, регуляторный путь, активируемый пептидом MtCLE35, блокирует развитие симбиотических клубеньков за счет подавления уровней экспрессии генов ключевых транскрипционных факторов, регулирующих программу симбиоза, но при этом не затрагивает самые ранние этапы, связанные с рецепцией Nod-факторов. В отличие от этого, согласно проведенному ранее транскриптомному анализу корней со сверхэкспрессией *MtCLE13* среди генов-мишеней пути, активируемого пептидом MtCLE13, выявлен ген NFP, кодирующий рецептор Nod-фактора [53]. Таким образом, пептиды MtCLE13 и MtCLE35, действуя через общий компонент рецепторного комплекса, рецепторную киназу MtSUNN, опосредуют подавление развития симбиотических клубеньков, но при этом действуют на разные этапы реализации симбиотической программы: MtCLE13 действует на самых ранних этапах взаимодействия, подавляя рецепцию Nod-факторов, тогда как MtCLE35 осуществляет блок программы симбиоза на чуть более поздних ее этапах, подавляя активность генов ключевых транскрипционных факторов, регулирующих развитие симбиотического клубенька и бактериальной инфекции.

3.2.5. Оптимизация методов трансформации гороха с помощью *Agrobacterium rhizogenes* и получение материала для транскриптомного анализа корней со сверхэкспрессией генов *PsCLE12*, *PsCLE13* и *PsCEP10*

На предыдущем этапе работы нами были получены конструкции для сверхэкспрессии генов гороха *PsCLE12*, *PsCLE13* и *PsCEP10*. Гены *PsCLE12* и *PsCLE13* являются ближайшими гомологами генов люцерны *MtCLE12* и *MtCLE13*, соответственно. Ген *PsCEP10*, как мы обнаружили ранее с помощью анализа доступных баз данных, активируется в ходе развития симбиотических клубеньков, а также его экспрессия подавляется при воздействии нитрата. Его ближайшим гомологом у люцерны является ген *MtCEP7*, продукт которого является стимулятором развития симбиотических клубеньков у люцерны [9] и, таким образом, выступает в роли антагониста пептидов *MtCLE12* и *MtCLE13* – ингибиторов клубенькообразования.

Конструкции для сверхэкспрессии генов гороха *PsCLE12*, *PsCLE13* и *PsCEP10* включают кодирующие области соответствующих генов контролем конститутивного промотора 35S в векторе назначения pB7WG2D, который также содержит кассету для экспрессии флуоресцентного маркера GFP, позволяющую производить отбор трансгенных корней по флуоресценции GFP после трансформации с помощью *A. rhizogenes*.

Полученные конструкции (*PsCLE12-oe*, *PsCLE13-oe* и *PsCEP10-oe*) были введены в штамм *A. rhizogenes* Arqua-1 и использованы для трансформации проростков гороха. В качестве контроля использовали конструкцию *GUS-oe*.

Для трансформации гороха проводили способом, позволяющим сохранить собственную корневую систему растения. Трёхдневные проростки гороха помещали на чашки Петри, содержащие среду Farhaeus, по четыре проростка на чашку. С помощью иглы от шприца с нанесенной суспензией агробактерий, осуществляли укол в верхней части корня, и дополнительно наносили суспензию агробактерий в область укола. Через примерно 7 дней в области укола наблюдали формирование каллуса, и на этом сроке растения переносили в горшки с вермикулитом (см. рисунок 20). Из каллусов в последствие формировались трансгенные корни, которые отбирали по флуоресценции репортерного белка GFP.



Рисунок 20 — Фотографии растений гороха, трансформированных *A. rhizogenes*. Стрелки указывают на каллус, образующийся в месте инокуляции агробактериями.

С помощью количественной ПЦР в реальном времени на матрице кДНК, полученной из трансгенных корней, было подтверждено наличие высоких уровней экспрессии генов *PsCLE12*, *PsCLE13* и *PsCEP10* в соответствующих вариантах (см. рисунок 21).

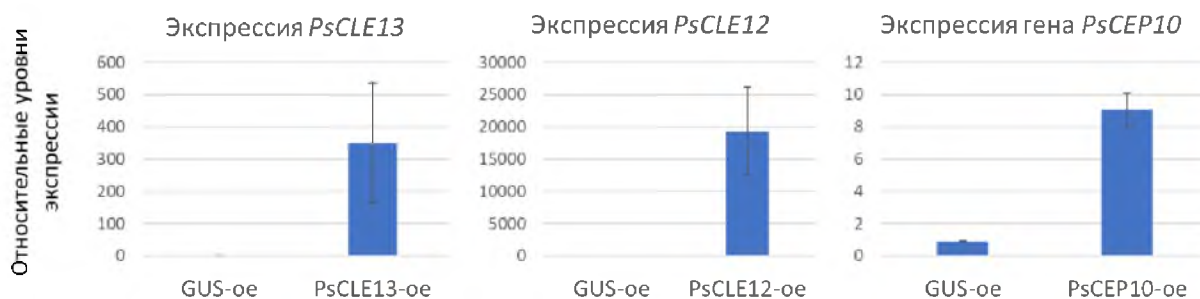


Рисунок 21 — Уровни экспрессии генов *PsCLE13*, *PsCLE12* и *PsCEP10* в трансгенных корнях со сверхэкспрессией соответствующих генов.

РНК, выделенную из трансгенных корней, содержащих конструкции *PsCLE12-oe*, *PsCLE13-oe*, *PsCEP10-oe* и *GUS-oe* (контроль), использовали для приготовления библиотек для транскриптомного анализа методом MACE.

3.3. Результаты, полученные на третьем (заключительном) этапе выполнения проекта

3.3.1. Анализ полиморфизма генов *PsSym29* и *PsNRLK1* у гороха

Число клубеньков и уровень микоризной колонизации корней у бобовых растений контролируется системой авторегуляции (autoregulation of nodulation (AON)). В данной системе важную роль играют пептиды CLE (CLAVATA3/EMBRYO SURROUNDING REGION-RELATED), которые синтезируются в корнях при образовании клубеньков и транспортируются в побеговую часть, являясь сигналом о статусе клубенькообразования. В побегах пептиды CLE связываются с рецепторной киназой, гомологичной CLAVATA1 (CLV1) *Arabidopsis thaliana* (L.) Heynh., что в дальнейшем приводит к ингибированию дальнейшего клубенькообразования в соответствии с пока неохарактеризованным механизмом.

У люцерны слабоусечённой (*Medicago truncatula* Gaertn.) киназа, гомологичная CLV1, кодируется геном *MtSUNN* (*Super Numeric Nodules*, *Medtr4g070970*). У гороха соответствующий ген носит название *PsSym29* [54]. В геноме *M. truncatula* присутствует не менее 10 генов, гомологичных *MtSUNN*, среди которых один, *Nodulation related LRR kinase 1* (*MtNRLK1*, *Medtr5g090100*), повышает свой уровень экспрессии при инокуляции ризобиями [55].

Полиморфизм генов, вовлечённых в AON, может оказывать влияние на урожай растений гороха, поскольку AON влияет на количество образующихся клубеньков, а это, в свою очередь, связано с количеством фиксированного азота. Для того, чтобы исследовать эту связь, мы изучили полиморфизм гена *PsSym29* и его гомолога *PsNRLK1*, который характеризуется клубенёк-специфичной экспрессией, и провели анализ ассоциаций с параметрами урожайности 99 генотипов гороха, выращенных в условиях инокуляции клубеньковыми бактериями и арбускулярно-микоризными грибами.

Гены *PsSym29* (*Psat7g183240*) и *PsNRLK1* (*Psat2g019520*) состоят из двух экзонов, разделённых интроном, и кодируют белки с 20 доменами LRR (leucine rich repeat), трансмембранный домен и киназный домен (рисунок 22). Белки *PsSym29* и *PsNRLK1* сходны на 57.6% по параметру «similarity» и на 44.1 % по параметру «identity». В соответствии с данными экспрессионного атласа гороха [43], максимальный уровень экспрессии *PsSym29* наблюдается в листьях и усиках, а максимальный уровень экспрессии *PsNRLK1* – в клубеньках (рисунок 23).

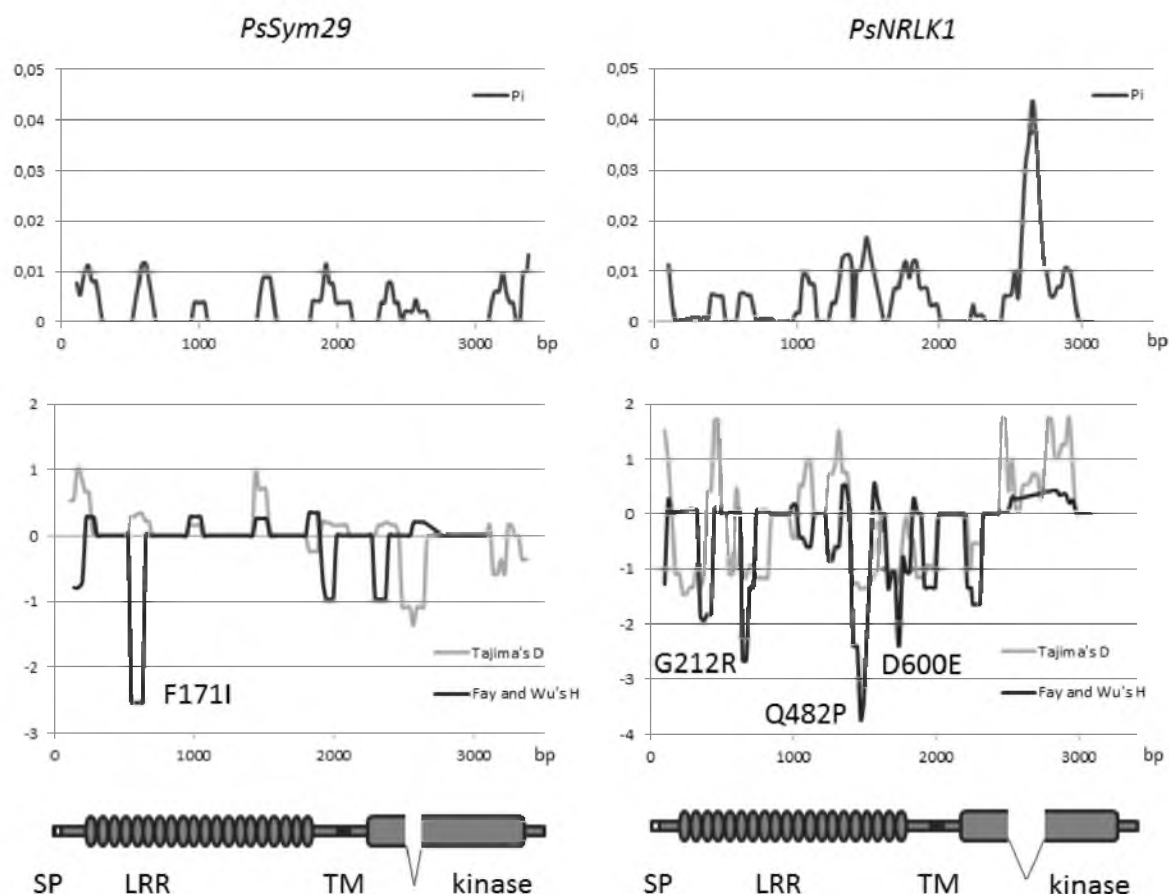


Рисунок 22 — Нуклеотидное разнообразие и значения критериев D Таджими и H Фэя и Ву, рассчитанные для последовательностей генов *PsSym29* и *PsNRLK1*, и доменная структура белков, кодируемых данными генами.

Указаны аминокислотные замены в сайтах, в которых выявлены сигнал положительного отбора. SP – сигнальный пептид, LRR – лейцин-богатые повторы, TM – трансмембранный домен, kinase – киназный домен.



Рисунок 23 — Экспрессия генов *PsSym29* и *PsNRLK1* по данным экспрессионного атласа гороха

В ходе работы кодирующие участки генов *PsSym29* и *PsNRLK1* были секвенированы на ДНК 99 генотипов гороха из коллекции ВИР, которые различаются по месту происхождения и таким образом представляют разнообразие вида горох посевной. Информация об использованных праймерах приведена в таблице А.7 Приложения. На основании анализа полиморфизма были определены значения нуклеотидного разнообразия (π) для каждого из генов (таблица 5). Ген *PsNRLK1* продемонстрировал более высокое нуклеотидное разнообразие, чем *PsSym29*. Нуклеотидное разнообразие, рассчитанное для синонимичных замен (π_s) оказалось сопоставимым для обоих генов, но нуклеотидное разнообразие, рассчитанное для несинонимичных замен (π_a), оказалось выше для *PsNRLK1* (таблица 5). Соотношение π_a / π_s также выше для *PsNRLK1*, чем для *PsSym29*, что отражает более сильное давление очищающего отбора, действующего на последовательность гена *PsSym29*. У *M. truncatula*, мутации в гене *MtSUNN* (ортологичном *PsSym29*) приводят к фенотипу суперклубенькообразования [2], а мутации в гене *MtNRLK1* не влияют на фенотип клубенькообразования [55], что также указывает на более высокую важность *MtSUNN* для растений люцерны по сравнению с геном *MtNRLK1*.

Таблица 5 — Нуклеотидное разнообразие генов *PsSym29* и *PsNRLK1*

Ген	π	π_s	π_a	π_a / π_s
<i>PsSym29</i>	$2,87 \cdot 10^{-3}$	$7,17 \cdot 10^{-3}$	$1,13 \cdot 10^{-3}$	0,158
<i>PsNRLK1</i>	$4,48 \cdot 10^{-3}$	$6,94 \cdot 10^{-3}$	$1,98 \cdot 10^{-3}$	0,285

π – нуклеотидное разнообразие, π_s – нуклеотидное разнообразие, рассчитанное для синонимичных замен, π_a – нуклеотидное разнообразие, рассчитанное для несинонимичных замен, π_a / π_s – соотношение частот синонимичных и несинонимичных замен.

С целью выявления специфических районов генов, которые могут содержать свидетельства давления отбора, был проведён анализ полиморфизма генов *PsSym29* и *PsNRLK1* методом «скользящего окна». Были определены значения нуклеотидного разнообразия (π) и критериев D Таджimy и H Фэя и Wu. Значение π для последовательности *PsNRLK1* значительно увеличивается в интронной части гена (что указывает на то, что интрон этого гена весьма полиморфен), в то же время значение π для гена *PsSym29* одинаково в экзонах и в интроне. Значения критерия Таджimy D для полных последовательностей генов (0.023 для *PsSym29* и -0,607 для *PsNRLK1*), а также

определённые в «скользящем окне» размером 100 п.н., не указывает на какие-либо отклонения от модели нейтральной эволюции. В то же время, анализ критерия Н Фэя и Ву позволил выявить признаки позитивного отбора в одном сайте гена *PsSym29* и в трёх сайтах гена *PsNRLK1* (рисунок 22). Таким образом, результаты анализа показывают, что гены *PsSym29* и *PsNRLK1* эволюционируют в настоящее время, и их новые аллели, возникшие относительно недавно в эволюционном смысле, замещают старые аллельные варианты. Выявленные полиморфные сайты с несинонимичными заменами находятся в первых экзонах генов, которые кодируют рецепторные домены белков, что указывает на возможность того, что рецепторы, кодируемые разными аллельными состояниями генов, могут иметь различную аффинность к своим лигандам (пептидам CLE).

