

На правах рукописи

Фесенко Игорь Александрович

**СИСТЕМНЫЙ АНАЛИЗ ПЕПТИДОМА РАСТЕНИЙ
НА ПРИМЕРЕ МХА *PHYSCOMITRIUM PATENS***

Специальность 1.5.3 - Молекулярная биология

**Автореферат
диссертации на соискание ученой степени
доктора биологических наук**

Москва - 2022

Работа выполнена в отделе молекулярной биологии и биотехнологии растений Федерального государственного бюджетного учреждения науки Институт биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук.

Научный консультант:

доктор химических наук, академик **Иванов Вадим Тихонович**

Официальные оппоненты:

Лисица Андрей Валерьевич

доктор биологических наук, академик Российской академии наук, Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича», главный научный сотрудник группы биобанкинга, руководитель Центра научно-практического образования ИБМХ

Лось Дмитрий Анатольевич

доктор биологических наук, профессор, член-корреспондент Российской академии наук, Федеральное государственное бюджетное учреждение науки Институт физиологии растений им. К.А. Тимирязева Российской академии наук, директор института

Макеев Всеволод Юрьевич

доктор физико-математических наук, профессор, член-корреспондент Российской академии наук, Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук, заведующий лабораторией системной биологии и вычислительной генетики

Ведущая организация:

Федеральное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет»

Защита состоится 15 июня 2022 г. в 11 часов на заседании диссертационного совета 24.1.037.01 при Федеральном государственном бюджетном учреждении науки Институте биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук по адресу: 117997, ГСП-7, Москва В-437, ул. Миклухо-Маклая, д. 16/10.

С диссертацией можно ознакомиться в библиотеке Института биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова РАН и на сайте www.ibch.ru

Автореферат разослан ____ _____ 2020г.

Ученый секретарь

диссертационного совета,

доктор физико-математических наук



Олейников В.А.

ОБЩАЯ ХАРАКТЕРИСТИКА ПРОБЛЕМЫ

Актуальность исследования. Поиск новых и изучение механизмов действия уже известных молекул, имеющих регуляторные функции, является одной из важнейших задач современной биологии. Пептиды идентифицированы у всех живых организмов и являются одними из наиболее известных регуляторов биологических процессов, включая рост и развитие, иммунные реакции, ответ на стрессовые факторы и множество других. Учитывая это многообразие, открытие новых, ранее неизвестных биологически активных пептидов имеет важное фундаментальное и практическое значение. Однако несмотря на то, что у растений количество открытых за последние 30 лет пептидных гормонов уже превысило количество классических фитогормонов, качественный и количественный состав, а также динамика образования и деградации пептидома (совокупности всех пептидов) растительной клетки мало изучены. Связано это со сложностью идентификации, выделения и проверки биологической активности нативных пептидов. Так, системные исследования биоразнообразия пептидов, которые начались в начале 1990-х годов, в первое время продемонстрировали скромный прогресс из-за ограничений доступных аналитических методов. При этом только «деградом», образующийся при расщеплении белков клеточными системами деградации, состоит из тысяч пептидов, часть из которых может быть биологически активной. Например, в последние годы появляется все больше данных о том, что биологически активные пептиды могут быть скрыты в последовательностях функциональных белков. Такие пептиды называются «криптидами» или криптопептидами.

В последние годы стремительное развитие методов масс-спектрометрического (МС) анализа в сочетании со взрывным ростом доступных баз генетических данных сделало возможным проведение системного анализа внутриклеточных и секретируемых пептидомов растений. Учитывая вышесказанное, такой системный анализ позволит выявить новые биологически активные пептиды, пролить свет на механизмы их возникновения в процессе эволюции растений, обнаружить новые роли пептидома в регуляции жизненно важных процессов в клетке. Однако на момент начала этой работы данные о составе пептидомов растений практически отсутствовали.

Практически все известные функциональные пептиды растений образуются при протеолизе специализированных белковых предшественников, а часть из них подвергается пост-трансляционным модификациям. Следует отметить, что работа по идентификации и выявлению функций различных пептидных семейств еще далека от завершения. Кроме

того, проведенные на животных моделях исследования показали, что транскрипционный ландшафт эукариотических клеток сложен и гетерогенен, а основную его часть составляют РНК, не кодирующие белки. Показано, что часть из них, такие как длинные некодирующие РНК, тем не менее могут кодировать пептиды, таким образом формируя неизученную часть пептидома, которая способна выполнять регуляторные функции. Выявление и анализ биологических функций таких пептидов является одной из актуальных задач современной биологии. Таким образом, системный анализ пептидома растений помимо открытия новых биологически активных пептидов и выявления неизвестных прежде механизмов регуляции клеточных процессов, является необходимым шагом для разработки способов использования таких пептидных регуляторов в сельском хозяйстве и производстве.

Целью данной работы являлся системный анализ пептидомов и идентификация новых функциональных пептидов растений на примере известного модельного объекта растительной биологии - мха *Physcomitrella (Physcomitrium patens)*.

Для выполнения вышеописанной цели были поставлены следующие **задачи**:

1. Проанализировать внутриклеточные и секретируемые пептидные пулы *P. patens* с использованием масс-спектрометрического анализа.
2. Изучить влияние стрессовых факторов на процесс формирования пептидомов растений на примере *P. patens*.
3. Выявить роль пептидов – продуктов деградации белков-предшественников в ответе на стрессовые факторы.
4. Изучить эволюцию коротких открытых рамок считывания в геноме *P. patens*, потенциально способных кодировать пептиды.
5. Идентифицировать пептиды, кодируемые короткими открытыми рамками считывания у *P. patens* при помощи транскрипционного и масс-спектрометрического методов анализа.
6. Проанализировать функции пептидов, кодируемых транскриптами длинных некодирующих РНК, на примере *P. patens*.

Научная новизна и практическая значимость работы. Несмотря на то, что первый функциональный пептид растений был открыт более 30 лет назад, полного анализа пептидных пулов тканей и секретома до недавнего времени не проводилось. В ходе выполнения диссертационной работы были предложены методы анализа пептидома растений и проведен системный анализ внутриклеточных и секретируемых пептидных

пулов растений на примере модельного организма - мха *P. patens*. При помощи разработанных подходов впервые были идентифицированы несколько тысяч эндогенных пептидов, среди которых фрагменты белков-предшественников, а также продукты трансляции коротких открытых рамок считывания в тканях и секрете *P. patens*.

Было показано, что пептидные пулы специфичны для разных тканей и зависят от факторов внешней среды. В данной работе впервые было доказано влияние стрессовых условий на формирование пептидных пулов растений и показано образование пептидов, обладающих антимикробной активностью из функциональных белков-предшественников. На основе полученных данных была предложена гипотеза об «активном управлении» пептидными пулами, согласно которой регуляция протеолиза белков-предшественников в стрессовых условиях напрямую влияет на состав "белкового деградомата" и приводит к образованию биологически активных элементов. Предложенная гипотеза в том числе помогает объяснить ряд обнаруженных ранее эффектов: таких как подавление активности протеасомного пути деградации фитопатогенами или повышенной экспрессии генов, участвующих в ответе на стресс при накоплении пептидов от хлоропластных белков у мутантов по определенным олигопептидазам. Это открывает новые направления в изучении роли пептидогенеза в ответе на стрессовые факторы внешней среды и защите растений от фитопатогенов в частности.

Использование в данной работе системного подхода, включающего анализ транскриптома, протеома и пептидома, позволило впервые у растений идентифицировать с помощью масс-спектрометрического анализа пептиды, кодируемые различными классами коротких открытых рамок считывания (КОРС). Показано, что наиболее многочисленной группой продуктов трансляции КОРС в пептидоме являются пептиды, кодируемые альтернативными рамками считывания, локализованными на аннотированных транскриптах мРНК. Согласно этим данным, знания о кодирующем потенциале мРНК являются неполными, и подчеркивают необходимость ре-аннотации растительных протеомов.

Также в диссертационной работе впервые для растений проведен детальный анализ эволюции пептидов, кодируемых малоизученным типом транскриптов - длинными некодирующими РНК (длнкРНК). Данный анализ проведен с использованием базы, содержащей транскриптомы тысячи видов растений. Показано, что тысячи коротких рамок считывания находятся под действием стабилизирующего отбора, однако большинство КОРС являются эволюционно молодыми и могут представлять собой материал для дальнейшей эволюции в функциональные гены. Были идентифицированы многочисленные

консервативные пептиды, кодируемые кОРС, содержащие регионы низкой сложности или трансмембранные домены, а также предложены подходы для их функционального анализа. Эти результаты закладывают основу для дальнейшей идентификации и изучения новых семейств функциональных пептидов, кодируемых кОРС у разных групп растений.

В рамках такого анализа впервые для растений были изучены функции пептидов, кодируемых длнкРНК и содержащих трансмембранные домены и регионы низкой сложности. Функциональный анализ таких пептидов, кодируемых длнкРНК, показал, что они участвуют в регуляции роста и дифференцировки тканей. Эти результаты указывают на то, что пептиды, кодируемые кОРС, являются функционально разнообразным компонентом растительного протеома и прежде неизвестным источником регуляторных молекул.

Апробация работы. Материалы работы были представлены в стендовых и устных докладах на российских и международных научных конференциях. По теме диссертации опубликованы 19 статей в рецензируемых отечественных и зарубежных журналах.

Структура и объем диссертации.

Диссертация состоит из введения, литературного обзора, материалов и методов, результатов и обсуждения, заключения, выводов и списка цитируемой литературы.

Диссертация изложена на 252 страницах и содержит 86 рисунков и 9 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Идентификация и анализ внутриклеточных пептидных пулов *P. patens*

На сегодняшний момент основным методом идентификации эндогенных пептидов в тканях и биологических жидкостях является масс-спектрометрический анализ. На первом этапе данного исследования мы использовали масс-спектрометрический анализ для изучения внутриклеточных пептидных пулов растений, используя две жизненные формы - гаметофоры и протонему модельного объекта *Physcomytrium patens* (Рис. 1). Для идентификации пептидов, являющихся продуктами деградации аннотированных предшественников, была использована база белков *P. patens* из Phytozome V12. Данный подход позволил нам идентифицировать пептиды размером от 5 до 78 а.о. В тканях гаметофоров был идентифицирован в общей сложности 4361 пептид, который образовался из 1859 белков-предшественников. Среди наиболее пептидогенных предшественников у гаметофоров мы идентифицировали белки хлоропластов, такие как A9SPD7 (субъединица IV реакционного центра фотосистемы I, 52 пептида) и A9U1R6 (белок пор внешней

оболочки, 42 пептида), а также шапероноподобные белки, такие как A9SU24 (предполагаемый LEA белок, 41 пептид).



Рисунок 1. Схема выделения внутриклеточных и секретируемых эндогенных пептидов из протонемы мха *Physcomitrium patens*.

В клетках протонемы мы идентифицировали 4332 пептида, которые являются фрагментами 1554 белков-предшественников. Основные пептидогенные белки-предшественники протонемы — это большая субъединица рибулозо-1,5-бисфосфаткарбоксилазы/оксигеназы, фактор элонгации 1-альфа, хлоропластная фруктозо-бисфосфат альдолаза и субъединица реакционного центра фотосистемы I. Таким образом, мы обнаружили, что внутриклеточные пептидные пулы растений включают в себя несколько тысяч пептидов - продуктов деградации аннотированных белков (“белковый деградом”; **Рис. 2**).

Мы обнаружили сравнительно низкий процент ароматических аминокислот в обоих наборах пептидов и более высокий уровень гидрофобности у пептидов, выделенных из тканей гаметофор. Можно предположить, что на композицию пептидных пулов оказывает влияние как состав белков в соответствующей ткани, так и процесс пробоподготовки, который, например, мог привести к снижению количества гидрофобных пептидов. Анализ наиболее пептидогенных белков-предшественников выявил участки длиной 15-25 аминокислот, которые были покрыты наборами структурно связанных пептидов (“пептидные лесенки”). Мы полагаем, что первичный протеолиз происходит в участках, расположенных между этими сегментами, предположительно внутри протеасом. После того, как первичные фрагменты высвобождаются из протеасомы, они далее деградируют на более мелкие пептиды с помощью amino- и карбоксипептидаз. Можно предположить, что процесс деградации белков происходит случайным образом и что увеличение

представленности белка в протеоме коррелирует с увеличением количества эндогенных пептидов, продуктов гидролиза этого белка, в пептидоме.

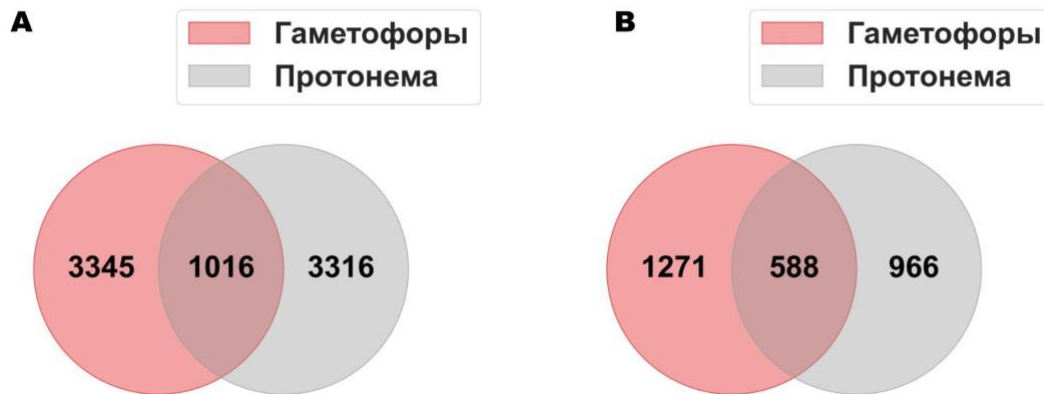


Рисунок 2. Диаграммы Венна, показывающие пересечение идентифицированных в протонеме и гаметофорах **А** - пептидов, **В** - белков-предшественников.

Для того, чтобы проверить эту гипотезу, мы выполнили количественный протеомный анализ протонемы и гаметофоров мха. Однако какой-либо линейной зависимости между количеством эндогенных пептидов от белка и его представленностью в протеоме клетки обнаружено не было. Полученные результаты могут быть объяснены несколькими причинами: 1) количество пептидов может быть связано со скоростью синтеза/деградации белков, а не с концентрацией белка; 2) существуют различия в активности протеолитических ферментов в зависимости от типа ткани; 3) деградация пептидов определенного состава зависит от типа ткани; 4) особенности пробоподготовки пептидов для масс-спектрометрического анализа.

Можно предположить, что в растительной клетке значительная доля пептидов может продуцироваться не протеасомой, а различными органельными протеазами. Чтобы проверить это предположение, мы далее сравнили аминокислоты в P1 позиции в нашем наборе пептидомных данных (**Рис. 3**) и в эндогенных пептидах, связанных с протеасомами в пептидомном датасете из клеток человека. В нашем наборе данных аминокислоты в позиции P1 не были похожи на таковые у пептидов, ассоциированных с протеасомой. Это может быть объяснено несколькими причинами: 1) после протеолитического расщепления не связанные с протеасомой пептиды подвергаются дальнейшей деградации олигопептидазами, формируя “пептидные лестницы”; 2) присутствие большого количества пептидов из органельных предшественников, которые выщепляются

специализированными протеазами; 3) особенностями выделения пептидов из растительной клетки.

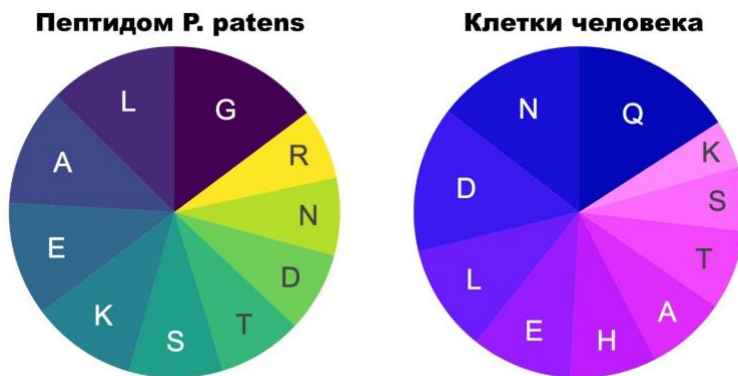


Рисунок 3. Частота встречаемости топ10 аминокислот на С-концах внутриклеточных пептидов мха в сравнении с пептидами человека, ассоциированными с протеасомами.

По всей видимости, внутриклеточные пептидные пулы растений являются сложной смесью пептидов, образующихся при гидролизе белков протеасомами, органельными протеазами и продуктами их дальнейшей деградации олигопептидазами.

