

Федеральное государственное бюджетное учреждение науки
ИНСТИТУТ БИООРГАНИЧЕСКОЙ ХИМИИ
им. академиков М.М. Шемякина и Ю.А. Овчинникова
Российской академии наук

На правах рукописи

Фесенко Игорь Александрович

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

Специальность 1.5.3 - Молекулярная биология

Диссертация на соискание ученой степени
доктора биологических наук

Научный консультант:
д.х.н., академик Иванов В.Т.

Москва-2022

Оглавление

| | |
|---|-----------|
| 1. ВВЕДЕНИЕ..... | 5 |
| 2. ОБЗОР ЛИТЕРАТУРЫ..... | 10 |
| 2.1 МЕХАНИЗМЫ ОБРАЗОВАНИЯ ПЕПТИДОВ В КЛЕТКАХ РАСТЕНИЙ И ЖИВОТНЫХ | 10 |
| 2.1.1 Пептиды – продукты гидролиза специализированных предшественников | 10 |
| 2.1.2 Пептиды - продукты деградации функциональных белков..... | 14 |
| 2.1.3 Пептиды - продукты трансляции коротких открытых рамок считывания | 17 |
| 2.2 РАЗНООБРАЗИЕ БИОЛОГИЧЕСКИ АКТИВНЫХ ПЕПТИДОВ РАСТЕНИЙ..... | 25 |
| 2.2.1 Пептиды, участвующие в иммунном ответе растений | 25 |
| 2.2.2 Пептиды - регуляторы процессов роста и развития у растений..... | 33 |
| 2.2.3 Функции пептидов - продуктов трансляции коротких открытых рамок считывания | 36 |
| 2.3 МЕТОДЫ АНАЛИЗА ПЕПТИДОМОВ | 43 |
| 2.3.1 Использование масс-спектрометрии для идентификации пептидов | 43 |
| 2.3.2 Методы анализа интерактомонов пептидов..... | 46 |
| 3. МАТЕРИАЛЫ И МЕТОДЫ..... | 50 |
| 3.1 МЕТОДИКИ РАБОТЫ С РАСТЕНИЯМИ..... | 50 |
| 3.1.1 Выращивание протонемы и гаметофоров <i>P. PATENS</i> | 50 |
| 3.1.2 Получение протопластов <i>P. PATENS</i> | 51 |
| 3.1.3 Индукция повреждений клеточной стенки раствором драйзелазы..... | 52 |
| 3.1.4 Выделение хлоропластов из <i>P. PATENS</i> | 52 |
| 3.2 МЕТОДИКИ ПОЛУЧЕНИЯ МУТАНТНЫХ РАСТЕНИЙ <i>P. PATENS</i> | 53 |
| 3.2.1 Создание гидовых РНК для получения нокаутов с помощью системы CRISPR/Cas9 | 53 |
| 3.2.2 Создание конструкций для сверхэкспрессии целевых пептидов..... | 54 |
| 3.2.3 Трансформация протопластов..... | 59 |
| 3.2.4 Анализ и отбор мутантных растений <i>P. PATENS</i> | 59 |
| 3.3 МЕТОДИКИ РАБОТЫ С ПЕПТИДАМИ | 60 |
| 3.3.1 Выделение нативных пептидов из протонемы, гаметофоров и протопластов <i>P. PATENS</i> | 60 |
| 3.3.2 Выделение нативных пептидов из секретома протонемы МХА | 61 |
| 3.3.3 Анализ антимикробной активности пептидов | 62 |
| 3.3.4 Синтез пептидов..... | 64 |

| | |
|---|-----------|
| 3.4 МЕТОДИКИ РАБОТЫ С БЕЛКАМИ..... | 64 |
| 3.4.1 ВЫДЕЛЕНИЕ БЕЛКОВ ИЗ КЛЕТОК ПРОТОНЕМЫ И ГАМЕТОФОРОВ | 64 |
| 3.4.2 ЭЛЕКТРОФОРЕЗ БЕЛКОВ В ДЕНАТУРИРУЮЩЕМ ГЕЛЕ | 65 |
| 3.4.3 ПОДГОТОВКА ОБРАЗЦОВ ДЛЯ МАСС-СПЕКТРОМЕТРИЧЕСКОГО АНАЛИЗА | 65 |
| 3.4.4 МЕЧЕНИЕ ТРИПТИЧЕСКИХ ПЕПТИДОВ С ПОМОЩЬЮ ИЗОБАРНЫХ МЕТОК iTRAQ..... | 66 |
| 3.4.5 ИММУНОБЛОТТИНГ | 66 |
| 3.5 МЕТОДИКИ РАБОТЫ С НУКЛЕИНОВЫМИ КИСЛОТАМИ..... | 67 |
| 3.5.1 ВЫДЕЛЕНИЕ ГЕНОМНОЙ ДНК | 67 |
| 3.5.2 ВЫДЕЛЕНИЕ ТОТАЛЬНОЙ РНК..... | 68 |
| 3.5.3 ОЧИСТКА ПОЛИ-А ФРАКЦИИ ИЗ ТОТАЛЬНОЙ РНК..... | 69 |
| 3.5.4 ЭЛЕКТРОФОРЕЗ ДНК В АГАРОЗНОМ ГЕЛЕ..... | 70 |
| 3.5.5. ВЫДЕЛЕНИЕ ПЛАЗМИДНОЙ ДНК..... | 70 |
| 3.5.6 ПЦР С РЕАКЦИЕЙ ОБРАТНОЙ ТРАНСКРИПЦИИ..... | 71 |
| 3.6 МЕТОДИКИ ТРАНСКРИПЦИОННОГО АНАЛИЗА | 72 |
| 3.6.1 ПРЯМОЕ СЕКВЕНИРОВАНИЕ РНК С ПОМОЩЬЮ НАНОПОРОВОГО СЕКВЕНИРОВАНИЯ | 72 |
| 3.6.2 ИДЕНТИФИКАЦИЯ И АНАЛИЗ ТРАНСКРИПТОВ | 74 |
| 3.7 МАСС-СПЕКТРОМЕТРИЧЕСКИЙ АНАЛИЗ..... | 75 |
| 3.7.1 ТАНДЕМНЫЙ ХРОМАТО-МАСС-СПЕКТРОМЕТРИЧЕСКИЙ АНАЛИЗ | 75 |
| 3.7.2 МОНИТОРИНГ МНОЖЕСТВЕННЫХ РЕАКЦИЙ | 76 |
| 3.7.3 АНАЛИЗ ДАННЫХ МАСС-СПЕКТРОМЕТРИЧЕСКОГО АНАЛИЗА | 77 |
| 3.7.4 КОЛИЧЕСТВЕННЫЙ АНАЛИЗ ПРОТЕОМНЫХ ДАННЫХ | 78 |
| 3.8 МЕТОДЫ АНАЛИЗА ИНТЕРАКТОМА..... | 79 |
| 3.8.1 ИММУНОПРЕЦИПИТАЦИЯ АНТИТЕЛАМИ К FLAG ЭПИТОПУ | 79 |
| 3.8.2 ПОЛУЧЕНИЕ КОНЪЮГАТА СТРЕПТАВИДИН-ПЕПТИД В БАКТЕРИАЛЬНОЙ СИСТЕМЕ ЭКСПРЕССИИ..... | 81 |
| 3.8.3 ПРОВЕДЕНИЕ АФФИННОЙ ХРОМАТОГРАФИИ..... | 81 |
| 3.9 ФЛУОРЕСЦЕНТНАЯ МИКРОСКОПИЯ | 82 |
| 3.10 МЕТОДЫ БИОИНФОРМАТИЧЕСКОГО АНАЛИЗА | 83 |
| 3.10.1 ПРЕДСКАЗАНИЕ ДЛИННЫХ НЕКОДИРУЮЩИХ РНК НА ОСНОВЕ ПРЕДСКАЗАННЫХ ТРАНСКРИПТОВ | 83 |
| 3.10.2 ПРЕДСКАЗАНИЕ КОРОТКИХ РАМОК СЧИТЫВАНИЯ С ВЫСОКИМ КОДИРУЮЩИМ ПОТЕНЦИАЛОМ..... | 83 |
| 3.10.3 ПРЕДСКАЗАНИЕ ВСЕХ КОРОТКИХ РАМОК СЧИТЫВАНИЯ НА ДЛИННЫХ НЕКОДИРУЮЩИХ РНК..... | 85 |
| 3.10.4 АНАЛИЗ КОНСЕРВАТИВНОСТИ ДЛИННЫХ РНК И КОРОТКИХ ОТКРЫТЫХ РАМОК СЧИТЫВАНИЯ..... | 85 |
| 3.10.5 АНАЛИЗ ЭВОЛЮЦИИ КОРОТКИХ ОТКРЫТЫХ РАМОК СЧИТЫВАНИЯ | 86 |
| 3.10.6 СТАТИСТИЧЕСКИЙ АНАЛИЗ..... | 87 |
| 3.10.7 ПРЕДСКАЗАНИЕ ДОМЕНОВ И МОТИВОВ В ПОЛИПЕПТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ | 87 |
| 3.10.8 ПРЕДСКАЗАНИЕ АНТИМИКРОБНОЙ АКТИВНОСТИ | 88 |
| 3.10.9 ПРЕДСКАЗАНИЕ 3D-СТРУКТУРЫ ПЕПТИДОВ | 88 |

| | |
|--|------------|
| 4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ..... | 88 |
| 4.1 Анализ нативных пептидомов модельного растения <i>P. PATENS</i>..... | 88 |
| 4.1.1 Идентификация и анализ внутриклеточных пептидных пулов | 88 |
| 4.1.2 Идентификация и анализ внеклеточных пептидных пулов..... | 99 |
| 4.1.3 Влияние стрессовых факторов на состав пептидных пулов растений | 108 |
| 4.1.4 Сравнение пептидомов различных организмов | 129 |
| 4.2 Идентификация продуктов трансляции коротких открытых рамок считывания в пептидоме <i>P. PATENS</i> | 134 |
| 4.2.1 Поиск коротких открытых рамок считывания (КОРС) с высоким кодирующим потенциалом в геноме <i>P. PATENS</i> | 135 |
| 4.2.2 Классификация КОРС с высоким кодирующим потенциалом | 139 |
| 4.2.3 Анализ консервативности “генных” коротких открытых рамок считывания..... | 143 |
| 4.2.4 Идентификация пептидов, кодируемых короткими открытыми рамками считывания | 146 |
| 4.2.5 Совместная экспрессия коротких рамок считывания и белков, которые расположены на одном транскрипте | 151 |
| 4.2.6 Альтернативный сплайсинг регулирует количество КОРС на транскриптах | 152 |
| 4.2.7 Анализ КОРС - потенциальных регуляторов белок-белковых взаимодействий | 156 |
| 4.2.8 Предсказание и классификация коротких рамок считывания на длинных некодирующих РНК..... | 159 |
| 4.2.9 Анализ консервативности КОРС, локализованных на длинных некодирующих РНК | 163 |
| 4.2.10 Анализ скоростей эволюции длинных некодирующих РНК и коротких открытых рамок считывания..... | 167 |
| 4.2.11 Структурные особенности пептидов, кодируемых длнкРНК | 173 |
| 4.2.12 Анализ экспрессии длнкРНК и трансляции коротких рамок считывания | 185 |
| 4.3 Биологически активные компоненты пептидомов растений | 191 |
| 4.3.1 Антимикробная активность компонентов пептидного пула | 191 |
| 4.3.2 Функции пептидов, кодируемых длнкРНК у растений | 201 |
| 4.3.3 Пептид FAMOSS - регулятор полярного роста у растений..... | 218 |
| 5. ЗАКЛЮЧЕНИЕ..... | 229 |
| 6. ВЫВОДЫ..... | 234 |
| 7. СПИСОК СОКРАЩЕНИЙ | 235 |
| 8. СПИСОК ЛИТЕРАТУРЫ..... | 238 |

1. ВВЕДЕНИЕ

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 [1–3]. 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 [4,5]. 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 криптическими пептидами [6].

В последние годы стремительное развитие методов масс-спектрометрического 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 [1]. 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 этих РНК тем не менее может кодировать пептиды, таким образом формируя неизученную часть пептидома, которая способна выполнять регуляторные функции [7]. 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 секрета 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 деградама" 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.

2. ОБЗОР ЛИТЕРАТУРЫ

2.1 Механизмы образования пептидов в клетках растений и животных

На сегодняшний момент известно, что пептиды образуются в клетке или при трансляции рибосомами с коротких рамок считывания, либо как продукт белкового метаболизма [8]. Большинство известных биологически активных пептидов (пептидных гормонов), идентифицированных у растений образуется при гидролизе протеазами нефункциональных, специализированных белковых предшественников [1,2]. В данной главе будут рассмотрены механизмы генерации пептидных пулов клеток (Рис. 2.1).

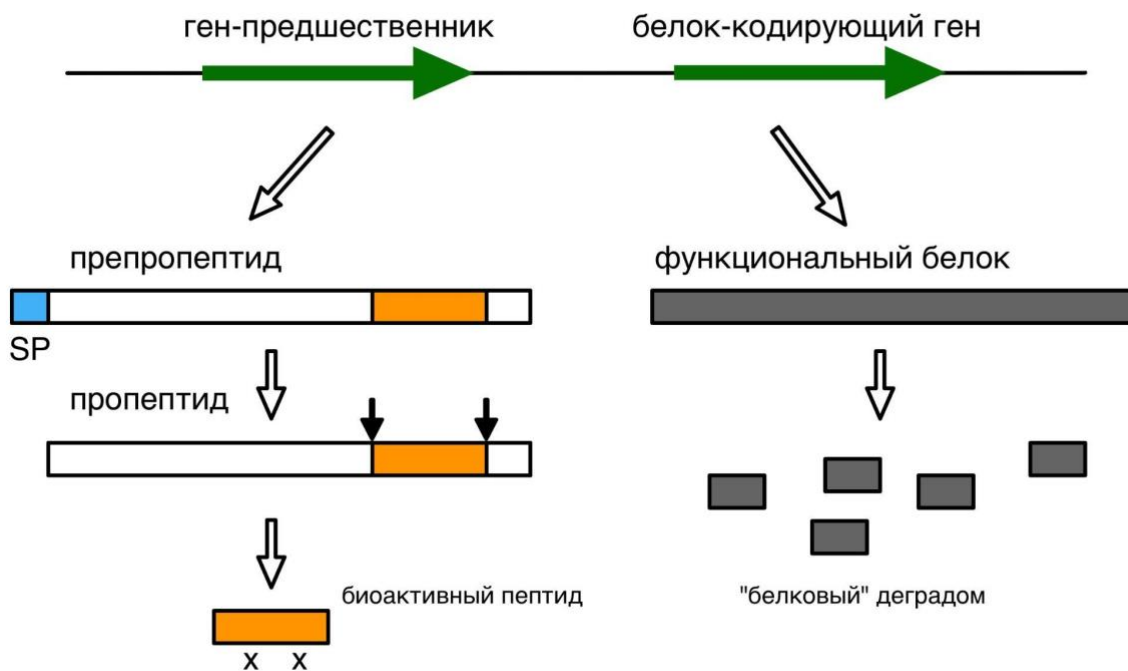


Рисунок 2.1 Механизмы генерации пептидных пулов растений при расщеплении белков предшественников.

2.1.1 Пептиды – продукты гидролиза специализированных предшественников

Гидролиз протеазами специализированных белков-предшественников является наиболее изученным на сегодняшний день механизмом образования биологически активных пептидов (например, пептидных гормонов). Гидролиз

нефункционального прекурсора продуцирует пептидную последовательность оптимальной длины для связывания с соответствующим рецептором (**Рис. 2.1**). Несмотря на то, что в геномах модельных растений идентифицировано много протеаз, функции большинства из них изучены слабо [9,10]. Из известных протеаз в формировании биологически активных пептидов наиболее изучена роль субтилизин-подобных протеаз (субтилаз). Субтилазы представляют из себя большое семейство в основном внеклеточных протеаз, включающее 56 генов у арабидопсиса [11], 86 у томатов [12] и 97 у винограда [13]. Наиболее изучена роль субтилазы S1P (SITE 1 PROTEASE)/SBT6.1 и SBT6.2, которые участвуют в биогенезе пептидов семейства ROOT GROWTH FACTOR (RGF)/GOLVEN (GLV)/CLE-Like (CLEL), контролирующей элонгацию клеток [1].

Частичный гидролиз N-конца прекурсора фитосульфокина PSK4 субтилазой SBT1.1 был показан *in vitro*. Показано, что субтилаза AtSBT6.1/S1P, которая локализована в эндоплазматическом ретикулуме (ЭПР) и аппарате Гольджи, процессирует прекурсор цистеин-богатого RALF23 (RAPID ALKALINIZATION FACTOR 23) пептида. Для гидролиза этой протеазе необходим канонический мотив RxxL/RxLx. Показано, что PRORALF23 процессируется субтилазой S1P в течение минут после индукции иммунного ответа элиситорами flg22 и elf18, участвуя в подавлении иммунного сигналинга [14,15]. Еще одним примером специфического протеолиза предшественника биоактивного пептида субтилазами является выщепление 14-а.о. пептида IDA (INFLORESCENCE DEFICIENT IN ABSCISSION, участвует в регуляции опадения цветков), которое осуществляется тремя субтилазами: SBT5.2, SBT4.13 и SBT4.12 [16]. Подобная избыточность может объясняться сложной регуляцией данного процесса, когда субтилазы, по сути, переводят неактивные пробелки в активные пептиды. В отличие от предыдущих примеров, у пробелка IDA не удалось обнаружить канонический мотив, по которому может происходить протеолиз, скорее, это некая протяженная область, в которой остатки P2-Pro и P4-Tyr являются наиболее важными.