В работе были выявлены несинонимичные замены в последовательностях генов (8 для *PsSym29* и 20 для *PsNRLK1*). Анализ ассоциаций был проведён после исключения из рассмотрения редких аллельных вариантов (для шести сайтов в *PsSym29* и семи в *PsNRLK1*) (Таблица А.8 Приложения). Согласно предсказанию, выполненному в программе SIFT [56], все аминокислотные замены, соответствующие выявленным полиморфным сайтам, не являются критичными для функционирования белка. В результате анализа не было выявлено статистически значимой ассоциации полиморфных сайтов с параметрами роста и урожайности растений 99 генотипов гороха, выращенных в условиях инокуляции клубеньковыми бактериями и грибами арбускулярной микоризы (Таблица А.8 Приложения). Также не было выявлено влияния полиморфных сайтов на прибавки, вызванные действием грибов АМ (Таблица А.9 Приложения).

3.3.2. Секвенирование генома мутанта гороха по гену *fas*, предположительно кодирующему компонент системы CLAVATA

Для выполнения данной задачи были выращены растения гороха посевного линии Штамбовый (несущей мутацию в гене *fas*). Из листьев одного растения были выделены ядра (с использованием модифицированной технологии, основанной на [57]), а затем из полученных ядер была выделена ДНК (с использованием модифицированного СТАВ-метода [58]). Качественный и количественный анализ выделенной геномной ДНК был проведен с использованием системы NanoDrop OneC (Thermo Scientific). В результате проведенного выделения конечная концентрация ядерной ДНК составила 231 нг/мкл. Секвенирование геномной ДНК мутантного генотипа Штамбовый с

использованием системы секвенирования нового поколения Illumina NovaSeq 6000 (две дорожки, ячейка SP, 2 x 150 нт).

В результате было получено 601 557 150 пар ридов для сорта Штамбовый (~ 1.2 миллиарда), прошедших контроль качества системой Illumina. Дополнительно анализ качества оценили в программе FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), которая также подтвердила высокое качество секвенирования. После фильтрации ридов на предмет возможных технических последовательностей и удаления участков с низкой достоверностью при помощи программы Trimmomatic (<http://usadellab.org/>), риды картировали на геном гороха линии Cameog [59] с использованием алайнера bowtie2 (<https://bowtie-bio.sourceforge.net/bowtie2>). Общий уровень совпадений при картировании ридов для генотипа Штамбовый составил 93.1%. Выявление нуклеотидных вариантов в геноме анализируемого генотипа сортов, включая небольшие делеции и инсерции (менее 10 по), провели с использованием программы GATK (<https://gatk.broadinstitute.org/>).

Приблизительная локализация гена *fas* известна [60]. В ходе анализе генома генотипа Штамбовый был изучен регион, фланкированный генетическими маркерами Рерсп и РК4, отдаленным друг на друга на расстояние 44.9 сМ. Данный регион содержит локус FAS и соответствуют участку от 474115077 нуклеотида до 547001760 нуклеотида 5-ой хромосомы гороха (III группа сцепления). В данном регионе было детектировано 54 нуклеотидных варианта в участках хромосомы, соответствующих аннотированным генам. Для одного гена *Psat5g277640* (транскрипционный фактор семейства NAC) было обнаружено, что четыре варианта приводят к аминокислотной замене в белке: 541653901_A/G, 541654040_T/A, 541654052_C/T и 541654102_G/A.

Проделанная работа является заделом для идентификации молекулярной природы гена *fas* у гороха посевного. В дальнейшей работе необходимо провести более точное картирование гена *fas* в геноме гороха, а также секвенировать геном родительского сорта Немчиновский для поиска мутации, являющейся причиной мутантного фенотипа *fas*.

3.3.3. Анализ транскриптома мутанта гороха по гену *fas* в условиях образования симбиоза с клубеньковыми бактериями и при обработке нитратом

Эффект от действия нитрата калия на клубенькообразование гороха «дикого типа» сорта Немчиновский и мутанта по гену *fas* Штамбовый оценили в вегетационном

эксперименте. Растения были выращены в сосудах с песком и при посадке были инокулированы штаммом клубеньковых бактерий *Rhizobium leguminosarum* bv. *viciae* RCAM1026. Один раз в неделю растения поливали питательным раствором, который содержал KNO₃ в концентрации 15 ммоль/л в варианте «обработка нитратом» и не содержал KNO₃ в варианте «контроль». Через 4 недели после посадки растения были извлечены из субстрата, для них было оценено число клубеньков, и одновременно корневые системы трёх растений из одного сосуда (одна биологическая повторность) были зафиксированы замораживанием в жидком азоте для последующего выделения РНК и транскриптомного анализа.

Было проведено статистическое сравнение числа клубеньков, образованных мутантным генотипом Штамбовый и генотипом «дикого типа» Немчиновский (рисунок 24). Для статистического анализа влияния фактора генотипа и фактора доступного азота на число клубеньков использовали двухфакторный дисперсионный анализ с последующим проведением пост-хок тестов на значимость факторов. Анализ показал, что сорт Штамбовый образует примерно на 20% меньше клубеньков ($t=-5.602$, $\text{Pr}(>|t|)=1.51\text{e-}07$) по сравнению с Немчиновским. Присутствие доступного азота в субстрате ингибирует клубенькообразование примерно на 30% для обоих сортов ($T=-9.134$, $\text{Pr}(>|t|)=3.08\text{e-}15$). Значимого взаимодействия двух факторов не выявлено. Без учета взаимодействия факторов значение $F_{3,112}=38.18$, что соответствует $p\text{-value}=2.2\text{e-}16$. Таким образом, можно сделать вывод, что мутация в гене *fas* негативно влияет на клубенькообразование, и при этом не оказывает влияния на ответ растения на повышенное содержание нитрата в субстрате.

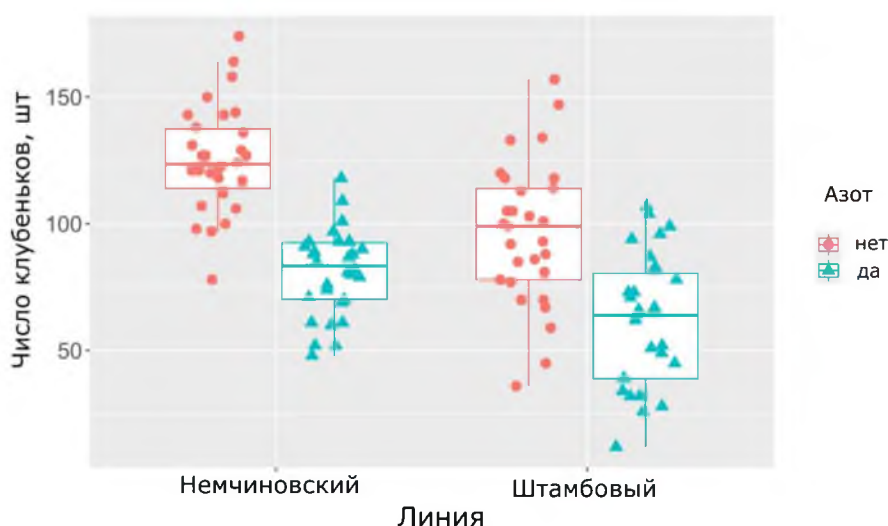


Рисунок 24 — Клубенькообразование растений генотипов Немчиновский и Штамбовый (*fas*) под влиянием нитрата в субстрате

Корневые системы, содержащие клубеньки, были гомогенизированы в замороженном виде на шаровой мельнице, и из полученного материала была выделена РНК и использована для приготовления библиотек для секвенирования по протоколу MACE. Библиотеки были секвенированы на приборе Illumina NovaSeq6000 в компании «Евроген» (Москва, Россия). Анализ данных секвенирования методом главных компонент продемонстрировал преобладание влияния на экспрессию генов мутации и в меньшей степени влияние KNO_3 (рисунок 25)

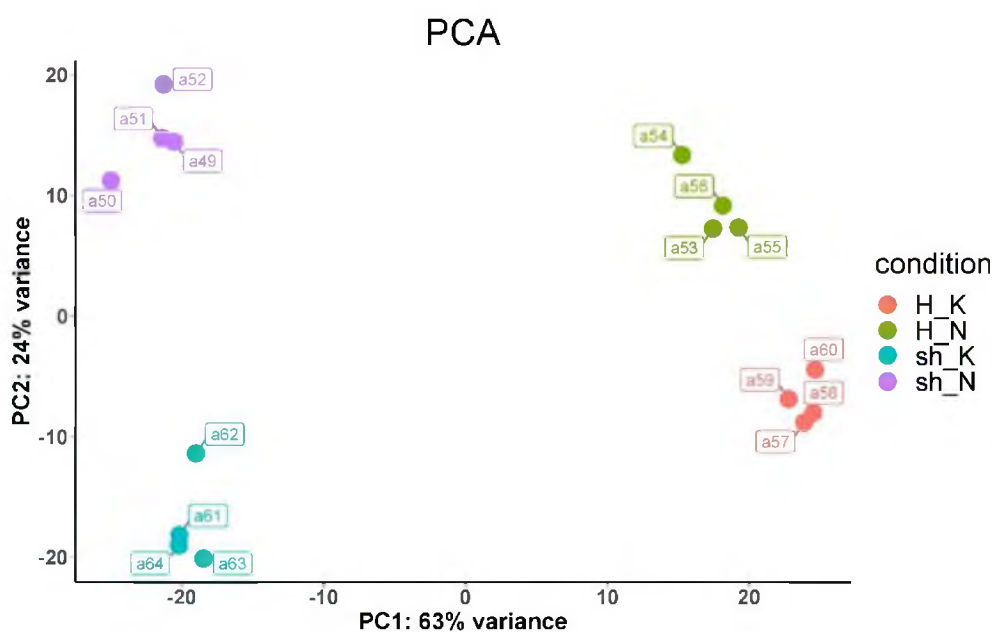


Рисунок 25 — Анализ распределения образцов методов главных компонент. H_K – Немчиновский, контроль; H_N – Немчиновский, KNO_3 ; sh_K – Штамбовый, контроль; sh_N – Штамбовый, KNO_3 .

При этом сорт Штамбовый демонстрирует более яркий ответ на воздействие KNO_3 : 979 генов с дифференциально повышенной экспрессией против 237 у Немчиновского (рисунок 26). Можно заключить, что генотипы являются контрастными по ответу на воздействие KNO_3 , особенно по генам с повышающейся экспрессией (рисунок 26).

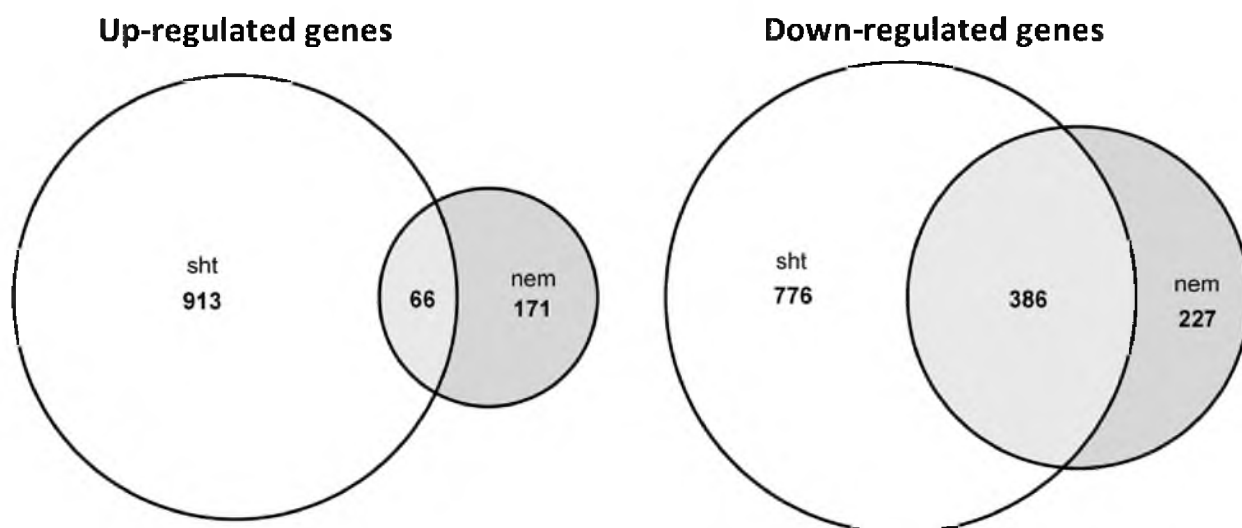


Рисунок 26 — Диаграммы Венна, демонстрирующие количество генов, дифференциально экспрессирующихся в ответ на воздействие KNO_3 у двух генотипов. Sht – Штамбовый. Nem – Немчиновский.

Анализ дифференциально экспрессирующихся генов методом теста на обогащение в терминах Gene ontology показал высокую степень различий в группах генов, отвечающих на воздействие KNO_3 у исследуемых генотипов. Наибольший контраст генотипы проявляют в различных биологических процессах, связанных с метаболизмом азота (эти процессы активируются только у Штамбового): «ответ на азотные соединения», «транспорт нитратов», «клеточный ответ на реактивные формы азота», «ответ на оксид азота», «транспорт нитрита». Кроме того, большие различия наблюдаются в процессах трансмембранного и ионного транспорта, клеточного ответа на гормональные стимулы, процессах гомеостаза, биогенеза клеточной стенки, ответом на осмотический стресс. Примечательно, что процессы, связанные с ответом на гиббереллин и биосинтез флавоноидов активируются только у Немчиновского (рисунок 27).

Более того, у Штамбового наблюдается понижение экспрессии генов, связанных с биосинтезом флавоноидов, ряда симбиотических генов и генов реакции сверхчувствительности (рисунок 28).

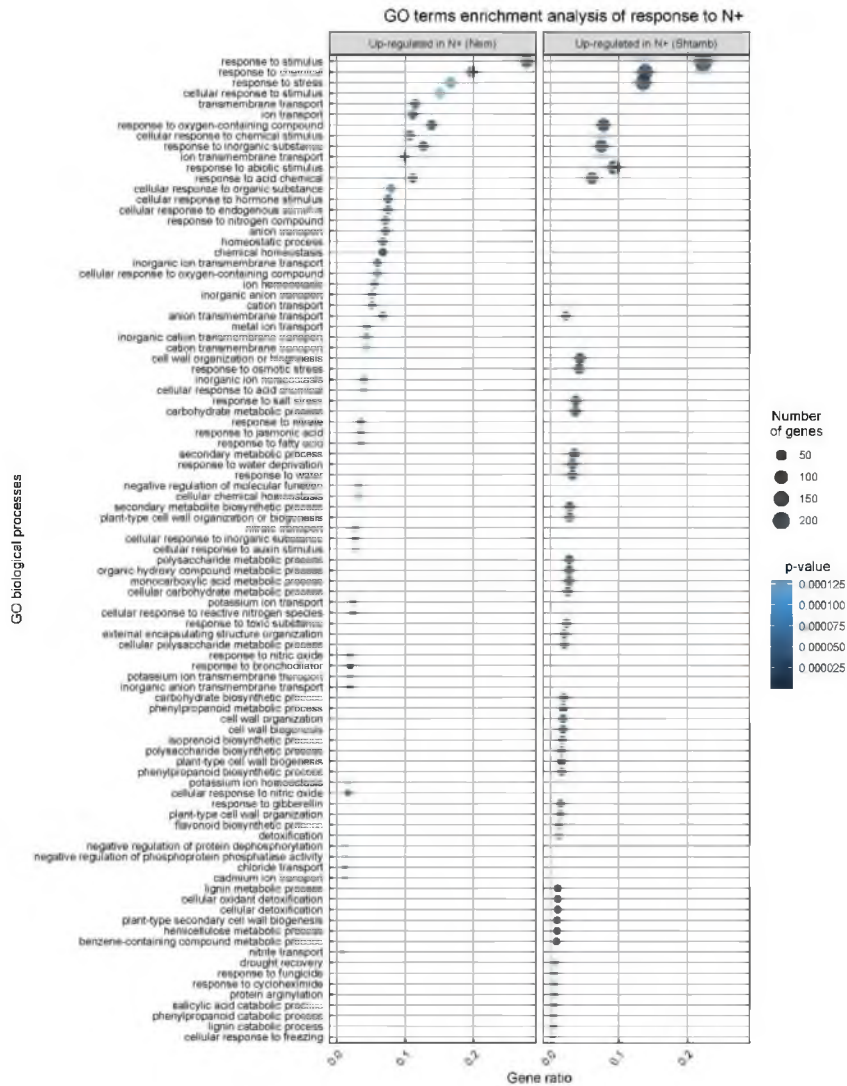


Рисунок 27 — Тест на обогащение в терминах Gene Ontology для генов с повышенной экспрессией в ответ на воздействие KNO₃.