Идентификация и анализ внеклеточных пептидных пулов *P. patens*

Ранее было показано, что секретом растений содержит десятки белков. Однако мало известно о составе секретируемых пептидных пулов и протеаз, ответственных за образование этих пептидных пулов. Исходя из этого, мы провели анализ пептидома секретома растений, используя в качестве модели культуральную среду, в которой выращивали протонему *P. patens*. Используя МС анализ, мы идентифицировали 802 эндогенных пептида, продуктов протеолиза 220 белков-предшественников. Из них 387 эндогенных пептидов были идентифицированы, по меньшей мере, в двух биологических повторях, что позволяет предположить существование стабильного пула пептидов в секретома мха. По сравнению с внутриклеточным пептидомом протонемы, наиболее пептидогенными белками-предшественниками в секретома были мембранные и секретируемые белки, такие как экспансин, пектинэстераза и ряд других. Мы обнаружили, что около 20% пептидов секретома были общими с пептидомами клеток (**Рис. 4**). С учетом формирования “пептидных лестниц”, количество общих пептидов возрастает до 40%. Большинство этих общих пептидов происходит из внутриклеточных белков, которые участвуют в различных метаболических процессах. Большая часть белков-предшественников секретируемых пептидов участвовали в фотосинтетических реакциях и модификации клеточной стенки (деградация пектина). Мы идентифицировали 12 белков

хлоропластной локализации, таких как psbA и psbB белок фотосистемы II, малые и большие субъединицы рибулозобисфосфат-карбоксилазы/оксигеназы.

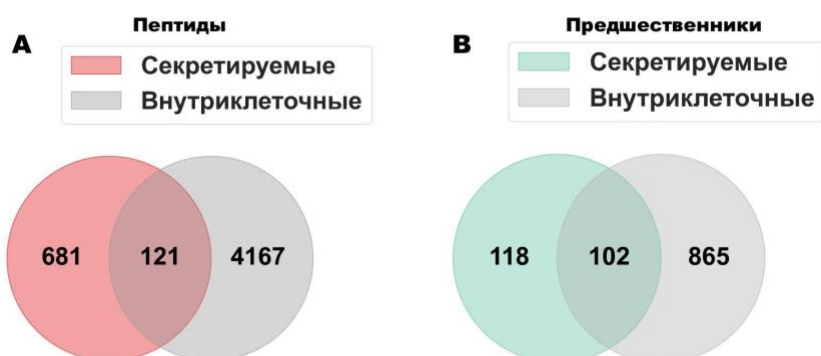
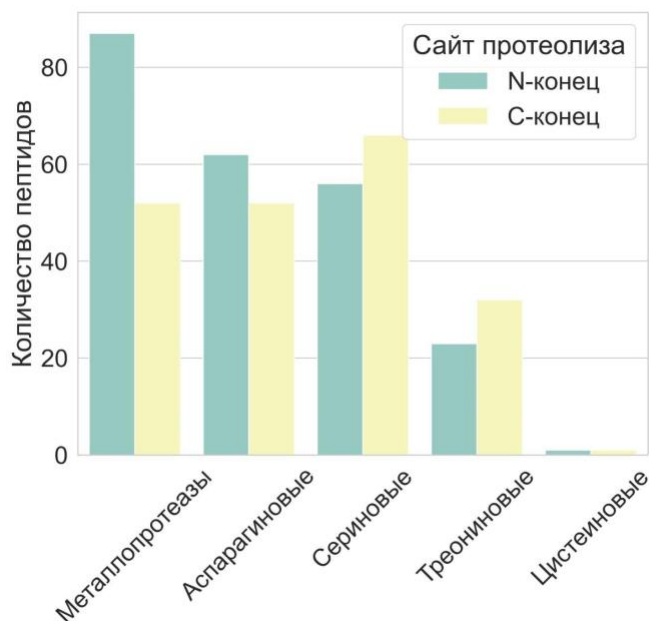


Рисунок 4. Диаграмма Венна, показывающая сравнение эндогенных пептидов (А) и их предшественников (В) между внутриклеточными и секретируемыми пептидными пулами (n = 3 независимых биологических повтора)

Далее мы *in silico* оценили типы протеаз, которые могут отвечать за формирование пептидных пулов в секрете растений. Используя программу ExPASy PeptideCutter, мы обнаружили, что примерно 50% N- или C-концов пептидов секретома содержат потенциальные сайты расщепления протеазами с трипсиноподобной или химотрипсиноподобной активностями. Также, согласно нашему анализу, существенный вклад в формирование пептидных пулов секретома вносят сериновые протеазы. Далее, мы использовали матрицы специфичности протеаз из базы данных MEROPS для *P. patens* и картировали потенциальные сайты расщепления на N- и C-концы идентифицированных эндогенных пептидов. Согласно этому анализу, сайты расщепления для всех пяти типов протеаз: сериновых (семейства S8, S9, S10, S14 и S26), аспарагиновых (семейство A1), металлопротеаз (семейства M1, M3, M14, M16, M24, M28 и M79), треонин- (семейство T1) и цистеиновых протеаз (семейства C13 и C65) были обнаружены в нашем наборе данных (**Рис. 5**). Наши результаты согласуются с недавними исследованиями пептидома млекопитающих, согласно которым протеазы серинового типа и металлопротеиназы являются наиболее распространенными классами протеаз, участвующих в формировании пептидомов жидкостей организма. Такие общие особенности пептидомов млекопитающих и растений могут указывать на определенный консервативный механизм их образования, детали которого еще только предстоит изучить. Мы предполагаем существование нескольких механизмов образования секретируемых пептидных пулов в растениях. Например, небольшая часть пептидома транспортируется транспортерами АТФ-связывающей кассеты (ABC) в эндоплазматический ретикулум (ER) и лизосомы у

животных. Кроме того, белки хлоропластов могут подвергаться гидролизу в специальных



везикулах или транспортироваться в вакуоль. Пептиды далее могут быть экспортированы в апопласт во время слияния вакуолярных пузырьков с плазматической мембраной.

Рисунок 5. Гистограмма, иллюстрирующая сопоставление N- и C-концов идентифицированных пептидов с матрицами специфичности различных типов протеаз из базы MEROPS.

Наконец, функциональные белки и/или специализированные предшественники биоактивных пептидов могут быть секретированы и подвергаться действию различных протеаз во внеклеточном пространстве.

Влияние стрессовых факторов на состав пептидных пулов растений

В целом ряде недавних исследований было показано, что стрессовые факторы влияют на механизмы деградации белков, а значит это может влиять на пептидомы тканей и секретома. Исходя из этого, мы использовали такие воздействия, как повреждение клеточной стенки; обработка стрессовыми гормонами - жасмоновой и салициловой кислотами для того, чтобы оценить их влияние на состав внутриклеточных и секретлируемых пептидных пулов мха *P. patens*.

Обработка ферментным препаратом драйзелазой, представляющим собой смесь природных ферментов из *Basidiomycetes* sp., имеющую ламинариазную, ксиланазную и целлюлазную активности, моделирует повреждение клеточной стенки и приводит к индукции иммунного ответа у растений. Мы далее использовали драйзелазу для получения протопластов *P. patens* из протонемы. Используя МС анализ, мы обнаружили, что пептидом протопластов содержит 20 427 пептидов размером от 6 до 78 а.о., которые являются фрагментами 3648 белков-предшественников и претерпевает значительные изменения в сравнении с пептидомом протонемы (**Рис. 6**). В пептидоме протопластов мы идентифицировали большое количество пептидов, полученных из хлоропластных белков, таких как большая

субъединица рибулозо-1,5-бисфосфат карбоксилазы, субъединица П-2 реакционного центра фотосистемы I и ряд других. Все эти данные указывают на интенсивную деградацию хлоропластных белков в протопластах.

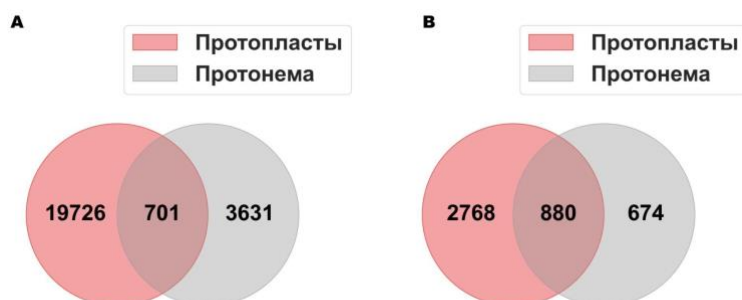


Рисунок 6. Диаграммы Венна, показывающие пересечение идентифицированных в протопластах и протонеме мха **А** - пептидов, **В** - белков-предшественников.

Однако в пептидоме нативных хлоропластов, выделенных из протопластов, мы идентифицировали только 82 уникальных пептида, которые были фрагментами 21 белка-предшественника, что резко контрастирует с результатами анализа клеточного пептидома. Эти данные подтверждают, что основные белки хлоропластов подвергаются гидролизу вне интактных хлоропластов и наблюдаемые нами явления связаны с экспортом предшественников из этих органелл.

Обработка протонемы 0,025% раствором драйзелазы, который не приводит к образованию протопластов, но вызывает повреждения клеточной стенки также увеличила количество нативных пептидов во внутриклеточном пептидоме примерно в 2,5 раза по сравнению с контролем. Этот результат свидетельствует о том, что повреждающее воздействие на клетку усиливает процессы протеолиза и генерацию большого количества эндогенных пептидов.

Используя транскрипционное профилирование, мы дополнительно проанализировали дифференциально экспрессируемые между протонемой и протопластами гены, чтобы выяснить, какие молекулярные механизмы могут быть ответственны за значительные изменения, наблюдаемые в пептидоме протопластов. Всего было идентифицировано 1936 генов, уровень транскрипции которых вырос в протопластах более чем в 4 раза ($P < 0.05$). Анализ терминов онтологии генов показал, что в протопластах увеличиваются уровни транскрипции генов, участвующих в ответах на различные стрессовые факторы. Например, протопласты имели повышенный уровень транскрипции генов алленоксидциклазы 2 (AOC2, Pp1s41_314V6), алленоксидциклазы 4 (AOC4, Pp1s135_3V6) и алленоксидсинтазы (AOS, Pp1s97_112V6), которые участвуют в биосинтезе стрессового гормона - жасмоновой кислоты. Поскольку синтез оксипинонов в растениях напрямую связан с ответом на

биотический стресс, данные транскрипционного профилирования представляют собой дополнительное доказательство того, что перестройка пептидома в протопластах может быть связана с индукцией иммунного ответа в ответ на обработку драйзелазой. Кроме того, наблюдалось значительное увеличение транскрипции генов Pp1s166_98V6 (активирующий белок протеасомы 200), Pp1s78_186V6 (субтилизин-подобная сериновая протеаза 2), Pp1s112_240V6 (субтилизин-подобная сериновая протеаза 3), Pp1s39_149V6 (гомолог митохондриальной протеазы FtsH3). Мы также отметили увеличение транскрипции Pp1s1_60V6 (AtTAP2), растительного аналога транспортера TAP, который переносит пептиды, генерируемые протеасомным комплексом, в эндоплазматический ретикулум, откуда они потом переносятся на молекулы главного комплекса гистосовместимости MHC класса I.

Ранее было показано, что *P. patens* отвечает на обработку стрессовым гормоном метилжасмонатом (МЖ) ингибированием роста и индукцией генов, связанных с патогенезом. Мы далее использовали MS анализ для идентификации внутриклеточных и внеклеточных эндогенных пептидов в контрольных и обработанных 0.4 мМ раствором метилжасмоната растениях *P. patens*. Принимая во внимание гетерогенность пулов пептидов, для дальнейшего анализа были отобраны только пептиды, идентифицированные как минимум в двух биологических повторах. Всего было идентифицировано 4533 эндогенных пептида, относящихся к 1000 белкам-предшественникам (Рис. 7).

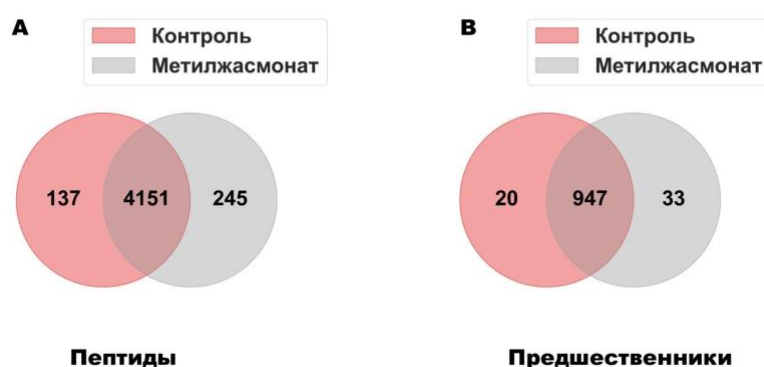


Рисунок 7. Диаграмма Венна, показывающая сравнение идентифицированных эндогенных пептидов (А) и их предшественников (В) с контрольными клетками протонемы и клетками протонемы, обработанными метилжасмонатом (n = 3 независимых биологических повтора).

Хотя значительная часть внутриклеточных пептидных пулов была идентична с контрольными и обработанными растениями, метилжасмонат индуцировал образование 245 (около 5,5% от общего количества пептидов) новых пептидов; 67 (27%) этих пептидов были получены из новых белков-предшественников в обработанных клетках. Таким образом, обработка метилжасмонатом привела к увеличению количества уникальных

пептидов в протонеме мха. Эти пептиды высвобождаются в результате деградации новых белков (протеолиза, индуцированного фитогормонами) и изменения характера деградации существующих предшественников. Последний процесс приводит к образованию "пептидных лестниц" из схожих пептидов у контрольных и обработанных метилжасмонатом образцов.

В отличие от внутриклеточных пептидных пулов, обработка метилжасмонатом не вызвала серьезных изменений в секретируемом пептидоме. Однако обработка метилжасмонатом приводила к образованию пептидов из различных групп белков, участвующих в стрессовом ответе. Например, мы обнаружили фрагменты белка PR10 (белка, связанного с патогенезом), белка теплового шока 70 кДа и фактора транскрипции MYC только в секретоме обработанных МЖ растений. Можно предположить, что пептиды от таких белков являются потенциальными индукторами или регуляторами иммунного ответа, как это было показано для фрагмента белка PR1 у томата.

Мы также изучили изменения в пептидоме протонемы мха, обработанной еще одним известным стрессовым гормоном - салициловой кислотой. Обработка протонемы 0.4 мМ салициловой кислотой (СА) вызывала более серьезные изменения в секретируемом пептидоме в сравнении с метилжасмонатом. Мы идентифицировали 830 эндогенных пептидов от 248 белков-предшественников после обработки СА. В пептидоме растений, обработанных СА мы наблюдали деградацию 134 новых белков-предшественников. Мы показали, что в контрольных и обработанных СА образцах десятки уникальных пептидов содержали сайты расщепления β 1c субъединицы протеасомы (треониновой протеазы). Пептиды, которые потенциально могли генерироваться протеасомами, происходили из белков-предшественников, участвующих в метаболизме углеводов, протеолизе и процессах окисления/восстановления. Однако мы не обнаружили значимых изменений распределения сайтов протеолиза в идентифицированных пептидных пулах после обработки СА.

Мы предложили несколько гипотез для объяснения резкого увеличения количества специфических пептидов, которые встречаются только в стрессовых пептидомах: 1) иммунный ответ, приводящий к специфической деградации клеточных белков; 2) увеличение скорости деградации белков, вызванное стрессом; и 3) деградация белков вследствие запуска механизмов запрограммированной гибели клеток. Рост числа пептидов в растениях, подвергшихся биотическому стрессу, также может быть связан с защитной или сигнальной ролью таких пептидов. Например, мы оценили антимикробный потенциал идентифицированных в протопластах пептидов и выявили 81 эндогенный пептид с высокой потенциальной антимикробной активностью. Индукция потенциальных антимикробных

пептидов в стрессовых условиях проливает свет на возможную функцию пептидных пулов, формирующихся в таких условиях у растений.

Сравнение пептидомов различных организмов

На сегодняшний день тысячи пептидов, различающихся по происхождению, количеству и возможным функциям, были идентифицированы в тканях и биологических жидкостях позвоночных, насекомых, грибов, растений и бактерий. Мы далее решили выявить некоторые общие тенденции в составе и физико-химических свойствах пептидных пулов разных организмов. Для этого мы сравнили полученные нами внутриклеточные и секретируемые пептидомы *P. patens* с наборами эндогенных пептидов, идентифицированных при анализе бактерий, дрожжей, хлопка, плазмы крови человека и пептидов, связанных с протеосомой у человека. В отобранных нами для анализа наборах данных средняя длина идентифицированных пептидов колеблется от 11 до 18 аминокислотных остатков и остается практически одинаковой для внутри- и внеклеточных пептидомов разных организмов.

Для того, чтобы оценить схожесть механизмов генерации пептидных пулов, мы сравнили С-концевые аминокислоты (P1 позиция сайта протеолиза) пептидов у разных организмов. Наш анализ показал, что лизин (K) и аргинин (R) были одними из наиболее представленных аминокислот в разных наборах данных (Рис. 4.8).



Рисунок 8. Круговые диаграммы, показывающие 10 наиболее представленных аминокислот на С-концах внутриклеточных пептидов из мха, человека, дрожжей, хлопка и внеклеточных пептидов из плазмы крови человека, мха и бактерий.

Это может указывать на то, что протеазы с трипсиноподобной активностью вносят значительный вклад в формирование нативных пептидомов у разных организмов. В целом,

вариабельность С-концевых аминокислот в разных наборах пептидомных данных может быть вызвана неспецифическим протеолизом, действием экзопептидаз, техническими особенностями метода выделения или особенностями МС анализа. Согласно целому ряду исследований, большинство известных растительных пептидных гормонов закодировано в С-концевой части белка-предшественника. В более ранних исследованиях нативного пептидома животных было показано, что идентифицированные пептиды неравномерно распределены по длине белка. Например, большая часть пептидома мозга мыши это N- или С-концевые фрагменты соответствующих предшественников. Чтобы определить, существует ли такая тенденция в пептидомах различных организмов, мы проанализировали распределение идентифицированных пептидов по длине белков предшественников. Для этого, мы рассчитали частоту встречаемости каждой аминокислоты из белков-предшественников в пептидоме и представили эти данные в виде графиков плотности, показывающих распределение вероятностей этих частот (**Рис. 9**).