Так же, как и IDA, CLE (CLV3/EMBRYO SURROUNDING REGION), пептиды процессируются с N – конца. Предполагается, что процессинг предшественников

этих пептидов зависит от консервативного остатка аргинина в позиции +1 и осуществляется секретруемыми сериновыми протеазами [17]. Показано, что эффективное узнавание происходит за четыре аминокислотных остатка до CLE домена и помимо консервативного аргинина требует консервативных остатков отрицательно заряженных аминокислот в позициях -2 и/или -3 [17]. Помимо сериновых протеаз, в процессинге активных CLE пептидов могут принимать участие мембранно-связанные Zn^{2+} -зависимые карбоксипептидазы. Показано, что SOL1 (SUPPRESSOR OF LLP1 1) карбоксипептидаза удаляет С-терминальный аргинин из CLE19 пробелка, продуцируя активный пептид [18]. Одной из интересных особенностей данной системы является возможная локализация каталитического домена SOL1 в эндосомах, что может указывать на место финального процессинга CLE19.

Согласно недавнему исследованию, пептиды CLEL6 и CLEL9 претерпевают целый ряд посттрансляционных модификаций, включая протеолитическое расщепление прекурсора субтилазами [19]. На первом этапе сигнальный пептид отщепляется при входе прекурсора в эндоплазматический ретикулум. После выхода из ЭР и при входе в аппарат Гольджи, прекурсор претерпевает первое протеолитическое расщепление, которое осуществляется субтилазой SBT6.1. В аппарате Гольджи пептиды подвергаются модификациям, таким как сульфатирование тирозина и гидроксилирование пролина. Предполагается, что при выходе из аппарата Гольджи или в апопласте клетки происходит финальный процессинг и выщепление активного пептида при помощи субтилазы SBT3.8. Работа этой протеазы может быть рН-зависимой и требовать кислого окружения, которое появляется при выходе пробелка из аппарата Гольджи. Внеклеточная локализация SBT3.8 подтверждает эту теорию. Такая сложная ступенчатая система процессинга CLE пептидов, по-видимому, является дополнительным уровнем регуляции их действия, когда быстрое формирование активного пептидного лиганда происходит под действием клеточных стимулов из уже предпроцессированных пептидов в определенном компартменте клетки. Помимо процессинга CLE пептидов, субтилаза SBT3.8 участвует в процессинге ещё одного биоактивного пептида -

фитосульфокина (PSK). Процессинг белка-прекурсора PSK протеазой SBT3.8 зависит от остатка аспарагиновой кислоты, следующей за сайтом протеолиза [20]. Помимо сериновых протеаз, важную роль в процессинге иммунных регуляторных пептидов играют цистеиновые протеазы. Показано, что пептид AtPER1 процессируется из 92-а.о. предшественника с помощью кальций-зависимой цистеиновой протеазы METACASPASE4 (MC4). Метакаспазы требуют низких концентраций ионов кальция Ca^{2+} для активности *in vitro* и разрезают субстрат после лизина или аргинина. Per1 был обнаружен в течение 30 секунд после повреждающего воздействия, достигал пика после 5 минут и детектировался в течение последующего часа [21,22]. Прекурсор PROPER1 остается связанным с тонопластом вакуоли до тех пор, пока вход ионов кальция при повреждающем воздействии на клетку не активирует цитозольную протеазу метакаспаза-4, которая выщепляет активный AtPER1. Связано это с тем, что концентрация ионов Ca^{2+} в цитозоле клетки слишком низка для активации MC4. Затем Per1 попадает в цитозоль, откуда может диффундировать в апопласт, где связывается с рецепторным комплексом ВАК1-PERP1/2 для активации защитных генов в окружающих клетках. По-видимому, растения используют такой консервативный механизм генерации иммунных сигнальных пептидов для того, чтобы быстро активировать защитную реакцию в ответ на повреждающие воздействия.

Более того, метакаспазы участвуют не только в активации AtPER. Известно, что метакаспаза-9 арабидопсиса участвует в выщеплении GRI (GRIM REAPER), который играет важную роль в активации клеточной смерти у арабидопсиса [23,24].

Помимо метакаспаз, еще одним примером цистеиновой протеазы, участвующей в процессинге активного пептида, являются папин-подобные протеазы (PLCPs). Показано, что PLCP участвует в процессинге пептида Zip1 у кукурузы. Тем не менее, точный механизм процессинга активных пептидных лигандов большинством протеаз остается неизвестным.

Протеолитическое расщепление необходимо при формировании всех известных сигнальных пептидов с пост-трансляционными модификациями. Дополнительные

пост-трансляционные модификации могут включать сульфатирование тирозина, гидроксирование пролина и арабинозилирование гидроксипролиновых остатков [25–27]. Установлено, что за сульфатирование тирозина ответствен фермент тирозил-сульфотрансфераза (tyrosylprotein sulfotransferase (TPST)), который заякорен в мембране аппарата Гольджи [28]. При этом ферменту необходим аспарат для узнавания субстрата. Сульфатирование тирозина является критическим для активности соответствующих пептидов [27]. Гидроксирование пролина - одна из ключевых модификаций пептидов, участвующих в регуляции процессов роста и развития растений. Данная модификация катализируется заякоренными в мембране эндоплазматического ретикулума и аппарата Гольджи ферментами – пролил-4-гидролазами (P4Hs). В геноме арабидопсиса обнаружены 13 P4Hs, часть из которых ответственна за модификацию гликобелков клеточной стенки, например, экстенсина [29]. Какие из пролил-4-гидролаз модифицируют сигнальные пептиды, а также их специфичность в отношении пролина в определенном контексте изучено слабо. Данная модификация необходима для проведения еще одной модификации регуляторных пептидов – гликозилирования. На первом этапе фермент аппарата Гольджи - гидроксипролин *O*-арабинозилтрансфераза (HPAT) переносит L-арабинозу на 4-гидроксил. В геноме арабидопсиса обнаружены три гена, кодирующие данный фермент, функционально замещающие друг друга [30,31].

2.1.2 Пептиды - продукты деградации функциональных белков

Первое упоминание внутриклеточных пептидов - продуктов деградации функциональных белков встречается в научной литературе в 50-х годах прошлого века [32,33]. В данный момент предполагается, что основным механизмом образования таких пептидов у животных и растений является протеолиз белков при помощи универсальной системы клеточной деградации - убиквитин-протеосомальной системы [34]. Было показано, что протеасомы ответственны за деградацию примерно 70% клеточных белков в животных клетках [35].

2.1.2.1 Эндogenous внутриклеточные пептиды животных

В целом ряде исследований было показано, что тысячи эндогенных пептидов - продуктов деградации функциональных белков - могут быть выделены из различных тканей животных, таких как мозги мышей [36] и модельной рыбы данио-рерио [37], клеточные линии человека [38,39], плазма крови [40,41] и спинномозговая жидкость человека [42], а также ряда других. С использованием различных биохимических подходов было показано, что протеосомальный путь деградации белков играет ключевую роль в формировании пула таких пептидов (Рис. 2.1). Длина пептидов, продуцируемая протеасомой, варьирует от 3 до 22 аминокислотных остатков, и представленность пептидов уменьшается с увеличением их длины, согласно логнормальному распределению [43]. Обработка клеточной линии НЕК293Т ингибиторами протеасомальной активности, такими как эпоксимицин, привела к значительному снижению количества детектируемых внутриклеточных пептидов [44]. Однако тестирование ряда обратимых и необратимых ингибиторов протеасомальной активности, таких как карфлизомиб, MG132, MG262, MLN2238, AM114 и clasto-lactacystin- β -lactone показало, что их влияние на пептидные внутриклеточные пулы может иметь различную направленность [45].

Использование технологии MAPP (mass spectrometry analysis of proteolytic peptides), которая основана на связывании и последующей идентификации выходящих из протеасомы продуктов деградации функциональных белков, позволило пролить свет на формирование внутриклеточных пептидных пулов [39]. При использовании этой методики было идентифицировано около пяти тысяч пептидов - продуктов деградации 1,004 уникальных белков. Показано, что идентифицированные пептиды происходили от активно-деградируемых белков, и их представленность не зависела от количества белка в клетке [39]. Пептидный пул был обогащен пептидами из С-концов белков-предшественников. Анализ сайтов расщепления пептидов выявил три четких паттерна, а именно: мотив, обогащенный отрицательно заряженными аминокислотами; мотив, обогащенный гидрофобными аминокислотами и обогащенный полярными остатками. Эти паттерны

соответствовали каталитической активности $\beta 1$ (каспазо-подобная) и $\beta 5$ (химотрипсин-подобная) субъединиц протеасомы [39].

Предполагается, что пептиды, образовавшиеся после работы систем деградации белков в клетке, подвергаются дальнейшему расщеплению при помощи олигопептидаз. Наиболее перспективной с точки зрения дальнейшего метаболизма внутриклеточных пептидов является thimet oligopeptidase (THOP1, EC 3.4.24.15; EP24.15). Было показано, что THOP1 участвует в регуляции активности нейропептидов, таких как брадикинин [46,47], нейротензин [48], опиоидные пептиды [49], ангиотензин [50] и гонадотропин-рилизинг-гормона [51]. Сверхэкспрессия THOP1 в клетках линии HEK293T привела к уменьшению представленности ряда внутриклеточных пептидов [52]. Нокаун THOP1 с помощью siRNA подтвердил участие этой олигопептидазы в метаболизме внутриклеточных пептидов в клетках HEK293T [53]. Пептидомный анализ различных областей мозга мышей, нокаутных по данной олигопептидазе, показал различия в вариабельности пептидных пулов и позволил идентифицировать пептиды - предполагаемые субстраты THOP1 [54]. Таким образом, THOP1 участвует в метаболизме пептидов, выходящих из протеасомы в клетках животных. Еще одной пептидазой, потенциально участвующей в метаболизме внутриклеточных пептидов, является нейролизин (peptidase neurolysin, Nln). Однако нокаут этой пептидазы у мышей привел к изменению представленности только двух десятков внутриклеточных пептидов в тканях мозга [55]. Таким образом, пептидаза нейролизин участвует главным образом в деградации определенных нейропептидов и оказывает небольшое влияние на пул эндогенных пептидов в клетке [56].

2.1.2.2 Эндогенные пептиды растений

Данные о пулах эндогенных пептидов в растительных клетках намного более скудны, чем для клеток животных. В клетках, составляющих зеленые ткани растений, наиболее представлены белки хлоропластов, деградация которых происходит при помощи специализированных протеаз и автофагии [57–59].

Пептиды, образованные при деградации белков в хлоропластах, утилизируются до аминокислот специализированными олигопептидазами OOP и PreP [60]. Это металлопротеазы, которые локализованы в строме хлоропластов и митохондриальном матриксе. Субстратом OOP являются пептиды размером 8-23 а.о., а PreP - пептиды размером 10-65 а.о. [61]. Нокаут этих олигопептидаз приводит к аккумуляции пептидов, выщепляющихся из N-конца и внутренней части хлоропластных белков. Двойные мутанты по *prep1 prep2* характеризуются хлорозом, уменьшением скорости роста и процесса дыхания в митохондриях [61]. В результате протеолиза пептидов олигопептидазами OOP и PreP образуются фрагменты размером 2-5 а.о. Эти фрагменты в последующем деградируют аминопептидазами M1, M17-10, M17-20 и M18 до свободных аминокислот [60]. Нокаутные по олигопептидазам PreP1/2- и OOP линии арабидопсиса аккумулялировали эндогенные пептиды, главным образом, продукты деградации хлоропластных белков [62]. В мутантных растениях были идентифицированы 180 пептидов размером от 6 до 30 а.о, образующиеся при расщеплении 95 белков предшественников. Эти пептиды не были обнаружены в мутантных растениях. Мутантные растения характеризовались пониженной скоростью роста. Транскрипционный анализ показал, что у мутантных растений была повышена экспрессия генов стрессового ответа, особенно тех, которые участвуют в ответе на биотический стресс [62].

2.1.3 Пептиды - продукты трансляции коротких открытых рамок считывания

Короткие открытые рамки считывания - это нуклеотидные последовательности между потенциальными старт и стоп кодонами размером от 20 до 300 нуклеотидов [63]. Однако в некоторых источниках верхний лимит размера kOCS расширен до 150 и даже 250 аа [64,65], в то время как у прокариот общепринятым верхним лимитом является 50 аа [66]. По-видимому, большинство коротких открытых рамок считывания в геноме не транслируются и не являются функциональными [65]. Поскольку миллионы нефункциональных коротких рамок считывания могут быть найдены в любом геноме, ввиду низкой статистической значимости, алгоритмы

предсказания генов обычно не учитывают кОРС при аннотации геномов для уменьшения ошибок при идентификации и классификации мРНК [67]. Поэтому предсказание и валидация кОРС, кодирующих функциональные пептиды или микробелки, является нетривиальной задачей [65,68]. Однако показано, что кОРС могут быть источником функциональных пептидов, регулирующих множество ключевых процессов в клетке [68–71], следовательно, их поиск и описание являются важнейшими задачами современной биологии.

На сегодняшний момент предложено несколько вариантов классификации кОРС в зависимости от их размера, местоположения, консервативности, потенциала транскрипции и трансляции и т.д. Согласно одному из предложенных недавно вариантов классификации [68], можно выделить пять основных классов кОРС. Самым многочисленным классом являются межгенные кОРС (“intergenic ORFs”; **Рис. 2.2**), которые находят в нетранскрибируемых участках генома. Этот тип кОРС, по-видимому, сформировался случайно и не кодирует функциональных продуктов. Таким образом, это просто участки ДНК между потенциальным старт-кодоном и одним из трех стоп-кодонов. Размер таких рамок связан с GC-составом генома (частота появления стоп-кодонов) и, в зависимости от вида, медиана длины составляет 22-23 кодона [68].

| Класс | Транскрипт | Медиана размера (aa) | Уровень трансляции | Консервативность | Особенности | Функции |
|-----------------|---------------|----------------------|--------------------|--------------------|--|---|
| Intergenic ORFs | Нет | 22 | Нет | Нет | Случайные AA | Нет |
| uORFs | 5' - НТФ | 22 | Низкий | Нет | • Неслучайные AA • Нет доменов | • Некодирующие • Регуляторная |
| lncORFs | длнкРНК | 24 | Низкий | Нет | • Неслучайные AA • Нет доменов | Некодирующие или кодирующие |
| Short CDSs | короткие мРНК | 79 | Высокий | В пределах класса | • "+"-заряженные AA • трансмембранные | • Кодирующие • Регуляторы белков |
| Short isoforms | изоформы мРНК | 79 | Высокий | В пределах царства | • каноничные AA • домены потери | • Кодирующие • Малые интерферирующие пептиды |
| Canonical ORFs | мРНК | 491 | Высокий | В пределах царства | • каноничные AA • множество доменов | • Кодирующие • Структурные, ферменты, регуляторные |

нетранслируемый регион

рамка считывания

другие кодирующие последовательности

ДНК

РНК сплайсинг

рибосомальный профайлинг

Рисунок 2.2 Классификация коротких открытых рамок считывания. Большинство открытых

рамок считывания в геномах - это кОПС в нетранскрибируемых областях (“intergenic ORFs”; светло-голубые). Открытые рамки считывания размером более 100 кодонов (“canonical ORFs”, фиолетовые) обычно транслируются в аннотированные белки с известными функциями. Также геном кодирует пептиды и микробелки размером от 10 до 100 аа, которые можно разделить на следующие классы: кОПС в 5’ - нетранслируемых регионах (НТР; “uORF”; бирюзовый) канонических мРНК; кОПС, которые находятся на длинных некодирующих РНК (“lncORF”; зеленые); короткие кодирующие последовательности (“short CDSs”), которые представляют собой аннотированные рамки считывания, локализованные на коротких мРНК (желтые); и короткие рамки считывания, которые образуются в результате альтернативного сплайсинга канонических мРНК (“short isoforms” розовый). Адаптировано из [68].

Остальные классы кОПС расположены на аннотированных транскриптах и потенциально могут быть транслированы. Например, кОПС найдены на мРНК - в нетранслируемых регионах или со сдвигом относительно основной белок-кодирующей рамки считывания (альтернативная рамка считывания). Таким образом, второй многочисленный класс кОПС включает в себя короткие рамки, обнаруженные в 5'-лидерном нетранслируемом регионе мРНК. Поскольку такие рамки располагаются на транскрипте до основной белок-кодирующей рамки, то их общепринятое обозначение - upstream ORFs (“uORF”; **Рис. 2.2**). У животных медианная длина таких рамок составляет около 22 кодонов, и они найдены примерно у 50% мРНК [68,72]. Появление старт кодона AUG в 5'-лидерной последовательности мРНК обычно имеет негативный эффект и отбраковывается в процессе эволюции при действии стабилизирующего отбора (purify selection) [73]. Действительно, кОПС редко находят в генах домашнего хозяйства, и существует значимая отрицательная корреляция между уровнем транскрипции гена и количеством кОПС в 5'-лидерной последовательности [74,75]. Интересно, что действие стабилизирующего отбора с одинаковой силой выбраковывает короткие рамки из 5'-лидерных последовательностей мРНК независимо от уровня их транскрипции [74]. Однако до сих пор плохо понятно, какое количество этих кОПС действительно может кодировать функциональные пептиды [76]. Одним из

способов оценки функциональности кодирующей последовательности может быть анализ ее эволюционной консервативности. Согласно недавнему исследованию кОПС в 5'-лидерных последовательностях генов позвоночных, насекомых и дрожжей их кодирующая часть менее консервативна, чем соответствующие старт-кодны [74]. Более того, соотношение числа несинонимичных и синонимичных замен (dN/dS) кодирующих областей uORFs было близко к единице. Это указывает на то, что они находятся под действием нейтрального отбора. Основываясь на метрике phyloCSF [77], которая на основе множественного выравнивания позволяет сделать вывод о потенциале кодирования конкретной последовательности, было показано, что только 0,44% кОПС (161/36655) из 5'-лидерных последовательностей мРНК человека может кодировать функциональные пептиды [74]. Учитывая данные масс-спектрометрического анализа, а также оценку эволюционной консервативности таких кОПС у разных групп организмов, можно сделать вывод о том, что основной их функцией является регуляция трансляции нижележащей рамки, и меньше одного процента из них могут кодировать функциональные пептиды. Тем не менее, принимая во внимание распространенность кОПС в 5'-лидерных последовательностях мРНК, число таких пептидов, функция которых неизвестна, может достигать нескольких сотен.