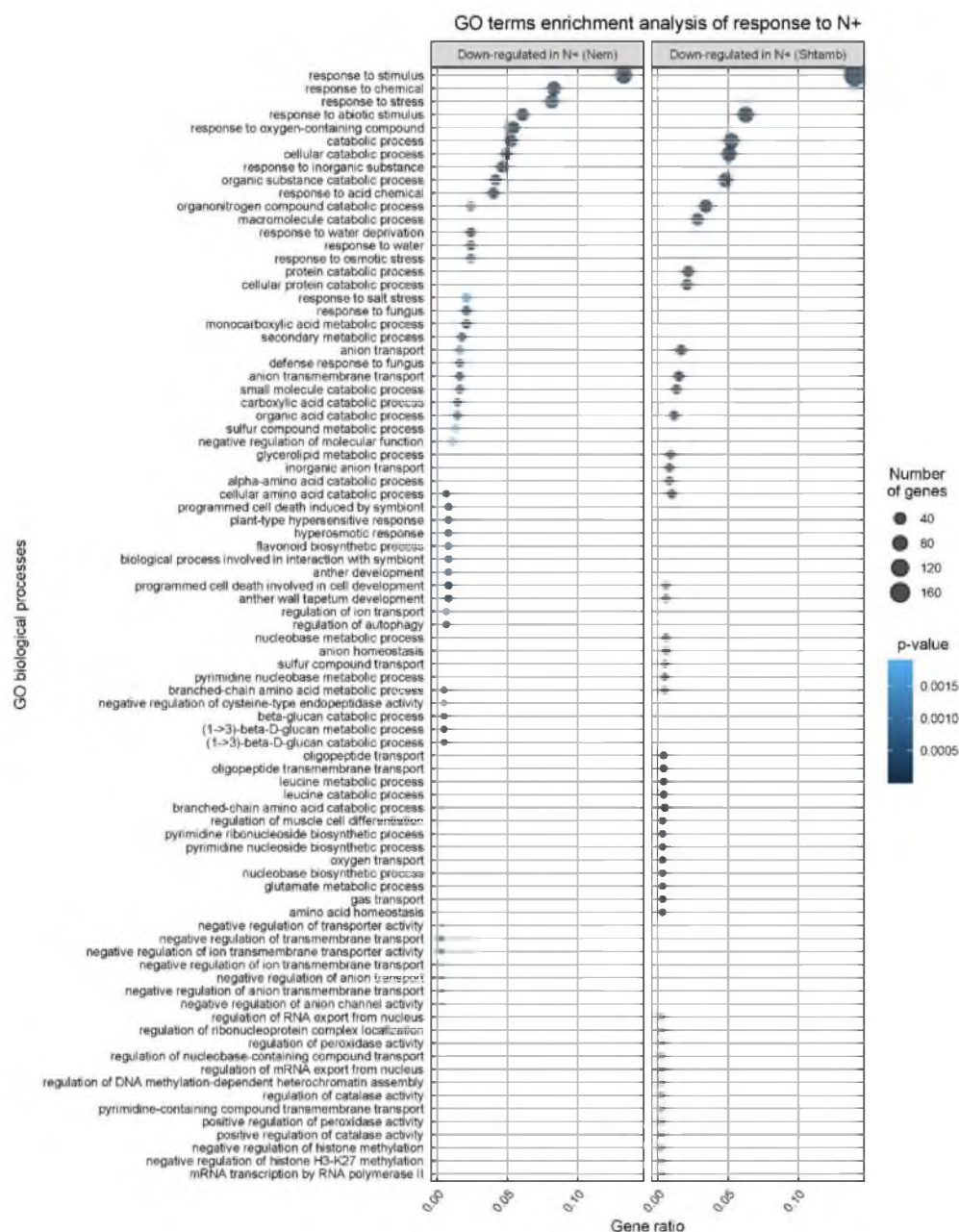


Рисунок 28 — Тест на обогащение в терминах Gene Ontology для генов с пониженной экспрессией в ответ на воздействие KNO₃.

Дополнительно мы изучили экспрессию известных *sym*-генов у исследуемых генотипов. В ходе анализа были выявлены гены, уровень экспрессии которых различается у двух генотипов. Это, в частности, гены, кодирующие уриказу, нодулин-25, флотиллин-подобный белок-4, лигаза LIN-1, нодулин-13, ранний нодулин-12А и ранний нодулин-75 (Рисунок 29).

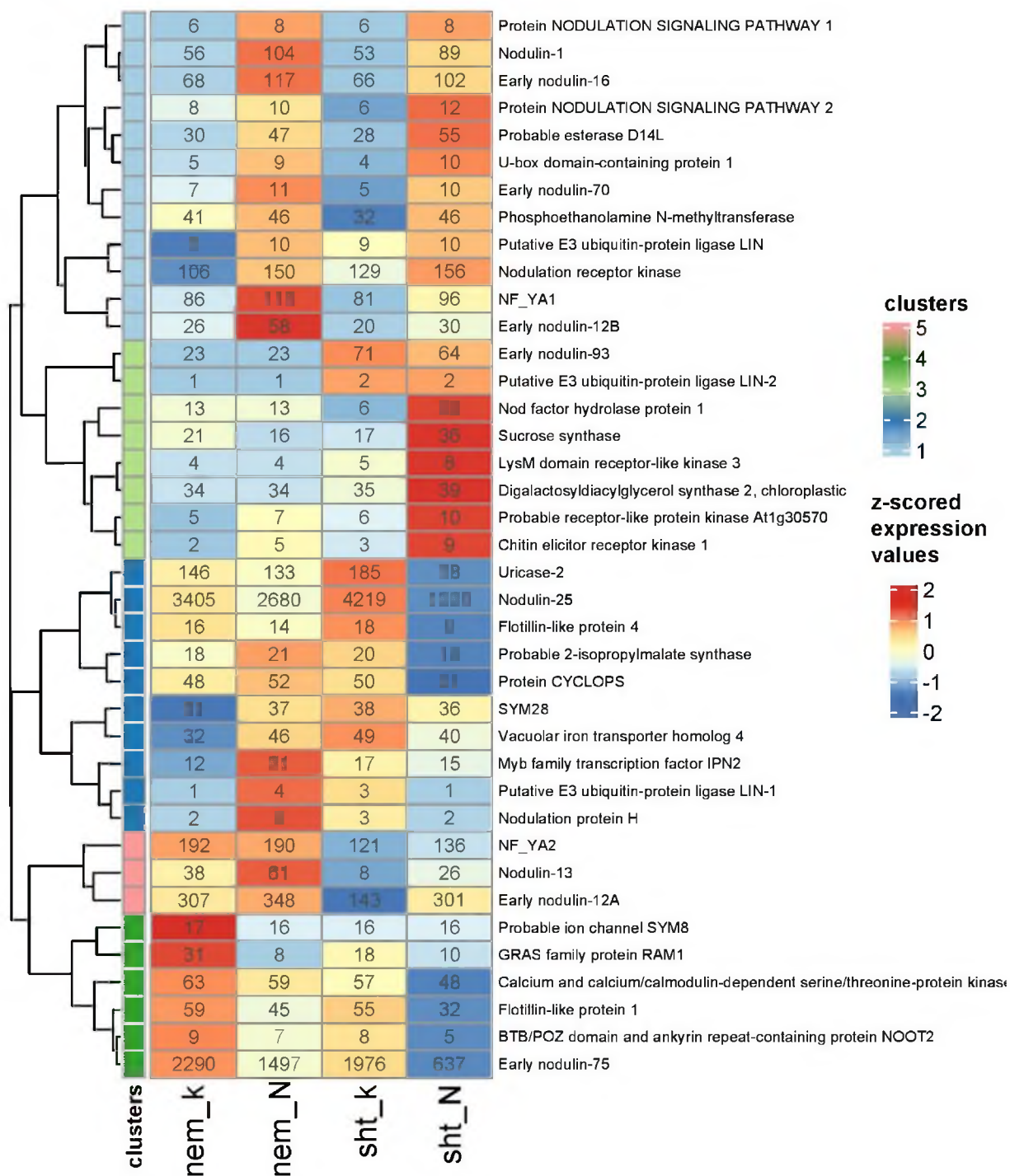


Рисунок 29 — Тепловая карта экспрессии известных sym-генов гороха посевного.

Различия между генотипами также выявлены в уровнях экспрессии генов, кодирующих NCR-пептиды. В частности, у Штамбового под воздействием нитрата наблюдается подавление экспрессии у 310 генов, кодирующих NCR-пептиды, в то время как у Немчиновского экспрессия снижена лишь у 117 NCR-генов.

Таким образом, полученные данные свидетельствуют о том, что мутация в гене *fas* негативно влияет на клубенькообразование, и в соответствии с этим экспрессия некоторых симбиотических генов у мутанта снижена. Примечательно, что экспрессия ключевого транскрипционного фактора NF-YA2 снижена у мутантной линии независимо от наличия или отсутствия нитрата в среде. Вероятно, ген *FAS* в норме является позитивным регулятором данного транскрипционного фактора. Тот факт, что Штамбовый более активно реагирует на внесение нитрата посредством изменения экспрессии генов может также свидетельствовать о том, что ген *FAS* в норме может быть негативным регулятором ответа растений на экзогенный нитрат.

3.3.4. Оценка уровня азотсодержащих соединений у растений со сверхэкспрессией гена *MtCLE35*

В ходе первого этапа реализации проекта нами было показано, что ген *MtCLE35* является нитрат-активируемым системным ингибитором развития симбиотических клубеньков у люцерны [45]. Экспрессия этого гена активируется как при обработке нитратом, так и при развитии симбиотических клубеньков, а усиление его работы в трансгенных корнях за счет сверхэкспрессии приводило к практически полному подавлению развития симбиотических клубеньков [45]. Мы предположили, что нитрат-активируемый ген *MtCLE35* может быть задействован не только в контроле развития симбиотических клубеньков, но и в других нитрат-регулируемых процессах, в частности, процессах накопления азота в растениях. Для проверки этого предположения мы провели метаболомный анализ корней и побегов растений люцерны дикого типа (линия R108) и трансгенных растений люцерны со сверхэкспрессией гена *MtCLE35* (содержат ген *MtCLE35* под контролем конститутивного промотора 35S (*p35S::MtCLE35*)). В работу были взяты трансгенные растения люцерны поколения T1 – потомки двух растений поколения T0 (семьи CLE35oe-2 и CLE35oe-3), по четыре растения из каждой семьи. Растения выращивали в течение двух недель в аэропонной установке, после чего корни и побеги растений замораживали по отдельности до последующего анализа. Из замороженного растительного материала часть (100 мг) использовали для выделения РНК для оценки уровня экспрессии гена *MtCLE35*, и часть материала (200 мг) брали для оценки содержания метаболитов.

Для подтверждения того, что у трансгенных растений из семей CLE35oe-2 и CLE35oe-3, взятых в анализ, наблюдается сверхэкспрессия гена *MtCLE35*, мы провели оценку уровня экспрессии этого гена в побеге с помощью ОТ-ПЦР в реальном времени. Мы показали, что у всех растений из семей CLE35oe-2 и CLE35oe-3 уровень экспрессии гена *MtCLE35* в 600-1400 раз превышает уровень его экспрессии у контрольных растений (исходная линия R108, дикий тип), что свидетельствует об эффективности работы конструкции, обуславливающей сверхэкспрессию (*p35S::MtCLE35*) (Рисунок 30).

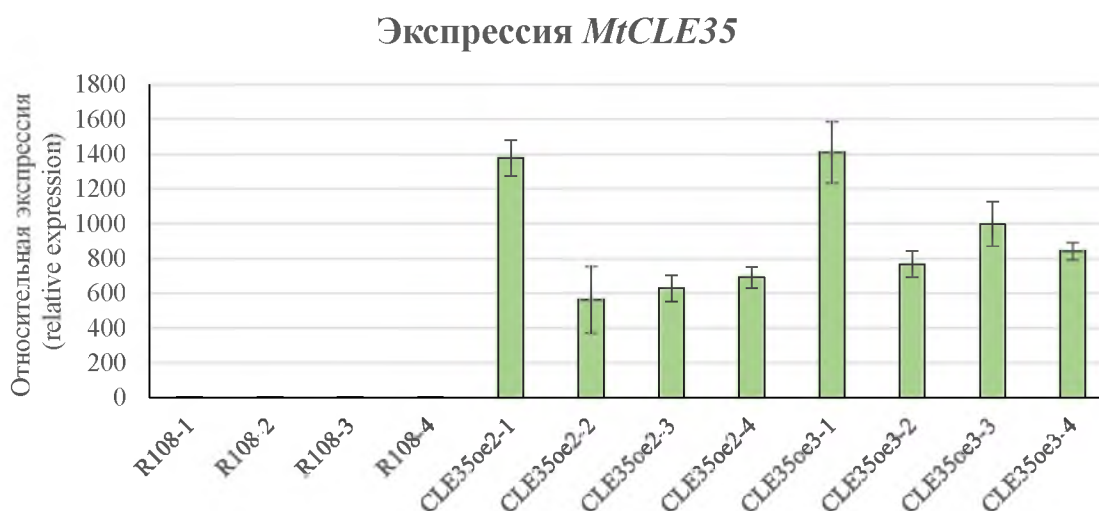


Рисунок 30 — Уровни экспрессии гена *MtCLE35* у растений дикого типа (R108) и у растений семей CLE35oe-2 и CLE35oe-3 со стабильной сверхэкспрессией гена *MtCLE35*.

На следующем этапе проводили анализ содержания метаболитов в образцах корня и побега у растений со сверхэкспрессией гена *MtCLE35* и у растений дикого типа. Были построены тепловые карты, иллюстрирующие содержание аминокислот и других азотсодержащих соединений в корнях и побегах растений дикого типа (R108) и растений двух семей со сверхэкспрессией *MtCLE35* (CLE35-OE-2 и CLE35-OE-3) (Рисунок 31). Можно отметить, что у растений обеих семей трансгенных растений наблюдается снижение уровня азотсодержащих соединений в побеге, по сравнению с линией R108. При этом в корнях, у растений со сверхэкспрессией *MtCLE35*, в особенности у растений семьи CLE35-OE-3, можно отметить повышенные уровни аминокислот и других азотсодержащих соединений (2-гидроксипиридина, первичных аминов, таких как этаноламин, фосфоэтаноламина), по сравнению с растениями дикого типа R108.