Рисунок 9. Распределение идентифицированных пептидов по длине предшественников в наборах пептидомных данных мха, человека, дрожжей, хлопка и бактерий. Позиция каждого идентифицированного пептида была нормализована к длине белка.

Действительно, сравнение различных наборов данных показало, что пептиды,

высвобождаемые с С- или N-конца предшественника, как правило, более представлены во внутриклеточных и внеклеточных пептидомах, чем мы могли бы ожидать в случае случайного расщепления белков. В целом, закономерности деградации предшественников из разных наборов данных схожи, с преобладанием С-концевых пептидов во внутриклеточных и N-концевых пептидов во внеклеточных пептидомах. Таким образом, недавний прогресс в масс-спектрометрическом анализе существенно расширил наши знания/представления о составе внутри- и внеклеточных пептидных пулов разных организмов. Однако необходим дальнейший прогресс в области пептидомного анализа,

связанный с улучшением детекции пептидов и уменьшением количества артефактов во время пробоподготовки.

Поиск коротких открытых рамок считывания (кОРС) с высоким кодирующим потенциалом

Помимо специфического и неспецифического протеолиза белков-предшественников, еще одним малоизученным источником нативных пептидов являются короткие открытые рамки считывания (кОРС, < 100 кодонов). Пептиды и микробелки, которые образуются при трансляции кОРС, а также продукты их протеолитической деградации могут существенно влиять на пептидомный ландшафт клетки и секретома. Однако данный источник пептидов ранее системно не анализировался у растений, и вклад кОРС в формирование пептидомов растений неизвестен. Исходя из этого, мы далее провели всесторонний анализ кОРС и продуктов их трансляции у растений, используя *P. patens* как модель, обладающую целым рядом преимуществ для такого рода исследований.

Существует несколько стратегий аннотации кодирующих кОРС в геноме исследуемого организма и создание совмещенной (белки+кОРС) базы данных последовательностей для последующего масс-спектрометрического анализа. Самым простым подходом является трансляция генома *in silico* в шести рамках считывания. Однако такой подход применим только к организмам с относительно небольшим и безинтронным геномом, поскольку полученная база может быть в десятки раз больше соответствующей белковой базы последовательностей, что приводит к большому количеству ложных идентификаций и экспоненциальному увеличению времени анализа масс-спектрометрических данных.

Для поиска потенциально транслирующихся коротких рамок считывания в геноме *P. patens* мы выбрали подход с предсказанием всех кОРС и последующим отбором тех из них, которые обладают кодирующим потенциалом. Для такого анализа мы выбрали разработанный ранее для анализа растительных геномов биоинформатический инструмент sORF finder, который ищет в геноме безинтронные кОРС, начинающихся со стартового кодона AUG и длиной от 30 до 300 нуклеотидов. Данный алгоритм основан на поиске различий в составе к-меров между кодирующими и не кодирующими последовательностями и последующим вычислением кодирующего потенциала кОРС и классификацией с использованием теоремы Байеса. Мы предсказали 638,439 кОРС с высоким кодирующим потенциалом (всего идентифицировано 6,706,696 коротких рамок), которые начинались с кодона AUG и были расположены во всех областях генома *P. patens* с помощью следующих

параметров: кодирующие последовательности - белок-кодирующие, не пересекающиеся с интронами, некодирующие - интроны из аннотации Phytozome V12, $P(\text{CDS})=0.3$. Длина кОПС с высоким кодирующим потенциалом значительно отличалась от тех, которые мы могли бы ожидать при случайной генерации кОПС в геноме *P. patens* (критерий «Манна-Уитни», $P < 10^{-15}$).

В результате предсказания потенциальных пептид-кодирующих коротких рамок считывания с помощью программы sORF finder в геноме *P. patens* были идентифицированы несколько сотен тысяч кОПС. Мы далее отбросили кОПС которые 1) являлись частью аннотированных белков; 2) не транскрибировались; 3) находились в повторяющихся регионах. В результате, были отобраны 70,095 уникальных кОПС, расположенных на 33,981 транскрипте различного типа, включая мРНК и длнкРНК из двух баз данных - SANTIAdb и GreeNC.

Мы классифицировали кОПС на основе их положения на транскриптах: 63,109 “генных кОПС” (расположены на транскриптах мРНК); 1241 “межгенных” кОПС (расположены на транскриптах из нашего набора данных и не аннотированных в текущей версии Phytozome V12); и 5745 “длнкРНК-кОПС” (расположены на длнкРНК из баз SANTIAdb, GreeNC, или предсказанных в нашем наборе транскриптов. “Генные кОПС” были классифицированы как 11 998 “upstream” кОПС, которые расположены в 5'-нетранслируемых регионах мРНК; 9443 “downstream” кОПС, которые расположены в 3'-нетранслируемых регионах мРНК; 36,732 “CDS”-кОПС, которые перекрываются с основной белок-кодирующей последовательностью мРНК, но со сдвигом рамки считывания +1 или +2; 3485 “interlaced” кОПС, которые перекрываются как с белок-кодирующей последовательностью, так и с нетранслируемыми регионами мРНК (**Рис. 10**). Ожидаемо, что полученный набор кОПС был обогащен “CDS”-кОПС (52%, точный тест Фишера, значение $P < 10^{-15}$), тогда как “upstream”, “downstream” и “interlaced” кОПС были недопредставлены (точный тест Фишера, $P < 10^{-15}$) по сравнению с набором случайных участков экзонов генов *P. patens*, который использовался в качестве отрицательного контроля. Длина кОПС в нашем наборе данных варьировала от 30 до 300 нуклеотидов. В среднем, “CDS”-кОПС (средний размер 22 кодона) были значительно короче, чем “upstream” кОПС (средний размер 35 кодонов; критерий Манна-Уитни $P < 10^{-15}$) и “downstream” кОПС (средняя длина 32 кодона, критерий Манна-Уитни $P < 10^{-15}$). Средний размер “interlaced” кОПС составляет 49 кодонов, и этот класс кОПС значительно длиннее, чем другие “генные” кОПС (критерий Манна-Уитни $P = 0,0021$).

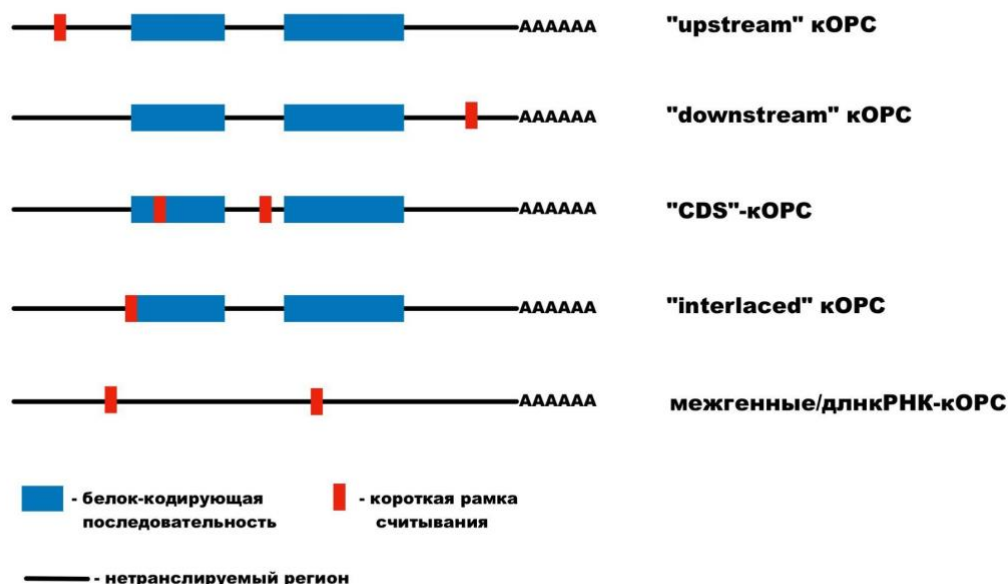


Рисунок 10. Предлагаемая классификация кОРС в соответствии с типом транскрипта и положением короткой рамки считывания: “upstream” кОРС (uORFs) и “downstream” кОРС (dORFs) из нетранслируемых областей (UTR) мРНК; “CDS”-кОРС, которые перекрываются с белок-кодирующими последовательностями, но расположены в альтернативных рамках считывания или являются короткими изоформами, образующимися в результате альтернативного сплайсинга; “interlaced” кОРС, которые перекрывают как кодирующую последовательность, так и UTR на одном и том же транскрипте; длинкРНК-кОРС и межгенные кОРС, которые расположены на транскриптах аннотированных как длинные некодирующие РНК или не содержащие белок-кодирующих последовательностей.

Далее мы проанализировали консенсусные последовательности Козак, окружающие стартовые кодоны транскрибирующихся кОРС. Консенсусная последовательность Козака играет важную роль в инициации трансляции эукариотических мРНК. В зависимости от наличия пурина в положении -3 и гуанина в положении +4 (где +1 означает “А” в кодоне “AUG”), контекст считается “сильным” (присутствуют оба варианта), “средним” (присутствует один вариант) или “слабым” (ни один не присутствует). Согласно этой оценке, 41,816 ($\approx 60\%$) предсказанных нами кОРС были в “сильном” и “среднем” контексте. Эти значения были существенно меньше, чем у аннотированных белок-кодирующих последовательностей из аннотации Phytozome V12 (87%, точный критерий Фишера $P < 2,2 \times 10^{-16}$).

Анализ консервативности “генных” коротких открытых рамок считывания

Считается, что консервативность кодирующей последовательности является маркером ее функциональности. Исходя из этого, мы далее оценили консервативность и степень эволюционного давления на “генные” кОРС из нашего набора данных используя подход,

основанный на поиске ортологичных последовательностей с помощью алгоритма TBLASTN ($E < 0,00001$, query coverage $> 60\%$). Для этого анализа были использованы транскриптомы десяти видов растений, такие как *Ceratodon purpureus*, *Sphagnum fallax*, *Marchantia polymorpha*, *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, *Selaginella moellendorffii*, *Spirodela polyrhiza*, *Volvox carteri*, *Chlamydomonas reinhardtii*. Используя указанные параметры фильтрации, мы идентифицировали 4848 консервативных, т.е. таких, для которых гомологичные последовательности были обнаружены как минимум в одном виде, “генных” коротких открытых рамок считывания. Ожидается, что наиболее представленными в этом наборе были “CDS”-кОПС - 3706 консервативных кОПС. Количество других консервативных “генных” кОПС составило: “upstream” кОПС - 318, “downstream” кОПС - 434, “interlaced” кОПС - 390. Оказалось, что “CDS”-кОПС были значительно представлены в наборе консервативных кОПС у двух ближайших видов мха - *C. purpureus* и *S. fallax* (точный тест Фишера, $P < 10^{-10}$). Мы также обнаружили, что доля консервативных “upstream” кОПС и “downstream” кОПС у более отдаленных видов становится выше по сравнению с исходным набором данных, что приводит к их значительной представленности в наборе консервативных кОПС (точный критерий Фишера, $P < 0.0005$), в то время как доля “CDS”-кОПС снижается (Рис. 4.11А).

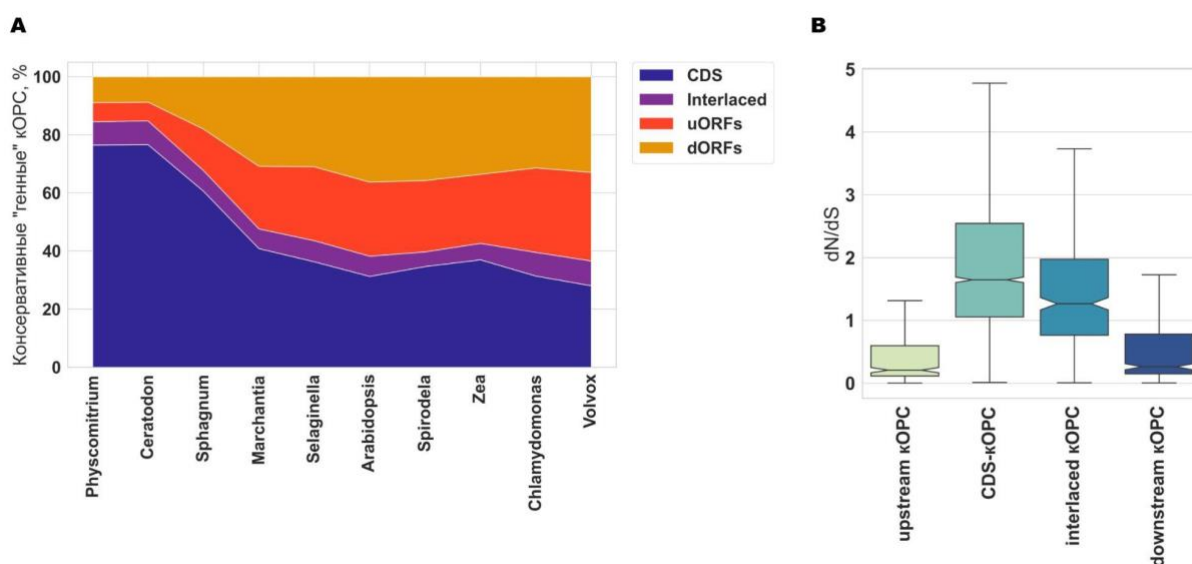


Рисунок 11. А - Доля каждого типа “генных” кОПС, имеющих возможных ортологов у десяти видов растений. В - Распределение соотношений dN/dS в разных типах “генных” кОПС, определенных на основе выравнивания с *C. purpureus*; $P < 0,0001$ по критерию суммы рангов Крускала–Уоллиса.

Мы также сравнили консервативность “генных” кОПС и 158 маленьких (<100 а.о.) белков без интронов из аннотации генома *P. patens*. Доли консервативных кОПС и этих белков

значительно различались (7,2% кОПС vs 86% маленьких белков), указывая на низкую консервативность последовательностей кОПС, идентифицированных в геноме *P. patens*.

Далее мы провели схожий анализ консервативности для случайных фрагментов экзонов, извлеченных из нетранслируемых регионов мРНК или областей белок-кодирующих последовательностей. В этом случае, мы получили прямо противоположные результаты - случайные участки белок-кодирующих последовательностей были наиболее консервативными у всех 10 растительных видов.

Для того, чтобы оценить давление эволюционного отбора на “генные” кОПС мы рассчитали соотношение синонимичных к несинонимичным заменам (показатель dN/dS), используя выравнивания с ближайшим видом мха *Ceratodon purpureus*. Этот анализ показал, что “CDS”-кОПС и “interlaced” кОПС имеют наибольшую долю коротких рамок с $dN/dS > 1$, что подразумевает положительный отбор для данных последовательностей (**Рис. 4.11В**).

В то же время, около 50% “upstream” кОПС и 40% “downstream” кОПС подвергались действию стабилизирующего отбора ($dN/dS \leq 0.2$). Таким образом, наш анализ показал, что действие на кОПС эволюционного отбора отличается от случайно выбранных последовательностей экзонов и зависит от местоположения короткой рамки с большими шансами быть зафиксированным у “upstream” кОПС и “downstream” кОПС. Также, согласно полученным нами результатам, значительная часть “CDS”-кОПС подвергается положительному отбору на аминокислотном уровне. По-видимому, с этим связано значительное снижение идентифицированных гомологичных последовательностей этого типа “генных” кОПС у отдаленных видов. Можно предположить, что данный процесс не является стохастическим, поскольку результаты, полученные для случайно выбранных экзонных фрагментов, прямо противоположны.

Идентификация пептидов, кодируемых короткими открытыми рамками считывания

На сегодняшний момент наиболее весомым доказательством трансляции кОПС считается идентификация соответствующих пептидов при помощи масс-спектрометрического (МС) анализа. Получение доказательств трансляции кОПС является важным шагом на пути к идентификации функциональных коротких открытых рамок считывания. Мы решили выяснить какое количество предсказанных нами кОПС можно обнаружить в пептидоме клеток *P. patens*. Принимая во внимание недостатки стандартного протеомного анализа для выделения и идентификации небольших белков или пептидов, мы использовали для поиска продуктов трансляции коротких открытых рамок считывания два набора данных: “пептидомный” — эндогенные пептиды, извлеченные из трех типов клеток мха - гаметофоров, протонемы и протопластов; и “протеомный” — триптические пептиды,

полученные при стандартном протеомном анализе тканей *P. patens*. Для анализа мы использовали поисковую программу MaxQuant и базу данных, содержащую 32,926 аннотированных белков *P. patens*, 85 хлоропластных белков, 42 митохондриальных белка и 70,095 кОПС. Это привело к идентификации 1177 пар теоретический пептид — настоящий спектр (peptide spectrum matches (PSMs)), идентифицированных с уровнем ложноположительных результатов (FDR) равным 1%, которые соответствовали 296 пептидам, кодируемым разными типами кОПС в наборе пептидомных данных, и 920 PSMs, соответствующим 532 кОПС в "протеомном" наборе данных.