Еще одним потенциальным источником функциональных пептидов являются длинные некодирующие РНК. Длинные некодирующие РНК – это обширная группа транскриптов длиной свыше 200 нуклеотидов, которые не содержат длинных рамок считывания и не обладают предсказанным кодирующим потенциалом [78,79]. Предполагается, что значительная часть таких транскриптов кэпирована и содержит поли-А хвост, а значит, может быть транслирована [80]. Короткие рамки считывания, расположенные на длнкРНК, являются третьим по распространенности классом кОПС (“lncORF”; **Рис. 2.2**) [68]. Медианная длина таких рамок около 24 кодонов, а эффективность трансляции схожа с uORFs. По крайней мере, 98% аннотированных длнкРНК содержат как минимум одну кОПС, а медиана количества рамок на один транскрипт равна 6. Хотя длнкРНК являются неконсервативными транскриптами [81], оказалось, что некоторые из них

кодируют высококонсервативные функциональные пептиды [69,82,83]. Теоретически, трансляция транскриптов с низким или средним уровнем экспрессии, что характерно для длинных некодирующих РНК, не приведет к существенным энергетическим затратам и действию отрицательного отбора на клетку [84]. Если пептид не токсичен для клетки, то такая аллель может закрепиться в популяции благодаря генетическому дрейфу. В том случае, если пептид обеспечивает селективные преимущества, он может закрепиться в популяции более быстро. Некоторые из таких микробелков/пептидов, обеспечивающих селективное преимущество, могут эволюционировать в *de novo* белки, которые продолжают эволюционировать под действием стабилизирующего отбора. Часть таких пептидов/микробелков может выпадать из действия отбора при изменившихся внешних условиях, в то время как другие пептиды могут становиться функционально активными. Таким образом, пептидомы клеток могут содержать некоторое количество пептидов - продуктов трансляции коротких рамок считывания, расположенных на некодирующих РНК [85].

Четвертый класс коротких открытых рамок считывания длиннее всех предыдущих классов кОРС (медиана 79 кодонов; “short CDS”; **Рис. 2.2**), и такие рамки являются основной кодирующей последовательностью на транскриптах, схожих по характеристикам с мРНК. Они эффективно транслируются [86] и классифицируются как каноничные открытые рамки считывания. Сотни таких кОРС были найдены в геномах различных организмов, но только небольшая часть из них была охарактеризована. Аминокислотный состав пептидов, кодируемых такими короткими рамками, указывает на их тропность к мембранам и функцию как регуляторов каноничных белков [68].

Наконец, еще один класс коротких рамок образуется при альтернативном сплайсинге и представляет собой изоформы каноничных белок-кодирующих транскриптов меньше 100 аа (“short isoforms”; **Рис. 2.2**). По всей видимости, это наименее представленный класс кОРС [68]. Медианная длина этого класса коротких рамок также составляет 79 кодонов. Аминокислотный состав пептидов от таких рамок ожидаемо схож с каноничными белковыми последовательностями.

Идентификация продуктов трансляции таких коротких рамок может быть затруднена из-за сходства с более длинными белковыми изоформами из того же локуса.

Было предложено несколько моделей трансляции коротких рамок считывания в эукариотических клетках (Рис. 2.3b, c) в зависимости от типа транскрипта, на котором они расположены [65].

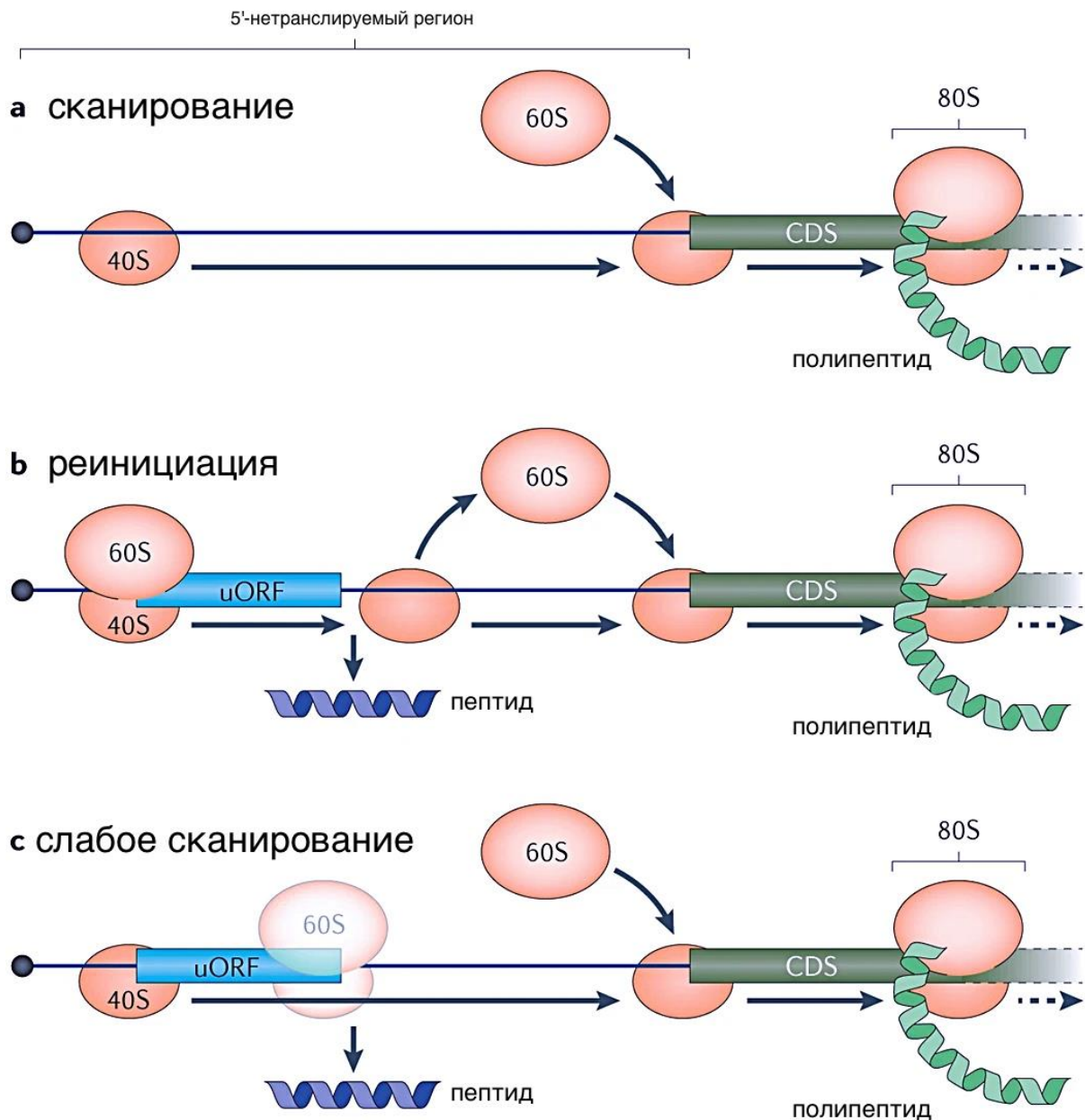


Рисунок 2.3 Модели трансляции коротких открытых рамок считывания. а | В стандартной модели сканирования эукариотической трансляции малая субъединица рибосомы 40S вместе с

преинициаторными факторами (не показаны) связывается с 5'-кэпом мРНК и сканирует транскрипт до тех пор, пока не будет распознан первый иницирующий кодон. Субъединица 60S и дополнительные факторы (не показаны) затем объединяются с субъединицей 40S, образуя рибосому, способную к элонгации (80S). **б** | Повторная инициация происходит, когда субъединица 40S ре-иницирует трансляцию в стартовом кодоне белок-кодирующей последовательности (CDS) после завершения трансляции короткой рамки считывания в 5'-лидерной последовательности ("uORF"). Трансляция "uORF" может приводить к продукции короткого пептида. **с** | При "слабом сканировании" малая субъединица рибосомы 40S может либо распознавать стартовый кодон короткой рамки считывания и инициировать трансляцию, либо инициировать трансляцию нижележащего стартового кодона белок-кодирующей рамки. Таким образом, часть рибосом будет инициировать трансляцию "uORF", а остальные будут игнорировать стартовый кодон "uORF" и продолжать сканирование по транскрипту, чтобы инициировать трансляцию в расположенном ниже стартовом кодоне белок-кодирующей последовательности. Адаптировано из [65].

Стандартная модель сканирования предполагает, что малая субъединица эукариотической рибосомы (40S) вместе с факторами инициации (43S комплекс) связывается с кэпом на 5' - конце транскрипта и движется по направлению к 3'-концу, связываясь с первым иницирующим кодоном [65]. Вкратце, после связывания инициаторного комплекса с иницирующим кодоном, 60S субъединица рибосомы и сопутствующие факторы формируют 80S рибосому, которая осуществляет трансляцию рамки считывания (ORF; **Рис. 2.3**).

По-видимому, трансляция коротких открытых рамок считывания на длинкРНК осуществляется таким образом. Однако механизм трансляции кОПС, расположенных в 5'-нетранслируемой области мРНК перед основной рамкой считывания, до конца неясен. Предполагается, что трансляция белкой рамки считывания может осуществляться с помощью механизмов ре-инициации или слабого сканирования [87]. В случае слабого сканирования пре-инициаторный 48S комплекс не иницирует трансляцию на коротких рамках считывания, а движется до стартового кодона белковой рамки считывания. В случае ре-инициации после трансляции кОПС в 5'- нетранслируемой области мРНК рибосома разделяется, однако 40S субъединица продолжает процесс сканирования, связываясь с

инициаторным кодоном белковой рамки и запуская процесс сборки 80S рибосомы. Оба этих механизма активно используют вирусы для трансляции их полицистронных РНК [88]. Однако существование подобных вариантов совместной трансляции кОРС и основной рамки считывания было также показано и у растений [89]. Следует отметить, что, помимо AUG, стартовыми кодонами кОРС могут служить CTG, GTG, и TTG [7,90].

Помимо 5'-нетранслируемых областей, транслирующиеся кОРС могут быть найдены в 3'-трейлерных последовательностях мРНК [69,91]. По-видимому, некоторые из этих пептидов могут быть функциональны. Так, в недавнем исследовании в клетках человека были идентифицированы два пептида, кодируемые короткими рамками, расположенными в 3'-UTR, которые были связаны с клеточной пролиферацией [69]. Биоинформатический анализ показал, что 82.4% белок-кодирующих мРНК человека содержит как минимум одну кОРС в 3' - нетранслируемой области, однако только небольшая часть из них может транслироваться в пептиды [91]. Медианный размер таких транслирующихся кОРС составляет 20аа. При этом такие кОРС имели высокую эффективность трансляции независимо от стартового кодона (ATG, CTG, GTG и TTC). Показано, что наличие транслирующихся кОРС в 3'-UTR усиливает трансляцию белок-кодирующей рамки считывания на соответствующей мРНК [91]. Причем количество транслирующихся кОРС прямо коррелирует с уровнем трансляции белковой рамки, но этот эффект не связан с аминокислотной последовательностью соответствующего пептида. Предполагается, что инициация трансляции коротких рамок считывания в 3'-UTR способствует привлечению факторов инициации трансляции, что напрямую влияет на уровень трансляции основной рамки считывания [91].

В последние годы появляется все больше данных о том, что существуют функциональные пептид-кодирующие рамки, которые перекрываются с белок-кодирующей областью, но со сдвигом рамки считывания [92]. В этом случае пептид имеет аминокислотную последовательность, отличную от белковой, и не является изоформой референсного для данной мРНК белка [93]. Для таких кОРС одним из возможных механизмов трансляции является слабое сканирование. Было показано,

что только у 37,4% мРНК человека стартовый кодон расположен в оптимальной Козак последовательности [94]. В таком случае рибосома может не инициировать трансляцию на стартовом кодоне белок-кодирующей рамки и транслировать следующие рамки считывания. У человека известны функциональные кОПС, пересекающие белок-кодирующую рамку считывания и транслирующиеся с помощью механизма слабого сканирования [95]. В целом, механизм слабого сканирования предоставляет возможность для трансляции альтернативных рамок считывания на мРНК.

2.2 Разнообразие биологически активных пептидов растений

До недавнего времени регуляцию процессов роста и развития растений, а также ответ на стрессовые факторы среды связывали исключительно с классическими фитогормонами, например, такими как ауксин или абсцизовая кислота. Однако в последнее время появляется все больше данных о том, что растения, как и животные, используют пептидные лиганды как сигнальные молекулы, регулирующие множество жизненно важных процессов. Функции таких биологически активных пептидов связаны с регуляцией роста меристем, образованием боковых корней, опадением цветков, регуляцией процессов оплодотворения и ответом на стрессовые факторы среды. Кроме того, пептидные лиганды обеспечивают связь между процессами роста и доступностью питательных соединений.

2.2.1 Пептиды, участвующие в иммунном ответе растений

Несмотря на отсутствие адаптивного иммунитета, система врождённого иммунитета растений позволяет им активно противостоять большому количеству фитопатогенов. Важную роль в регуляции иммунных реакций у растений играют пептиды. Помимо антимикробных пептидов, в последние годы были обнаружены сигнальные и регуляторные пептиды, участвующие в тонкой настройке амплитуды защитного сигнала. Эти пептиды скрыты в структуре специализированных

предшественников или функциональных белков и узнаются специализированными рецепторами на поверхности клеток.

2.2.1.1 Антимикробные пептиды

Антимикробные пептиды (АМП) являются важным компонентом иммунного ответа растений разных филогенетических групп, часто являясь первой линией защиты против фитопатогенов. Обычно АМП имеют размер 20-50 аминокислотных остатков и обладают рядом схожих особенностей, таких как положительный заряд и амфипатичность [96]. Предполагается, что амфипатичность антимикробных пептидов позволяет положительно заряженной части молекулы взаимодействовать с отрицательно заряженной мембраной, таким образом, нарушая избирательную проницаемость или полностью разрушая мембраны [97].

АМП растений делят на несколько семейств, основываясь на схожести аминокислотных последовательностей, наличии остатков цистеина и дисульфидных связей и определенной третичной структуры [97]. К основным семействам цистеин-богатых антимикробных пептидов растений относят дефензины, тионины, гевеин- и ноттин-подобные пептиды, липид-переносящие белки и ряд других. Кроме того, известен ряд антимикробных пептидов, обогащенных другими аминокислотами – глицин-богатые пептиды (Pg-AMP1), глицин- и гистидин-богатые шеферины и пептиды менее 10 аа, такие как Cn-AMP1 и Cr-ACP1 [98].

К наиболее изученным консервативным антимикробным пептидам относятся дефензины, обнаруженные в большинстве филогенетических групп растений, а также в клетках животных (Stotz *et al.* 2009). Дефензины взаимодействуют с сфинголипидами и фосфолипидами грибов или бактерий, нарушая целостность мембран и запуская каскады реакций, приводящие к гибели клеток. Помимо этого, дефензины также могут участвовать в ответе на абиотический стресс и ингибировать рост корней [99].

Катионные АМП, относящиеся к классу тионинов, характеризуются размером в 45-48 аа с тремя-четырьмя дисульфидными связями. Они были обнаружены у однодольных и двудольных растений в семенах, листьях и корнях [97]. Гидрофобные тионины токсичны для бактерий, грибов, растительных и животных клеток и предположительно взаимодействуют с мембранами, вызывая их лизис [100].

Гевеин-подобные пептиды — это основные молекулы размером 29-45 аа с 3-5 дисульфидными связями, которые обладают сходством с гевеином - лектином из латекса гевеи. Гевеиновый домен обладает способностью связываться с хитином, благодаря чему гевеин-подобным АМП приписывается роль в защите растений от фитопатогенных грибов и насекомых. Как и ряд других цистеин-богатых АМП, гевеин-подобные пептиды процессируются из предшественников, содержащих сигнал экспорта.