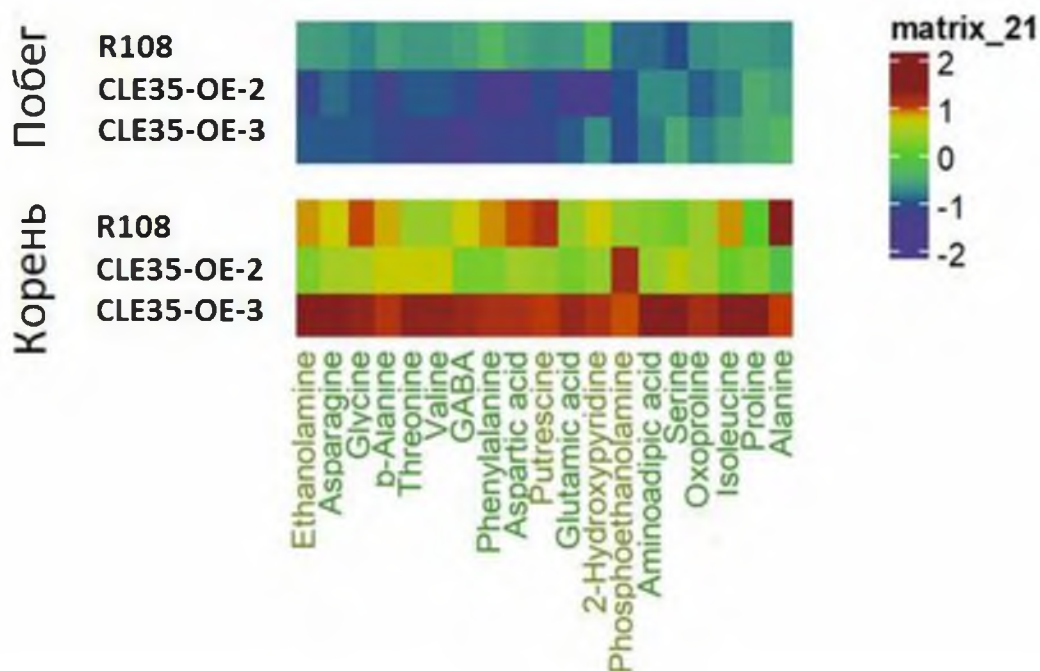


Рисунок 31 — Тепловая карта, иллюстрирующая содержание аминокислот и других-азотсодержащих соединений в корнях и побегах растений дикого типа (R108) и растений двух семей со сверхэкспрессией (CLE35-OE-2 и CLE35-OE-3).

Кроме того, данные по содержанию метаболитов в образцах были проанализированы с помощью дискриминантного анализа методом OPLS-DA (Orthogonal Partial Least Squares Discriminant Analysis, дискриминантный ортогональный анализ методом частных наименьших квадратов). Статистическая модель OPLS-DA выявила отклонения в распределении пула аминокислот и сахаров между корнем и побегом у растений со сверхэкспрессией гена *MtCLE35*, по сравнению с растениями дикого типа линии R108 (Рисунок 32, 33).

Анализ обогащения выявленных наборов метаболитов (MSEA, Metabolite Set Enrichment Analysis) показал, что у растений CLE35-OE содержание аминокислот увеличено в корнях, но снижено в побеге, по сравнению с линией дикого типа R108 (Рисунок 33). При этом распределение сложных сахаров, пентоз и гексоз демонстрирует обратную картину: их количество снижено в корнях у растений CLE35-OE и увеличено в побеге, по сравнению с линией R108.

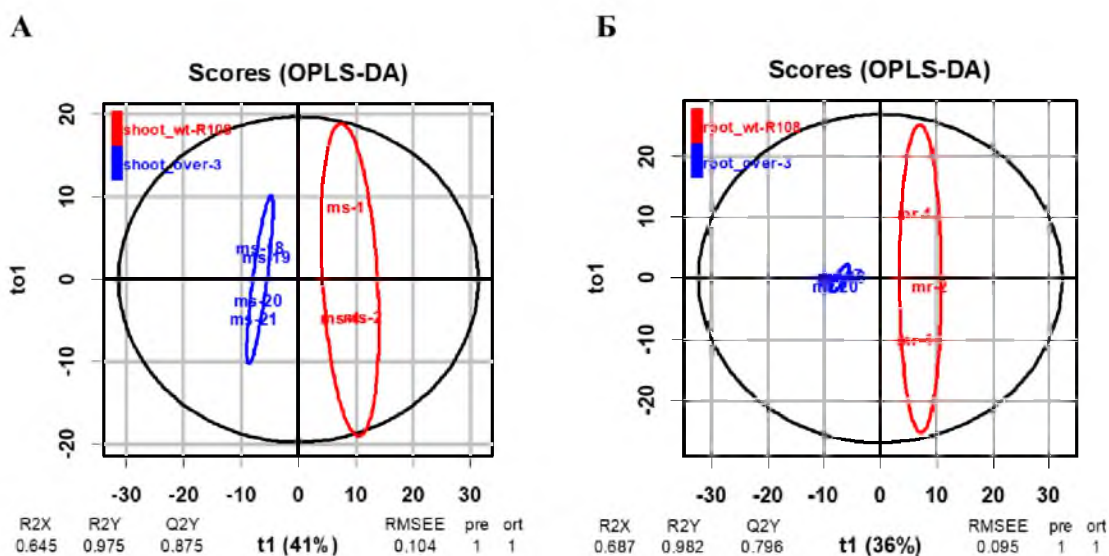


Рисунок 32 — График счетов, построенный по результатам анализа данных метаболомного анализа методом OPLS-DA для корней (А) и побегов (Б) растений дикого типа образцы (выделены красным) и растений CLE35oe-3.

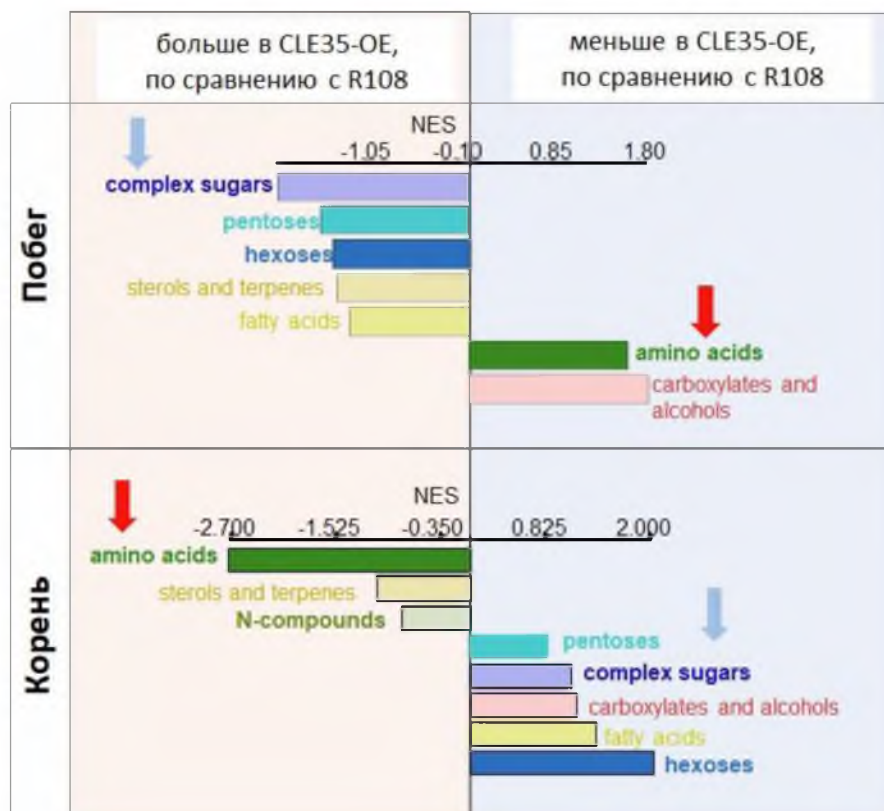


Рисунок 33 — Анализ обогащения наборов метаболитов (MSEA), различающихся в образцах побега и корней со сверхэкспрессией *MtCLE35* (CLE35-OE) и в образцах линии дикого типа R108.

Таким образом, результаты метаболомного анализа показали, что сверхэкспрессия гена *MtCLE35* приводит к общему увеличению содержания аминокислот и других азотсодержащих соединений в корнях, что может быть обусловлено усилением процессов поглощения азота и синтеза аминокислот. При этом наблюдаемое снижение уровня аминокислот в побеге у растений CLE35-OE, по сравнению с линией дикого типа R108, связано, вероятно, с нарушением транспорта аминокислот из корней в побег, который, как можно предположить на основании полученных нами данных, контролируется регуляторными путями, активируемыми MtCLE35. Результаты нашей работы расширяют представления о роли гена *MtCLE35* у люцерны и указывают на его возможное участие в контроле процессов поглощения азота и синтеза аминокислот в корнях, в также в регуляции транспорта аминокислот из корней в побег.

ЗАКЛЮЧЕНИЕ

В ходе выполнения работ нами были получены следующие основные результаты:

В геноме гороха идентифицировано 15 генов, кодирующих пептиды CEP, и показано, что гены *PsCEP10*, *PsCEP14* и *PsCEP15* экспрессируются на относительно высоких уровнях в симбиотических клубеньках на 10 день после инокуляции.

В геноме гороха идентифицировано 47 генов, кодирующих пептиды CLE, из которых четыре гена (*PsCam040702 (PsCLE13)*, *PsCam041632 (PsCLE12)*, *PsCam040984 (PsCLE12-like)* и *PsCam041632 (PsNIC-like)*) являются ближайшими гомологами генов CLE, вовлеченных в регуляцию развития клубеньков у других бобовых растений. Показано, что экспрессия генов *PsCam040702 (PsCLE13)*, *PsCam041632 (PsCLE12)*, *PsCam040984 (PsCLE12-like)* и *PsCam041632 (PsNIC-like)* усиливается в развивающихся примордиях симбиотических клубеньков у гороха. Экспрессия генов *PsCam040702 (PsCLE13)*, *PsCam041632 (PsCLE12)* и *PsCam041632 (PsNIC-like)* усиливается при обработке нитратом (10 mM KNO₃)

Показано, что экспрессия генов *MtCLE34* и *MtCLE35* усиливается в развивающихся примордиях симбиотических клубеньков и при обработке нитратом у люцерны (10 mM KNO₃). Показано, что уровни экспрессии генов *MtCEP (MtCEP1, MtCEP2, MtCEP5, MtCEP11)* снижаются при обработке нитратом (10 mM KNO₃). Обнаружено, что уровни экспрессии генов *MtCEP16* и *MtCEP17* значительно увеличены в клубеньках люцерны, по сравнению с корнем.

Созданы конструкции для сверхэкспрессии генов люцерны *MtCLE34*, *MtCLE35*, *MtCEP17* *MtCEP16*, а также генов гороха *PsCam040702 (PsCLE13)*, *PsCam041632 (PsCLE12)*, *PsCam040984 (PsCLE12-like)* и *PsCam041632 (PsNIC-like)*.

В ходе анализа разнообразия последовательностей четырех генов *MtCLE12*, *MtCLE13*, *MtCLE34*, *MtCLE35* у 262 линий *M truncatula* показано, что в гене *MtCLE13* присутствуют синонимичные замены, тогда как в остальных генах выявлены как синонимичные, так и несинонимичные замены. У четырех линий (линии HM101 (линия A17), HM256, HM038 и HM058 из базы NapMap) в последовательности гена *MtCLE34* обнаружен стоп-кодон, наличие которого должно приводить к синтезу укороченного нефункционального продукта без функционального домена CLE (т.е. эти линии являются природными мутантами по гену *MtCLE34*). В отличие от люцерны, гомологичные гены гороха посевного *PsCLE12*, *PsCLE13* и *PsCLE34* у 99 генотипов гороха не являются полиморфными (в пределах кодирующих частей генов).

При помощи подхода GWAS (Genome-wide association study) выявлена область генома (на хромосоме 2), расположенная перед кодирующей последовательностью гена *MtCLE34*, в которой SNP (однонуклеотидные полиморфизмы) демонстрируют ассоциацию с количеством клубеньков, образующихся на растениях.

При помощи транскриптомного анализа корней гороха после обработки нитратом выявлены более 100 дифференциально экспрессирующихся генов, связанных с ответом растения на нитрат. При помощи транскриптомного анализа трансгенных корней люцерны, сверхэкспрессирующих ген *MtCLE35*, показано, что искусственное повышение экспрессии данного гена приводит к повышению в корнях растений экспрессии генов, ответственных за метаболические процессы и белковый метаболизм, а также кодирующих компоненты клеточной стенки. Также повышение экспрессии *MtCLE35* приводит к практически полному подавлению клубенькообразования и снижению экспрессии ключевых генов-регуляторов формирования клубеньков (*NIN*, *NSP1*, *NF-YA1/HAP2* и *ERN1*), однако не затрагивает гены, ответственные за самые ранние этапы сигнальных взаимодействий с клубеньковыми бактериями.

Оптимизированы методы трансформации гороха с помощью *Agrobacterium rhizogenes* и получен материал для транскриптомного анализа корней со сверхэкспрессией генов *PsCLE12*, *PsCLE13* и *PsCEP10*.

Проанализирован полиморфизм генов *PsSym29* и *PsNRLK1*, кодирующих рецепторы CLE-пептидов, у гороха посевного. Не выявлено ассоциаций аллельного состояния данных генов и значений хозяйственно-значимых признаков у растений, выращенных в условиях инокуляции клубеньковыми бактериями.

Секвенирован геном мутанта гороха по гену *fas*, предположительно кодирующему компонент системы CLAVATA. Секвенирован транскриптом корневой системы мутанта гороха по гену *fas* и исходного генотипа при инокуляции клубеньковыми бактериями и при воздействии нитрата, проанализирована роль гена *FAS* в клубеньковом симбиозе и в ответе растений гороха на нитрат в субстрате.

С помощью метаболомного анализа проведена оценка уровня азотсодержащих соединений у растений люцерны со сверхэкспрессией гена *MtCLE35*. Показано, что ген *MtCLE35* задействован не только в контроле развития симбиотических клубеньков, но и в других нитрат-регулируемых процессах, в частности, процессах накопления азота в растениях. Результаты работы расширяют представления о роли гена *MtCLE35* у люцерны и указывают на его возможное участие в контроле процессов поглощения азота и синтеза аминокислот в корнях, в также в регуляции транспорта аминокислот из корней в побег.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Okamoto S., Ohnishi E., Sato S., Takahashi H., Nakazono M., Tabata S., et al. Nod factor/nitrate-induced CLE genes that drive HAR1-mediated systemic regulation of nodulation // *Plant Cell Physiol.* — 2009. — № 50(1). P. 67–77.
2. Schnabel E., Journet E.-P., de Carvalho-Niebel F., Duc G., Frugoli J. The *Medicago truncatula* SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length // *Plant Mol Biol.* — 2005. — № 58(6). P. 809–822.
3. Searle I.R., Men A.E., Laniya T.S., Buzas D.M., Iturbe-Ormaetxe I., Carroll B.J., et al. Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase // *Science* (80-). — 2003. — № 299(5603). P. 109–112.
4. Reid D.E., Ferguson B.J., Hayashi S., Lin Y.-H., Gresshoff P.M. Molecular mechanisms controlling legume autoregulation of nodulation // *Ann Bot.* — 2011. — № 108(5). P. 789–795.
5. Nishida H., Suzuki T. Nitrate-mediated control of root nodule symbiosis // *Curr Opin Plant Biol.* — 2018. — № 44. P. 129–136.
6. Reid D.E., Ferguson B.J., Gresshoff P.M. Inoculation- and Nitrate-Induced CLE Peptides of Soybean Control NARK-Dependent Nodule Formation // *Mol Plant-Microbe Interact.* — 2011. — № 24(5). P. 606–618.
7. Lim C.W., Lee Y.W., Hwang C.H. Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression // *Plant Cell Physiol.* — 2011. — № 52(9). P. 1613–1627.
8. Imin N., Mohd-Radzman N.A., Ogilvie H.A., Djordjevic M.A. The peptide-encoding CEP1 gene modulates lateral root and nodule numbers in *Medicago truncatula* // *J Exp Bot.* — 2013. — № 64(17). P. 5395–5409.
9. Laffont C., Ivanovici A., Gautrat P., Brault M., Djordjevic M.A., Frugier F. The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically // *Nat Commun.* — 2020. — № 11(1). P. 1–13.
10. Jing R., Vershinin A., Grzebyta J., Shaw P., Smýkal P., Marshall D., et al. The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis // *BMC Evol Biol.* — 2010. — № 10(1). P. 44.
11. Macas J., Neumann P., Navrátilová A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. // *BMC Genomics* [Internet]. — 2007. — № 8. P. 427. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18031571>
12. Kreplak J., Madoui M.-A., Cápál P., Novák P., Labadie K., Aubert G., et al. A reference genome for pea provides insight into legume genome evolution // *Nat Genet.* — 2019. — № 51(9). P. 1411–1422.
13. Young N.D., Debelle F., Oldroyd G.E.D., Geurts R., Cannon S.B., Udvardi M.K., et al. The *Medicago* genome provides insight into the evolution of rhizobial symbioses // *Nature.* — 2011. — № 480(7378). P. 520–524.