Известно, что идентификация неканоничных продуктов трансляции при анализе данных масс-спектрометрических экспериментов требует особых подходов. Для того, чтобы повысить достоверность полученных результатов и выбрать кОПС для дальнейшего функционального анализа, мы отфильтровали полученные идентификации, используя очень жесткие критерии: минимальное значение апостериорной ошибки меньше 0,01 и Andromeda score выше 60. В результате мы получили набор “достоверно” транслирующихся 46 кОПС: 17 в гаметофорах, 29 в протонеме и 14 в протопластах (**Рис. 12А**). Примерно у 63% достоверно транслирующихся кОПС (29 кОПС) стартовый кодон был в «сильном» и «среднем» контексте Козак, что аналогично результатам, полученным для всех предсказанных кОПС (~60%). Этот результат предполагает, что инициация трансляции может немного различаться для коротких рамок и белков. Длина транслирующихся кОПС варьировалась от 11 до 100 аминокислот, и они были значимо длиннее нетранслируемых (критерий “Манна-Уитни” $P=5.33 \times 10^{-6}$). Длина транслирующихся “upstream” кОПС значительно отличалась от “CDS”-кОПС (критерий Манна-Уитни $P = 0,003$), но не отличалась от “downstream” кОПС (критерий “Манна-Уитни” $P=0,11$) (**Рис. 12В**). Согласно полученным данным, “генные” кОПС были наиболее представленной группой “достоверно” транслирующихся кОПС. Например, “CDS”-кОПС составляли около 41,3% (19 кОПС) всех рамок (**Рис. 12С**), “upstream” кОПС (6 кОПС, 13%) и “downstream” кОПС (9 кОПС, 19,6%). В целом, мы обнаружили, что только пять транслирующихся кОПС являются консервативными (11% от всех) и четыре из них находились под действием негативного отбора ($dN/dS < 0.5$). МС анализ выявил совместную трансляцию семи “достоверно” транслирующихся кОПС (пять “CDS”-кОПС) и белок-кодирующей рамки, что указывает на то, что эти транскрипты являются полицистронными. В некоторых случаях мы детектировали продукты трансляции кОПС и белок-кодирующей рамки одновременно в клетках одного и того же типа (например, Pp3c18_sORF1035), в то время как другие демонстрировали контрастные паттерны трансляции (например, Pp3c1_sORF1909). Эти

результаты указывают на то, что трансляция кОПС и белков, расположенных вместе в одном и том же локусе, может быть тканеспецифичной.

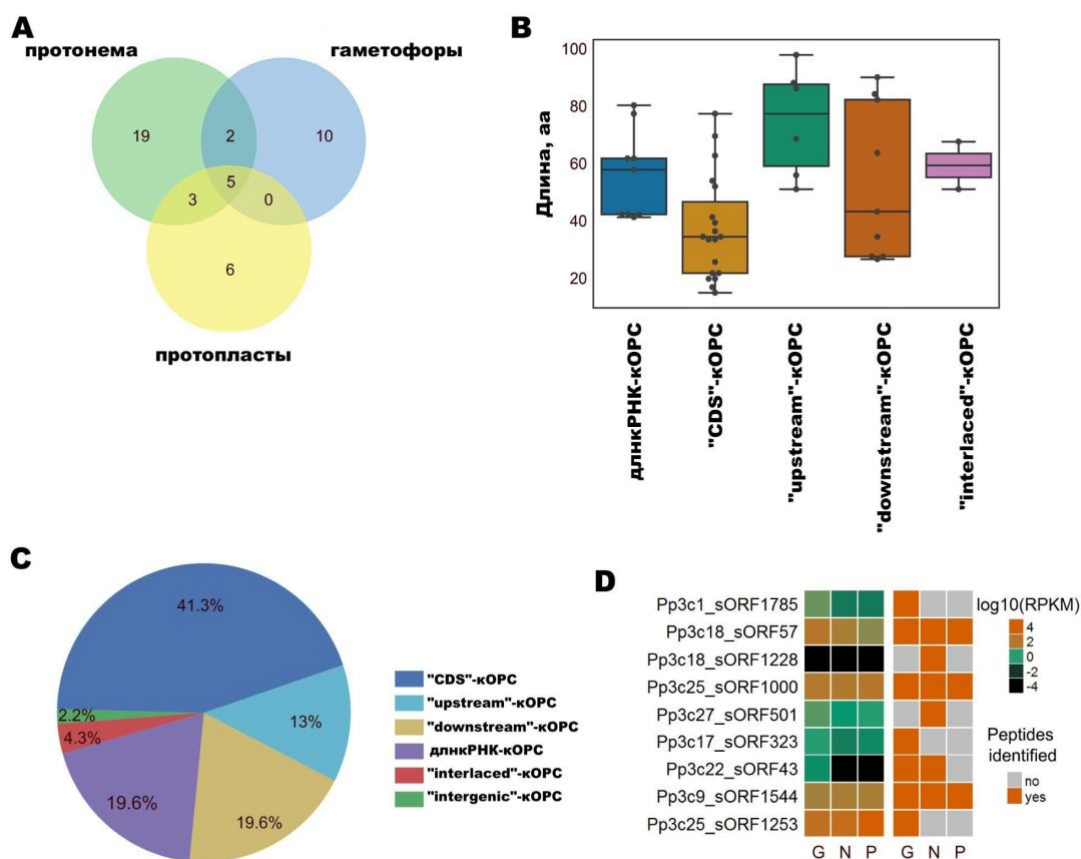


Рисунок 12. **A** - Диаграмма Венна, показывающая распределение идентифицированных транслируемых “достоверных” кОПС среди трех типов клеток мха. **B** - Распределение длины различных групп транслирующихся “достоверных” кОПС. **C** - Диаграмма, показывающая процентное соотношение разных типов транслирующихся кОПС. **D** - Тепловая карта, показывающая уровни экспрессии (\log_{10} [RPKM]) для длнкРНК (слева), несущих кОПС (длнкРНК-кОПС), и бинарная тепловая карта, показывающая доказательства трансляции (определяется по тому, был ли пептид идентифицирован [коричневый] или нет [серый] в МС данных) для соответствующих длнкРНК-кОПС (справа) в трех тканях мха: гаметофоры (G), протонема (N) и протопласты (P).

Кроме того, была обнаружена “достоверная” трансляция девяти кОПС, расположенных на длнкРНК. Уровень транскрипции некоторых днРНК (согласно нашим предыдущим данным и экспрессионному атласу Phytozome V12) и признаки трансляции соответствующих длнкРНК-кОПС представлены на **Рис. 12D**. Три из этих “длнкРНК-кОПС” - Pp3c18_sORF57 (40 а.о.), Pp3c9_sORF1544 (41 а.о.) и Pp3c25_sORF1000 (61 а.о.), были общими для всех трех типов клеток и подтверждены несколькими уникальными эндогенными пептидами.

Мы пришли к выводу, что анализ пулов эндогенных пептидов вместо триптических пептидов имеет целый ряд достоинств с точки зрения идентификации продуктов трансляции кОПС: (1) стандартные протеомные подходы не годятся для выделения и

анализа небольших и низкопредставленных эндогенных пептидных молекул; и (2) пептиды от кОРС короче стандартных белков, и маловероятно, что в одном протеомном эксперименте будет обнаружено более одного триптического фрагмента для одной короткой рамки. Более того, пептидомные подходы теоретически можно использовать для идентификации полноразмерных продуктов трансляции кОРС в клетке. Интересно отметить, что мы не наблюдаем какого-либо значительного перекрытия между кОРС, идентифицированными с помощью протеомного и пептидомного подходов. Таким образом, эти результаты демонстрируют преимущества использования взаимодополняющих подходов для более полной идентификации продуктов трансляции кОРС.

Предсказание и классификация коротких рамок считывания на длинных некодирующих РНК

Используя программу sORF finder, мы обнаружили 5745 транскрибируемых “длнкРНК-кОРС”, которые обладали высоким кодирующим потенциалом, а 9 из них относились к “достоверно” транслирующим коротким рамкам. Однако наш подход, основанный только на анализе кОРС с предсказанным высоким кодирующим потенциалом, не позволяет системно оценить долю длнкРНК, которая является источником пептидов или микробелков. Между тем, показано, что у человека транскрипты аннотированные как длнкРНК, кодируют функциональные пептиды, играющие ключевые роли во многих клеточных процессах. Следовательно, поиск функциональных “длнкРНК-кОРС” у растений представляет значительный интерес. Кроме того, изучение эволюции длнкРНК может пролить свет на появление *de novo* новых белок-кодирующих генов. Поскольку случайная трансляция кОРС, расположенных на длнкРНК, потенциально может быть вредной для клетки, было высказано предположение, что первичный отбор образующихся пептидов направлен против способных агрегировать друг с другом (гипотеза «не навреди»). Еще одна гипотеза образования новых генов (“ТМ-first”) предполагает, что гидрофобные пептиды, которые могут образовываться при трансляции некодирующих АТ-богатых последовательностей генома, эволюционируют в малые трансмембранные (ТМ) белки. Поскольку системный анализ кОРС, кодируемых длинными некодирующими РНК, ранее не проводился у растений, мы дополнительно провели всесторонний анализ этого типа кОРС, используя длнкРНК *P. patens*.

Для этого мы объединили в один набор аннотированные длнкРНК *P. patens*, которые были опубликованы ранее: 1498 длнкРНК из базы данных SNTATdb 2.0, 9416 длнкРНК из GreeNC, 3018 длнкРНК из аннотации NCBI (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Physcomitrella_patens/100), 1512

днкРНК из статьи Lang et al., 2018 г. и 4648 днкРНК из статьи Simopoulos et al., 2019 г.. Транскрипты длинных некодирующих РНК из разных источников были картированы на геном *P. patens* v3.3 и объединены в 9168 локусов. В качестве контрольного набора мы использовали 16,178 транскриптов мРНК *P. patens*, кодирующих аннотированные функциональные белки из Phytozome v12. Этот набор мРНК также включал 252 мРНК, кодирующих белки размером менее 100 а.о. («малые ОРС»), которые содержат функциональные домены (средний размер = 81 а.о.). Длины и ГЦ состав транскриптов из нашего набора днкРНК значительно отличались от мРНК (критерий Манна-Уитни, $P < 10^{-15}$). Кроме того, днкРНК содержали значительно меньше экзонов (медиана=1), чем мРНК (медиана=6; критерий Манна-Уитни, $P < 0,000001$).

Далее, используя транскрипты днкРНК, мы предсказали на них все короткие открытые рамки считывания, сформировав два набора кОРС: начинающиеся с кодона «AUG» (90,057 коротких рамок) и альтернативных кодонов «UUG» и «CUG» (85,215 коротких рамок). Затем два набора предсказанных кОРС были объединены и отфильтрованы. Во-первых, были отброшены кОРС, которые имели значительное сходство с аннотированными белками более 200 а.о. из Phytozome V12, а также «малыми ОРС» из нашего контрольного набора мРНК (BLASTP, $E < 0,00001$; процент идентичности $\geq 80\%$); во-вторых были отфильтрованы кОРС, которые на любой из цепей перекрывались более чем на 50% с белок-кодирующими последовательностями функциональных генов. После проведенной фильтрации полученный набор кОРС был разбит на несколько типов в соответствии с аннотацией соответствующих транскриптов (**Рис. 13А**). Короткие рамки считывания, которые продемонстрировали неполное сходство с аннотированными белками *P. patens* (BLASTP, значение $E < 0,00001$), были обозначены как «неклассифицированные». Протеом *P. patens* (Phytozome v12) включает 7028 предсказанных белков размером менее 100 а.о. («малый белок»), многие из которых не имеют функциональной аннотации и могут не являться действительно кодирующей рамкой считывания. В том случае, если транскрипт днкРНК, согласно аннотации Phytozome v12, имел предсказанный «малый белок», соответствующая ему кОРС была обозначена как «малый белок»-кОРС (**Рис. 13А**). Короткие рамки считывания, которые не соответствовали категориям «малый белок» и «неклассифицированные», были обозначены как «новые». Примерно 49% «новых», ~99% «малых белков» и ~76% «неклассифицированных» кОРС начинались с кодона «AUG». Процент кОРС с альтернативными стартовыми кодонами составил ~33% «UUG» и ~18% «CUG» в «новых», ~16% «UUG» и ~8% «CUG» в «неклассифицированных» и ~0,45% «CUG» и ~0,55% «UUG» в «малых белках»-кОРС. Средние размеры предсказанных на длинных

некодирующих РНК кОРС значительно отличались от среднего размера случайных коротких рамок считывания для генома *P. patens* (медиана=13 кодонов, GC = 45,9%; медианный тест Мура $P < 10^{-15}$; **Рис. 13В**).

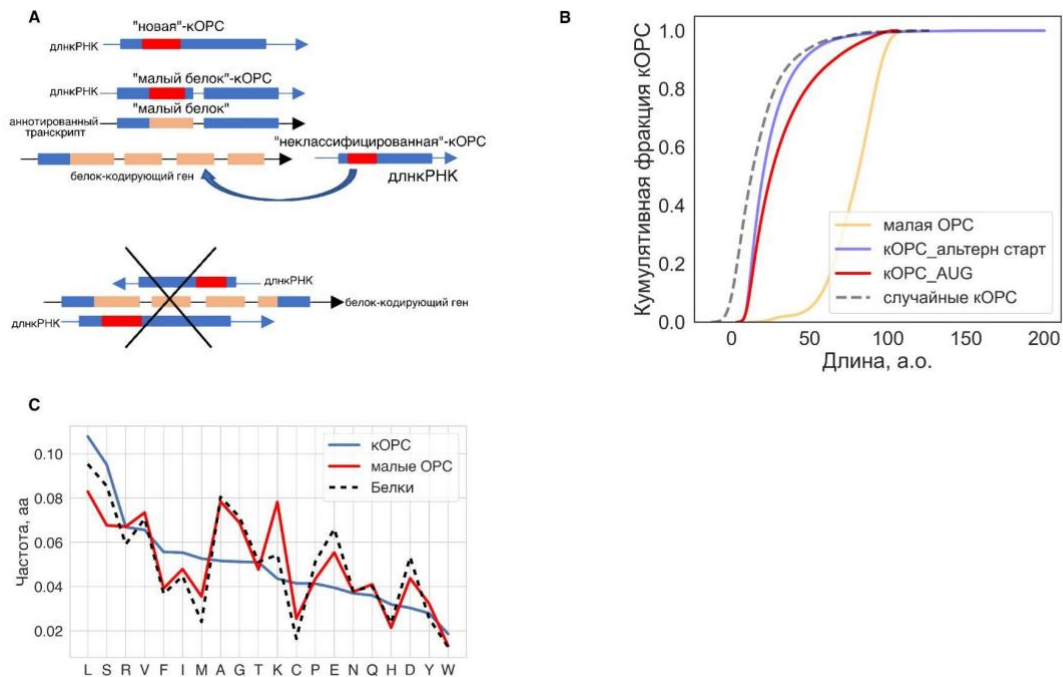


Рисунок 13. А- Предсказание и классификация кОРС; В - Кумулятивное распределение длин различных рамок считывания: «случайные кОРС» - короткие рамки считывания, которые теоретически могут появляться в геноме в зависимости от частоты появления стоп-кодонов; медиана длин кОРС, которые начинались со стартового кодона "AUG" составляли 25 а.о. и 20 а.о. для кОРС, начинающихся с альтернативных стартовых кодонов; ~26% кОРС, которые начинались со стартового кодона "AUG" имели длину более 40 а.о.; С - Сравнение аминокислотного состава функциональных белков (в том числе, малых ОРС, <100 а.о.) и пептидов, кодируемых кОРС, начинающихся с кодона "AUG"; кОРС были обогащены лейцином (хи-квадрат $P < 10^{-15}$), изолейцином (хи-квадрат $P < 10^{-15}$), фенилаланином (хи-квадрат $P < 10^{-15}$) в сравнении с функциональными белками.

Мы также обнаружили, что, в среднем, предсказанные кОРС имели менее богатый ГЦ-состав, чем кОРС, кодирующие белки (критерий Колмогорова-Смирнова, $P < 10^{-20}$). Аминокислотный состав предсказанных нами пептидов, кодируемых кОРС, отличался от функциональных белков, особенно содержанием некоторых гидрофобных аминокислот (**Рис. 13С**). Повышенные частоты метионина и цистеина, а также пониженные частоты аланина, глутамата и аспартата схожи с составом кОРС из длнкРНК млекопитающих.

Анализ консервативности кОРС, локализованных на длинных некодирующих РНК

Ранее было показано, что более 70% длнкРНК животных не имеют гомологов у видов, которые разделились более 50 миллионов лет назад. Таким образом, на нуклеотидном

уровне длнкРНК являются намного менее консервативными, чем мРНК. Используя 478 транскриптомов из проекта “1000 транскриптомов растений”, с помощью алгоритма BLASTN ($E < 0,00001$), мы показали, что около 45% локусов длнкРНК содержали хотя бы одну консервативную область в транскриптомах 41 вида мха. Мы обнаружили, что количество обнаруженных гомологичных последовательностей длнкРНК резко снижается в более отдаленных растительных таксонах, в то время как мРНК были гораздо более консервативны (Рис. 14А). Как и ожидалось, консервативные области в транскриптах мРНК были длиннее, чем в длнкРНК, со средней длиной 753 и 168 нуклеотидов соответственно (Рис. 14В). Таким образом, наши результаты показали, что у растений эволюционная консервативность длнкРНК на нуклеотидном уровне существенно ниже, чем у мРНК.