Липид-переносящие белки (LTP) имеют размер 70 и 90 аа, относясь к большому семейству защитных белков растений. Они делятся на два класса: LTP1s (MW=9 kDa) и LTP2s (MW=7 kDa), которые отличаются по размеру и аминокислотному составу [101]. Липид-переносящие белки ингибируют рост грибов и некоторых бактериальных патогенов и являются одним из основных компонентов защитных реакций у растений. Липид-переносящие белки были идентифицированы в большом количестве видов, включая сельскохозяйственно-важные, такие как ячмень, кукуруза, шпинат, виноград, пшеница, лук [97]. Они синтезируются в виде белков-предшественников, содержащих сигнальный пептид и “зрелую” последовательность с 8 остатками цистеина.

Помимо антимикробных пептидов, кодируемых специализированными генами, ряд исследований указывает на то, что последовательности функциональных белков могут кодировать биоактивные пептиды, в том числе, обладающие антимикробными свойствами [102,103]. Потенциально антимикробные последовательности могут быть расположены внутри одного или нескольких регионов полипептидной цепочки белка. Высвобождаясь при гидролизе белков, эти пептиды могут оказывать ингибирующее действие на ряд фитопатогенов. Для

предсказания антимикробного потенциала пептидов используют *in silico* подходы, которые основаны на различных алгоритмах машинного обучения. На основании аминокислотной последовательности белков и их физико-химических свойств, таких как заряд и гидрофобность, можно предсказать потенциальные антимикробные мотивы внутри длинных полипептидных цепей белков [102,104]. Показано, что такие пептиды могут играть важную роль в защитных реакциях у растений и животных. К примеру, в клетках животных биоактивный пептид брадикинин процессируется из высокомолекулярного белка киногенина путем протеолитической деградации, а антимикробные пептиды слизистой оболочки кожи лягушек выщепляются из консервативного препропептида в результате определенного сигналинга прогормона [105]. Растительные белки также могут служить источником антимикробных пептидов. К примеру, в клетках сои было обнаружено, что энзиматический протеолиз фермента D-мио-инозитол-3-фосфат синтазы приводит к генерации антимикробного пептида, который обеспечивает резистентность к спорам азиатской ржавчины (*Phakopsora pachyrhizi*) [106]. Интересно отметить, что трансформация клеток *Nicotiana tabacum*, *Solanum tuberosum* и *Brassica rapa* антимикробными пептидами лягушки *esculentin-1*, *dermaseptin SI* и *hCAP18/LL37* способствовала приобретению устойчивости к патогенным грибам и бактериям [107,108]. Дальнейший поиск и изучение антимикробных пептидов, которые являются фрагментами белков и генерируются путем протеолитической деградации, необходим для понимания пластичности генома и принципов биологического разнообразия.

2.2.1.2 Пептиды – регуляторы иммунного ответа (фитоцитокнины)

Регуляторные иммунные пептиды, образующиеся в результате повреждающего воздействия на клетку, сигнализируют окружающим клеткам об опасности. Установлено, что такие пептиды играют значительную роль в запуске и регуляции амплитуды иммунного ответа у растений [21]. На сегодняшний день известно около 15 различных семейств пептидов, участвующих в ответе на биотический и абиотический стрессы. Все они выщепляются из белков-предшественников

размером свыше 50 а.о. Среди наиболее изученных пептидов растений, участвующих в ответе на биотический стресс, можно отметить системин, гидроксипролин-богатый системин и элиситорные пептиды растений (PEPs).

Системин является первым сигнальным пептидом, обнаруженным у растений. Данный 18-а.о. пролин-богатый пептид выщепляется из неактивного 200-а.о. предшественника и регулирует ответ на повреждения у томатов [109]. Прекурсор PROSYSTEMIN локализуется в цитоплазме и найден у растений семейства Пасленовые, за исключением табака. Было показано *in vitro*, что субтилизин-подобная протеаза фитаспаза гидролизует просистемин, приводя к выщеплению системина [110]. Согласно недавнему исследованию, системин узнается рецептором SYR1, относящимся к классу рецептор-подобных LRR киназ, что приводит к активации системного ответа, заключающегося в продукции ингибиторов протеаз, индукции образования активных форм кислорода и биосинтезе этилена [111]. Также системин стимулирует вход ионов кальция в клетку, фосфорилирование митоген-активированных киназ MAPK и усиление биосинтеза жасмоновой кислоты.

Одной из самых распространённых и наиболее хорошо изученной среди различных растительных таксонов является группа элиситорных пептидов – PEPs. AtPEP1 является лигандом двух рецепторов – PEP РЕСЕРТОР1 и 2 (PERP1/2), связывание с которыми при помощи ко-рецептора Bak1 запускает каскад иммунных реакций [112]. Было показано, что синтетический PEP повышает устойчивость растений к полубиотрофным бактериям и к некротрофным грибам [113]. PEP1 регулирует биосинтез иммунных сигналов, таких как жасмоновая и салициловая кислоты, синтез активных форм кислорода и вход ионов кальция в клетку. Также было показано, что PEP1 усиливает защитный ответ, индуцируемый консервативными микробными сигналами. Недавно было показано, как происходит выщепление активного 23-а.о. пептид арабидопсиса AtPEP1 из 92-а.о. прекурсора PROPEP1, прикрепленного к вакуолярной мембране клетки, при помощи метакаспазы 4 (МС-4; Рис. 2.4).

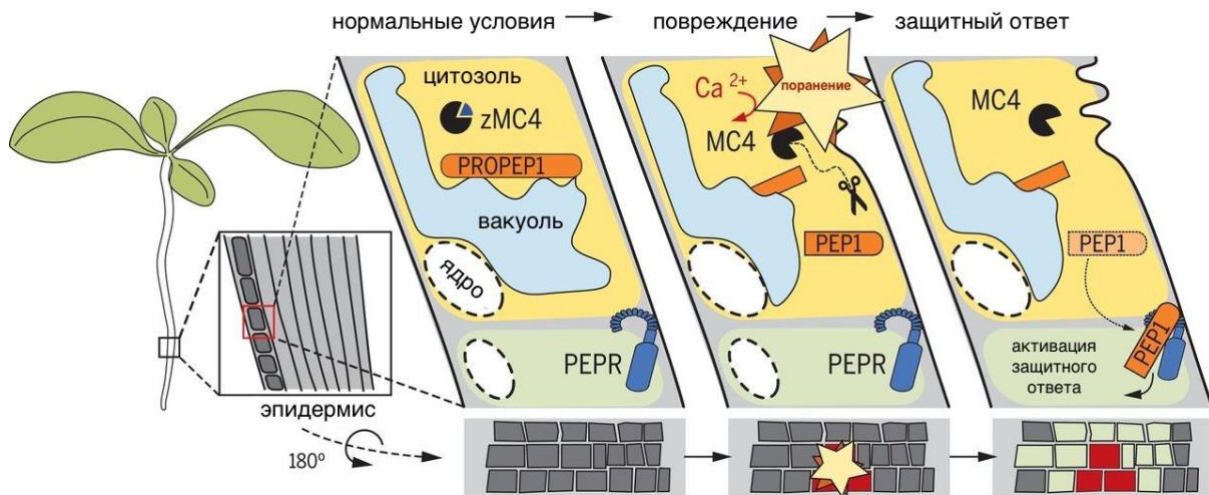


Рисунок 2.4 Схема работы фитогормона PEP. Неповрежденные эпидермальные клетки корня содержат неактивный зимоген MC4 (zMC4) в цитозоле и PROPEP1, прикрепленный к вакуолярной мембране. Повреждение вызывает приток ионов Ca^{2+} в цитозоль, активируя MC4, который расщепляет PROPEP1 и высвобождает PEP1. Затем PEP1 может диффундировать в соседние клетки и связываться с рецептором PEPR, чтобы активировать защитный ответ (внизу, светло-зеленые клетки). Высвобождение PEP1 происходит только в клетках, утративших целостность плазматической мембраны (внизу, красные клетки). Адаптировано из [22].

Обнаружено, что при повреждении клетки происходит вход ионов кальция, они же, в свою очередь, активируют MC-4, которая в обычных условиях неактивна [22]. По-видимому, этот механизм не является универсальным – такая активация происходит в листьях, но не в корнях. Кроме того, у арабидопсиса обнаружено 9 метакаспаз и 8 генов PROPEP, что может говорить о сложной системе контроля активации данного класса пептидов.

Оказалось, что не все сигнальные пептиды можно отнести к специфическим семействам, индуцирующим специфический иммунный ответ, такой как синтез активных форм кислорода и активацию синтеза стрессовых фитогормонов. У кукурузы был обнаружен 17-а.о. сигнальный пептид Zip1, который высвобождается протеазой PLCPs из 137-а.о. прекурсора PROZIP1 после действия салициловой кислоты [114]. Данный пептид активирует синтез генов, отвечающих на действие салициловой кислоты, усиливая таким образом сигнал данного гормона [114]. Синтетический Zip1 ингибирует развитие биотрофных грибов, однако повышает чувствительность к некротрофным при заражении.

Ряд недавних исследований показал, что функциональные белки растений могут содержать скрытые пептидные сигналы - криптиды, которые играют важную роль в иммунном ответе у растений. Такие пептиды играют роль сигнальных молекул, сигнализирующих о повреждениях клетки - DAMP (damage-associated molecular patterns). Первый такой криптический пептид, участвующий в регуляции иммунного ответа у растений, был обнаружен у представителей семейства бобовые [115]. Инсептин (ICDINGVCVDA) высвобождается из γ -субъединицы хлоропластной АТФ-синтазы в желудочно-кишечном тракте гусениц *Spodoptera frugiperda*. Инсептины – это группа пептидов размером 11-13 аа, каждый из которых содержит дисульфидную связь. Asp-3, Asp-20, Cys-8 и аланин на С-конце считаются необходимыми для биоактивности инсептинов. Инсептин образуется в кишечнике гусеницы *Spodoptera frugiperda* и затем с оральным секретом попадает обратно в растения [116]. Этот пептид активирует прямой и непрямой защитный ответ, который включает синтез стрессовых гормонов, а также летучих компонентов (индолы и метил салицилат), привлекающих природных врагов *S. frugiperda*. Кроме того, инсептин активирует синтез этилена и других защитных гормонов растений – салициловой и жасмоновой кислот. Обработка синтетическим инсептином также приводит к продукции летучих органических соединений, таких как индол и метил-салицилат, привлекающих насекомых - естественных врагов *S. frugiperda* и, таким образом, обеспечивается непрямой защитный ответ. Вместе с тем, инсептины индуцируют экспрессию ингибиторов протеаз и синтез коричной кислоты, которая является предшественником в многих защитных метаболитов у растений. Несмотря на то, что пептид также был обнаружен в оральном секрете гусениц *S. frugiperda* при поедании растений табака и кукурузы, индукция защитного ответа наблюдалась только у растений родов вигна и фасоль, принадлежащих к семейству бобовых. Предполагается, что инсептины являются продуктами неспецифического гидролиза, на что указывает биологическая активность целого спектра пептидов разной длины, выщепляющихся из данного участка АТФ синтазы. Предполагается, что рецептором инсептинов является

лейцин-богатая рецептор-подобная киназа INR, недавно обнаруженная у вигны [117].

В 2010 году было показано, что 789-а.о. субтилизин-подобная протеаза G1y1a18g48580 из сои является прекурсором для 12-а.о. пептида GmSubPep [118]. Данный 12-а.о. пептид находится в РА (protease-associated) домене пептидазного региона S8 последовательности внеклеточной субтилизин-подобной протеазы (ген G1y1a18g48580), уникальной для бобовых. Аргинин в позиции 10 и гистидин в позиции 12 необходимы для биологической активности пептида. Предполагается, что субтилаза высвобождается в апопласт растений при атаке патогена и GmSubPep усиливает экспрессию генов, участвующих в защитных реакциях у растений. Так, обработка наномолярными концентрациями синтетического пептида усиливает экспрессию генов Суo93A1 и achs (синтез фитоалексинов), Chib-1b (хитиназа), PDR12 (АТФ-связывающий транспортер). Интересно, что обработка GmSubPep растений томата, табака и арабидопсиса не вызывала индукции генов защитного ответа, указывая на специфичность данного пептидного сигнала [118].

Еще одним примером пептида, прекурсором которого является белок с установленной функцией, является CAPE1 [119]. Этот 11-а.о. пептид выщепляется из 159-а.о. белка PR-1b (PATHOGENESIS-RELATED PROTEIN 1b) у томата при поранении [119] и обнаружен с помощью методов масс-спектрометрии в пептидоме листьев, обработанных метил-жасмонатом. Данный пептид индуцирует иммунный ответ к вредителям и бактериальным фитопатогенам, а гомологичный пептид CAPE-PR1 увеличивает устойчивость растений арабидопсиса к солевому стрессу [120]. В отличие от GmSubPep, CAPE1-подобные пептиды с характерным консервативным пролин-богатым мотивом P_xGN_{xxxxxx}P_Y были найдены у *Vitaceae*, *Solanoideae*, *Fabaceae*, *Brassicaceae*, and *Nicotianoideae*.

Существуют доказательства того, что негистоновые ядерные высококонсервативные белки амфотерины могут участвовать в ответе на биотический стресс. Было показано, что AtHMGB3 высвобождается в апопласт и участвует в активации иммунного ответа [121]. Этот сигнал схож с тем, который индуцируется пептидами PEP.

Таким образом, с каждым годом растет количество идентифицированных биоактивных пептидов, являющихся продуктами специфического гидролиза функциональных белков. Это указывает на то, что функциональные белки, содержащие криптоические пептидные последовательности, являются дополнительным источником регуляторных молекул в условиях биотического стресса.

2.2.2 Пептиды - регуляторы процессов роста и развития у растений

За период времени, прошедший с момента обнаружения первого пептидного гормона растений [122], исследователи доказали, что, подобно животным, растения используют пептидные лиганды как сигнальные молекулы для регуляции роста и развития (**Рис. 2.3**), а также ответа на внешние стимулы [25,123,124]. Эти пептидные гормоны могут быть поделены на две большие категории: пептиды с пост-трансляционными модификациями и цистеин-богатые пептиды. Основными пост-трансляционными модификациями являются сульфирование тирозина, гидроксилирование пролина и добавление остатков арабинозы к гидроксипролину. К пептидным семействам с пост-трансляционными модификациями относят C-TERMINALLY ENCODED PEPTIDE (CEP), CLAVATA3/EMBRYO SURROUNDING REGION RELATED (CLE) и ROOT GROWTH FACTOR (RGF)/GOLVEN (GLV)/CLE-LIKE (CLEL). Размер цистеин-богатых пептидов намного больше, чем пептидов с пост-трансляционными модификациями, они положительно заряжены и содержат от 4 до 16 цистеиновых остатков. К семействам цистеин-богатых пептидов относятся RALF (RAPID ALKALINIZATION FACTOR), EPIDERMAL PATTERNING FACTOR1 (EPF1), EPF2, TAPETUM DETERMINANT1 (TPD1), STOMAGEN/EPF-LIKE9 (EPFL9) и EARLY NODULIN.

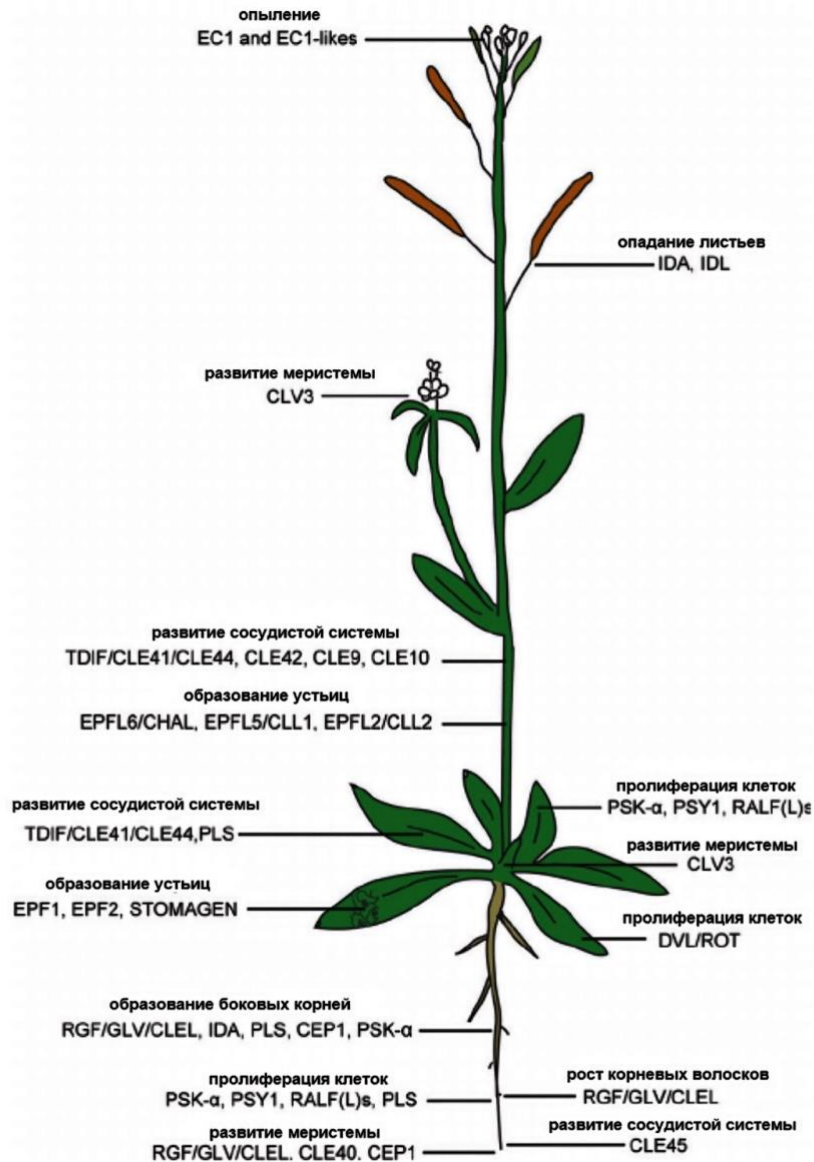


Рисунок 2.5 Разнообразие пептидных гормонов у растений. Адаптировано из [2]

Появление большинства пептидных семейств связано с одним из наиболее значимых событий в земной биосфере - выходом растений на сушу, который сопровождался появлением множества специфических инноваций. Необходимо отметить несколько ключевых пептидных семейств, которые появились у наземных растений и регулируют ключевые этапы роста и развития. Например, одной из ключевых адаптаций к наземному существованию является появление устьиц [125]. У арабидопсиса семейство пептидов EPFL регулирует плотность расположения устьиц, причем EPF1, EPF2 и EPFL6 [126,127] являются

негативными регуляторами, а STOMAGEN/EPFL9 - позитивным [128]. Ожидается, что данное семейство пептидов не было обнаружено у одноклеточных водорослей [129]. Тем не менее, ортологи данных пептидов были найдены у несосудистых растений, таких как печеночники (*Marchantia polymorpha*), плауны (*Selaginella moellendorffii*) и мхи *Physcomitrella* (*Physcomitrium patens*). Примечательно, что, например, у *P. patens* устьица появляются только в момент развития спорофита [125].