14. Oh E., Seo P.J., Kim J. Signaling peptides and receptors coordinating plant root development // *Trends Plant Sci.* — 2018. — № 23(4). P. 337–351.
15. Notaguchi M., Okamoto S. Dynamics of long-distance signaling via plant vascular tissues // *Front Plant Sci.* — 2015. — № 6. P. 161.
16. Ohyama K., Ogawa M., Matsubayashi Y. Identification of a biologically active, small, secreted peptide in Arabidopsis by in silico gene screening, followed by LC- MS- based structure analysis // *Plant J.* — 2008. — № 55(1). P. 152–160.
17. Limpens E., Ramos J., Franken C., Raz V., Compaan B., Franssen H., et al. RNA interference in *Agrobacterium rhizogenes*- transformed roots of Arabidopsis and *Medicago truncatula* // *J Exp Bot.* — 2004. — № 55(399). P. 983–992.
18. FÄHRAEUS G. The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique // *Microbiology.* — 1957. — № 16(2). P. 374–381.
19. Livak K.J., Schmittgen T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method // *methods.* — 2001. — № 25(4). P. 402–408.
20. Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data // *Bioinformatics [Internet].* — 2014. — № 30(15). P. 2114–2120. Available from: <https://doi.org/10.1093/bioinformatics/btu170>
21. Tang H., Krishnakumar V., Bidwell S., Rosen B., Chan A., Zhou S., et al. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula* // *BMC Genomics [Internet].* — 2014. — № 15(1). P. 312. Available from: <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-312>
22. Kim D., Langmead B., Salzberg S.L. HISAT: a fast spliced aligner with low memory requirements // *Nat Methods.* — 2015. — № 12(4). P. 357–360.
23. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., et al. The Sequence Alignment/Map format and SAMtools // *Bioinformatics [Internet].* — 2009. — № 25(16). P. 2078–2079. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp352>
24. Perteua M., Perteua G.M., Antonescu C.M., Chang T.-C., Mendell J.T., Salzberg S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads // *Nat Biotechnol.* — 2015. — № 33(3). P. 290–295.
25. Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 // *Genome Biol.* — 2014. — № 15(12). P. 550.
26. Kang Y., Sakiroglu M., Krom N., Stanton- Geddes J., Wang M., Lee Y., et al. Genome-wide association of drought- related and biomass traits with HapMap SNPs in *Medicago truncatula* // *Plant Cell Environ.* — 2015. — № 38(10). P. 1997–2011.
27. Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. TASSEL: software for association mapping of complex traits in diverse samples // *Bioinformatics.* — 2007. — № 23(19). P. 2633–2635.
28. Zhukov V.A., Zhernakov A.I., Sulima A.S., Kulaeva O.A., Kliukova M.S., Afonin A.M., et al. Association study of symbiotic genes in pea (*Pisum sativum* L.) cultivars grown in symbiotic conditions // *Agronomy.* — 2021. — № 11(11). P. 2368.
29. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and

- maximum parsimony methods // *Mol Biol Evol.* — 2011. — № 28(10). P. 2731–2739.
30. Librado P., Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data // *Bioinformatics.* — 2009. — № 25(11). P. 1451–1452.
 31. Smith T., Heger A., Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy // *Genome Res.* — 2017. — № 27(3). P. 491–499.
 32. Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., et al. Twelve years of SAMtools and BCFtools // *Gigascience.* — 2021. — № 10(2). P. giab008.
 33. Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features // *Bioinformatics.* — 2014. — № 30(7). P. 923–930.
 34. Alexa A., Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology (2.46.0)[Computer software]. Bioconductor version: Release (3.14) // 2022.
 35. Wickham H. ggplot2: elegant graphics for data analysis // springer; 2016.
 36. Slowikowski K., Schep A., Hughes S., Lukauskas S., Irisson J.-O., Kamvar Z.N., et al. Package ggrepel // Autom position non-overlapping text labels with 'ggplot2'. — 2018. — .
 37. Hastie T., Tibshirani R., Sherlock G., Eisen M., Brown P., Botstein D. Imputing missing data for gene expression arrays // 1999. — .
 38. Gu Z., Eils R., Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data // *Bioinformatics.* — 2016. — № 32(18). P. 2847–2849.
 39. Thevenot E.A. ropls: PCA, PLS (-DA) and OPLS (-DA) for multivariate analysis and feature selection of omics data // R Packag version. — 2016. — № 1(0).
 40. Korotkevich G., Sukhov V., Budin N., Shpak B., Artyomov M.N., Sergushichev A. Fast gene set enrichment analysis // *BioRxiv.* — 2021. — P. 60012.
 41. Zhou P., Silverstein K.A.T., Gao L., Walton J.D., Nallu S., Guhlin J., et al. Detecting small plant peptides using SPADA (small peptide alignment discovery application) // *BMC Bioinformatics.* — 2013. — № 14(1). P. 335.
 42. Boschiero C., Dai X., Lundquist P.K., Roy S., de Bang T.C., Zhang S., et al. MtSSPdb: The Medicago truncatula Small Secreted Peptide Database // *Plant Physiol.* — 2020. — № 183(1). P. 399–413.
 43. Alves-Carvalho S., Aubert G., Carrère S., Cruaud C., Brochot A.L., Jacquin F., 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 [Internet].* — 2015. — № 84(1). P. 1–19. Available from: <http://doi.wiley.com/10.1111/tpj.12967>
 44. Mortier V., Den Herder G., Whitford R., Van de Velde W., Rombauts S., D'haeseleer K., et al. CLE peptides control Medicago truncatula nodulation locally and systemically // *Plant Physiol.* — 2010. — № 153(1). P. 222–237.
 45. Lebedeva M., Azarakhsh M., Yashenkova Y., Lutova L. Nitrate-induced CLE peptide systemically inhibits nodulation in Medicago truncatula // *Plants.* — 2020. — № 9(11). P. 1456.
 46. de Bang T.C., Lundquist P.K., Dai X., Boschiero C., Zhuang Z., Pant P., et al. Genome-

- wide identification of Medicago peptides involved in macronutrient responses and nodulation // *Plant Physiol.* — 2017. — № 175(4). P. 1669–1689.
47. Imin N., Patel N., Corcilius L., Payne R.J., Djordjevic M.A. CLE peptide tri-arabinylation and peptide domain sequence composition are essential for SUNN-dependent autoregulation of nodulation in *Medicago truncatula* // *New Phytol.* — 2018. — № 218(1). P. 73–80.
 48. Roux B., Rodde N., Jardinaud M., Timmers T., Sauviac L., Cottret L., et al. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing // *Plant J.* — 2014. — № 77(6). P. 817–837.
 49. Le Marquer M., Bécard G., Frei dit Frey N. Arbuscular mycorrhizal fungi possess a CLAVATA3/embryo surrounding region-related gene that positively regulates symbiosis // *New Phytol.* — 2019. — № 222(2). P. 1030–1042.
 50. Mens C., Hastwell A.H., Su H., Gresshoff P.M., Mathesius U., Ferguson B.J. Characterisation of *Medicago truncatula* CLE34 and CLE35 in nitrate and rhizobia regulation of nodulation // *New Phytol.* — 2021. — № 229(5). P. 2525–2534.
 51. Stanton-Geddes J., Paape T., Epstein B., Briskine R., Yoder J., Mudge J., et al. Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula* // *PLoS One.* — 2013. — № 8(5). P. e65688.
 52. Якоби Л.М., Кукалев А.С., Ушаков К.В., Цыганов В.Е., Проворов Н.А., Борисов А.Ю., et al. Полиморфизм форм гороха посевного по эффективности симбиоза с эндомикоризным грибом *Glomus* sp. в условиях инокуляции ризобиями // СЕЛЬСКОХОЗЯЙСТВЕННАЯ БИОЛОГИЯ СЕРИЯ: БИОЛОГИЯ РАСТЕНИЙ СЕРИЯ: БИОЛОГИЯ ЖИВОТНЫХ. — 2000. — (3). P. 94–102.
 53. Gautrat P., Laffont C., Frugier F. Compact root architecture 2 promotes root competence for nodulation through the miR2111 systemic effector // *Curr Biol.* — 2020. — № 30(7). P. 1339–1345.
 54. Roy S., Liu W., Nandety R.S., Crook A., Mysore K.S., Pislariu C.I., et al. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation // *Plant Cell.* — 2020. — № 32(1). P. 15–41.
 55. Laffont C., De Cuyper C., Fromentin J., Mortier V., De Keyser A., Verplancke C., et al. MtNRLK1, a CLAVATA1-like leucine-rich repeat receptor-like kinase upregulated during nodulation in *Medicago truncatula* // *Sci Rep.* — 2018. — № 8(1). P. 1–9.
 56. Ng P.C., Henikoff S. SIFT: Predicting amino acid changes that affect protein function // *Nucleic Acids Res.* — 2003. — № 31(13). P. 3812–3814.
 57. Zerpa-Catanho D., Zhang X., Song J., Hernandez A.G., Ming R. Ultra-long DNA molecule isolation from plant nuclei for ultra-long read genome sequencing // *STAR Protoc.* — 2021. — № 2(1). P. 100343.
 58. Allen G.C., Flores-Vergara M.A., Krasynanski S., Kumar S., Thompson W.F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide // *Nat Protoc [Internet]*. — 2006. — № 1(5). P. 2320–2325. Available from: <https://doi.org/10.1038/nprot.2006.384>
 59. Tayeh N., Aluome C., Falque M., Jacquin F., Klein A., Chauveau A., et al. Development

of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high-density, high-resolution consensus genetic map // *Plant J* [Internet]. — 2015. — № 84(6). P. 1257–1273. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26590015>

60. Синюшин А.А., Гостимский С.А. Фасциация у гороха посевного: основные закономерности морфогенеза // *Онтогенез*. — 2006. — № 37(6). P. 449–457.

ПРИЛОЖЕНИЕ А

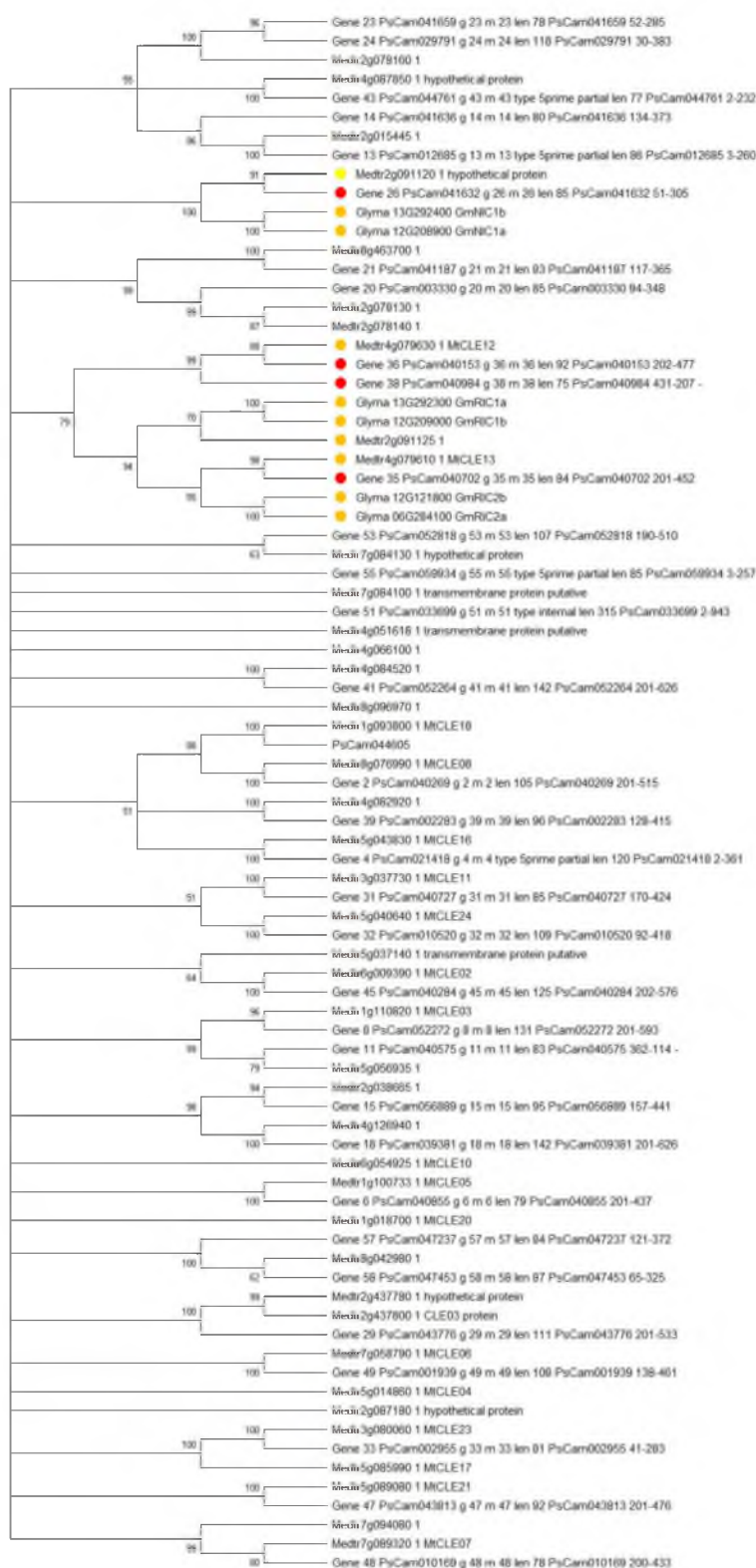


Рисунок А.1 — Филогенетическое дерево, построенное на основании аминокислотных последовательностей MtCLE и их ближайших гомологов у гороха.

Красными кружками выделены ближайшие гомологи CLE, задействованных в регуляции клубенькообразования у других бобовых растений (отмечены желтыми кружками).