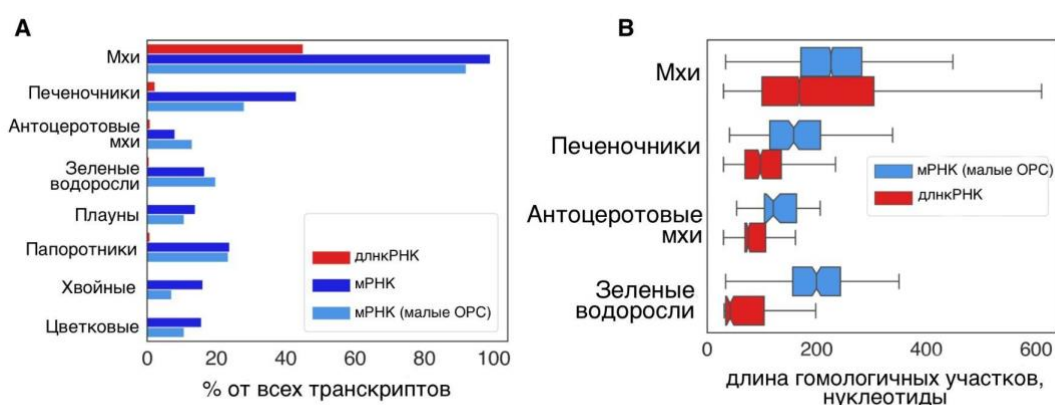


Рисунок 14. А- Сравнение консервативности длнкРНК и мРНК, выполненное с использованием транскриптомов из разных растительных таксонов; В - Распределение длин консервативных областей длнкРНК и мРНК, кодирующих белки меньше 100 а.о.; Показаны медиана, квартили, 5-й и 95-й процентиля.

Затем мы проанализировали консервативность пептидов/микробелков, кодируемых предсказанными нами кОРС, с помощью алгоритма TBLASTN ($E < 0.001$). Мы обнаружили, что 15,167 кОРС, начинающихся с кодона “AUG”, и 9425 кОРС, начинающихся с альтернативных кодонов, показали, по крайней мере, одно совпадение с транскриптами из 41 вида мха. Пептиды, кодируемые консервативными кОРС, были значительно длиннее, чем неконсервативные (критерий Манна-Уитни, $P < 0,00001$). Мы также обнаружили возможные ортологи сотен пептидов, кодируемых кОРС, и в других растительных таксонах (Рис. 15). Таким образом, почти половина локусов нашего набора длнкРНК (4078 из 9168) содержала, по крайней мере, одну консервативную кОРС. Мы использовали алгоритм k-means для кластеризации консервативных кОРС в зависимости от количества возможных ортологов у разных филогенетических линий. Всего мы обнаружили три характерных

кластера: (i) высокий уровень консервативности, ортологи найдены в различных отделах, включая мхи, плауны и папоротники (n=645; кластер 1); (ii) гомологичные последовательности найдены у многих видов мхов, а также идентифицируются в отделах печеночников и антоцеротовых мхов (n=1423; кластер 2); (iii) кОРС консервативны у небольшого количества видов мха (n=22,524; 83% имеют ортологи только у одного вида; кластер 3).

	Зеленые водоросли	Мхи	Печеночники	Антоцеротовые	Плауны	Папоротники	Хвойные	Цветковые	
новые		99	19607	481	213	196	182	145	118
малый белок		523	3067	1005	651	681	670	598	529
неклассифицированные		398	1722	765	499	535	523	465	391
		1020	24396	2251	1363	1412	1375	1208	1038

	Зеленые водоросли	Мхи	Печеночники	Антоцеротовые	Плауны	Папоротники	Хвойные	Цветковые
кОРС	1020	24396	2251	1363	1412	1375	1208	1194
Белки	15653	16162	16095	15802	15879	15921	15776	15577
% от малых ОРС	89	94	93	91	92	91	91	89

Рисунок 15. Количество возможных ортологов кОРС и белков, которые были идентифицированы с помощью алгоритма TBLASTN у разных таксонов растений; «% от малых ОРС» показывает процент малых ОРС, имеющих ортологи в разных классах растений.

В отличие от первого и второго кластеров, «новые» кОРС были значительно шире представлены в третьем, в сравнении с другими типами кОРС (критерий хи-квадрат $P < 10^{-15}$; Рис. 16).

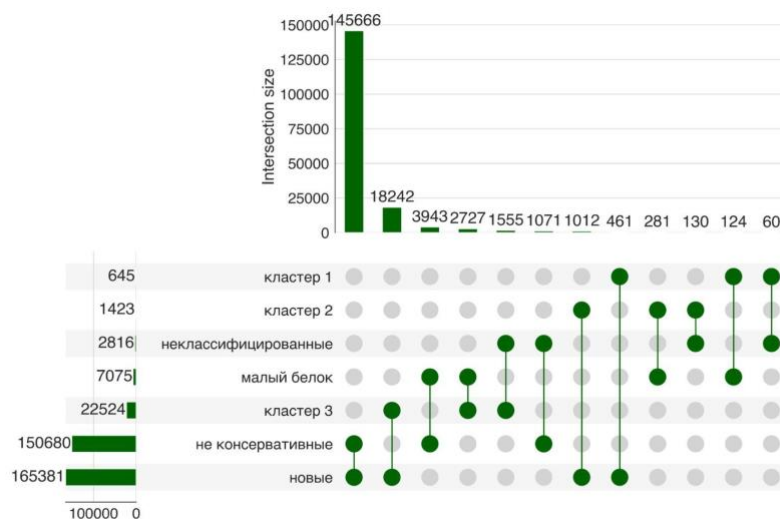


Рисунок 16. График UpSet показывающий пересечение трех консервативных кластеров и разных типов кОРС, локализованных на длнкРНК.

Таким образом, количество возможных ортологов кОРС быстро снижалось при анализе отдаленных таксонов, и, в частности, наблюдалось быстрое исчезновение «новых» кОРС. Как и ожидалось, аннотированные белки (включая малые ОРС) были более консервативны, чем короткие рамки считывания, при этом количество возможных ортологов консервативных белков лишь немного снижалось в отдаленных таксонах растений.

Наш анализ показал, что только относительно небольшая часть длнкРНК является потенциальными псевдогенами или антисмысловыми транскриптами генов, кодирующих белок. В свою очередь, «новые» кОРС, по-видимому, являются основным источником изменчивости и могут участвовать в возникновении генов *de novo* из некодирующих областей. Образующиеся *de novo* «орфанные» гены принадлежат только одному таксону и составляют до 30% генов у некоторых эукариот. Полученные нами результаты согласуются с предыдущими данными, описывающими эволюционную консервативность кОРС в геномах животных.

Анализ скоростей эволюции длинных некодирующих РНК и кОРС

Поскольку особенности эволюции длинных некодирующих РНК и расположенных на них кОРС являются малоизученными, мы проанализировали силу действующего на транскрипты длнкРНК и кодирующие последовательности отбора. Используя двухпараметрическая модель Кимуры (*Kd*), мы показали, что хотя растительные длнкРНК, как правило, гораздо менее консервативны на нуклеотидном уровне, чем мРНК, некоторые из них содержат консервативные области со значениями *Kd*, сравнимыми с транскриптами, кодирующими белки меньше 100 а.о. Это наблюдение согласуется с результатами сравнительного анализа кодирующих и некодирующих РНК у животных. Затем мы оценили скорость эволюции и провели статистические тесты для обнаружения признаков отрицательного отбора в предсказанных кОРС. На основании анализа аминокислотных выравниваний мы пришли к выводу, что значения *Kd* у кОРС были статистически неотличимы от таковых для белок-кодирующих последовательностей, в том числе, малых ОРС (критерий суммы рангов Краскела-Уоллиса, $P = 0,20$; **Рис. 17А**). Мы далее проанализировали скорости эволюции и давление отбора на 4022 кОРС, которые находились в консервативных участках длнкРНК (на основе нуклеотидных выравниваний между *P. patens* и *Physcomitrium* sp.). Такие участки, перекрывающие более 80% длины кОРС, обнаружены примерно в 45% консервативных локусов длнкРНК. Для того, чтобы проверить, содержат ли эти области аминокислотные последовательности на которые действует отбор и оценить направление этого отбора, мы далее рассчитали соотношение несинонимичных к синонимичным заменам (*dN/dS*) у кОРС.

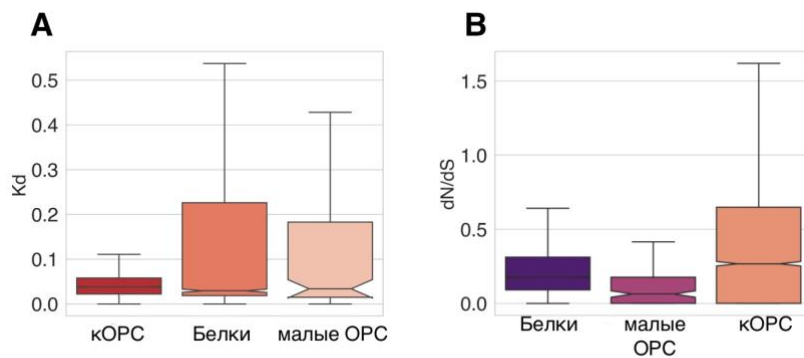


Рисунок 17. А - распределение значений Ka , у кОРС и белковых ОРС; $P=0,2$ по критерию суммы рангов Краскела-Уоллиса. В - распределение отношений dN/dS у кОРС ($n=4022$) и функциональных белков ($n=8203$); $P<0,0001$ по критерию суммы рангов Краскела-Уоллиса.

Как и ожидалось, последовательности, кодирующие белок (включая малые ОРС), имели более сильные признаки отрицательного отбора в сравнении с кОРС (критерий суммы рангов Краскела-Уоллиса, $P<0,0001$; **Рис. 17В**). В целом, ~76% малых ОРС продемонстрировали признаки сильного негативного отбора ($dN/dS < 0,20$) по сравнению с ~45% кОРС (1771 коротких рамок, $dN/dS < 0,20$). Наши результаты показали, что примерно половина эволюционно консервативных кОРС кодирует пептиды/микропротеины, специфичные только для определенной филогенетической ветви. В свою очередь, кОРС с сопоставимыми уровнями значений dN и dS могут подвергаться отбору на уровне нуклеотидов.

Можно предположить, что продукты трансляции кОРС могут быть материалом для дальнейшего отбора и играть важную роль в адаптации организма к изменяющимся условиям окружающей среды. Около 12% (507 кОРС) проанализированных выравниваний между *P. patens* и *Physcomitrium* sp. имели значения $dN/dS > 1$. Напротив, в наборе малых ОРС не было обнаружено последовательностей с $dN/dS > 1$. Для дальнейшего анализа был использован алгоритм HyPhy BUSTED, который позволяет ответить на вопрос, находилась ли данная последовательность по действием положительного отбора, по крайней мере, в одном сайте, по крайней мере, на одной филогенетической ветви. Мы показали, что ~12% как кОРС, так и малых ОРС содержали свидетельства эпизодического диверсифицирующего отбора ($P<0,05$).

Таким образом, наши данные свидетельствуют о существовании группы кОРС, кодирующих маленькие белки (<100 а.о.), которые в целом консервативны, и разных групп кОРС, которые поддерживаются путем отбора в группах сравнительно близкородственных организмов (специфических для вида или филогенетической линии). Полученные нами результаты также указывают на то, что некоторые из кОРС расположены в областях

днкРНК, которые подвергаются давлению отбора на нуклеотидном уровне и могут быть функционально важны для РНК-РНК и/или РНК-белковых взаимодействий. Тем не менее, поскольку кодирующие последовательности могут существовать также и на функциональных РНК, мы не можем отбросить возможность того, что кОПС в этих областях также кодируют пептиды. Хорошо известными примерами РНК, сочетающих кодирующие и некодирующие функции, являются растительные при-миРНК, которые участвуют в регуляции генов, а также кодируют функциональные пептиды. Кроме того, некоторые кОПС могут участвовать в регуляции экспрессии днкРНК, связывая соответствующие транскрипты с рибосомами и запуская механизм нонсенс-опосредованного распада РНК. Такие позиционно-специфичные кОПС могут играть регуляторные роли, сходные с функциями многих “upstream” кОПС и “downstream” кОПС у мРНК.

Структурные особенности пептидов, кодируемых днкРНК

Большое количество белков, особенно у эукариот, содержат регионы низкой сложности (РНС) различной длины. При поиске гомологичных последовательностей, например с использованием алгоритма BLAST, такие участки или маскируются и исключаются из дальнейшего поиска, либо их вклад в оценку значимости полученных выравниваний существенно снижается. Мы предположили, что, когда речь идет о коротких последовательностях, исключение регионов низкой сложности при поиске гомологичных последовательностей может привести к недооценке количества консервативных и потенциально функциональных пептидов, кодируемых днкРНК. Для того, чтобы изучить этот вопрос, мы проанализировали рамки, начинающиеся с кодона “AUG”, на наличие РНС с использованием инструмента SEG. Мы обнаружили, что примерно 10% коротких рамок (7831 кОПС), содержали предсказанные РНС (средняя длина=14 а.о.).

Мы сравнили состав РНС у коротких рамок считывания и малых ОПС. Было обнаружено, что РНС в 79 малых ОПС были значительно обогащены лизином и аланином по сравнению с кОПС (хи-квадрат, значение $P < 0,00001$; **Рис. 18А**). В свою очередь, регионы низкой сложности у кОПС были значительно обогащены неполярными аминокислотами, такими как лейцин, фенилаланин и изолейцин (точное значение критерия Фишера $P < 0,00001$). Далее мы повторно проанализировали консервативность таких кОПС без фильтрации регионов низкой сложности (TBLASTN, значение $E < 0,001$, параметр SEG = «No»). По таким параметрам поиска были идентифицированы 2095 консервативных кОПС, в сравнении с 1520, идентифицированными при фильтрации РНС. Например, богатый пролином 47-а.о. пептид, кодируемый «новой» кОПС, был изначально отнесен к группе

“неконсервативных”, но по новым данным имел возможных ортологов в 43 видах растений из различных отделов, включая мхи, печеночники и папоротники (Рис. 18В).

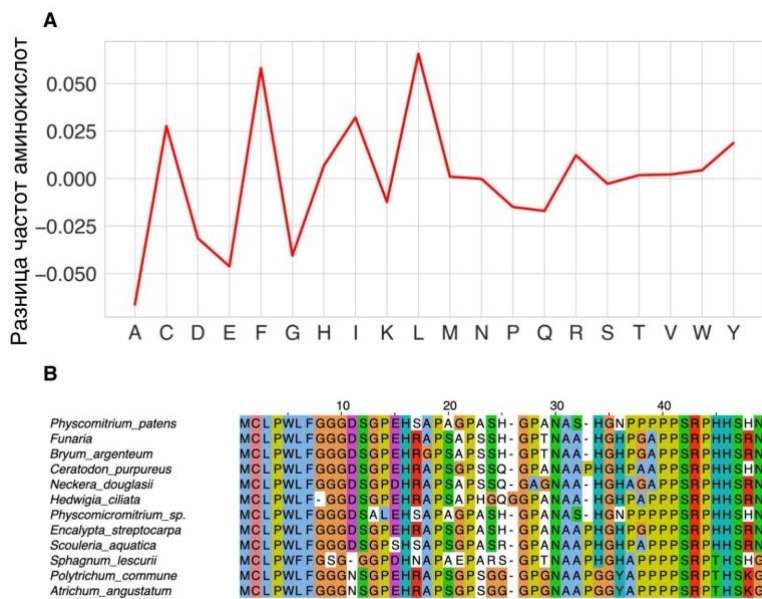


Рисунок 18. А - частоты аминокислот входящих в состав регионов низкой сложности у кОРС; показано отличие от РСН, входящих в состав малых ОРС. КОРС, были значительно обогащены лейцином, фенилаланином и изолейцином в сравнении с малыми ОРС (точный критерий Фишера $P < 0,00001$). В - Аминокислотное выравнивание нового 47-аа пептида, содержащего РНС, и отобранных ортологов.

Наши результаты свидетельствуют о том, что количество функциональных консервативных пептидов с РНС может быть существенно выше, чем считалось ранее.

Короткие открытые рамки считывания, особенно АТ-богатые, могут быть склонны к образованию трансмембранных (ТМ) доменов, которые могут также служить сигналами сортировки белков. Используя два алгоритма: ТМНМ 2.0 и SignalP-5.0, было предсказано 9472 потенциально секретируемых кОРС. Кроме того, 4978 кОРС содержали предсказанные трансмембранные домены (ТМ-кОРС), т.е. потенциально могли являться маленькими трансмембранными белками. Предполагается, что возникающие *de novo* новые, адаптивные гены образуются в АТ-богатых областях генома. Согласно нашим данным, те локусы длнкРНК, которые кодировали ТМ-кОРС имели более низкий ГЦ состав (критерий Колмогорова-Смирнова, $P < 10^{-20}$; Рис. 19А). Кроме того, ГЦ-состав соответствующих ТМ-кОРС был значительно ниже, чем у соответствующих локусов длнкРНК. Это говорит о том, что эти кОРС расположены в АТ-богатых участках соответствующих длинных некодирующих РНК (критерий Колмогорова-Смирнова, $P < 10^{-20}$; Рис. 19В). Также, мы обнаружили, что неконсервативные ТМ-кОРС значительно менее богаты ГЦ, чем консервативные (Рис. 19С). Этот результат может указывать на отбор против сильно гидрофобных кОРС, возможно, из-за их способности к агрегации внутри клетки.