Еще одной важной инновацией, связанной с выходом растений на сушу, является появление проводящих тканей, позволяющих перемещать воду и питательные вещества. В развитии проводящих тканей большую роль играют семейство CLE пептидов, особенно TDIF (CLE41/CLE44), CLE42 и CLE45 пептиды [130,131]. Это семейство пептидов возникло раньше появления развитой проводящей системы, поскольку ортологи некоторых CLE пептидов обнаружены у бриофитов. Однако прямых ортологов CLE пептидов, участвующих в регуляции роста проводящих тканей, у них найдено не было [132]. Было показано, что у мха *P. patens* пептиды CLAVATA регулируют плоскость деления стволовых клеток при формировании гаметофоров из протонемы [133]. Это говорит об их важной роли в переходе от протонемы (двумерный тип роста) к трехмерным структурам - гаметофору. У плаунов, имеющих ксилему и флоэму, не найдено ортологов TDIF (CLE41/CLE44) пептидов, следовательно, участвовать в регуляции развития сосудистой системы у этой растительной клады могут другие члены этого семейства [132]. Предполагается, что гены, кодирующие CLE пептиды, могли появиться ещё у зеленых водорослей. Например, предполагаемые ортологи были найдены у *Chlamydomonas reinhardtii* [132,134]. Рецепторы CLE пептидов были идентифицированы и изучены у различных видов растений, в том числе, ортологи CLV1 и CLV2 найдены в геномах *S. moellendorffii* and *M. polymorpha* [135,136].

Еще одним известным примером пептидного семейства, регулирующего рост и развитие растений, являются RALF пептиды. Предполагается, что роль этих пептидов заключается в регуляции размера клеток в процессе полярного роста, в частности, они регулируют рост корней [137]. Обработка синтетическим пептидом

RALF1 проростков арабидопсиса приводила к ингибированию роста гипокотилия. Ингибирование гипокотилия также было обнаружено при обработке синтетическими RALF19, 22, 23, 24, 31, 33 и 34 [138]. Напротив, рост гипокотилия был усилен у растений арабидопсиса с нокдауном RALF1 пептида. Показано, что пептид RALF1 регулирует экспрессию ферментов, моделирующих клеточную стенку, таких как PROLINE RICH PEPTIDE1 (PRP1), PROLINE RICH PEPTIDE2 (PRP2), HYDROXYPROLINE-RICH GLYCOPROTEIN2 (HRPG2) и XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE22 (XTH22)/TOUCH4 (TCH4).

Пептиды данного семейства и соответствующие рецепторы были идентифицированы у множества растительных видов [137,139]. Интересно, что ортологи пептидов данного семейства были обнаружены у растений, не имеющих корневой системы, таких как мох *P. patens* и печеночник *M. polymorpha* [136,140]. Однако роль пептидов RALF у современных бриофитов изучена слабо. Например, у мутантов *M. polymorpha* по одному из рецепторов RALF - MrTHESEUS был нарушен процесс роста и развития ризоидов [141]. Это указывает на то, что функции данных пептидов изначально не были связаны с регуляцией функции корневой системы. Вероятно, процесс элонгации клеток при полярном росте, также как и мониторинг целостности клеточной стенки, — это ключевые процессы, контролируемые пептидами RALF.

2.2.3 Функции пептидов - продуктов трансляции коротких открытых рамок считывания

Известно, что в количество коротких открытых рамок считывания в одном геноме достигает нескольких миллионов. Однако очень немногие из них транслируются и являются функциональными [68]. На данный момент количество известных функциональных пептидов или микробелков, кодируемых кОРС, существенно ниже числа тех, для которых была подтверждена транскрипция и трансляция. Следует отметить, что транскрипция и/или трансляция некодирующих участков ДНК может приводить к появлению функциональных молекул. Например,

случайным образом транслирующиеся кОРС могут продуцировать биохимически активные пептиды. Так, в работе Neme et al. (2017) было показано, что экспрессия случайных последовательностей размером 65-аа у *E.coli* в 25% случаев приводит к ускорению роста клеток и в 52% процентах случаев к ингибированию [142]. Интересно, что в некоторых случаях транскрипция самой РНК влияла на рост бактерий сильнее, чем трансляция соответствующих микробелков, подтверждая предположение о том, что активная РНК может предшествовать активным пептидам/белкам при *de novo* эволюции генов. В другой работе, выполненной на арабидопсисе, изучали эффект трансляции случайных 16-а.о. циклических пептидов на рост и развитие растений арабидопсиса [143]. Из 2000 растений, трансформированных соответствующими генетическими конструкциями, 12 показали изменения в фенотипе, подтверждая активность соответствующих пептидов. Среди выявленных пептидов обнаружены влияющие на фотосинтез, цветение и ответ на красный свет. Данные эксперименты показали, что транскрипция и/или трансляция некодирующих участков ДНК может приводить к появлению функциональных молекул.

Действительно, кОРС, кодирующие функциональные пептиды, были обнаружены как на мРНК, так и на целом ряде транскриптов, традиционно считавшихся некодирующими, таких как при-микроРНК [144], длинные некодирующие РНК [68,145], кольцевые РНК [146], теломеразная РНК человека [147], рРНК митохондрий [148].

2.2.3.1 Функциональные пептиды, кодируемые первичными транскриптами микроРНК

МикроРНК (миРНК) — это класс некодирующих молекул длиной 22 нуклеотида, которые участвуют в пост-транскрипционной регуляции мРНК [149]. В ряде исследований было показано, что предшественники микроРНК могут содержать короткую рамку считывания на 5'-конце транскрипта. Эти рамки могут потенциально кодировать пептиды (микроРНК-кодируемые пептиды, miPEP) [144]. Трансляция пептидов с при-миРНК транскриптов и их функциональная роль были подтверждены у арабидопсиса [150,151], люцерны [150,152], винограда [153].

Было установлено, что сверхэкспрессия или добавление синтетического пептида приводит к увеличению транскрипции соответствующей миРНК и, как следствие, изменению представленности регулируемых ею мРНК [154]. Размер известных miPEP варьирует от 7 до 50-а.о. Однако механизм регуляции транскрипции предшественника миРНК пока неизвестен. Возможно, пептиды действуют как транс-факторы, усиливая в ядре транскрипцию соответствующих при-миРНК. Однако связывание miPEP с белками комплекса, иницирующего транскрипцию, не обнаружено. Тем не менее, биологические функции miPEP у растений были изучены наиболее подробно среди всех пептидов, транслируемых с некодирующих РНК. Так, miPEP171b (20aa) и miPEP165a (18 aa), которые были найдены у люцерны и арабидопсиса, регулируют ветвление корней при сверхэкспрессии или экзогенном применении [150]. В результате происходит уменьшение ветвления корневой системы и усиливается рост главного корня. Обработка синтетическими miPEP172с у сои увеличивает количество клубеньков [152]. У винограда экзогенное применение vvimiPEP171d1 усиливает экспрессию vvi-MIR171d, что приводит к стимуляции корнеобразования. Еще один хорошо изученный функциональный пептид - miPEP858a, размер которого 44-а.о., регулирует экспрессию генов фенилпропаноидного пути и сигналинга ауксина [151]. Растения с редактированной последовательностью miPEP858a, а также сверхэкспрессирующие данный пептид, имели нарушения в развитии и накапливали флавоноиды в тканях.

2.2.3.2 Функциональные пептиды, кодируемые длнкРНК

Еще одним широко известным источником пептидов, кодируемых короткими открытыми рамками считывания, являются транскрипты длинных некодирующих РНК. Функциональная роль таких пептидов была наиболее полно изучена у человека и животных. Например, анализ тканеспецифичной транскрипции длинных некодирующих РНК, совмещенный с предсказанием эволюционно консервативных коротких открытых рамок считывания, позволил

идентифицировать два пептида – dwarf open reading frame (DWORF) и myoregulin (MLN), участвующих в регуляции сокращения сердечной мышцы [82,155].

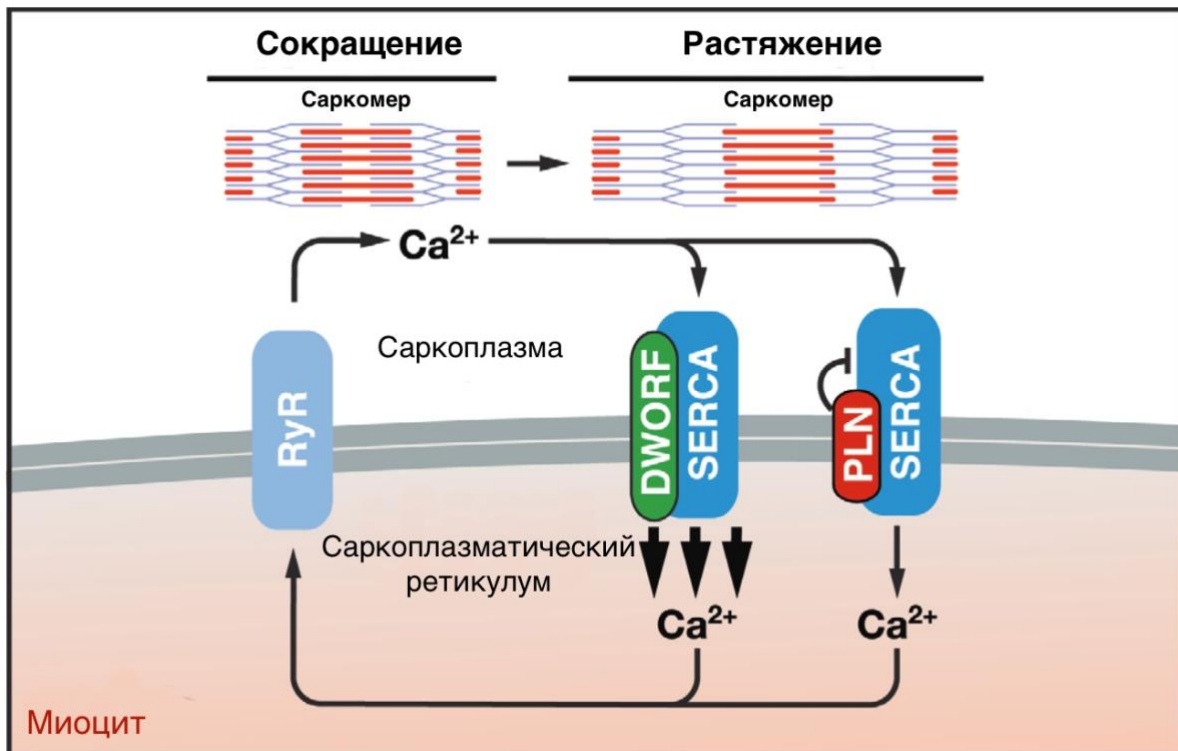


Рисунок 2.6 Модель, описывающая функции 34-а.о. пептида DWORF, кодируемого длнкРНК. DWORF локализуется на мембране саркоплазматического ретикулума, где он усиливает активность транспортера ионов кальция - SERCA, вытесняя его ингибиторы (фосфоламбан, сарколипин и миорегулин). Адаптировано из [82].

Пептид HOXB-AS3, супрессирующий рост опухоли рака прямой кишки, также был обнаружен при анализе экспрессии антисенс-некодирующих РНК в различных тканях методами секвенирования второго поколения с последующим предсказанием коротких рамок на выявленных транскриптах [71]. Ещё один пептид - SPAR - участвует в регуляции работы комплекса mTORC1 и влияет на регенерацию мышц [156]. Установлено, что SPAR - это трансмембранный 90-аа микробелок, который взаимодействует с субъединицами V-АТФ-азного комплекса лизосом. Взаимодействие с V-АТФ-азой регулирует образование комплекса с белком Ragulator, активирующего mTORC1 киназу в ответ на изменение пула свободных аминокислот. Нокаут по этому микробелку у мышей характеризовался

ускоренной регенерацией сердечной мышцы [156]. Функциональные пептиды кодируемых кОПС, которые локализованы на длнкРНК, также участвуют в регуляции клеточной пролиферации [157], пролиферации опухолей [158], апоптозе и образовании аутофагосом [147], активности митохондриальной дыхательной цепи [159], презентации антигенов [160], половом размножении [161].

В сравнении с животными, у растений функциональная роль пептидов, кодируемых длнкРНК, практически не изучена. Только несколько пептидов, кодируемых транскриптами, не имеющими рамок считывания длиной более 100 аа, были функционально охарактеризованы у растений к настоящему времени. Это 36-аа POLARIS (PLS; [162]), 53-аа ROTUNDIFOLIA4 (ROT4; [163]), 51-а.о. ROT18/DLV1 [164], EARLY NODULIN GENE 40 (ENOD40; 12-, 24-аа; [165]), 25-аа KISS OF DEATH (KOD; [166]), 84-аа BRICK1 (BRK1; [167]), 97-аа Zm-908p11 [168], 89-аа Zm-401p10 [169], 10-аа OSIP108 [170].

Первыми обнаруженными у растений функциональными пептидами, которые кодируются короткими рамками считывания, расположенными на транскрипте, предварительно аннотированом как длнкРНК, являются ENOD40 [165]. Оказалось, что это полицистронный транскрипт, который кодирует два пептида, размером 12 и 24 аминокислоты. Показано, что пептиды взаимодействуют с сахарозосинтазой, регулируя ее активность в клетке и контролируя таким образом образование клубеньков у бобовых.

Короткая рамка считывания, кодирующая функциональный 36-а.о. пептид POLARIS, расположена на коротком 500 нт транскрипте, который индуцируется ауксином и репрессируется этиленом [162,171]. Нокаутные по данному пептиду растения арабидопсиса имели измененный фенотип: укороченные корни и сниженное количество жилок в листьях. Кроме того, корни нокаутных растений были гиперчувствительны к цитокинину, но менее чувствительны к действию ауксина. Эктопическая экспрессия данного пептида приводит к увеличению количества жилок на листе и снижению эффекта ингибирования роста корней при обработке цитокинином. Предполагается, что функции PLS связаны с эффектом на динамику цитоскелета и сигналинг ауксина, что в свою очередь, влияет на рост и

развитие корневой системы. Однако белки-партнеры данного пептида не были обнаружены.

Показано, что пептиды Brk1 и ROT4 также связаны с регуляцией морфогенеза. У кукурузы рецессивная мутация в гене Brk1 приводила к дефектам в развитии эпителия листа [172]. ROT4 регулирует полярный рост клеток и развитие листьев [163]. У арабидопсиса были найдены еще два пептида, кодируемые короткими рамками считывания и участвующие в органогенезе. Это пептиды ROT18/DLV1 и KOD. Таким образом, несмотря на практически отсутствующие данные о механизмах действия и скудность информации о функциях таких пептидов у растений, можно отметить, что они вовлечены в регуляцию целого ряда важнейших физиологических процессов.

2.2.3.3 Функциональные пептиды, кодируемые kOPC в мРНК

Короткие рамки считывания, потенциально кодирующие функциональные пептиды, были найдены в различных регионах мРНК - 5'-лидерных последовательностях, 3'-трейлерных последовательностях, а также они могут пересекаться с кодирующей областью, но со сдвигом рамки считывания относительно белок-кодирующей последовательности.

Короткие рамки считывания, расположенные в 5'- нетранслируемых регионах мРНК (upstream open reading frame, uORF), представляют собой многочисленный класс потенциально транслирующихся kOPC [68]. Примерно 50% аннотированных мРНК как у человека, так и у мыши содержат одну или две uORF [173], в то время как у растений данный показатель составляет ~30% [174]. Однако ряд недавних исследований показал, что помимо AUG, трансляция uORF может инициироваться с альтернативных старт-кодонов, таких как CUG, UUG и GUG [175,176]. Таким образом, количество транслирующихся uORF может быть выше, чем считалось ранее. Показано, что действие стабилизирующего отбора является одной из сил, регулирующих появление старт кодонов AUG в 5' - лидерной последовательности мРНК [74].