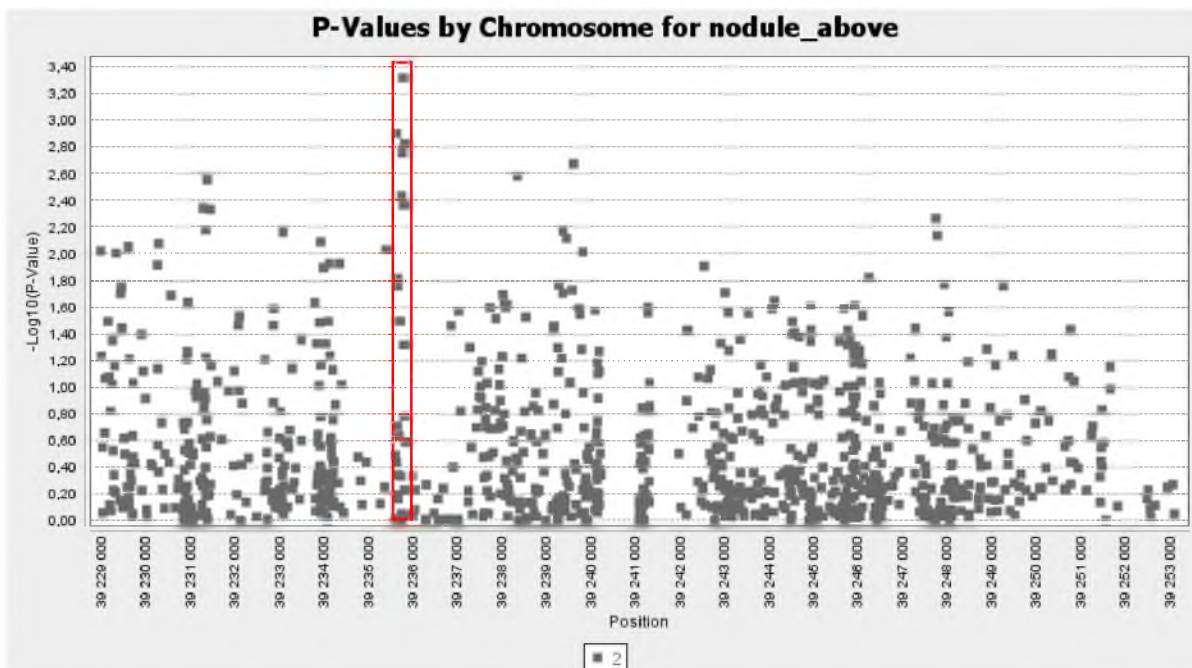


Рисунок А.3 — График Manhattan для *MtCLE34* и *MtCLE35* в хромосоме 2 по nodule_above.

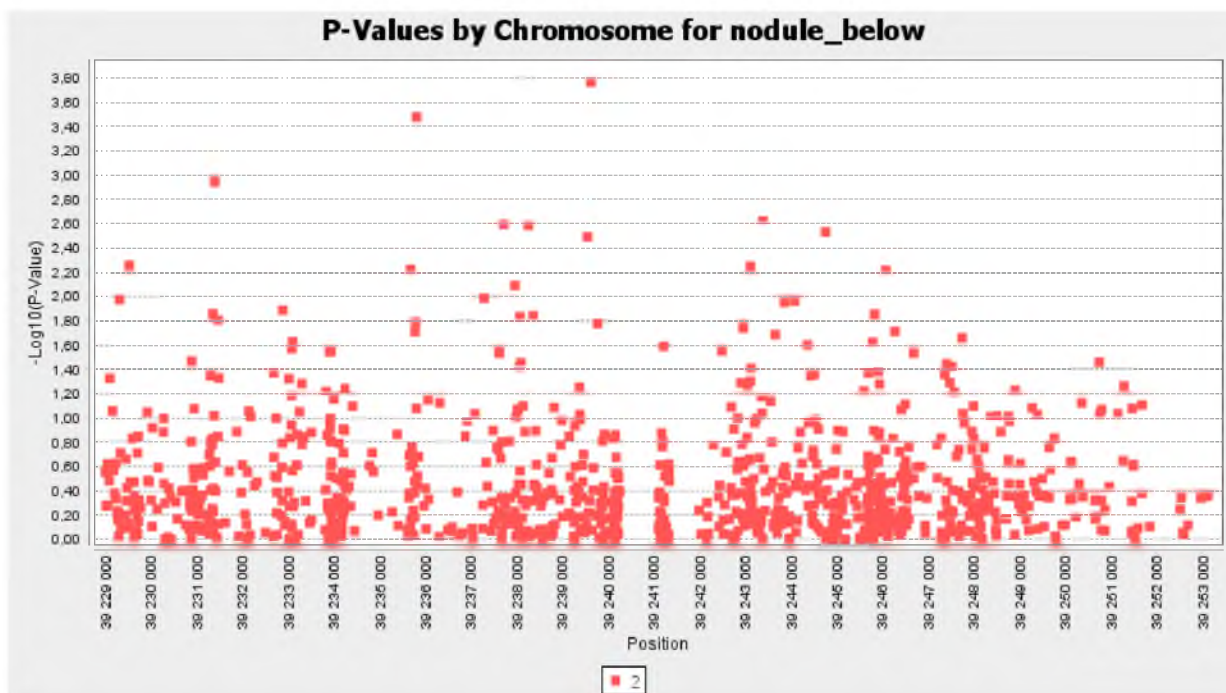


Рисунок А.4 — График Manhattan для участка хромосомы 2, включающего гены *MtCLE34* и *MtCLE35*, в хромосоме 2 по nodule_below.

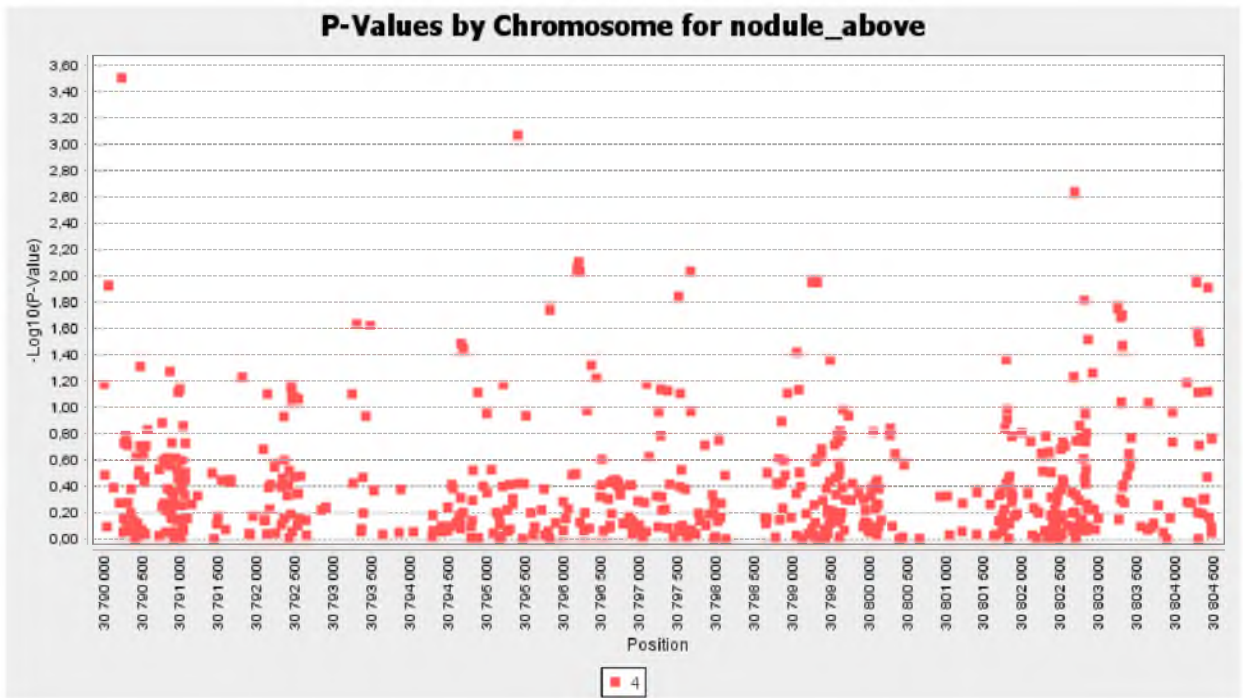


Рисунок А.5 — График Manhattan для участка хромосомы 4, включающего гены *MiCLE12* и *MiCLE13* по признаку «количество клубеньков в верхней части корневой системы».

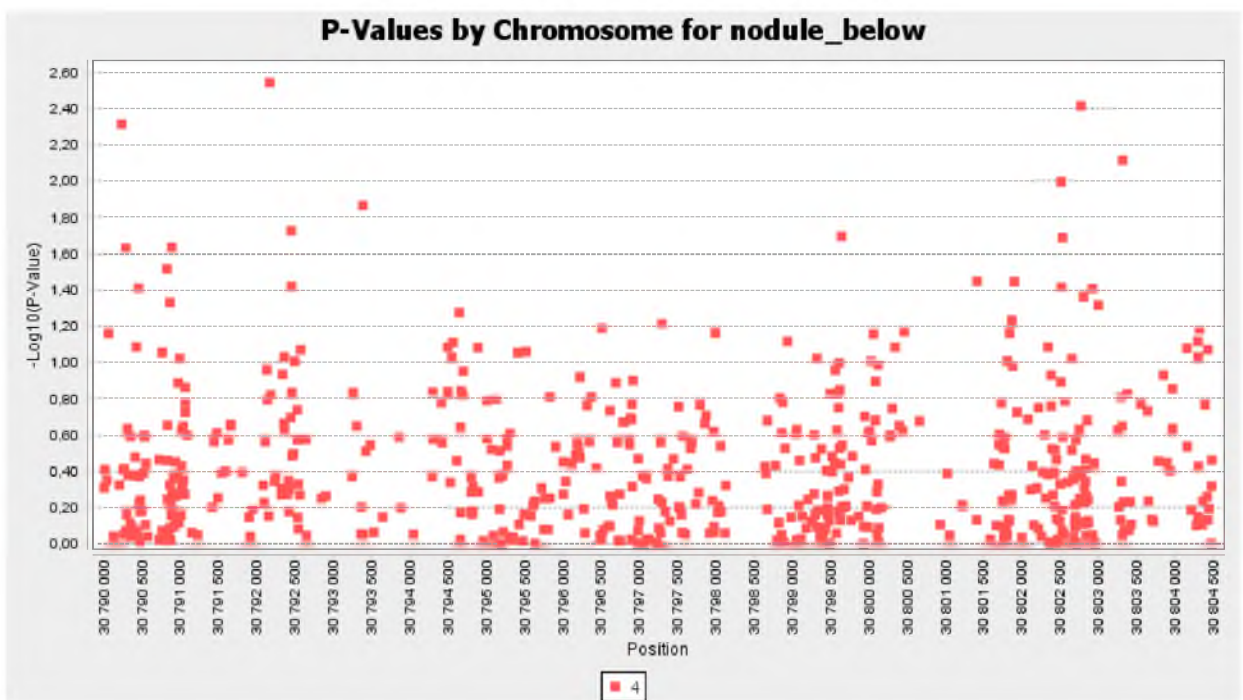


Рисунок А.6 — График Manhattan для участка хромосомы 4, включающего гены *MiCLE12* и *MiCLE13* по признаку «количество клубеньков в нижней части корневой системы».

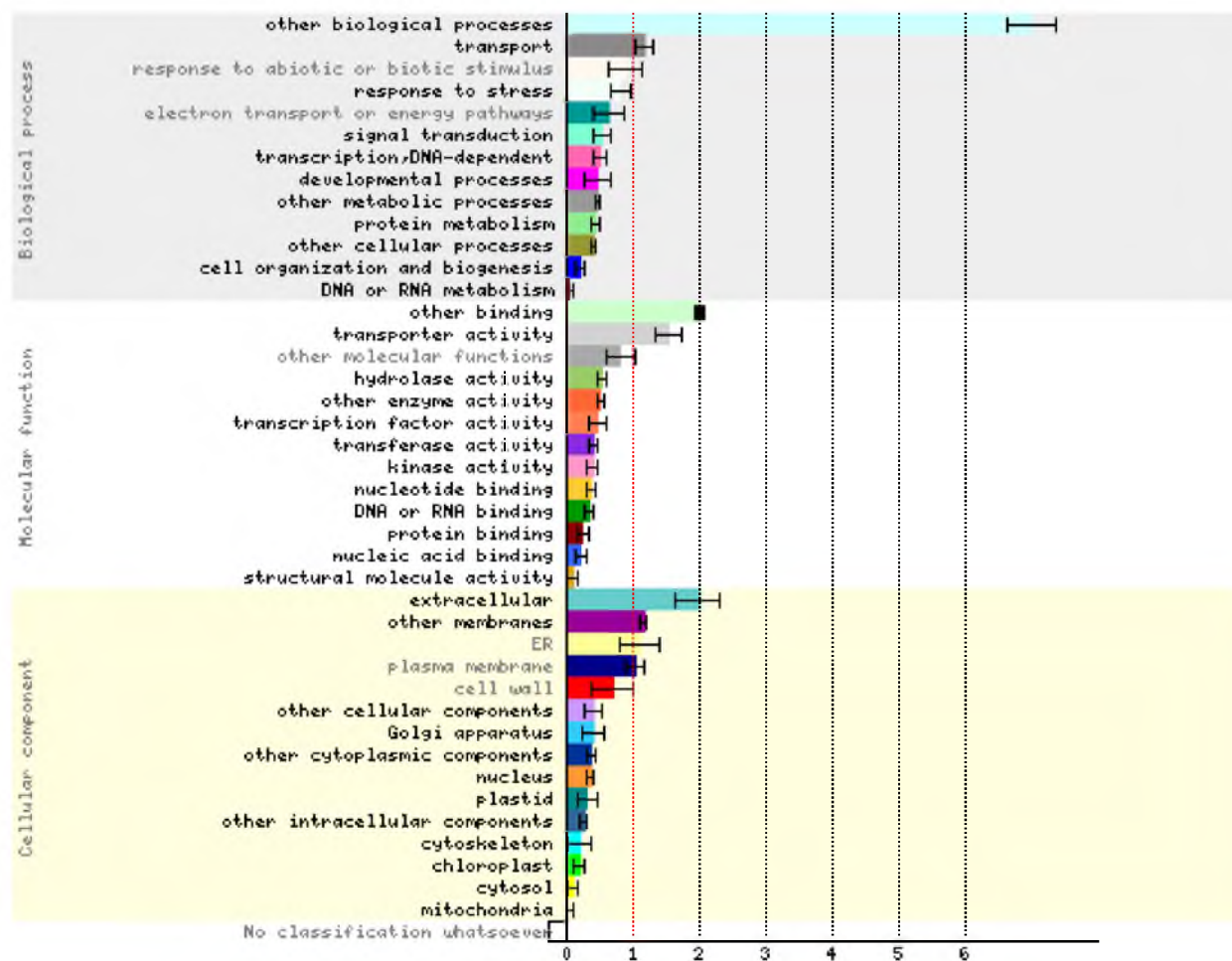


Рисунок А.7 — Гистограмма, иллюстрирующая функциональную аннотацию (Gene Ontology, UniProt) продуктов генов, экспрессия которых подавляется в корнях со сверхэкспрессией *MtCLE35*. Жирным шрифтом выделены процессы, для которых отмечено статистически значимое обогащение продуктами генов с пониженной экспрессией в образцах *MtCLE35_oe* ($p < 0.05$)

Таблица А.1 — ID транскриптов генов *CLE*, выявленные в транскриптоме гороха (Alves-Carvalho, 2015) с указанием их ближайших гомологов у люцерны *M. truncatula* (номенклатура по данным Hastwell et al., 2015; de Bang et al. 2017). Жирным шрифтом выделены ближайшие гомологи, задействованных в регуляции клубенькообразования у люцерны и у других бобовых растений.