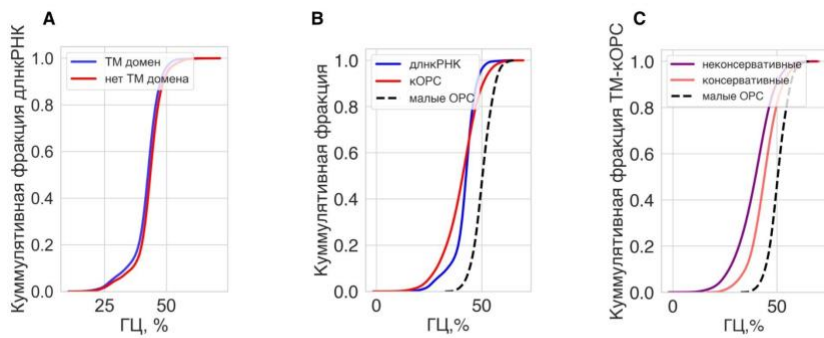


Рисунок 19. А - Анализ ГЦ-состава локусов длинкОРС с/без ТМ-кОРС; В - Анализ ГЦ-состава локусов длинкОРС с ТМ-кОРС, без ТМ-кОРС и малыми ОРС; С – Анализ ГЦ-состава консервативных и неконсервативных ТМ-кОРС и малых ОРС.

Процесс элонгации кодирующей рамки был предложен в качестве модели эволюции генов, возникающих *de novo*. Примерами таких случаев из полученных нами данных является “новая” консервативная ТМ-кОРС, кодирующая 51 а.о. пептид, содержащий специфический консервативный мотив [P***R*R***LR] на С-конце, который является общим со множеством неохарактеризованных небольших белков из базы данных RefSeq (Рис. 20А). Другим примером такой кОРС, элонгация которой у ортологов происходила на С-конце, является потенциально секретрируемый микробелок из 66 аминокислот (Рис. 20В).

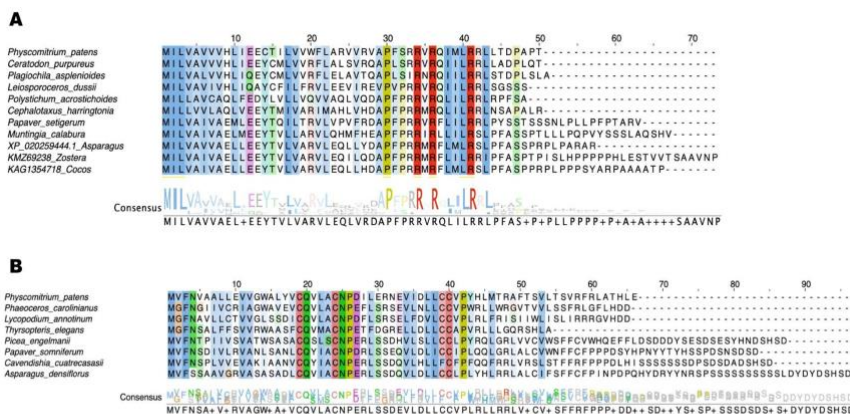


Рисунок 20. А - Множественное выравнивание выбранных ортологов «новой» 50 а.о. ТМ-кОРС; В - Множественное выравнивание ортологов секретрируемого 66-а.о. микробелка, кодируемого длиной не кодирующей РНК.

Эти примеры указывают на возможную эволюцию кОРС путем мутаций в стоп-кодоне, что приводит к элонгации транскрибируемой части, которая при этом является неконсервативной. Таким образом, мы обнаружили ряд новых консервативных кОРС, содержащих предсказанные сигналы экспорта и/или трансмембранные домены, что указывает на их возможную роль в межклеточных коммуникациях. Поскольку мутационный процесс благоприятствует переходам Г/Ц в А/Т, длинкОРС могут преимущественно продуцировать пептиды, содержащие трансмембранные домены. Новые пептиды, содержащие ТМ домены, могут

быть защищены в мембранах от деградации или взаимодействий с цитоплазматическими белками, которые токсичны для клетки.

Анализ экспрессии длнкРНК и трансляции коротких рамок считывания

Ранее было показано, что эволюционно молодые и/или гены специфичные только для определенного таксона, а также длнкРНК в среднем короче консервативных генов и экспрессируются на более низком уровне. Для того, чтобы проанализировать экспрессию нашего набора длнкРНК и сравнить ее с мРНК, мы провели прямое нанопоровое секвенирование полиА(+) фракций РНК, выделенных из протонемы (n = 3) и гаметофоров (n = 4) *P. patens*. Уровень транскрипции длнкРНК, содержащих консервативные кОРС, был значительно выше, чем уровень экспрессии длнкРНК содержащими консервативные на нуклеотидном уровне области, а также неконсервативные кОРС. Однако мы не обнаружили существенных различий между уровнями транскрипции длнкРНК, содержащих кОРС разного уровня консервативности (кластеры 1,2 и 3). Таким образом, хотя уровень экспрессии длнкРНК ожидаемо ниже, чем у мРНК, уровень транскрипции длнкРНК может быть тесно связан с консервативностью локализованных на них кОРС, как ранее было показано для белков. Следовательно, длнкРНК с консервативными кОРС можно рассматривать как мРНК с предсказанным низким кодирующим потенциалом.

Затем мы использовали масс-спектрометрический анализ для идентификации транслируемых кОРС. Выявление транслирующихся кОРС с использованием масс-спектрометрии является сложной задачей из-за низкой экспрессии и быстрой деградации пептидов, кодируемых длнкРНК. Используя два поисковых алгоритма - PEAKS 8.0 и MaxQuant, мы получили доказательства трансляции 195 кОРС, в том числе 56 из них были идентифицированы обеими поисковыми системами (**Рис. 21А**). Ожидаемо, что приблизительно 44% транслируемых кОРС принадлежали к классу «новых» кОРС (**Рис. 21В**).

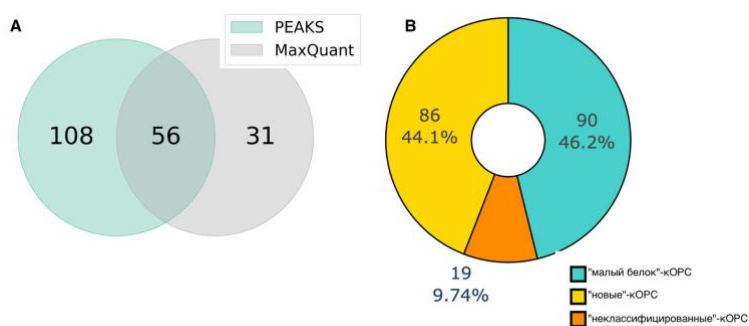


Рисунок 21. А - Диаграмма Венна, показывающая сравнение количества идентификаций транслирующихся кОРС двумя поисковыми системами — PEAKS 8.0 и MaxQuant; В - Круговая диаграмма, показывающая количество транслирующихся кОРС разного типа.

Интересно, что около 31% «новых» транслируемых кОПС были неконсервативны, что свидетельствует о сильной изменчивости пептидов/микробелков, кодируемых длнкРНК.

Пептиды с антимикробной активностью в секрете *P. patens*

Согласно нескольким исследованиям, функциональные белки растений и животных содержат аминокислотные последовательности, потенциально обладающие антимикробной активностью. Мы предположили, что стрессовые условия могут приводить к выщеплению из функциональных белков биоактивных пептидов, например, обладающих антимикробной активностью. Исходя из этого, мы протестировали антимикробную активность секретомов, обработанных метилжасмонатом (МЖ), на бактериях *E. coli* и *B. subtilis*, используя метод серийных разведений. Мы обнаружили, что секретомы, обработанные 0.4 мМ МЖ в течение часа, ограничивали рост бактерий (бактериостатический эффект) по сравнению с необработанными образцами и культуральной средой с добавлением разных концентраций МЖ (Рис. 22). Интересно, что при добавлении в ходе обработки коктейля ингибиторов протеаз, предотвращающий протеолитическую деградацию белков, мы обнаружили значительное снижение бактериостатического эффекта. Это свидетельствует о возможной роли пептидов, образующихся при расщеплении белков, в качестве быстро высвобождаемых антимикробных агентов.

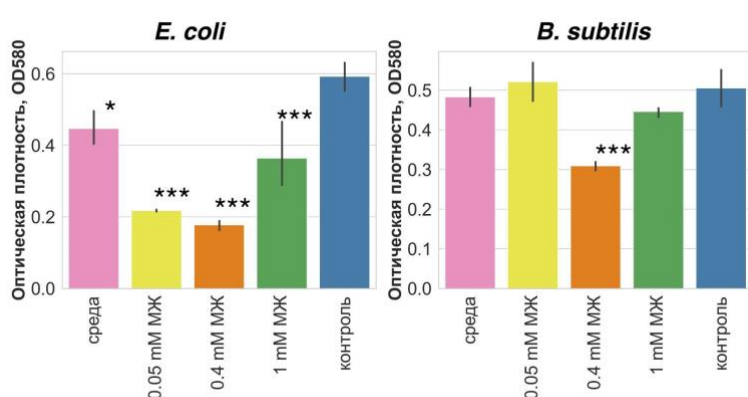


Рисунок 22. Анализ антибактериальной активности секретомы *P. patens* обработанного разными концентрациями метилжасмоната. Гистограмма показывает оптическую плотность культур *E. coli* и *B. subtilis* после 24-часовой инкубации с контролем, обработанными метилжасмонатом образцами секретомы и культуральной средой с 0.4 мМ МЖ. Столбцы ($M \pm SD$) представляют

результаты трех независимых экспериментов, проведенных в трех повторностях. * $P < 0.05$; ** $P < 0.005$; *** $P < 0.0005$. Значение P рассчитывали с помощью теста Тьюки (показано сравнение образцов с контролем).

Затем мы оценили потенциальную антимикробную активность пептидов, идентифицированных в клетке и секрете, используя три программы: iAMPpred, CAMP и

ADAM. Согласно полученным результатам, примерно 3,5% всех пептидов, идентифицированных нами в секрете, могут обладать антимикробными свойствами на основании предсказаний всех трех сервисов.

Для проверки биологической активности предсказанных антимикробных пептидов мы отобрали восемь (пять из клеточного пептидома и три из секрета) на основании следующих критериев: физико-химические свойства, высокий предполагаемый антимикробный потенциал и изменение содержания при обработке метилжасмонатом по сравнению с контролем. Мы исследовали противомикробную активность этих пептидов *in vitro* методом серийных разведений с бактериями *E. coli* и *B. subtilis*. Было обнаружено, что минимальная ингибирующая концентрация (МИК) для двух пептидов - LVQIGTKIVGVGRNYAAH (фрагмент белка, содержащего домен фумарилацетоацетат гидролазы) из клеточного пептидома и INIINAPLQGFKIA (фрагмент предсказанного белка) из клеточного секрета - составляла 64 и 16 мкг/мл соответственно. Таким образом, полученные нами результаты показали, что некоторые пептиды - продукты деградации белков обладают выраженной антимикробной активностью. Более того, обработка стрессовыми гормонами, такими как метилжасмонат, может увеличить количество потенциальных АМП и представленность таких пептидов в клетке и секрете.

Функции пептидов, кодируемых длнкРНК у растений

Согласно полученным нами результатам, пептидомы растений, помимо продуктов деградации белков, содержат сотни продуктов трансляции кОПС. Чтобы изучить функции этих пептидов у растений, мы исследовали влияние сверхэкспрессии и нокаута четырех “длнкРНК-кОПС”: *PSEP25*, *PSEP18*, *PSEP1* и *PSEP3*. Их трансляция была надежно подтверждена методом масс-спектрометрического анализа, а стартовые кодоны находятся в «сильном» и «среднем» контексте Козак. Мы получили несколько независимых мутантных линий *P. patens* для каждой из этих “длнкРНК-кОПС” и проанализировали изменения в фенотипе, а также изменения в протеоме.

Мы обнаружили, что нокаут 61 а.о. микробелка *PSEP25* приводил к снижению скорости роста и изменению архитектуры нитей протонемы на среде без глюкозы, но с добавлением тартрата аммония (**Рис. 23**). Мутантные линии со сверхэкспрессией *PSEP25* продемонстрировали небольшое снижение скорости роста по сравнению с диким типом и почти полное отсутствие морфологических различий в структуре протонемных филаментов. Интересно, что сравнительный количественный протеомный анализ с изобарными метками iTRAQ растений дикого типа и мутантных линий выявил всего 14 дифференциально экспрессирующихся белков.

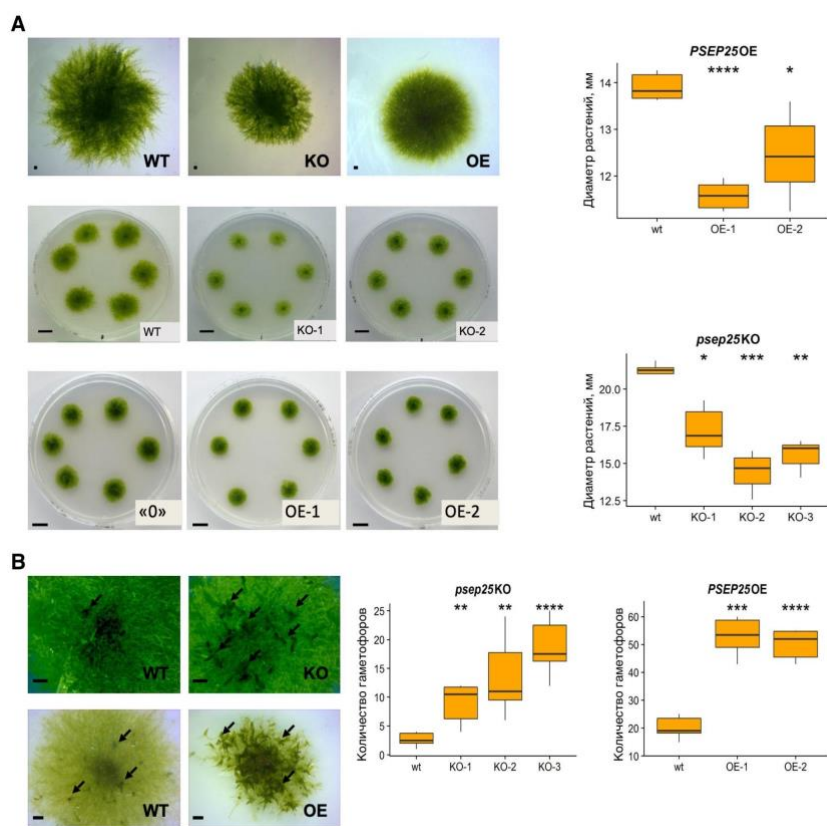


Рисунок 23. Морфология растений дикого типа (WT) и мутантных линий по микробелку *PSEP25*. **А** - Фенотипы нокаутных (*psep25* KO) и сверхэкспрессирующих (*PSEP25* OE) линий. Показаны различия в диаметре растений дикого типа и мутантных линий; “0” - линий со сверхэкспрессией пустой плазмиды. **В** - Количество гаметофоров у дикого типа и мутантных линий. Масштабная линейка для чашек Петри - 1 см. Графики показывают сравнение количества гаметофоров у растений дикого типа и мутантных линий. *P*-значение рассчитывали с помощью непарного *t*-критерия Стьюдента. (***) $P < 0,0001$, (***) $P < 0,001$, (**) $P < 0,01$, (*) $P < 0,05$.

Таким образом, несмотря на выраженные изменения в морфологии растений, по-видимому, данный микробелок связан со специфическими регуляторными каскадами и слабо влияет на глобальную экспрессию генов у *P. patens*.

Далее мы проанализировали фенотипы мутантных по 40 а.о. неконсервативному пептиду *PSEP18* растений *P. patens*. Нокаут данного пептида приводил к небольшому уменьшению диаметра растений мха на среде с глюкозой и без тартрата аммония. Линии со сверхэкспрессией *PSEP18* OE продемонстрировали значительное снижение скорости роста по сравнению с диким типом на среде с глюкозой. Количественный сравнительный протеомный анализ нокаутных линий выявил всего 8 меняющихся белков. Таким образом, полученные результаты позволяют предположить, что эволюционная консервативность может быть маркером функциональности “днкРНК-кОПС”.

Пептиды и белки, богатые пролином, играют важную роль в каскадах передачи сигналов, развитии растений и устойчивости к стрессу. Наш МС-анализ подтвердил трансляцию семи “днкРНК-кОПС”, которые были обогащены пролинами, включая 57-аминокислотный микробелок, который мы обозначили *PSEP3* (~30% пролинов) и его паралог. Согласно данным нанопорового секвенирования, кодирующая *PSEP3* днкРНК активно транскрибировалась как в протонеме, так и в гаметофорах *P. patens* (Рис. 24А). Поиск с

отключенной фильтрацией (SEG="no") позволил выявить ортологи *PSEP3* у 27 видов мхов и одного вида печеночника (Рис. 24В).