Несмотря на низкую эффективность инициации трансляции большинства uORF [177], продукты трансляции этого класса kORC были найдены у разных групп организмов [69,80,178,179]. Однако, по-видимому, большинство kORC этого класса кодируют короткие неконсервативные пептиды, участвующие в регуляции трансляции нижележащей белок-кодирующей рамки [72,74]. Показано, что такая регуляция играет важную роль при реализации программ развития [180,181], а также при ответе на стресс [182,183]. Например, регуляция метаболических путей у растений может быть связана с трансляцией uORF. У арабидопсиса мРНК *AdoMetDC* (аденозилметионин декарбоксилазы) содержит две консервативные перекрывающиеся короткие рамки длиной 12 (рамка 1) и 156 нуклеотидов (рамка 2) в 5'-нетранслируемом регионе. Когда уровень полиаминов в клетке низкий, короткий пептид, транслирующийся с рамки 1, блокирует трансляцию рамки 2, что обеспечивает трансляцию соответствующего белка, участвующего в биосинтезе полиаминов. В случае высокого уровня полиаминов в клетке трансляция рамки 2 репрессирует синтез белка, что приводит к уменьшению биосинтеза полиаминов [184]. Помимо полиаминов, схожие механизмы используются у растений для контроля метаболизма сахарозы и витамина С [185,186]. В данных исследованиях было показано, что uORF, участвующие в регуляции метаболических путей, большей частью консервативны и регулируют концентрацию соответствующих метаболитов в клетке. В данном случае важен процесс трансляции коротких пептидов, а не образующийся пептид. Такой механизм регуляции экспрессии белок-кодирующих генов был найден как у эукариотических, так и прокариотических организмов [187].

Примером функционального микробелка, который кодируется короткой рамкой, расположенной в 5'-UTR гена *AtTTM3* (RIPHOSPHATE TUNNEL METALLOENZYME 3), является 65-а.о. микробелок *AtCDC26*. В данном случае ~1200 нуклеотидный транскрипт является бицистронной мРНК, которая кодирует белок *AtTTM3* и микробелок *AtCDC26*. Интересно, что трансляция белок-кодирующей рамки *TTM3* увеличивает связывание транскрипта с полисомами и, таким образом, косвенно влияет на трансляцию *CDC26* [188]. Согласно данным

интерактивных экспериментов, микробелок AtCDC26 участвует в формировании комплекса APC/C. Интересно, что функциональной связи между AtCDC26 и AtTTM3 обнаружено не было, хотя известно, что неорганические полифосфаты способствуют выходу из клеточного цикла и связь между трансляцией этих двух рамок может существовать. Однако биохимических взаимодействий между ними обнаружено не было [188].

Помимо коротких рамок в 5'-лидерной последовательности мРНК, трансляция kOPC в 3'-трейлерных последовательностях мРНК была подтверждена с помощью рибосомального профайлинга и масс-спектрометрического анализа [69,189,190]. Однако функции соответствующих пептидов пока не изучены.

2.3 Методы анализа пептидомов

2.3.1 Использование масс-спектрометрии для идентификации пептидов

Пептидомика — это специфический раздел протеомики, который занимается идентификацией пептидов, присутствующих в биологическом образце. Термин “пептидом” как совокупность всех пептидов в образце впервые был определен в 2001 году [191]. В отличие от протеомного анализа, пептидомика является менее развитой областью. На сегодняшний момент идентификация белков и нативных пептидов, экстрагированных из тканей и биологических жидкостей, практически полностью основана на использовании масс-спектрометрического (МС) анализа. Однако существуют принципиальные различия между аналитическими методами идентификации белков и пептидов [192]. Методы идентификации белков в протеоме основаны на энзиматическом расщеплении белков с помощью специфических ферментов, таких как трипсин. Для анализа нативного пептидома подобные этапы не требуются, и после выделения образцы непосредственно можно идентифицировать с помощью масс-спектрометрии. Однако представленность эндогенных пептидов в биологических образцах низка, в то время как большое количество белков, липидов и солей влияет на эффективность их идентификации [193]. Например, часть низкомолекулярных пептидов, особенно при высоких

концентрациях белка, может быть потеряна в процессе осаждения растворителем и ультрафильтрации [194].

Пептидомный анализ обычно включает следующие этапы: сбор образцов, экстракция пептидов, фракционирование, МС анализ, идентификация пептидов и сбор данных [192]. При этом клеточные и внеклеточные пептидные пулы могут представлять совокупность пептидов, образующихся в тканях и биологических жидкостях как в нативных условиях, так и в результате артефактов пробоподготовки. Кроме того, методы пробоподготовки [40] и некоторые особенности МС анализа могут способствовать преимущественной идентификации наиболее представленных пептидов с определенными физико-химическими свойствами.

Таким образом, одной из важнейших задач пептидомного анализа является предотвращение деградации образцов с образованием большого количества пептидных артефактов. Это особенно важно для анализа тканей, где может происходить активная и быстрая деградация белка после гибели клеток. Для уменьшения неспецифического протеолиза были разработаны различные подходы, такие как быстрый нагрев микроволновым излучением [195], добавление ингибиторов протеаз или органических растворителей. Проблема с использованием ингибиторов протеазной активности заключается в их потенциальном влиянии на состав нативного пептидома ткани или биологических жидкостей. Так, для некоторых образцов, таких как моча, которые достаточно стабильны при криогенном хранении, большинство исследований проводят без ингибирования протеазной активности [196].

Поскольку эндогенные пептиды имеют низкую представленность и широкий динамический диапазон, для последующей очистки нативных пептидов от различных контаминантов используются различные стратегии, а именно: осаждение белков органическими растворителями [197], дифференциальную солюбилизацию [198], ультрафильтрацию [199], твердофазную экстракцию [200] и т.д. Осаждение органическими растворителями представляет собой наиболее быстрый и экономичный метод удаления белков из образцов.

Из-за сложности и неоднородности пептидных образцов, фракционирование проб перед МС анализом часто является необходимым этапом. Использование разных способов фракционирования - например, твердофазная экстракция с использованием различных типов стационарных фаз, таких как обращенная фаза (RP, reverse phase), сильный катионный обменник (SCX, strong cation exchange) или хроматография гидрофильного взаимодействия (HILIC, Hydrophilic interaction chromatography), а также их сочетаний для двумерного разделения - активно используется в пептидомике [40]. При этом, использование различных сорбентов для фракционирования пептидных образцов напрямую влияет на состав идентифицируемых пептидов [41]. Несмотря на то, что многомерное разделение с использованием различных стационарных фаз может значительно улучшить разрешающую способность пептидного анализа, применение обращенно-фазовой твердофазной экстракции остается наиболее популярным подходом в пептидных исследованиях, потому что требует меньше времени и обладает хорошим выходом. При этом очень гидрофильные короткие пептиды могут быть потеряны при разделении на стационарной фазе C18 [201].

Недавние исследования пептида клеток НК-2, обработанных TGF- β 1, показали, что индексы GRAVY, указывающие на гидрофобность пептидной последовательности [202], были значительно ниже нуля у значимо меняющихся пептидов, что позволяет предположить, что большинство из них - гидрофильные пептиды [203]. Таким образом, физико-химические свойства идентифицированных в результате масс-спектрометрического анализа пептидов могут как отражать биологические тенденции, так и быть артефактом подготовки образцов и последующего анализа. В этой связи наше представление о пептидомах, основанное на МС-анализе, может быть смещено в сторону пептидов с определенными физико-химическими свойствами. При этом профилирование пептидома с помощью масс-спектрометрического анализа может быть улучшено путем оптимизации параметров фрагментации, таких как диссоциация, индуцированная столкновением (CID, Collision-Induced Dissociation), высокоэнергетическая столкновительная диссоциация (HCD, Higher Energy Collision Dissociation), диссоциация с переносом

электронов (ETD, Electron-transfer dissociation). Использование различных методов фрагментации, а также их комбинаций, может предоставить дополнительную информацию, необходимую для определения последовательности и структуры пептидов [204,205]. Например, метод диссоциации с переносом электронов позволяет проводить более интенсивную фрагментацию и, следовательно, получить дополнительную информацию для некоторых пептидов.

2.3.2 Методы анализа интерактомоов пептидов

Пептидные гормоны и фитоцитокнины действуют как лиганды соответствующих рецепторов, активируя каскады последующих реакций, включая фосфорилирование белков, и индуцируют экспрессию соответствующих генов [206,207]. Результаты функционального анализа большинства известных криптидов или микробелков, кодируемых кОРС, показали, что чаще всего их биологическая роль заключается в регуляции активности клеточных белков при прямом взаимодействии. Таким образом, анализ интерактомоов является необходимым этапом изучения молекулярных механизмов их биологического действия. На сегодняшний момент существует несколько подходов для определения белковых партнеров микробелков и пептидов. Одним из наиболее известных методов выявления бинарных белковых взаимодействий является дрожжевая двугибридная система [208]. В этом методе кодирующая последовательность предполагаемого белка-интерактора сливается с ДНК-связывающим доменом транскрипционного фактора в составе вектора таким образом, чтобы в клетках дрожжей был экспрессирован слитый белок. Транскрипционный фактор контролирует экспрессию репортерного белка (например, фермента, катализирующего специфическую реакцию). Кроме этого, конструируются векторы для трансфекции, у которых белок интереса сливается с доменом, привлекающим РНК-полимеразу того же транскрипционного фактора. Если белки взаимодействуют между собой, то при их совместной экспрессии два домена транскрипционного фактора будут сближены и транскрипция репортерного гена будет запущена. Недостатками метода являются: возможность выявления

только бинарных взаимодействий, отличие условий в дрожжевой клетке от условий в исследуемом объекте, невозможность учесть локализацию белков, трудности при работе с крупными и мембранными белками [209]. Тем не менее, помимо белок-белковых взаимодействий, данный способ используется для изучения интерактомоов пептидов. Например, с помощью метода дрожжевой двугибридной системы было показано, что пептид Zm908p11, кодируемый кОРС у кукурузы, взаимодействует с белком профилином [168].

Однако наиболее часто для определения возможных белков-партнеров пептидов/микробелков используют методы аффинной очистки, основанные на специфическом сродстве взаимодействующих макромолекул. Для этого создаются мутантные организмы, эндогенно экспрессирующие белок (пептид) интереса, слитый со специфической полипептидной последовательностью, так называемым “тэгом”. Для тэгов известно взаимодействие с высоким сродством с другой молекулой или химической группой, которая может быть иммобилизована на твердом носителе. Для анализа интерактома лизат ткани мутанта, экспрессирующего слитый с тэгом пептид, пропускают через соответствующий носитель. На сорбенте иммобилизуются химерные молекулы и взаимодействующие с ними белки, которые затем элюируются и идентифицируются с помощью масс-спектрометрического анализа. В качестве такого “тэга” могут выступать пептидные последовательности – эпитопы коммерчески доступных антител (FLAG, HA, c-тус), более крупные белки (зелёный флуоресцентный белок (GFP)), ферменты (глутатион-S-трансфераза (GST)), небольшие белки (стрептавидин (СА), белок, связывающий мальтозу (MBP)), полигистидиновые последовательности. Спектр иммобилизованных молекул и групп, таким образом, включает антитела (для пептидных эпитопов и GFP), биотин и его аналоги (для стрептавидина), глутатион (для GST), амилозу (для MBP) и Ni-аффинный сорбент (для полигистидина) [210,211]. При этом, если аффинная очистка основана на использовании пары антиген-антитело и происходит по принципу хроматографии в объёме, то такой вариант метода называют коиммунопреципитацией. Поскольку методом идентификации

взаимодействующих с пептидом белков является масс-спектрометрический анализ, это позволяет определять взаимодействия любого типа (бинарные, тройственные и т. д.), и дает возможность детекции пост-трансляционных модификаций (ПТМ) и проведения количественного анализа. К недостаткам метода коиммунопреципитации относятся: высокий уровень неспецифических взаимодействий, возможность изменения свойств целевого полипептида после присоединения “тэга”, сложности с выявлением временных и слабых белок-белковых взаимодействий [211]. С помощью метода коиммунопреципитации в ряде исследований функций кОПС человека и животных были выявлены белки-партнёры микробелков SPAR (партнёры – субъединицы V-АТФазы) [156], NoBody (партнёры – субъединицы декэпирующего комплекса) [212], CYREN (партнёры – белки, осуществляющие негомологичную репарацию ДНК) [213] и ряда других. Метод аффинной очистки может быть дополнен стабилизацией взаимодействий между белками интерактома и микробелком с помощью добавления кросс-сшивающих реагентов, взаимодействующих с аминокислотными остатками близко расположенных белков и соединяющих их ковалентно. Таким образом повышается устойчивость слабо связанных комплексов, способствующая их обнаружению, а также облегчается работа с мембранными белками. Однако необходимо помнить о правильном контроле для исключения из дальнейшего рассмотрения высокопредставленных белков, которые с высокой вероятностью образуют сшивки с целевым белком [214].

Таким образом, при помощи методов аффинной очистки были идентифицированы белки-партнёры для целого ряда микробелков (Табл. 1). Недостатком существующих методов анализа интерактома является наличие контаминантов, которые представляют собой мажорные белки клетки и существенно осложняют дальнейший анализ.

Таблица 1. Белки-партнёры и биологические функции некоторых микробелков млекопитающих. КоИП – коиммунопреципитация, АО – аффинная очистка.

| Микробелок | Объект | Метод: тип тэга | Белковые партнёры | Функция | Источник |
|------------|---------|--------------------------|---|---|----------|
| NoBody | человек | КоИП: FLAG, с- тус | Субъединиц ы декэпирующ его комплекса (EDC4, EDC4, Dcp1A, Dcp1B, Dcp2) | Дестабилиза ция Р- телец и снижение эффективн ости нонсенс- опосредова нного распада | [212] |
| SPAR | мышь | КоИП: FLAG | Субъединиц ы v-АТРазы | Ингибиров ание mTORC1- киназного комплекса | [156] |
| P155 | мышь | АО: биотин | Шаперон HSC70 | Уменьшен ие эффективн ости презентаци и антигена у АПК, воспаления | [160] |
| DWOLF | мышь | КоИП: GFP, HA | Субъединиц ы SERCA- АТРазы | Усиление работы SERCA- АТРазы за счёт снятия ингибиров ания | [82] |
| CYREN | человек | КоИП: FLAG | Ku70/80 гетеродимер | Ингибиров ание NHEJ- репарации во время S и G2 фаз клеточного цикла | [213] |

Для решения этой проблемы был создан репозиторий контаминантов под названием CRAPome – база данных, полученная при статистическом анализе идентификаций в отрицательных контрольных пробах в интерактомных экспериментах [215]. Используя соответствующий ресурс (<https://reprint-arms.org>), можно проанализировать полученные данные и отредактировать результаты полученных экспериментальных идентификаций, очистив их от основных контаминантов. К сожалению, база данных на настоящий момент доступна только для *Homo sapiens* и *Saccharomyces cerevisiae*. Это отчасти объясняет, что, несмотря на все преимущества интерактомных методов при идентификации белковых партнёров микробелков, они редко используются у растений.

3. МАТЕРИАЛЫ И МЕТОДЫ

3.1 Методики работы с растениями

3.1.1 Выращивание протонемы и гаметофоров *P. patens*

Растения *Physcomitrella* (*Physcomitrium patens* ‘Gransden 2004’, Freiburg) двух жизненных форм - протонемы и гаметофоров - культивировали в климокамере при следующих условиях: температура - 24°C, влажность - 50%, световой период 16/8 ч день/ночь. Протонему выращивали на питательной среде BCDAT с небольшими модификациями. BCDAT среда содержала 1 mM MgSO₄, 10 mM KNO₃, 45 M FeSO₄, 1.8 mM KH₂PO₄ [pH 7], микроэлементы (0.22 μM CuSO₄, 0.19 μM ZnSO₄, 10 H₃BO₃, 0.10 /μM Na₂MoO₄, 2 /μM MnCl₂, 0.23μM CoCl₂, and 0.17 /μM KI), 5 mM тартрат аммония, 0.5% (w/v) глюкозы и 0.8% (w/v) агара. Для приготовления питательной среды смешивали Ca(NO₃)₂*7H₂O – 0,8 г, MgSO₄*7H₂O – 0,25 г, FeSO₄*7H₂O – 12,5 мг, сток микроэлементов - 1 мл, фосфатный буфер – 1 мл., тартрат аммония – 0,5 г, глюкозу безводную – 5 г, агар - 1,5 г и доводили объем до 1 л. Сток микроэлементов готовили, смешивая CuSO₄*5H₂O – 55 мг, ZnSO₄*7H₂O – 55 мг, H₃BO₃ – 614 мг, MnCl₂*4H₂O – 389 мг, CoCl₂*6H₂O – 55 мг, KI – 28 мг, (NH₄)₂MoO₄*2H₂O – 25 мг и доводили дистиллированной водой до 1 л. Приготовленную питательную среду автоклавировали при 1 атм. в течении 20 мин и разливали по 10 мл в 9 см чашки

Петри в стерильных условиях. Для размножения протонемы использовали полупогруженный способ выращивания, когда биомасса находится на агаризованной среде и покрыта небольшим количеством стерильной воды.

Для экспериментов с фенотипическими анализами роста протонемы на каждую чашку Петри помещали 6 небольших участков колоний, равномерно располагая их на поверхности агара. Наблюдения за ростом колоний выполняли в течение 5 дней.

Для проведения протеомных и пептидомных экспериментов протонему размножали в жидкой питательной среде VCDAT. Для этого, небольшое количество протонемы помещали в колбы с 500 мл питательной среды. Культивирование проводили на настольном горизонтальном шейкере при 120 об/мин в течение 5 дней.