N	ID (по Alves-Carvalho, 2015) - ближайший гомолог <i>Medicago truncatula</i>
1	PsCam002283 - MtCLE40
2	PsCam002955 - MtCLE23
3	PsCam003330 - MtCLE31
4	PsCam005544 - MtCLE49
5	PsCam006947 - MtCLE49
6	PsCam008069 - MtCLE04
7	PsCam010169 - MtCLE07
8	PsCam010324 - Medtr3g087870
9	PsCam010520 - MtCLE24
10	PsCam012685 - MtCLE29
11	PsCam017409 - MtCLE43
12	PsCam021418 - MtCLE16
13	PsCam029791 - MtCLE33
14	PsCam030931 - MtCLE27
15	PsCam033699 - MtCLE09
16	PsCam039381 - MtCLE44
17	PsCam040153 - MtCLE12
18	PsCam040269 - MtCLE08
19	PsCam040284 - MtCLE02
20	PsCam040575 - MtCLE45
21	PsCam040702 - MtCLE13
22	PsCam040727 - MtCLE11
23	PsCam040855 - MtCLE05
24	PsCam040884 - MtCLE28
25	PsCam040984 - MtCLE12

26	PsCam041187 - MtCLE53
27	PsCam041559 - MtCLE46
28	PsCam041591 - Medtr7g012530
29	PsCam041632 - MtCLE34
30	PsCam041636 - MtCLE29
31	PsCam041659 - MtCLE33
32	PsCam043776 - MtCLE37
33	PsCam043813 - MtCLE21
34	PsCam044605 - MtCLE18
35	PsCam044761 - MtCLE42
36	PsCam046649 - MtCLE52
37	PsCam047237 - MtCLE51
38	PsCam047453 - MtCLE51
39	PsCam047545 - MtCLE27
40	PsCam051179 - MtCLE27 (= PsCam047545)
41	PsCam052264 - MtCLE41
42	PsCam052272 - MtCLE03
43	PsCam052818 - MtCLE09
44	PsCam052913 - Medtr2g013960 (интрон)
45	PsCam052951 - CLE-пептид без явных гомологов в геноме люцерны и NCBI.
46	PsCam053544 - CLE-пептид без явных гомологов в геноме люцерны.
47	PsCam056889 - MtCLE30
48	PsCam059704 - MtCLE46
49	PsCam059958 - Medtr5g013950

Таблица А.2 — Дифференциально экспрессирующиеся гены, экспрессия изменяется в корнях при обработке нитратом, выявленные при использовании в качестве референса сборки генома гороха линии Frisson (ФГБНУ ВНИИСХМ). Розовым цветом отмечены гены, экспрессия которых увеличивается при обработке нитратом, зеленым цветом – гены, экспрессия которых снижается при обработке нитратом.

GID	C_1	C_2	C_3	N_1	N_2	N_3	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
novel_gene_828_607e8030	0	0	0	80,69069	86,13333	69,7141	39,42301921	8,704939457	1,291812	6,738552	1,60E-11	2,48E-08
evm.TU.contig_1242.150	0	0	0	19,92363	111,0995	9,440451	23,41059782	7,947660296	1,588549	5,003095	5,64E-07	0,000531
evm.TU.contig_988.60	3,99643	0	2,363438	47,8167	78,64347	68,98791	33,63465977	4,840591691	0,80912	5,982539	2,20E-09	3,05E-06
evm.TU.contig_526.57	360,678	566,7918	263,1295	7978,416	13973,57	11561,65	5784,039237	4,817436553	0,394524	12,21075	2,72E-34	1,20E-30
evm.TU.contig_1214.180	422,623	399,3618	225,3145	2322,099	4250,492	3201,039	1803,488186	3,225778606	0,250737	12,86518	7,07E-38	6,21E-34
evm.TU.contig_1652.243	483,568	513,2691	433,297	1844,928	3219,389	3243,158	1622,934893	2,539800009	0,207612	12,2334	2,06E-34	1,09E-30
evm.TU.contig_1206.40	1205,92	940,0785	415,9651	3465,715	5634,867	3787,8	2575,058037	2,331166874	0,480576	4,850774	1,23E-06	0,001046
evm.TU.scaffold_1368.125	122,89	175,6643	222,951	730,2009	843,8569	894,6643	498,3712869	2,239377404	0,232134	9,64691	5,07E-22	1,48E-18
evm.TU.scaffold_1169.478	356,681	458,374	352,9401	1495,268	1914,906	1563,484	1023,60898	2,094422714	0,166692	12,56465	3,30E-36	2,18E-32
evm.TU.contig_379.46	336,699	345,8391	511,2905	1246,223	1581,608	2180,018	1033,612899	2,066045312	0,233343	8,854109	8,44E-19	2,22E-15
evm.TU.contig_396.114	243,782	389,7552	237,9195	858,7083	1413,086	1186,592	721,6405336	1,994234964	0,245046	8,138216	4,01E-16	8,14E-13
evm.TU.contig_413.6	62,9438	89,20453	64,60065	255,0224	338,2918	249,8089	176,6453343	1,968289465	0,284816	6,910733	4,82E-12	9,08E-09
evm.TU.scaffold_810.350	170,847	72,736	103,9913	331,7284	410,6937	434,987	254,1639544	1,754384066	0,315989	5,552042	2,82E-08	3,55E-05
evm.TU.contig_2621.12	3281,07	3507,796	2740,801	8081,023	10961,4	11648,06	6703,3594	1,687958629	0,152476	11,07034	1,75E-28	6,58E-25
evm.TU.contig_1224.116	57,9482	60,3846	67,7519	184,2935	176,0116	177,9162	120,7176786	1,528724632	0,285311	5,358104	8,41E-08	9,64E-05
evm.TU.contig_2415.14	5717,89	5264,439	5002,611	13056,95	15744,92	15360,34	10024,52594	1,466168847	0,106928	13,71171	8,64E-43	1,14E-38
evm.TU.scaffold_1092.14	7801,03	8193,093	7993,936	18847,75	21911,57	22368,79	14519,36138	1,396072802	0,095557	14,60988	2,43E-48	6,41E-44
evm.TU.scaffold_392.222	3331,03	3332,132	2688,017	7059,937	8128,989	7502,254	5340,392388	1,279613977	0,118026	10,84178	2,18E-27	7,19E-24
evm.TU.scaffold_2192.262	458,59	708,1467	306,4592	1018,097	1228,336	1051,521	795,1918004	1,167090968	0,260483	4,48048	7,45E-06	0,005307
evm.TU.contig_1649.364	1698,48	2014,65	2483,186	3540,428	4315,405	5459,486	3251,939579	1,103412821	0,181153	6,091067	1,12E-09	1,64E-06
evm.TU.contig_541.44	558,501	558,5576	441,963	1057,945	1240,819	1033,366	815,1919796	1,09754013	0,160683	6,830481	8,46E-12	1,39E-08
evm.TU.scaffold_1705.131	449,598	285,4545	991,0685	129,5036	218,4541	132,1663	367,7075708	-1,852087809	0,400977	-4,61894	3,86E-06	0,002825
evm.TU.scaffold_6316.27	6,99375	12,3514	109,506	0	0	0	21,47518776	-7,922919974	1,660654	-4,77096	1,83E-06	0,001381
evm.TU.scaffold_6303.52	7,99286	4302,403	2329,562	0	0	0	1106,659686	-13,60357072	2,849194	-4,77453	1,80E-06	0,001381

Таблица А.3 — Дифференциально экспрессирующиеся гены, экспрессия которых повышается в корнях при обработке нитратом, выявленные при использовании в качестве референса сборки генома гороха (*Pisum sativum_v1a* <https://urgi.versailles.inra.fr/download/pea/>). Приведены первые 25 генов с наименьшим значением p-value.

GID	baseMea						log2FoldC						
	C_1	C_2	C_3	N_1	N_2	N_3	n	hange	lfcSE	stat	pvalue	padj	
Psat6g231040.1	7903,033	8230,899	8076,642	19135,55	22345,7	22645,5	14722,89	1,405384	0,088998	15,79125	3,57E-56	1,18E-51	
Psat7g123960.1	334,7564	536,6196	249,4199	7674,325	13417,09	11105,61	5552,97	4,847953	0,333515	14,53593	7,17E-48	1,18E-43	
Psat1g167560.1	5683,906	5236,479	4716,948	11614,5	14582,61	12862,25	9116,114	1,320739	0,126566	10,43518	1,71E-25	1,41E-21	
Psat2g178360.1	325,8164	419,1914	313,1518	1354,527	1718,518	1377,272	918,0795	2,075901	0,199431	10,40911	2,25E-25	1,49E-21	
Psat4g048680.1	0	0	0	654,8705	797,2635	669,1259	353,5433	11,87645	1,203687	9,866723	5,80E-23	3,19E-19	
Psat2g077040.1	415,2172	391,8825	223,4551	2296,028	4173,544	3157,032	1776,193	3,225581	0,335686	9,608914	7,33E-22	3,45E-18	
Psat4g098200.1	0	0	0	388,1451	353,3749	355,5183	182,8397	10,92585	1,211945	9,015131	1,97E-19	7,21E-16	
Psat7g030360.1	2881,687	3155,542	2406,076	6956,754	9678,754	10262,34	5890,192	1,67212	0,194364	8,603019	7,76E-18	2,56E-14	
Psat5g118080.1	0	0	0	343,3591	243,0228	324,4466	151,8047	10,65843	1,241061	8,588158	8,84E-18	2,65E-14	
Psat1g168000.6	0	0	0	149,2866	205,8254	368,5251	120,6062	10,32671	1,347768	7,662082	1,83E-14	4,02E-11	
Psat4g178120.2	0	0	0	151,2771	110,3522	133,6807	65,88498	9,454073	1,287707	7,341791	2,11E-13	4,09E-10	
Psat7g183200.2	0	0	0	9892,724	10768,64	0	3443,56	28,2878	3,906772	7,240708	4,46E-13	8,10E-10	
Psat1g222080.2	0	2,730889	0	1978,545	117,7916	997,1855	516,0421	10,42231	1,455344	7,161404	7,99E-13	1,25E-09	
Psat1g222080.4	0	0	0	360,2783	85,55393	290,4845	122,7194	10,35139	1,453505	7,121677	1,07E-12	1,53E-09	
Psat7g228600.2	281,1159	210,2784	207,7188	1514,761	2246,721	864,2274	887,4704	2,725179	0,394183	6,913491	4,73E-12	6,00E-09	
Psat5g089800.1	57,61387	60,07955	66,09234	202,0345	190,9465	187,8755	127,4404	1,655769	0,24043	6,886696	5,71E-12	6,98E-09	
Psat3g164640.1	297,0094	308,5904	445,3365	1107,706	1450,697	1979,919	931,5432	2,108947	0,307779	6,852147	7,27E-12	8,27E-09	
Psat7g011960.1	62,58058	88,75389	64,51872	254,7824	336,0162	248,5738	175,8709	1,966901	0,303069	6,489953	8,59E-11	9,44E-08	
Psat2g126320.2	0	0	0	78,62426	71,9149	70,81462	36,8923	8,616541	1,34223	6,419569	1,37E-10	1,45E-07	
Psat5g207280.1	242,3756	387,7862	237,6177	858,8954	1403,58	1180,725	718,4968	1,991766	0,316345	6,296188	3,05E-10	3,14E-07	
Psat7g195480.12	1,986685	0	0	120,4245	464,967	134,4033	120,2969	8,480741	1,356583	6,251546	4,06E-10	4,06E-07	
Psat0s3458g0040.2	0	0	0	873,8241	37,19736	232,6766	190,6163	10,98602	1,762498	6,233214	4,57E-10	4,31E-07	
Psat7g018600.2	18,87351	12,289	3,147254	111,4673	680,7117	591,8079	236,3828	5,346683	0,889033	6,01404	1,81E-09	1,61E-06	
Psat2g171360.2	168,8682	278,5507	197,4902	841,9763	590,1982	695,1395	462,0372	1,726804	0,289316	5,968574	2,39E-09	2,08E-06	
Psat1g014800.3	0	0	0	101,5149	73,15481	46,96888	36,93976	8,61726	1,445702	5,960604	2,51E-09	2,13E-06	

Таблица А.4 — Дифференциально экспрессирующиеся гены, экспрессия которых снижается в корнях при обработке нитратом, выявленные при использовании в качестве референса сборки генома гороха (*Pisum sativum*_v1a <https://urgi.versailles.inra.fr/download/pea/>). Приведены первые 25 генов с наименьшим значением p-value.

GID	log2Fold						baseMean	Change	lfcSE	stat	pvalue	padj
	C_1	C_2	C_3	N_1	N_2	N_3						
Psat5g283040.1	2004,565	1264,402	1695,583		0	0	1,445196	827,6659-11,3825	1,0658	-10,6797	1,27E-26	1,39E-22
Psat2g148560.1	400,317	375,4972	185,688		0	0	0	160,2504 -10,821	1,307742	-8,27458	1,29E-16	3,54E-13
Psat5g112240.1	175,8216	155,6607	117,2352		0	0	0	74,78625-9,72155	1,290125	-7,53536	4,87E-14	1,00E-10
Psat7g235200.2	164,8949	165,2188	104,6462	11,94293	9,919297	16,61976		78,87364-3,46994	0,479621	-7,23475	4,66E-13	8,10E-10
Psat6g072360.1		0 3985,732	7278,813		0	0	0	1877,424 -28,138	3,906789	-7,20232	5,92E-13	9,76E-10
Psat4g088480.1	168,8682	176,1423	86,5495		0	0	0	71,92668-9,66458	1,357068	-7,12166	1,07E-12	1,53E-09
Psat7g219200.3	650,6394	547,5432	894,6071	187,1058	121,5114	162,5846		427,3319 -2,1497	0,308126	-6,97669	3,02E-12	4,15E-09
Psat3g024360.2	201,6485	158,3915	219,521	15,9239	11,15921	28,18133		105,8043-3,36009	0,483795	-6,94527	3,78E-12	4,98E-09
Psat5g037320.1	1642,989	2435,953	2214,88	422,9786	658,3933	677,7971		1342,165-1,83862	0,268118	-6,85751	7,01E-12	8,25E-09
Psat0s3597g0040.3	66,55395	117,4282	62,15828		0	0	0	41,02341-8,85346	1,41748	-6,24592	4,21E-10	4,09E-07
Psat2g126320.1	52,64715	58,71411	89,69675		0	0	0	33,50967-8,56608	1,423755	-6,01654	1,78E-09	1,61E-06
Psat4g178120.1	288,0693	297,6669	257,2881	76,63378	85,55393	113,4479		186,4433-1,60374	0,270259	-5,93411	2,95E-09	2,38E-06
Psat5g118080.2	338,7298	495,6563	443,7629	100,5196	59,51578	36,85251		245,8395-2,70286	0,459755	-5,87892	4,13E-09	3,24E-06
Psat0s3108g0040.2	66,55395	62,81044	43,27475		0	0	0	28,77319-8,34298	1,431105	-5,82975	5,55E-09	4,26E-06
Psat7g014800.8	85,42746	45,05967	56,65058		0	0	0	31,18962 -8,462	1,453731	-5,82089	5,85E-09	4,39E-06
Psat1g100480.3	907,9151	663,606	604,2729	0,995244	18,59868		0	365,898-6,84236	1,23227	-5,55265	2,81E-08	1,93E-05
Psat7g260960.2		0 94,21566	0	0	0	0	0	15,70261-21,6142	3,909731	-5,52832	3,23E-08	2,13E-05
Psat0s2052g0040.8	32,7803	191,1622	74,74729		0	0	0	49,78164-9,13337	1,651234	-5,53124	3,18E-08	2,13E-05
Psat2g137240.1		0 92,85022	0	0	0	0	0	15,47504 -21,595	3,909774	-5,52333	3,33E-08	2,15E-05
Psat6g064720.2		0 91,48477	0	0	0	0	0	15,24746-21,5744	3,909819	-5,51801	3,43E-08	2,17E-05
Psat0s471g0160.1		0 90,11933	0	0	0	0	0	15,01989-21,5543	3,909866	-5,5128	3,53E-08	2,20E-05
Psat6g180000.3		0 88,75389	0	0	0	0	0	14,79231-21,5342	3,909913	-5,50758	3,64E-08	2,22E-05
Psat2g182960.2	220,522	31,40522	39,34068		0	0	0	48,54466 -9,0995	1,793962	-5,0723	3,93E-07	0,000188
Psat2g015880.9	10,92677	583,0448	68,45279		0	0	0	110,4041-10,2836	2,043916	-5,03132	4,87E-07	0,000226
Psat3g144040.1	180,7883	17,75078	66,87916		0	0	0	44,23638-8,96594	1,799589	-4,98222	6,29E-07	0,000276

Таблица А.5 — Дифференциально экспрессирующиеся гены, экспрессия которых повышается в корнях люцерны со сверхэкспрессией MtCLE35, отсортированные по откорректированному значению p-value (padj). Приведены первые 25 генов с наименьшим значением p-value. Желтым цветом выделен ген *MtCLE35*.