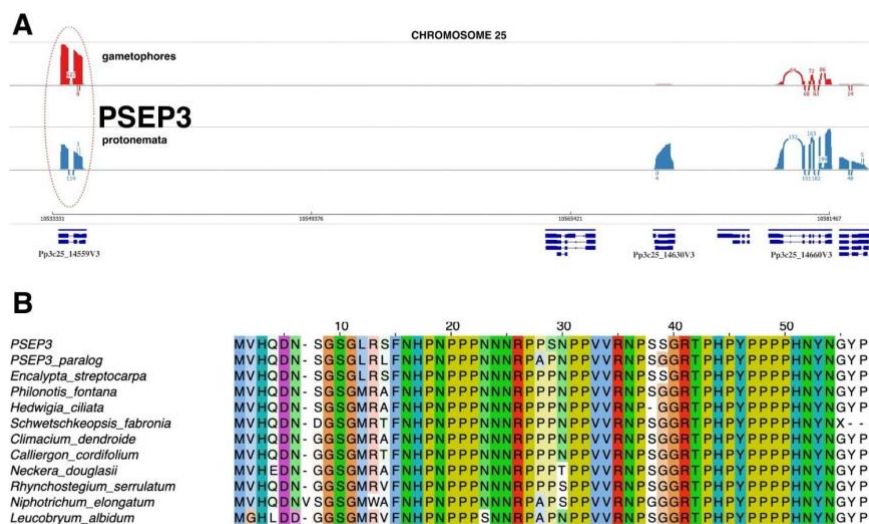


Рисунок 24. А - график Сашими, показывающий уровень транскрипции *PSEP3* и окружающего региона на хромосоме 25; В - множественное попарное выравнивание выбранных ортологов *PSEP3*, а также его паралога.

Для дальнейшего функционального анализа мы получили линии с нокаутом данного пептида, однако получение линий со сверхэкспрессией было затруднено из-за сильного ингибирования роста и гибели клеток в мутантных линиях. Исходя из этого, мы создали мутантные линии, в которых кодирующая кОПС находилась под регуляцией индуцибельного промотора (система индукции β -эстрадиолом, *PSEP3* OE). У линий с нокаутом по *PSEP3* (*psep3* KO) наблюдалось снижение скорости роста и изменение ветвления протонемных филаментов (Рис. 25А).

В отличие от растений дикого типа, у которых бледно-зелёная диффузная сеть каулонемных филаментов окружала центральную зону (состоящую в основном из нитей хлоронемы), нокаутные линии продуцировали короткие боковые филаменты на среде без глюкозы и тартрата аммония. Индукция экспрессии *PSEP3* приводила к существенному увеличению количества мертвых клеток протонемы (Рис. 25В), которые мы детектировали с помощью красителя флуоресцеин диацетат через 48 часов после добавления эстрадиола (Рис. 25С). Мы также обнаружили значительное увеличение уровня активных форм кислорода в клетках протонемы мутантной линии со сверхэкспрессией *PSEP3* через 24 часа после индукции (ANOVA, $P < 0,001$; Рис. 25D).

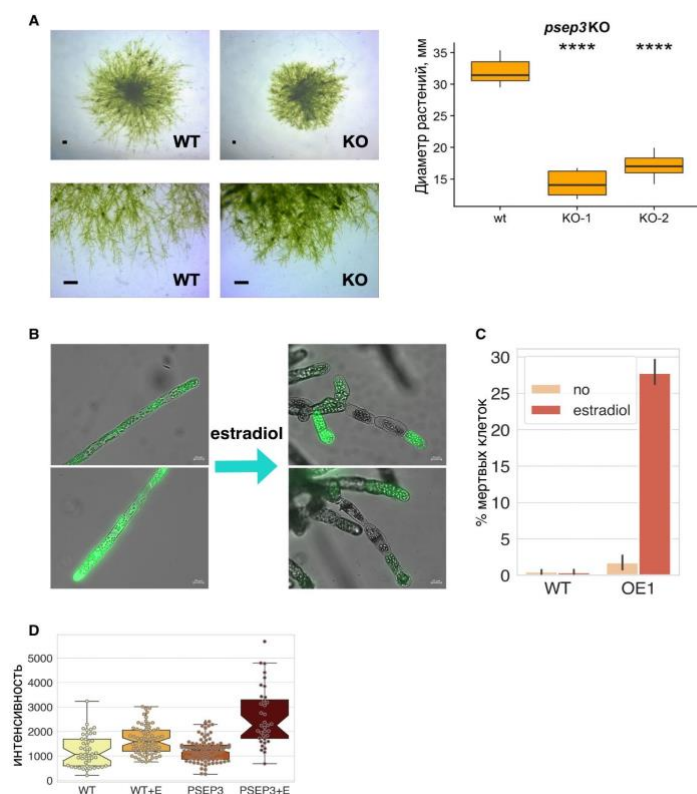


Рисунок 25. А - Фенотипы линий *psep3* KO и *PSEP3* OE. Показан диаметр растений мха дикого типа и нокаутных линий. *P*-значение рассчитывали с помощью непарного *t*-критерия Стьюдента. (****) $P < 0,0001$, (***) $P < 0,001$, (*) $P < 0,05$. В, С - влияние индукции сверхэкспрессии *PSEP3* на жизнеспособность клеток, измеренное с помощью красителя флуоресцеин диацетат (FDA). “WT” - дикий тип; “KO” - нокаут *PSEP3*; “OE1” - линия со сверхэкспрессией *PSEP3*; Д - Разница в интенсивности флуоресценции красителя dichloro-dihydro-fluorescein diacetate (DCFH-DA) при сравнении растений дикого типа и линий со сверхэкспрессией *PSEP3*, обработанных эстрадиолом ($P < 0,001$ двухфакторный дисперсионный анализ с последующим тестом Тьюки).

Для изучения активности эндогенного промотора соответствующей длнкРНК мы использовали систему CRISPR-Cas9 для слияния в одной рамке считывания репортерной последовательности гена GUS с последовательностью *PSEP3* непосредственно в геноме *P. patens*. Анализ мутантных линий показал экспрессию микробелка в активно делящихся “точках роста” гаметофоров, что может указывать на важную роль данного микробелка в регуляции процессов развития у *P. patens* (Рис. 26).

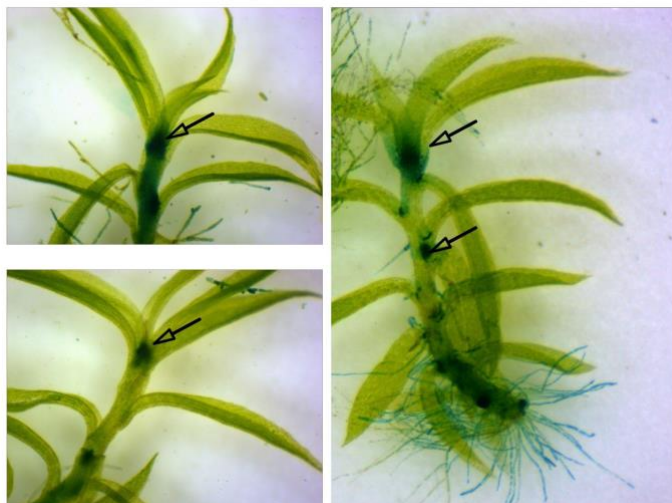


Рисунок 26. Анализ пространственной экспрессии микробелка *PSEP3-GUS* в тканях *P. patens*. Стрелками показаны некоторые области экспрессии *PSEP3* у гаметофоров, которые находятся в регионах, содержащих активно делящиеся клетки.

Таким образом, мы показали, что пептиды от кОРС, содержащие регионы низкой сложности, могут играть важную роль в регуляции клеточных процессов у растений.

Ряд недавних исследований показал, что пептиды, кодируемые кОПС и содержащие трансмембранные домены, адаптированы к тому, чтобы быть предшественниками новых функциональных генов. Для того, чтобы понять возможные функции таких пептидов у растений, мы сконцентрировались на изучении функций 41-аа пептида *PSEPI*, который мы, основываясь на характерном фенотипе, назвали *FAMOSS* («FAst-growing MOSS»). Мы провели поиск ортологов *FAMOSS* в различных растительных таксонах, используя транскриптомы из проекта “1000 транскриптомов растений”. Поиск с использованием алгоритмов TBLASTN ($E < 0,001$) и программы HMMER ($E < 0,01$) выявил возможные ортологи данного пептида у печеночников, мхов и папоротников, зеленых водорослей, голосеменных, покрытосеменных (Рис. 27А и В). Таким образом, ортолог пептида *FAMOSS* присутствовал у последнего общего предка наземных растений и стрептофитовых водорослей и сохранился в различных линиях наземных растений. Высокий уровень экспрессии соответствующего транскрипта свидетельствует о том, что этот пептид является неаннотированным высококонсервативным малым белком. Для дальнейшего изучения взаимосвязи между консервативными аминокислотными остатками и структурными свойствами *FAMOSS* мы предсказали его трехмерную структуру, используя недавно разработанный алгоритм машинного обучения AlphaFold2 (Рис. 27С).

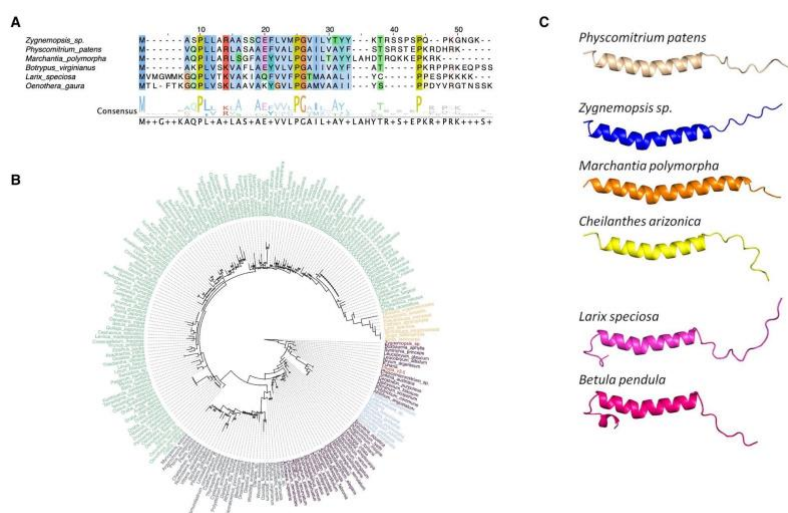


Рисунок 27. А - Выравнивание пептида *FAMOSS* и ортологических кОПС из разных отделов растений. В - Филогенетическое дерево идентифицированных ортологов *FAMOSS*. Мхи - бордовый; печеночники - синий; водоросли - красный; папоротники - серый; голосеменные - оранжевый; покрытосеменные - зеленый. С - Структура пептида *FAMOSS* и ортологов из разных отделов растений, предсказанные с помощью алгоритма AlphaFold2.

Известно, что протонема мхов состоит из двух морфологически различных типов клеток: хлоронемы и каулонемы. Мы обнаружили, что сверхэкспрессия пептида *PSEPI* приводила к формированию значительного количества более длинных каулонемных клеток в сравнении с клетками дикого типа и нокаутами (*psep1* KO; Рис. 28А). Также наблюдалась

небольшая, но значительная разница в скорости роста между линиями дикого типа и мутантными линиями *psep1KO*, выращенными на твердой среде и в жидкой культуре без глюкозы (Рис. 28В).

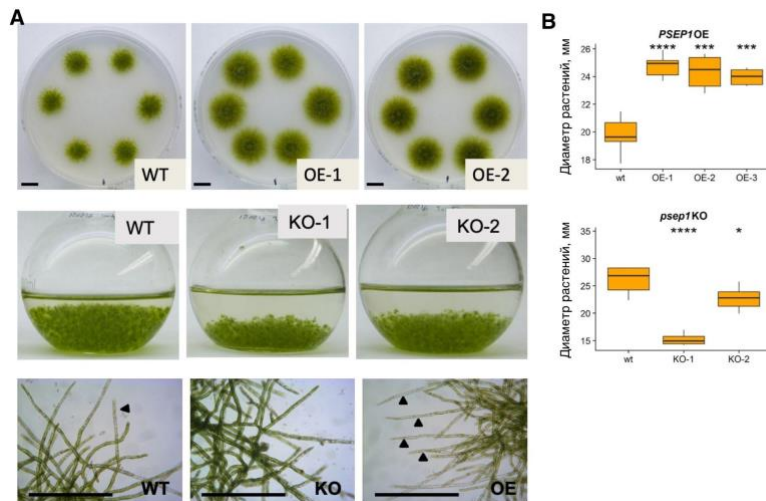


Рисунок 28. Морфология растений *P. patens* дикого типа (WT), нокаутных линий (КО-1 и КО-2), выращенных в жидкой питательной среде, и линий со сверхэкспрессией (OE-1 и OE-2) микробелка. Показано увеличение количества каулонемных нитей у линий со сверхэкспрессией. Графики показывают различия в диаметре растений мха дикого типа и мутантных линий. *P*-значение рассчитывали с помощью непарного *t*-критерия Стьюдента. (***) $P < 0,0001$, (*) $P < 0,05$.

Соотношение каулонемных и хлоронемных клеток было значительно выше в линии со сверхэкспрессией *PSEPI* и ниже в нокаутной линии по сравнению с диким типом (критерий хи-квадрат, $P < 10^{-15}$). Эти результаты указывают на связь между экспрессией *FAMOSS* и развитием протонемы у *P. patens*. Сравнительный количественный протеомный анализ выявил 118 дифференциально регулируемых белков в линии со сверхэкспрессией и 75 в нокаутной линии. Интересно, что представленность некоторых белков, которые участвуют в реакции на стресс, значительно снизилась в мутантных линиях. Кроме того, мы также идентифицировали группу белков, связанных с везикулярным транспортом, представленность которых снижалась в нокаутной линии. Мы далее идентифицировали возможных белковых партнеров, с которыми взаимодействует *FAMOSS*, используя два подхода: метод белковой преципитации (“pull-down assay”) и ко-иммунопреципитацию (“co-IP”). Помимо большого количества рибосомальных и хлоропластных белков, которые наиболее часто являются контаминантами интерактомных экспериментов, мы обнаружили в интерактомах высококонсервативные белки - малые ГТФазы Rab-типа. Известно, что белки ГТФазы Rab-типа регулируют слияние внутриклеточных везикул с плазматической мембраной, тем самым стимулируя интенсивность их транспорта к апексу клетки и увеличивая скорость полярного роста. Используя краситель SynaptoGreen C4, мы действительно обнаружили, что интенсивность везикулярного транспорта была

значительно выше в апикальных клетках линии со сверхэкспрессией *FAMOSS* в сравнении с диким типом и нокаутными линиями (ANOVA, $P < 0,001$; Рис. 29).

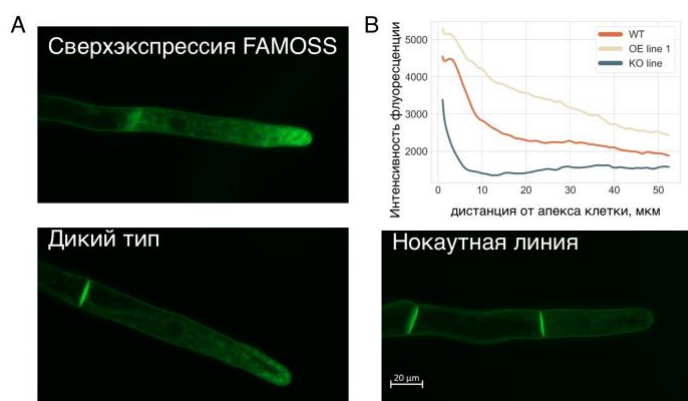


Рисунок 29. А - Анализ интенсивности везикулярного транспорта с помощью красителя SynaptoGreen C4 в апикальных клетках дикого типа и мутантных линий. В - Профиль флуоресценции SynaptoGreen C4 в апикальных клетках “WT” - дикого типа, “OE line 1” - линии со сверхэкспрессией, “KO line” - нокаутной линии.

Это согласуется с быстрым ростом каулонемных нитей в сверхэкспрессирующей линии. Таким образом, пептид *FAMOSS* можно считать важным компонентом регуляции везикулярного транспорта у растений и дальнейшие исследования помогут более точно прояснить роль данного пептида.

ЗАКЛЮЧЕНИЕ

Пептиды играют ключевую роль в регуляции множества процессов, однако системные исследования биоразнообразия пептидов, начавшиеся в начале 1990-х годов, продемонстрировали скромный прогресс из-за ограничений доступных аналитических методов. С развитием методов масс-спектрометрического анализа стала возможной подробная инвентаризация клеточных и секретируемых пептидомов разнообразных организмов. В рамках данной работы мы сконцентрировались на анализе пептидных пулов растений, показав, что ткани модельного объекта *P. patens* содержат тысячи эндогенных пептидов, основная часть которых является “белковым деградомом”. Нами было установлено отсутствие корреляции между содержанием белка-предшественника, уровнем его транскрипции и количеством эндогенных пептидов в пептидоме. Этот факт, а также проведенный нами сравнительный анализ пептидомов подтверждает, что формирование пептидных пулов является более детерминированным процессом, чем представлялось ранее, и что консервативные белки могут генерировать стабильные пулы пептидов из сходных областей у разных организмов. Далее, проанализировав влияние стрессовых условий на состав внутриклеточных и секретируемых пептидов, мы продемонстрировали значительные изменения в клеточных пулах в ответ на действие различных факторов. Например, полученные нами результаты показывают, что обработка стрессовыми гормонами, а также повреждение клеточной стенки приводят к увеличению общего

количества эндогенных пептидов в клетке и секрете, в том числе, пептидов с антимикробной активностью. Эти данные свидетельствуют о важной роли пептидов, образующихся из функциональных белков как источника биоактивных молекул. Мы предполагаем, что в этом случае фрагменты “белкового деградомы” могут быть источником пептидов с антимикробной активностью для первичного ответа на стрессовые факторы, например, атаку фитопатогенов.