Гаметофоры *P. patens* культивировали в чашках Петри на твердой питательной среде VCD. Состав среды был аналогичен таковой при выращивании протонемы, за исключением тартрата аммония. В каждую чашку переносили 3-4 веточки гаметофора немного погружая их в агар. В одной чашке размещали 7 групп гаметофоров и инкубировали в климокамере в течение 2 месяцев.

3.1.2 Получение протопластов *P. patens*

Пятидневные нити протонемы собирали шпателем с поверхности агара и удаляли лишнюю влагу. Около 1 г ткани протонемы помещали в 14 мл 0,5% (вес/объем) раствора драйзелазы (Sigma-Aldrich, Сент-Луис, Миссури, США) в 0,48 М маннитоле и инкубировали 60 мин при постоянном покачивании в темноте. Затем суспензию фильтровали через стальную сетку 100 мкм (Sigma-Aldrich, США) и полученные протопласты инкубировали в растворе драйзелазы еще 15 минут. Затем протопласты осаждали центрифугированием в пластиковых пробирках объемом 50 мл с использованием бакет-ротора при 100 g в течение 5 мин. Далее протопласты дважды промывали 0,48 М маннитом с центрифугированием в тех же условиях и снова осаждали. Для экстракции пептидов и выделения РНК полученные протопласты замораживали в жидком азоте. В экспериментах по

трансформации подсчитывали количество протопластов с помощью камеры Горяева и отбирали необходимое количество.

3.1.3 Индукция повреждений клеточной стенки раствором драйзелазы

Собранную с чашек Петри 5-тидневную протонему помещали в 0,025% масс/об и 0,0025% масс/об растворы драйзелазы в 0,48 М маннитоле и инкубировали 60 мин при покачивании. В качестве контроля также инкубировали ткань протонемы в 0,48 М маннитола. Затем протонему осаждали центрифугированием в пластиковых пробирках объемом 50 мл с использованием бакет-ротора при 100 g в течение 5 мин и удаляли супернатант. Далее ткань дважды промывали 0,48 М маннитом с центрифугированием в тех же условиях и снова осаждали. Полученную протонему замораживали для выделения пептидома.

3.1.4 Выделение хлоропластов из *P. patens*

Хлоропласты выделяли из протопластов *P. patens*. Для этого протопласты ресуспендировали в буфере А (50 мМ HEPES-КОН, рН 7,5, 330 мМ сорбита, 2 мМ ЭДТА и 0,4 мМ фенилметилсульфонилфторида) и фильтровали через двойной слой бумаги Miracloth (Calbiochem Behring, La Jolla, CA, США). Затем фильтрат центрифугировали при 1200 g в течение 3 мин в пластиковых пробирках объемом 50 мл с использованием бакет-ротора. Осадок ресуспендировали в небольшом объеме буфера А и фракционировали с помощью центрифугирования в бакет-роторе при 3800 g в течение 10 мин в ступенчатом градиенте перколла 10%-40%-85% (Sigma-Aldrich, США) в пластиковых пробирках объемом 15 мл. Неповрежденные хлоропласты между слоями 40% и 85% перколла собирали, промывали буфером А и центрифугировали при 1200 g в течение 3 мин в пластиковых пробирках объемом 15 мл (Falcon) в бакет-роторе. Полученный осадок хлоропластов использовали для экстракции нативных пептидов.

3.2 Методики получения мутантных растений *P. patens*

3.2.1 Создание гидовых РНК для получения нокаутов с помощью системы CRISPR/Cas9

Для получения независимых нокаутных линий мха по 4 коротким открытым рамкам считывания (*PSEP1*, *PSEP3*, *PSEP18* и *PSEP25*) нами использована система редактирования CRISPR-Cas9 [216]. Для проведения трансформации были использованы три плазмиды кодирующие (1) белок Cas9, (2) гидовую РНК (sgRNA) и плазмиду (3), несущую ген устойчивости растений к антибиотику G418. Плазмиды с Cas9 (pACT-Cas9) и устойчивостью к G418 (pBNRF) любезно предоставлены Fabien Nogue (Institut Jean-Pierre Bourgin, INRA Centre de Versailles-Grignon). Генетические конструкции были получены сотрудниками лаборатории Генетической инженерии ФГБУ ФНКЦ НИИ Физико-химической медицины (Харлампиева Д.Д.)

Векторы с гидовыми РНК создавали на основе плазмиды pTZ57R/T, в которых гидовая РНК находилась под контролем промотора U6 *P. patens*. Для подбора гидовых 20-нуклеотидных последовательностей РНК, предшествующих РАМ-мотиву *S. pyogenes* Cas9 (NGG), использовали сервис CRISPR DESIGN (<http://crispr.mit.edu>). Все выбранные гидовые РНК находились вблизи старт кодонов кОПС для максимально эффективного сдвига рамок считывания (Табл. 2). На первом этапе мы создали плазмиду, содержащую экспрессионную кассету гидовых РНК. Эта кассета состоит из промотора U6 из *P. patens* и нуклеотидного каркаса tracrRNA [216] с двумя сайтами рестриктазы BbsI между ними. Кассета была синтезирована из олигонуклеотидов и клонирована в вектор ТА pTZ57R/T (Thermo Fischer Scientific, США). Полученная плазида была названа pBV.

Таблица 2. Нуклеотидные последовательности гидовых РНК (sgRNA), использованных для получения нокаутных линий.

| № | sORF | Последовательность sgRNA |
|---|------|-----------------------------|
| | | |

| | | |
|---|--------|-----------------------|
| 1 | PSEP1 | tggtccagcctttacttgca |
| 2 | PSEP3 | gcaccaagataacagcggctc |
| 3 | PSEP18 | gcaagctttcacggacacgca |
| 4 | PSEP25 | gctgctcaagttcattccca |

Плазмиду pBB с направляющей кассетой экспрессии РНК линейаризовали расщеплением BbsI. Олигонуклеотиды содержали совместимые выступающие части и гидовую последовательность из 20-меров (на PSEP1, PSEP3, PSEP25, PSEP18). Гибридизированные олигонуклеотиды (3GTop-3GBottom и т.д.) лигировали в расщепленную плазмиду, получая конечную полную экспрессионную кассету гидовой РНК. Полученные плазмиды были названы pBB-3-1 для *PSEP1*, pBB-3 для *PSEP3*, pBB-25 для *PSEP25* и pBB-18 для *PSEP18*. Для трансформации мха плазмиды очищали с помощью набора Qiagen Plasmid Maxi Kit (Qiagen, Германия).

3.2.2 Создание конструкций для сверхэкспрессии целевых пептидов

Для создания линий со сверхэкспрессией *PSEP1*, *PSEP18* and *PSEP25* кодирующие последовательности были амплифицированы с праймерами PEP4f, PEP4r, pep25FXho, pep25RNhe, pep18FXho и pep18RNhe с использованием геномной ДНК в качестве матрицы (Табл. 3). Далее ампликоны были клонированы в вектор pPLV27 (GenBank JF909480) с использованием системы клонирования без лигазы (ligation-independent cloning (LIC) [217]. Полученные плазмиды были названы pPLV-Нра-4FR (*PSEP1*), pPLV-Нра-25FR (*PSEP25*), and pPLV-Нра-18FR (*PSEP18*). Система сверхэкспрессии с индуцибельным промотором была создана на основе вектора pPGX8 [218]. Кодировующую последовательность микробелка *PSEP3* амплифицировали на основе геномной ДНК и клонировали в вектор между

LexA:Pm35S промотором и Треа3А терминатором. Полученная плазмида была названа рPGX8-MCS3.

Таблица 3. Используемые в исследованиях олигонуклеотидные последовательности

| Название | Нуклеотидная последовательность | Назначение |
|-----------------|--|---------------------------------------|
| 1f | GAGCTCGAATTCGTCCATTGAA GCAGACGTGTTGCGACAG | Создание экспрессионной кассеты sgRNA |
| 2r | ATCTACACCCATCGTCGCTAAC CTGTCGCAACACGTCTGC | - |
| 3f | TAGCGACGATGGGTGTAGATGT GATGTGATGTGATGGTGT | - |
| 4r | GACGCCGCCGTGGAAGAACCAC ACCATCACATCACATCAC | - |
| 5f | GGTTCTTCCACGGCGGCCTT GCGGTGGCGGAGAAGG | - |
| 6r | TGCCGCTCCTTCGGGATATCCCC TTCTCCGCCACCGCAAG | - |
| 7f | GATATCCCGAAGGAGCGGCAGC GGGAGAGCACAAGCAGAAAG | - |
| 8r | GGACCCACTCACTCACTGCACC CTTTCTGCTTGTGCTCTC | - |
| 9f | TGCAGTGAGTGAGTGGGTCCAG CTGGGTGGCTGGCCGAGT | - |
| 10r | CTCGAAACCCGGTCGCGTCCAC TCGGCCAGCCACCCAGCT | - |
| 11f | GACGCGACCGGGTTTCGAGGGG GCGGGGGAGAAAAGGGATGGA G | - |

| | | |
|------------|---|---|
| 12r | CATTCCATGTGGGTTATATCCCT CGCTCCATCCCTTTTCTCC | - |
| 13f | GGATATAACCCACATGGAATGG AGGTGGGTGTGAAGGCGGGT | - |
| 14r | GTAAGTCCTCCACCTTCCTATAT ACCCGCCTTCACACCCACCT | - |
| 15f | TAGGAAGGTGGAGGACTTACAA CCCATGGGTCTTCGAGAAG | - |
| 16r | TTTCTAGCTCTAAAACAGGTCTT CTCGAAGACCCATGGGT | - |
| 17f | CCTGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAG | - |
| 18r | TTTCAAGTTGATAACGGACTAG CCTTATTTTAACTTGC | - |
| 19f | TCCGTTATCAACTTGAAAAAGT GGCACCGAGTC | - |
| 20r | GACGAGCTCAAAAAAAGCACCG ACTCGGTGCCACT | - |
| bbF | GAGCTCGAATTCGTCCATTG | - |
| bbR | GACGAGCTCAAAAAAAGCAC | - |
| 3GTop | CCATGTGGTCCAGCCTTTACTTG CA | Создание гидовой последовательности из 20-меров PSEP1 |
| 3GBottom | AAACTGCAAGTAAAGGCTGGAC CAC | - |
| per3_1_Top | CCATGCACCAAGATAACAGCGG CTC | Создание гидовой последовательности из 20-меров PSEP3 |

| | | |
|--------------------|--|--|
| pep3_1_B ottom | AAACGAGCCGCTGTTATCTTGG TGC | - |
| pep25_2_ Top | CCATGATGTCAGCTCCGCTGTTC TC | Создание гидовой последовательности из 20-меров PSEP25 |
| pep25_2_ Bottom | AAACGAGAACAGCGGAGCTGA CATC | - |
| pep18_1_ Top | CCATGCAAGCTTTCACGGACAC GCA | Создание гидовой последовательности из 20-меров PSEP18 |
| pep18_1_ Bottom | AAACTGCGTGTCCGTGAAAGCT TGC | - |
| PEP4f | TAGTTGGAATAGGTTTCATGGTC CAGCCTTTACTTGCAAG | Для создания линий со сверхэкспрессией PSEP1 |
| PEP4r | AGTATGGAGTTGGGTTCTCATTT CCGGTGATCTCGTTTA | - |
| pep18FX ho | AGTCCTCGAGATGCAAGCTTTC ACGGACAC | Для создания линий со сверхэкспрессией PSEP18 |
| pep18RN he | AGTCGCTAGCTCAGGAGGATGG ACGCCAGC | - |
| pep25FX ho | AGTCCTCGAGATGGTGCAATCC AAGCAAGG | Для создания линий со сверхэкспрессией PSEP25 |
| pep25RN he | AGTCGCTAGCCTACTTGGATTCT TCCTTAGG | - |
| h1 | AGCTGCGCCGATGGTTTCTACA A | Идентификация вставки гигромицин- устойчивости |

| | | |
|--------|------------------------|---|
| h2 | GCGCGTCTGCTGCTCCATACAA | - |
| p1 | GATAGCTTGGGCTGCAGGT | Идентификация вставки целевых рамок со сверхэкспрессией |
| p2 | ACGGACGAAAGCTGGGAGG | - |
| seqF | CCAGGGTTTGAAGGTAGTCAAC | Идентификация нокаутов PSEP1 технологией CRISPR |
| seqR | CAGTGGAGCGACTCGTCGAG | - |
| seqF3 | CGTATTCTTACTCGGCGGCA | Идентификация нокаутов PSEP3 технологией CRISPR |
| seqR3 | TCCACAGGAGATGAGTCCGA | - |
| seqF18 | TCTCGGCACAAACTCACAG | Идентификация нокаутов PSEP18 технологией CRISPR |
| seqR18 | CACAGAACAGCAAGGTGG | - |
| seqF25 | AAAGAGGGATTGCGACGGAG | Идентификация нокаутов PSEP25 технологией CRISPR |
| seqR25 | ATTACTCACTGGGGCGACCT | - |

Выделение и очистку плазмид проводили Qiagen Plasmid Maxi Kit (Qiagen, Germany). Линеаризировали плазмиды pPLV-Нра-4FR, pPLV-Нра-25FR и pPLV-Нра-18FR рестриктазой SacI и pPGX8-MCS3 - KpnI. Верифицировали нуклеотидные последовательности используя капиллярное секвенирование с

помощью BigDye Terminator Cycle Sequencing Kit (v. 3.1) and AbiPrism 3730xl (Applied Biosystems, USA).

3.2.3 Трансформация протопластов

Трансформация протопластов мха выполнялась по методике [219] с использованием полиэтиленгликоля. Для каждой трансформации использовали 7 чашек Петри 5-дневной культуры протонемы. Итоговое количество протопластов для каждого варианта в опыте - 0,8 млн. клеток в 15 мл фальконе. Перед трансформацией фалькон с суспензией протопластов центрифугировали при 250 g в течение 5 минут и сливали надосадочную жидкость. Осадок протопластов ресуспендировали в 600 мкл MMg (0,4 M маннитол, 15mM MgCl₂·6H₂O, 4mM MES) и инкубировали в течение 20 минут при комнатной температуре. Далее добавляли по 10 мкг целевых плазмид и аккуратно перемешивали. В каждый фалькон вносили 700 мкл PEG/Ca, осторожно переворачивали пробирки до гомогенного состояния суспензии и инкубировали 30 минут при комнатной температуре. Добавляли 3 мл раствора W5 (154 mM NaCl, 5mM KCl, 125mM CaCl₂, 2 mM MES). Осаждали протопласты 5 мин при 250 g и удаляли супернатант. Ресуспендировали протопласты в 5 мл предварительно нагретой агаризованной среде для посева и высевали суспензию клеток на 2 чашки Петри с питательной средой и целлофаном. Заворачивали чашки Петри пленкой, закрывали фольгой и помещали в климокамеру. Через сутки удаляли фольгу и культивировали клетки в течение 5 дней от момента посева.

3.2.4 Анализ и отбор мутантных растений *P. patens*

3.2.4.1 *Сверхэкспрессионные линии*

На пятый день культивирования целлофан с трансформированными протопластами переносили на чашки Петри с питательной средой и селективным агентом. Для получения линий со сверхэкспрессией использовали антибиотик гигромицин в концентрации 30 мкг/мл. После этого, чашки Петри инкубировали в климокамере

в течение 21 дня. Отобранные колонии высевали в 24 луночные плашки с агаризованной питательной средой и инкубировали две недели. Далее колонии переносили в новую плашку с агаризованной питательной средой и гигромицином в той же концентрации. Колонии, прошедшие вторую селекцию, отсеивали на новую питательную среду без селективного агента. Через 2-3 недели отбирали растительный материал и выделяли ДНК. Наличие вставки в отобранных колониях определяли ПЦР с использованием праймеров на ген устойчивости к гигромицину (h1, h2) и праймеров, фланкирующих короткую рамку целевых пептидов (p1, p2; **Табл. 3**). Результаты амплификации анализировали электрофорезом в 1,5% агарозном геле. В результате скрининга мутантов отбирали растения, в которых присутствовали оба локуса.

3.2.4.2 Нокаутные линии

Отбор мутантных растений с нокаутами целевых рамок осуществляли схожим способом с некоторыми модификациями. Перенесенный на 5 день целлофан с трансформированными протопластами инкубировали в селективной среде с антибиотиком G418 (50 мкг/мл) в течение 1 недели. Это обусловлено необходимостью лишь временной экспрессии плазмиды pBNRF. Далее целлофан с протопластами переносили на новые чашки Петри с агаризованной питательной средой. Колонии пересаживали в 24-луночный планшет по мере появления. При достижении максимального размера колонии, часть биомассы отбирали для выделения ДНК. Нокаутные линии идентифицировали секвенированием соответствующего участка с использованием специфических праймеров.