GID	GUS_n_1	GUS_n_2	GUS_n_3	CLE35_n_1	CLE35_n_2	CLE35_n_3	baseMean	log2Fold Change	lfcSE	stat	pvalue	padj	Gene product
Medtr5g037950	44,32102	69,66223	76,43289	679,32947	905,38983	413,42418	364,7599	3,370619519	0,324091	10,40023	2,47E-25	2,19E-22	thioredoxin
Medtr4g108150	389,7193	481,489	635,0535	1639,602	3302,4465	2314,9486	1460,543	2,265706366	0,269237	8,415273	3,92E-17	8,51E-15	Lipid transfer protein
Medtr2g091125	15,28311	4,097778	7,548928	139,32011	253,806	111,15383	88,53496	4,337909185	0,531203	8,166192	3,18E-16	5,65E-14	Clavata3/ESR (CLE) gene family member
Medtr4g059670	84,05711	68,97927	46,23718	227,97837	304,27035	309,64281	173,5275	2,095689298	0,276751	7,572475	3,66E-14	4,70E-12	root cap/late embryogenesis related
Medtr3g088565	87,11373	62,8326	85,86905	255,61211	230,05807	234,7841	159,3783	1,62874287	0,22557	7,220575	5,18E-13	5,40E-11	1-aminocyclopropane-1-carboxylate oxidase
Medtr4g029190	38,20778	38,24593	68,88397	184,22494	417,07302	302,83747	174,9122	2,630080459	0,375681	7,00083	2,54E-12	2,52E-10	class III peroxidase
Medtr6g029330	255,2279	281,3808	352,9124	702,35759	1211,1444	830,25104	605,5457	1,620385564	0,237594	6,81999	9,10E-12	8,25E-10	transmembrane protein putative
Medtr5g004980	191,0389	211,7186	236,8476	485,89328	529,87568	424,1993	346,5956	1,160297194	0,17271	6,718178	1,84E-11	1,61E-09	transmembrane protein putative
Medtr7g093260	206,322	212,4015	231,1859	421,41455	528,39144	447,45087	341,1944	1,098462662	0,165273	6,646346	3,00E-11	2,51E-09	actin-97
Medtr6g015905	50,43426	73,07705	31,13933	200,34462	284,97516	249,529	148,2499	2,23388069	0,337398	6,620912	3,57E-11	2,95E-09	alpha-1,4-glucan-protein synthase [UDP-forming]-like protein
Medtr3g117060	136,0197	206,9378	175,5126	388,02378	822,27207	819,47592	424,707	1,962764028	0,308526	6,361738	1,99E-10	1,48E-08	salicylic acid carboxyl methyltransferase
Medtr3g028540	48,90595	61,46668	80,20736	185,37635	244,90053	193,95209	135,8015	1,692394812	0,268663	6,299331	2,99E-10	2,16E-08	transmembrane protein putative
Medtr8g075920	24,45298	23,90371	25,47763	85,204035	112,80267	95,841821	61,28047	1,989155321	0,325112	6,118378	9,45E-10	6,45E-08	gibberellin-regulated family protein
Medtr7g010580	611,3244	721,209	551,0717	1070,8075	1435,2655	1309,4601	949,8564	1,015933954	0,169306	6,000591	1,97E-09	1,26E-07	gibberellin-regulated family protein
Medtr7g084770	551,7203	674,0846	607,6887	1089,23	1729,1461	1303,789	992,6098	1,164940048	0,194324	5,994833	2,04E-09	1,29E-07	xyloglucan endotransglucosylase/hydrolase family protein
Medtr5g075190	16,81142	15,70815	16,98509	65,630135	87,570491	73,724477	46,07163	2,197034151	0,376405	5,836889	5,32E-09	3,14E-07	hypothetical protein
Medtr4g117660	122,2649	127,0311	102,8541	255,61211	376,99839	272,21345	209,4957	1,355252711	0,232805	5,821396	5,84E-09	3,41E-07	copper amine oxidase enzyme domain protein
Medtr4g094815	322,4736	514,2712	474,6388	834,76927	1295,7464	1262,957	784,1427	1,366630487	0,236456	5,779651	7,49E-09	4,32E-07	caffeoyl-CoA 3-O-methyltransferase
Medtr7g059290	157,416	192,5956	149,0913	314,33381	479,41133	459,36021	292,0347	1,323165005	0,229379	5,768468	8,00E-09	4,60E-07	pectate lyase-like protein
Medtr4g044463	45,84933	28,00149	32,08294	133,56308	201,8574	108,88538	91,7066	2,086305872	0,362979	5,74773	9,04E-09	5,14E-07	STIG1
Medtr1g033130	181,869	292,3082	238,7348	508,9214	1006,3185	682,23497	485,0645	1,615442641	0,282056	5,727385	1,02E-08	5,70E-07	pollen protein Ole E I-like protein
Medtr2g103170	55,0192	88,10224	100,0233	261,36914	650,09958	269,94501	237,4264	2,266076048	0,398125	5,691864	1,26E-08	6,93E-07	plant/F18G18-200 protein
Medtr2g030835	27,5096	33,46519	45,29357	112,83778	241,93204	138,94228	99,99674	2,196837943	0,386268	5,687341	1,29E-08	7,10E-07	root cap/late embryogenesis related
Medtr5g011210	30,56622	25,9526	34,91379	86,355441	141,00333	196,22053	85,83532	2,225719571	0,392173	5,675352	1,38E-08	7,57E-07	expansin A10
Medtr1g028610	79,47217	117,4696	140,5988	284,39725	445,27368	303,40458	228,436	1,597277434	0,28165	5,671144	1,42E-08	7,70E-07	arabinogalactan peptide-like protein

Таблица А.6 — Дифференциально экспрессирующиеся гены, экспрессия которых снижается в корнях со сверхэкспрессией MtCLE35, отсортированные по откорректированному значению p-value (padj). Приведены первые 25 генов с наименьшим значением p-value.

GID	GUS_n_1	GUS_n_2	GUS_n_3	CLE35_n_1	CLE35_n_2	CLE35_n_3	baseMean	log2Fold Change	lfcSE	stat	pvalue	padj	Gene product
Medtr0416s0030	2792,224	2343,929	3535,729	0	0	2,8355568	1445,786	-10,926	0,682638	-16,0056	1,17E-57	2,48E-53	leginsulin related MtN11/16/17 family
Medtr2g069245	539,4938	542,9556	443,4995	0	7,4212281	10,208005	257,263	-6,25578	0,395338	-15,8239	2,13E-56	2,27E-52	Nodule-specific Glycine Rich Peptide
Medtr5g045250	519,6257	480,806	648,2642	41,450612	19,295193	29,489791	289,8219	-4,17368	0,268076	-15,569	1,18E-54	8,39E-51	cytochrome P450 family protein
Medtr4g036990	411,1157	380,4104	387,8262	57,570294	51,948597	48,204466	222,8459	-2,91427	0,198691	-14,6673	1,04E-48	5,55E-45	transmembrane protein putative
Medtr2g103570	1381,593	1249,822	1235,193	411,0519	417,07302	402,64907	849,5638	-1,65129	0,115359	-14,3144	1,78E-46	7,58E-43	microsomal signal peptidase subunit
Medtr1g074410	950,6095	836,6298	401,9804	1,1514059	5,9369825	3,9697796	366,713	-7,65394	0,54339	-14,0855	4,66E-45	1,65E-41	transmembrane protein putative
Medtr1g051120	460,0216	333,286	394,4315	34,542177	16,326702	20,98312	209,9318	-4,045	0,293607	-13,7769	3,51E-43	1,07E-39	auxin-induced SNG4-like protein
Medtr6g078100	314,8321	219,9141	322,7167	13,816871	16,326702	13,610673	150,2028	-4,31091	0,32933	-13,0899	3,76E-39	1,00E-35	Kunitz type trypsin inhibitor / miraculin
Medtr5g056815	282,7375	269,0874	332,1528	5,7570294	0	3,4026682	148,8563	-6,47666	0,497439	-13,02	9,42E-39	2,23E-35	MtN27 family
Medtr4g120950	790,1368	639,9364	967,2064	4,6056235	0	0	400,3142	-9,33943	0,753546	-12,394	2,82E-35	6,00E-32	pathogenesis-related protein bet V I family protein
Medtr4g066070	3810,079	3895,621	2384,518	0	0	0	1681,703	-14,1733	1,198924	-11,8217	3,02E-32	5,84E-29	Nodule-specific Glycine Rich Peptide
Medtr3g055450	6666,493	5906,265	2714,783	0	0	0	2547,923	-14,7728	1,30598	-11,3116	1,15E-29	2,04E-26	NMS32/34 protein putative
Medtr0174s0090	2127,409	2199,824	1445,62	0	0	0	962,1421	-13,3677	1,196467	-11,1726	5,55E-29	9,09E-26	Nodule-specific Glycine Rich Peptide
Medtr0093s0090	1645,991	1369,341	1920,259	0	0	0	822,5984	-13,1412	1,192079	-11,0238	2,93E-28	4,46E-25	leginsulin related MtN11/16/17 family
Medtr4g130800	1629,18	1423,978	1315,401	0	0	0	728,093	-12,9643	1,188156	-10,9113	1,02E-27	1,36E-24	plastocyanin-like domain protein
Medtr6g093180	209,3786	199,4252	223,637	0	0	5,1040023	106,2575	-6,3631	0,583184	-10,911	1,02E-27	1,36E-24	beta-amyrin synthase
Medtr4g130780	1592,5	1473,834	1187,069	0	0	0	708,9005	-12,9263	1,191063	-10,8527	1,94E-27	2,42E-24	early nodulin-like protein
Medtr0240s0040	3855,929	4308,814	1715,494	0	0	0	1646,706	-14,1431	1,313468	-10,7677	4,89E-27	5,78E-24	Nodule Cysteine-Rich (NCR) secreted peptide
Medtr5g094210	466,1349	331,2371	497,2856	81,749818	60,85407	63,516473	250,1297	-2,65014	0,246685	-10,743	6,39E-27	7,16E-24	Lipid transfer protein
Medtr8g023090	201,7371	197,3763	202,8774	0	0	3,9697796	100,9934	-6,62688	0,62389	-10,6219	2,36E-26	2,51E-23	Defensin related
Medtr5g018770	359,1531	296,406	418,0219	73,689977	57,885579	66,919141	212,0126	-2,42643	0,230778	-10,5141	7,44E-26	7,54E-23	CAP cysteine-rich secretory protein antigen 5
Medtr6g007160	1135,535	1156,256	889,8299	0	0	0	530,2702	12,508	1,191713	-10,4958	9,03E-26	8,74E-23	transmembrane protein putative
Medtr5g037795	1445,782	1732,677	691,6705	0	0	0	645,0217	-12,791	1,224472	-10,4462	1,53E-26	1,41E-22	Nodule Cysteine-Rich (NCR) secreted peptide
Medtr2g030470	1078,988	806,5794	1052,132	0	0	0	489,6165	-12,3913	1,193203	-10,3849	2,90E-25	2,47E-22	stress up-regulated Nod 19 protein
Medtr0330s0050	1192,083	1124,84	693,5577	0	0	0	501,7468	-12,4278	1,204996	-10,3135	6,12E-25	5,01E-22	Nodule Cysteine-Rich (NCR) secreted peptide

Таблица А.8 — Значения p-value с коррекцией FDR для результатов F-теста на достоверность эффекта влияния SNV (single nucleotide variation) на параметры роста и урожайности растений в двухфакторном дисперсионном анализе (two-way ANOVA)

SNV	PW, Shoot weight, g	PL, Shoot length, cm	PN, Pod number	SN, Seed number per plant	SW, Seed weight per plant, g	TSW, Thousand seed weight, g	NtS, Seed nitrogen content, mg/g	PhS, Seed phosphorus content, mg/g
NRLK1.19SL	0.8975	0.8975	0.8975	0.8975	0.8975	0.9013	0.8975	0.8975
NRLK1.30SP	0.5255	0.9871	0.8602	0.5255	0.5255	0.3465	0.9871	0.9871
NRLK1.212RG	0.8226	0.8226	0.8226	0.1208	0.8226	0.2551	0.8226	0.8226
NRLK1.356LQ	0.7879	0.0188	0.8651	0.7879	0.7879	0.7879	0.6018	0.6721
NRLK1.443GR	0.6138	0.0670	0.7739	0.6138	0.6292	0.6138	0.6138	0.6138
NRLK1.875FY	0.6113	0.6113	0.6113	0.6113	0.6765	0.6113	0.6113	0.9087
NRLK1.890GE	0.2631	0.2728	0.2365	0.2631	0.2816	0.2599	0.2728	0.3020
Sym29.16FS	0.2778	0.1792	0.8113	0.3173	0.4688	0.8113	0.8113	0.2778
Sym29.43PH	0.4068	0.7399	0.8814	0.8814	0.5393	0.1745	0.7858	0.8814
Sym29.53AGV	0.4022	0.7395	0.8792	0.8792	0.5575	0.2034	0.7608	0.8792
Sym29.171IF	0.7670	0.6475	0.6475	0.8005	0.8005	0.6475	0.6475	0.6676
Sym29.S618P	0.4857	0.3378	0.4857	0.3378	0.4857	0.4857	0.3378	0.4857
Sym29.T623I	0.5105	0.4421	0.7061	0.6451	0.5105	0.8590	0.4021	0.5105

Таблица А.9 Значения p-value с коррекцией FDR для результатов тестов Likelihood-Ratio Tests на достоверность эффекта влияния SNV (single nucleotide variation) на прибавки, вызванные действием арбускулярной микоризы (AM)

SNV	PW, Shoot weight, g	PL, Shoot length, cm	PN, Pod number	SN, Seed number per plant	SW, Seed weight per plant, g	TSW, Thousa nd seed weight, g	NtS, Seed nitroge n con- tent, mg/g	PhS, Seed phosph orus content, mg/g
NRLK1.19SL	0.3444	0.7264	0.7264	0.7264	0.4034	0.3444	0.4034	0.3444
NRLK1.30SP	0.4384	0.8388	0.8003	0.8388	0.6873	0.4384	0.6873	0.4384
NRLK1.212RG	0.8517	0.8517	0.8517	0.8861	0.8517	0.3277	0.8517	0.8861
NRLK1.356LQ	0.1350	0.8911	0.7722	0.8412	0.1350	0.1209	0.1350	0.5528
NRLK1.443GR	0.2186	0.9221	0.9221	0.9221	0.2186	0.0901	0.1646	0.9221
NRLK1.875FY	0.2499	0.2499	0.4094	0.4488	0.2499	0.4453	0.7828	0.4094
NRLK1.890GE	0.3769	0.9540	0.8308	0.9591	0.4788	0.3769	0.4788	0.4788
Sym29.16FS	0.8836	0.8998	0.8998	0.8836	0.8836	0.6663	0.8836	0.8836
Sym29.43PH	0.1674	0.8405	0.7943	0.9994	0.3782	0.1674	0.7943	0.7943
Sym29.53AGV	0.9035	0.7123	0.5304	0.5304	0.9035	0.5304	0.7123	0.7123
Sym29.171IF	0.8495	0.8495	0.8495	0.8495	0.8495	0.8495	0.8495	0.8495
Sym29.S618P	0.5022	0.9898	0.5022	0.9898	0.5853	0.5022	0.9898	0.5022
Sym29.T623I	0.7992	0.7992	0.5987	0.8889	0.8042	0.7992	0.7992	0.5987