В данном исследовании мы также изучили еще одну практически неизвестную часть пептидома растений, экспериментально подтвердив трансляцию сотен коротких открытых рамок считывания (кОРС), расположенных на длинных некодирующих РНК (длнкРНК) или мРНК. Мы показали, что использование пептидомных подходов имеет целый ряд преимуществ для идентификации таких пептидов, особенно, транслируемых с длнкРНК. Хотя согласно целому ряду исследований пептиды, кодируемые кОРС, могут играть фундаментальную роль в ключевых клеточных процессах, они обычно не учитываются при аннотации геномов. Полученные нами данные значительно расширяют картину представлений о происхождении, эволюции и функциях таких пептидов у растений. Например, было показано, что кОРС, расположенные в нетранслируемых областях мРНК, быстро эволюционируют и, скорее, играют регуляторные, а не кодирующие функции. В то же время, мы обнаружили десятки транслируемых кОРС, которые перекрывались с белок-кодирующими последовательностями аннотированных генов. Согласно нашим данным, эволюция таких кОРС может зависеть от их расположения относительно доменной структуры белок-кодирующей последовательности. Поскольку недавно было показано, что такие альтернативные рамки считывания имеют важные биологические функции у животных, наше исследование, продемонстрировав недооцененность кодирующего потенциала мРНК растений, подчеркивает необходимость ре-аннотации растительных протеомов.

Подавляющее большинство транскриптов (98–99%) в эукариотической клетке представляет собой некодирующие РНК. Среди этих РНК выделяется гетерогенный класс разнообразных длинных некодирующих РНК (длнкРНК), который традиционно определяется как транскрипты длиной более 200 нуклеотидов без заметного белок-кодирующего потенциала. Ранее было показано, что только небольшая часть длнкРНК выполняет специфические функции, прежде всего, в ремоделировании хроматина и регуляции экспрессии генов. Однако согласно предыдущим исследованиям, длнкРНК часто связаны с рибосомами, что приводит к важному вопросу: все ли длнкРНК на самом деле некодирующие? В ходе данной работы мы попытались ответить на вопрос о вкладе кОРС, локализованных на длнкРНК

(“днкРНК-кОРС”), в формирование пептидома растительной клетки. Наш анализ выявил несколько возможных сценариев возникновения и последующей эволюции кОРС. Во-первых, некоторые из идентифицированных консервативных кОРС являются *bona fide* небольшими орфанными функциональными белками или предшественниками секретируемых пептидов. В этом случае соответствующие транскрипты были ошибочно идентифицированы как длинные некодирующие РНК или представляют собой бифункциональные транскрипты. Этот вывод согласуется с идентификацией нами большого количества ранее неизвестных, специфичных только для определенных таксонов или высококонсервативных секретируемых кОРС. Во-вторых, большая фракция “днкРНК-кОРС” либо появилась только у одного вида, либо имеет ортологов только у близких видов, что предполагает их случайное возникновение и исчезновение в растительных геномах в процессе эволюции. Мы обнаружили, что некоторые из этих кОРС могут транслироваться в пептиды, при том, что уровень их экспрессии низкий. Такие транслируемые, но слабо консервативные пептиды могут не иметь четко определенной функции, но служить предшественниками функциональных генов. В-третьих, существует подмножество “днкРНК-кОРС”, которые консервативны на уровне нуклеотидов, но не имеют признаков отбора на уровне белков. Возможные функции и эволюция таких кОРС представляют интерес и требуют дальнейшего выяснения. Таким образом, мы идентифицировали многочисленные и ранее неизвестные транслируемые кОРС и подтвердили их экспрессию с помощью анализа транскриптома и протеома. Следовательно, еще одним не изученным ранее компонентом пептидома, помимо пептидных гормонов и продуктов “белкового деградума”, являются пептиды, транслирующиеся с днкРНК.

Поскольку ранее функции пептидов, кодируемых днкРНК, не были изучены у растений, мы провели функциональный анализ четырех таких пептидов. Согласно нашим результатам, нокаут выбранных пептидов не был летальным для мутантных линий, но снижал скорость их роста по сравнению с растениями дикого типа. Напротив, мы обнаружили, что растения со сверхэкспрессией пептидов, кодируемых днкРНК, демонстрируют больше фенотипических различий по сравнению с растениями дикого типа и нокаутами. Мы наблюдали как значительное увеличение скорости роста и числа облиственных побегов при экспрессии некоторых пептидов, так и снижение скорости роста. Различия между линиями дикого типа и мутантными линиями часто проявлялись только при определенных условиях роста – на твердых или жидких средах с/без глюкозы. Эти данные указывают на точную регуляцию трансляции пептидов, кодируемых днкРНК. В свете этих находок мы предположили, что пептиды, кодируемые днкРНК, могут быть

важны для адаптации к определенным условиям внешней среды. Так, проведенный нами анализ 41-а.о. пептида с предсказанным трансмембранным доменом показал, что он взаимодействует с ГТФ-азами Rab-типа, по-видимому, являясь ранее неизвестным консервативным участником механизмов регуляции везикулярного транспорта у растений. Наши результаты согласуются с функциональными исследованиями “днкРНК-кОРС” у животных, показывая общность принципов эволюции таких кодирующих единиц.

Таким образом, в ходе проведенного нами системного исследования пептидома растений мы обнаружили три большие фракции пептидов: “деградом”, образованный при расщеплении неспециализированных белков-предшественников; пептиды, кодируемые короткими открытыми рамками считывания, и продукты их деградации; биоактивные пептиды, образующиеся при специфическом расщеплении специализированных предшественников (Рис. 30).

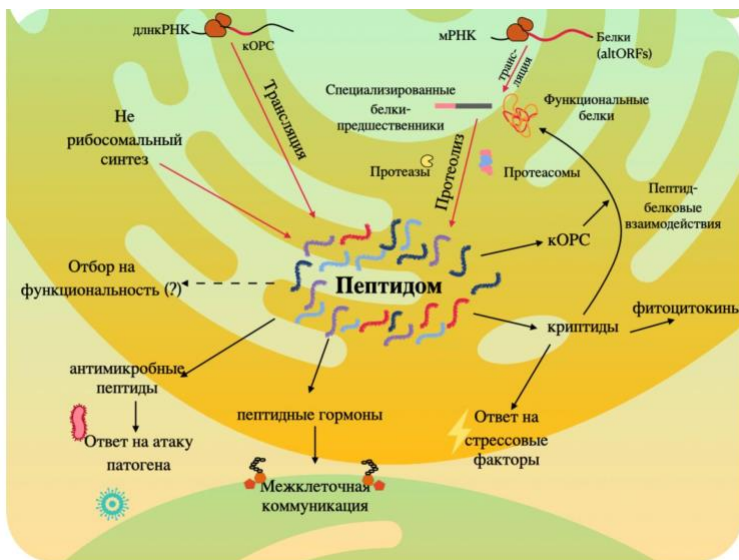


Рисунок 30. Общая схема путей образования и функций пептидов у растений.

ВЫВОДЫ

1. Впервые проведен системный анализ пептидома растений на примере модельного объекта - *Physcomitrium patens*; идентифицированы тканеспецифичные пептидные пулы, состоящие из тысяч пептидов.
2. На основе полученных данных предложена модель ответа растений на биотический стресс на уровне пептидома путем образования пептидов с антимикробной активностью из функциональных белков-предшественников.

3. Проведен анализ различных классов коротких открытых рамок считывания, потенциально способных кодировать пептиды в геноме *P. patens*. Впервые в пептидоме растений идентифицированы продукты трансляции коротких открытых рамок считывания разных типов.
4. Для растений впервые проведен эволюционный анализ и изучена роль коротких открытых рамок считывания, локализованных на длинных некодирующих РНК, в процессе образования новых генов.
5. Впервые показано, что растительные геномы содержат множество не аннотированных ранее транскрибирующихся кОПС, кодирующих пептиды/микробелки, большая часть которых специфична для определенного вида или филогенетической линии.
6. Проведен функциональный анализ пептидов, кодируемых длинными некодирующими РНК. Впервые показана важная роль таких пептидов в регуляции роста и развития растений.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ:

Статьи в международных рецензируемых журналах

1. A. Mamaeva, A. Knyazev, A. Glushkevich, **I. Fesenko**, Quantitative proteomic dataset of the moss *Physcomitrium patens* PSEP3 KO and OE mutant lines. *Data in Brief*. 40 (2022), p. 107715.
2. I. Lyapina, V. Ivanov, **I. Fesenko**, Peptidome: Chaos or Inevitability. *Int. J. Mol. Sci.* 22(23), 13128 (2021)
3. **I. Fesenko**, S. A. Shabalina, A. Mamaeva, A. Knyazev, A. Glushkevich, I. Lyapina, R. Ziganshin, S. Kovalchuk, D. Kharlampieva, V. Lazarev, M. Taliansky, E. V. Koonin, A vast pool of lineage-specific microproteins encoded by long non-coding RNAs in plants. *Nucleic Acids Res.* 49, 10328–10346 (2021).
4. M. Breygina, E. Klimenko, E. Shilov, A. Podolyan, A. Mamaeva, V. Zgoda, **I. Fesenko**, Hydrogen peroxide in tobacco stigma exudate affects pollen proteome and membrane potential in pollen tubes. *Plant Biol.* 23, 592–602 (2021).
5. I. Lyapina, A. Filippova, S. Kovalchuk, R. Ziganshin, A. Mamaeva, V. Lazarev, I. Latsis, E. Mikhailchik, O. Panasenko, O. Ivanov, V. Ivanov, **I. Fesenko**, Possible role of small secreted peptides (SSPs) in immune signaling in bryophytes. *Plant Mol. Biol.* 106, 123–143 (2021).
6. A. Knyazev, A. Glushkevich, **I. Fesenko**, Direct RNA sequencing dataset of SMG1 KO mutant *Physcomitrella* (*Physcomitrium patens*). *Data in Brief*. 33, 106602 (2020).
7. A. Mamaeva, M. Taliansky, A. Filippova, A. J. Love, N. Golub, **I. Fesenko**, The role of chloroplast protein remodeling in stress responses and shaping of the plant peptidome. *New Phytol.* 227, 1326–1334 (2020).
8. **I. Fesenko**, R. Azarkina, I. Kirov, A. Kniazhev, A. Filippova, E. Grafkskaia, V. Lazarev, V. Zgoda, I. Butenko, O. Bukato, I. Lyapina, D. Nazarenko, S. Elansky, A. Mamaeva, V. Ivanov, V. Govorun, Phytohormone treatment induces generation of cryptic peptides with antimicrobial activity in the Moss *Physcomitrella patens*. *BMC Plant Biol.* 19, 9 (2019).

9. I. Lyapina, A. Filippova, **I. Fesenko**, The Role of Peptide Signals Hidden in the Structure of Functional Proteins in Plant Immune Responses. *Int. J. Mol. Sci.* 20(18), 4343 (2019).
10. **I. Fesenko**, I. Kirov, A. Kniazev, R. Khazigaleeva, V. Lazarev, D. Kharlampieva, E. Grafkskaia, V. Zgoda, I. Butenko, G. Arapidi, A. Mamaeva, V. Ivanov, V. Govorun, Distinct types of short open reading frames are translated in plant cells. *Genome Res.* 29, 1464–1477 (2019).
11. A. Filippova, I. Lyapina, I. Kirov, V. Zgoda, A. Belogurov, A. Kudriaeva, V. Ivanov, **I. Fesenko**, Salicylic acid influences the protease activity and posttranslation modifications of the secreted peptides in the moss *Physcomitrella patens*. *J. Pept. Sci.* 25, e3138 (2019).
12. I. Kirov, M. Gilyok, A. Knyazev, **I. Fesenko**, Pilot satellitome analysis of the model plant, *Physcomitrellapatens*, revealed a transcribed and high-copy IGS related tandem repeat. *Comp. Cytogenet.* 12, 493–513 (2018).
13. **I. Fesenko**, R. Khazigaleeva, V. Govorun, V. Ivanov, Analysis of Endogenous Peptide Pools of *Physcomitrella patens* Moss. *Methods Mol. Biol.* 1719, 395–405 (2018).
14. **I. A. Fesenko**, I. V. Kirov, A. A. Filippova, Impact of noncoding part of the genome on the proteome plasticity of the eukaryotic cell. *Russ. J. Bioorganic Chem.* 44, 397–402 (2018).
15. R. A. Khazigaleeva, **I. A. Fesenko**: Biologically active peptides encoded by small open reading frames. *Russ. J. Bioorganic Chem.* 43(6), 617–624 (2017).
16. **I. Fesenko**, R. Khazigaleeva, I. Kirov, A. Kniazev, O. Glushenko, K. Babalyan, G. Arapidi, T. Shashkova, I. Butenko, V. Zgoda, K. Anufrieva, A. Seredina, A. Filippova, V. Govorun, Alternative splicing shapes transcriptome but not proteome diversity in *Physcomitrella patens*. *Sci. Rep.* 7, 2698 (2017).
17. R. A. Khazigaleeva, S. V. Vinogradova, V. L. Petrova, **I. A. Fesenko**, G. P. Arapidi, A. M. Kamionskaya, V. M. Govorun, V. T. Ivanov: Antimicrobial activity of endogenous peptides of the moss *Physcomitrella patens*. *Russ. J. Bioorganic Chem.* 43(3), 248–254 (2017).
18. **I. Fesenko**, A. Seredina, G. Arapidi, V. Ptushenko, A. Urban, I. Butenko, S. Kovalchuk, K. Babalyan, A. Knyazev, R. Khazigaleeva, E. Pushkova, N. Anikanov, V. Ivanov, V. M. Govorun, The *Physcomitrella patens* Chloroplast Proteome Changes in Response to Protoplastation. *Front. Plant Sci.* 7, 1661 (2016).
19. **I. A. Fesenko**, G. P. Arapidi, A. Y. Skripnikov, D. G. Alexeev, E. S. Kostryukova, A. I. Manolov, I. A. Altukhov, R. A. Khazigaleeva, A. V. Seredina, S. I. Kovalchuk, R. H. Ziganshin, V. G. Zgoda, S. E. Novikova, T. A. Semashko, D. K. Slizhikova, V. V. Ptushenko, A. Y. Gorbachev, V. M. Govorun, V. T. Ivanov, Specific pools of endogenous peptides are present in gametophore, protonema, and protoplast cells of the moss *Physcomitrella patens*. *BMC Plant Biol.* 15, 87 (2015).

Избранные тезисы докладов конференций:

1. Фесенко И.А. Moss proteomics and peptidomics. New insight in the old story. Peptides in the stress adaptation process. Устный доклад, Новосибирск, 25-29 июня 2012, международная конференция «POSTGENOMIC TECHNOLOGY FOR BIOMEDICINE»
2. Фесенко И.А. Системный анализ пептидогенеза растений. Пептидомика стресса для клеток мха *Physcomitrella patens*. Устный доклад. Казань, 20-25 ноября 2012, международная конференция “ПОСТГЕНОМ-2012”

3. Igor Fesenko et.al. Specific pools of endogenous peptides are present in gametophore, protonema and protoplast cells of the moss *Physcomitrella patens*. Устный доклад. Praha, Czech Republic, 17-19 июня 2013. Международная конференция “MOSS-2013”
4. Identification of small open reading frame (sORF) in genome of the moss *Physcomitrella patens*. Устный доклад. Устный доклад. Международная конференция «MOSS-2014» 25.09-28.09.2014, Пекин, Китай
5. Середина А.В., Фесенко И.А., Хазигалеева Р.А., Пушкова Е.Е., Арапиди Г.П., Урбан А.С., Ковальчук С.И., Бутенко И.О., Говорун В.М. Сравнительный количественный анализ хлоропластных протеомов клеток мха *Physcomitrella patens*. Устный доклад. СИМПОЗИУМ «БЕЛКИ И ПЕПТИДЫ» 12–17 июля 2015, Академгородок, Новосибирск
6. Igor Fesenko et.al. Identification of peptide encoding small open reading frames (sORFs) in the genome of the moss *Physcomitrella patens*. Устный доклад. Международная конференция “3rd European Workshop on Peptide Signalling and Activity in Plants”, 3-5 сентября 2015, Гент, Бельгия
7. Igor Fesenko et.al. Studying functions of sORF-encoded peptides in the Moss, *Physcomitrella patens*. Устный доклад. Лидс, Великобритания, 2-3 сентября 2016. Международная конференция “MOSS-2016”
8. Igor Fesenko, Ilya Kirov, Andrey Knyazev, Regina Azarkina, Anna Mamaeva, Vadim Ivanov and Vadim Govorun. Distinct types of small open reading frames in plants: trends in evolution and peptide coding capacity. Устный доклад. Международная конференция “PLANT PEPTIDES & RECEPTORS 2018, the 6th European Workshop on Peptide and Receptor Signaling in Plants”, 5-7.09.18 Антекера, Испания
9. Igor Fesenko. Elucidating the function of peptides encoded by sORFs in model plant – the moss *Physcomitrella patens*. Устный доклад. Международная конференция “II Объединенный научный форум: VI Съезд биохимиков России и IX Российский симпозиум «Белки и пептиды»”, 1–6 октября 2019, Дагомыс, Сочи