3.3 Методики работы с пептидами

3.3.1 Выделение нативных пептидов из протонемы, гаметофоров и протопластов *P. patens*

Для экстракции внутриклеточных пептидов из тканей *P. patens* использовали следующий буфер: 1 М уксусной кислоты в 10% ацетонитриле и 10 мкл/мл

коктейля ингибиторов протеаз (Sigma-Aldrich, США). Измельченный в жидком азоте растительный материал помещали в охлажденный экстракционный раствор. Суспензию центрифугировали при 10000 g в течение 10 минут при 4° С. Супернатант переносили в чистую пробирку и снова центрифугировали при 10000 g в течение 5 минут при 4°С, после чего осадок отбрасывали. Все операции проводились на льду. После центрифугирования супернатант немедленно наносили на хроматографическую колонку для гель-фильтрации для разделения и фракционирования пептидов. Гель-фильтрацию проводили на колонке XK 26/40 GE Healthcare Life Science, заполненной сорбентом Sephadex G25. Элюирование проводили буферным раствором 0,1 М уксусной кислоты при скорости потока 2 мл/мин. Пики детектировали на приборе АКТА Pure 25 (GE Healthcare) на длине волны 280 нм. Фракции, содержащие пептиды, лиофилизировали и ресуспендировали в 500 мкл 50% ацетонитрила-0,1% трифторуксусной кислоты. Далее суспензии озвучивали в течение 5 минут на приборе Elmasonic S 100 (Elma; Germany) и вакуумно концентрировали на приборе Savant SVC 200H SpeedVac Concentrator (Savant) до полного выпаривания жидкости. Далее проводили восстановление дисульфидных связей в образцах. Для этого выпаренный осадок растворяли в водном растворе, содержащем хлорацетамид (CAA), трис(2-карбоксиэтил)фосфин (ТСЕР) и 8.5 pH Tris-HCl, затем нагревали в термостате до 90° С в течение 10 минут и следом охлаждали. Далее к образцам добавляли трифторуксусную кислоту до конечной концентрации 1% и наносили на микроколонки с четырьмя дисками C18 (Empore) для обессоливания. Пептиды элюировали 50% раствором ацетонитрила в 0,1% трифторуксусной кислоте с последующим масс-спектрометрическим анализом всех приготовленных образцов.

3.3.2 Выделение нативных пептидов из секрета протонемы мха

Секретируемые пептиды экстрагировали из 400 мл культуральной среды. Культуральную среду фильтровали через 0,22 мкм мембранный фильтр (Millipore, USA), лиофилизировали и ресуспендировали в 700 мкл 5% водного раствора ацетонитрила, содержащего 0,1% трифторуксусной кислоты. Следом суспензию

центрифугировали при 10000 g в течение 10 минут при 4° C, а полученный супенатант переносили в чистые пробирки и центрифугировали при 5000 × g в течение 5 минут при 4°С, осадок отбрасывали. Пептиды выделяли из культуральной среды путем твердофазной экстракции на обращенно-фазовых картриджах DSC-18 (Discovery DSC-18, Supelco, USA) с использованием 50% раствора ацетонитрила в качестве элюирующего. Элюированные пептиды концентрировали в SpeedVac и ресуспендировали в 5% растворе ацетонитрила, содержащем 0,1% трифторуксусной кислоты. Пул пептидов далее выделяли с помощью твердофазной экстракции на наконечниках ZipTip (Millipore, USA) с использованием 50% раствора ацетонитрила в качестве элюирующего. Элюированные пептиды снова концентрировали в SpeedVac и ресуспендировали в 5% растворе ацетонитрила, содержащем 0,1% трифторуксусной кислоты, с последующим масс-спектрометрическим анализом всех приготовленных образцов.

3.3.3 Анализ антимикробной активности пептидов

Анализ антимикробной активности синтезированных пептидов на фитопатогенных бактериях: грам (-) бактерии *X. arboricola* 3004, выделенной из ячменя (Россия, 1999 год), и грам (+) бактерии *C. michiganensis* pv. *michiganensis*, выделенной из томата (США, 2007 год), проводили методом серийных разведений. Анализ антимикробной активности был выполнен сотрудниками группы биоинженерии растений ФИЦ “Биотехнологии” РАН (Виноградова С. В.).

Бактерии выращивали на агаризованной питательной среде Luria-Bertani (LB) до единичных колоний при 28°С в течение 48 часов, а затем готовили суспензию с оптической плотностью 0,4. Для эксперимента использовали 1 мкл суспензии, содержащей 10⁵ КОЕ/мл бактерий *X. arboricola* 3004 и *C. michiganensis* pv. *michiganensis* и 25 мкл суспензии этих бактерий, содержащей 10⁵ КОЕ/мл и 10⁶ КОЕ/мл *X. arboricola* 3004 и *C. michiganensis* pv. *michiganensis*. Синтезированные пептиды растворяли в стерильной воде mQ и наносили на 96-луночные планшеты в концентрации 7 или 100 мкМ. В качестве отрицательного контроля использовали жидкую питательную среду LB. В качестве положительного контроля

использовали жидкую питательную среду LB, инокулированную 1 мкл и 25 мкл бактерии. Эксперимент проводили в трех повторностях. Планшеты инкубировали при 28°C без перемешивания. Оптическую плотность измеряли каждые 24 часа в течение 3 суток с помощью планшетного фотометра Multiskan FC, Thermo scientific (длина волны 620 нм).

Антимикробную активность секрета определяли в 96-луночных полипропиленовых микротитровальных планшетах (Corning, кат. № CLS3879, Corning, США) стандартным методом микротитровальных разведений в соответствии с общепринятой практикой (Wiegand et al., 2008). Анализ антимикробной активности секрета был выполнен сотрудниками лаборатории Генетической инженерии ФГБУ ФНКЦ НИИ Физико-химической медицины (Графская Е.Н.)

Два бактериальных штамма — *Escherichia coli* (K-12 substr. MG1655) и *Bacillus subtilis* (168НТ) — до использования хранились в 10% глицерине при температуре -70 °С. Бактериальные культуры наращивали в среде LB в течение 16 ч при 37°C и аликвоты по 100 мкл культуры засеивали в 9,9 мл среды Бульон Мюллера-Хинтона (МНВ, Difco; Becton Dickinson Diagnostics) (2 г/л говяжьего экстракта, 17,5 г/л кислотного гидролизата казеина, 1,5 г/л крахмала). Культуры инкубировали на шейкере до конечной концентрации 1×10^8 КОЕ/мл с последующим серийным разведением в $2 \times$ питательной среде МНВ для получения тест-культур в концентрации 1×10^6 КОЕ/мл. Секретом получали из жидкой культуры мха путем фильтрования через 0,22 мкм мембранный фильтр (Millipore, USA) с последующей лиофилизацией. 50 мл лиофилизированного секрета подвергали диализу против физиологического раствора (0,9%) в течение 16 ч при 25°C с использованием мембран CelluSep H1 (MFPI, США). В каждую лунку планшета для микротитрования помещали аликвоты по 50 мкл культуральной суспензии с последующим добавлением 50 мкл разведенного секрета. В качестве контроля использовали как чистую жидкую среду, так и среду с добавлением 400 мкМ метилжасмоната. После инкубации при 37 °С в течение 24 часов чувствительность к противомикробным препаратам мониторили путем измерения оптического

поглощения при 570 нм с помощью микропланшетного ридера. Для визуализации результатов использовали разведение образцов 1/32.

Метод серийных разведений в стерильных 96-луночных планшетах для сбора (Corning 3870) в жидкой среде МНВ использовали для определения минимальной ингибирующей концентрации (МИК) пептидов против *E. coli* K-12 MG1655 и *B. subtilis* 168НТ. В каждую лунку добавляли разведенный пептид (50 мкл) в среде МНВ в соответствующей концентрации (от 128 до 0,25 мкг/мл) вместе с 50 мкл бактериальной суспензии в концентрации 10^6 КОЕ/мл микробных клеток. Опыт проводили в трехкратной повторности. Жидкую среду МНВ без добавления пептида использовали в качестве отрицательного контроля. Планшеты инкубировали при 37°C в течение 20 ч без перемешивания. МИК пептидов оценивали визуально и путем измерения спектрофотометрического поглощения при 570 нм для определения концентрации, ингибирующей видимый рост бактерий.

3.3.4 Синтез пептидов

Пептиды для исследований синтезировались компанией Shanghai Ruifu Chemical Co., Ltd. (Shanghai, China), а также сотрудниками лаборатории Генетической инженерии ФГБУ ФНКЦ НИИ Физико-химической медицины (Лацис И.А.). Чистота лиофилизированных пептидов составляла > 95%, их молекулярный вес был подтвержден с помощью масс-спектрометрического анализа. Синтезированные пептиды растворялись в стерильной воде и хранились при – 80 °С.

3.4 Методики работы с белками

3.4.1 Выделение белков из клеток протонемы и гаметофоров

Белки экстрагировали фенольным методом [220]. Протонему или гаметофору *P. patens* растирали в жидком азоте, добавляли буфер для экстракции

(500 мМ Трис-НСl рН 8,0, 50 мМ ЭДТА, 700 мМ сахараза, 100 мМ КСl, 1 мМ фенилметилсульфонилфторид, 1 мМ ДТТ) и инкубировали 10 мин на льду. Добавляли равный объем ледяного трис-НСl (рН 8,0)-насыщенного фенола, перемешивали и инкубировали в течение 10 мин. Смесь центрифугировали 10 мин, 5500 g, 4°C, фенольную фазу отбирали, повторно смешивали с буфером для экстракции и центрифугировали. Белки осаждали из конечной фенольной фазы четырьмя объемами ледяного 0,1 М ацетата аммония в метаноле в течение ночи при -20°C. Осадок промывали ледяным 0,1 М ацетатом аммония в метаноле трижды и ледяным ацетоном один раз (после каждой промывки - центрифугирование 5 мин, 5500 g, 4°C), а затем сушили.

3.4.2 Электрофорез белков в денатурирующем геле

Белки растворяли в буфере для нанесения (Bio-Rad, USA) и разделяли в денатурирующих условиях (ДСН-На-ПААГ) в 15 % и 20 % полиакриламидных гелях (толщина гелей 1 мм, 10 лунок для образцов) [221]. Электрофорез проводили в ячейке Mini-PROTEAN 3 (Bio-Rad) в течение 1 часа при 200 В. Для определения молекулярной массы полипептидов использовали маркеры молекулярного веса Precision Plus Protein Standards (Bio-Rad, Калифорния, США). Гели окрашивали как описано в [222].

3.4.3 Подготовка образцов для масс-спектрометрического анализа

Белок растворяли в буфере, содержащем 8М мочевины, 2М тиомочевину и 10 мМ Трис. Содержание белка определяли по Брэдфорд (Bio-Rad, Калифорния, США). К 100 мкг белка последовательно добавляли 5 мМ DDT (30 минут при 50 °С) и 10 мМ йодацетамида (20 минут в темноте при комнатной температуре). Белок разбавляли в 5 раз в 50 мМ бикарбонате аммония, добавляли 1 мкг трипсина (Promega, США) и инкубировали при 37 °С в течение ночи. Реакцию останавливали добавлением 1% трифторуксусной кислоты. Триптические пептиды очищали на картриджах Empore octadecyl C18 (Supelco, США) и высушивали в вакуумном концентраторе Savant SVC 200H SpeedVac Concentrator (Savant).

3.4.4 Мечение триптических пептидов с помощью изобарных меток iTRAQ

Для проведения количественного протеомного анализа триптические пептиды метили при помощи набора для относительной и абсолютной квантификации (iTRAQ, Applied Biosystems, Калифорния, США) в соответствии с руководством производителя. Для реакции отбирали по 20 мкг триптических пептидов каждого образа, меченые образцы смешивали и очищали при помощи катионно-обменной хроматографии (SCX, Supelco, США). Для увеличения количества идентификаций вместо очистки триптических пептидов сначала проводили фракционирование при помощи катионно-обменной хроматографии (SCX, Supelco, США), после чего каждую фракцию дополнительно очищали на дисках Empore octadecyl C18 (Supelco, США) и высушивали в вакуумном концентраторе Savant SVC 200H SpeedVac Concentrator (Savant).

3.4.5 Иммуноблоттинг

К протонеме мха, растертой в жидком азоте, добавляли лизирующий буфер (50 мМ Tris-HCl pH 8, 150 мМ NaCl, 1% Triton X-100, 0,1% SDS, 1 мМ ЭДТА, 1 мМ PMSF) в соотношении 15 мг ткани - 100 мкл буфера и инкубировали в течение 30 минут на льду при перемешивании. Затем лизаты центрифугировали при 13 000g 15 мин., 4 °C для удаления дебриса. В супернатантах определяли концентрацию белка методом BCA (Thermo Scientific). Объёмы растворов, содержащие 240 мкг белка, отбирали для осаждения белка ацетоном.

К раствору добавляли в 4 раза больший объём охлаждённого (-20 °C) 100% ацетона, перемешивали и инкубировали при -20 °C в течение 1,5-2 часов. Затем суспензию центрифугировали 13 000g, 4 °C, 10 мин., к осадку добавляли 200 мкл ацетона (-20 °C), перемешивали и центрифугировали при тех же условиях. Промывку повторяли ещё один раз. Осадок высушивали при комнатной температуре, добавляли к нему 30 мкл двухкратного буфера нанесения для денатурирующего электрофореза по Лэммли и инкубировали при 95 °C в течение 15 мин. Растворы центрифугировали при 13 000g, 20 °C, 5 мин.

Готовили 20% полиакриламидные гели по Лэммли. Образцы загружали по 7,5 мкл в лунку. После электрофореза гель инкубировали в буфере переноса в течение 10 минут: 192 мМ Глицин, 25 мМ Tris, 20% метанол. Одновременно фрагмент PVDF-мембраны (заранее погружённую в 100% метанол в течение 5 минут) инкубировали с буфером переноса вместе с фрагментами фильтровальной бумаги. Блоттинг проводили в полусухой системе Bio-Rad при 25 V в течение 11 минут. Мембрану блокировали желатином (Sigma-Aldrich, USA) в течение 1 часа. Затем мембрану инкубировали с первичными антителами, аффинно-очищенными в лаборатории (кроличьи антитела, разведение 1:800) 2 часа при перемешивании при комнатной температуре и 2 часа с раствором вторичных антител (против кроличьих антител, ферментативная метка - щелочная фосфатаза (Sigma-Aldrich), разведение 1:6000). Хромогенное окрашивание производили с использованием субстрата NBT/BCIP. Сыворотка, содержащая первичные антитела, была получена у кролика на синтетический пептид сотрудниками Лаборатории мембранных и биоэнергетических систем ИБХ РАН (Павлюков М.С.). Далее она была аффинно-очищена на иминобиотин-агарозе (Thermo Scientific) с иммобилизованным на ней рекомбинантным химерным белком стрептавидин-FAMOSS.

3.5 Методики работы с нуклеиновыми кислотами

3.5.1 Выделение геномной ДНК

Геномную ДНК выделяли из 0,2 мг свежесобранной ткани с использованием набора DU-250 (Биолабмикс, Россия). Гомогенизировали 20 мг растительной ткани в 600 мкл буфера для лизиса и инкубировали 10 минут при 24 °С. Центрифугировали лизат 30 с при 1000 об/мин и переносили супернатант на колонку. Центрифугировали 30 с при 10000 об/мин и удаляли фильтрат. Наносили на колонку 500 мкл буфера для промывки и центрифугировали 30 с при 10000 об/мин. Удаляли фильтрат и наносили на колонку 500 мкл буфера для промывки. Центрифугировали 30 с при 10000 об/мин и удаляли фильтрат. Для полного удаления буфера для промывки повторно центрифугировали колонку 3 мин при 10000 об/мин. Переносили колонку в новую микроцентрифужную пробирку 1.5 мл.

На центр фильтра колонки наносили 60 мкл буфера для элюции и инкубировали 3 мин при температуре 24°C). Центрифугировали 2 мин при 10000 об/мин. Целостность выделенной геномной ДНК проверяли с помощью гель-электрофореза в 1% агарозном геле.

3.5.2 Выделение тотальной РНК

Тотальная РНК выделялась с использованием TRIzol™ Reagent (Ambion, USA). Около 0,3 г культуральной ткани мха замораживали в жидком азоте и растирали в ступке до состояния пудры, которую затем переносили в 15 мл пробирки. К растертому растительному материалу добавляли реагент тризол (TRIzol Reagent ambion) из расчёта 1 мл на 0,1 г растительной ткани. Образцы тщательно перемешивали и инкубировали в течение 5 минут при комнатной температуре, после чего осаждали нерастворившиеся остатки в центрифуге с угловым ротором на скорости 12000 g в течение 5 минут при 4°C. Надосадочную жидкость переносили в новые 15 мл пробирки и добавляли к ней смесь хлороформа с изоамиловым спиртом (24:1) из расчёта 200 мкл на каждый 1 мл тризола, после чего 2-3 минуты аккуратно перемешивали суспензию. Далее центрифугировали образцы на скорости 12000 g в течение 15 минут при 4°C и переносили верхнюю фазу в 1,5 мл пробирки. Затем добавляли к образцам изопропанол из расчёта 500 мкл на каждый 1 мл тризола и тщательно перемешивали, после чего в течение 20 минут инкубировали при -20°C. Затем снова центрифугировали образцы на скорости 12000 g в течение 10 минут при 4°C для осаждения РНК. Надосадочную жидкость отбирали и добавляли к осадкам этанол 75% из расчёта 0,5 мл на каждый 1 мл тризола, немного перемешивали и центрифугировали на скорости 7500 g в течение 5 минут при 4°C. Далее тщательно отбирали всю жидкость из пробирок и подсушивали осадки РНК на льду до 10 минут. Затем растворяли РНК в воде, очищенной от нуклеаз, и хранили при -70°C.

Качество и количество РНК оценивались с помощью Thermo Scientific™ Nanodrop One при длинах волн 260/280 нм и 260/230 нм и фореа в агарозном геле с окрашиванием бромистым этидием. Точная концентрация РНК определялась

набором Quant-iT™ RNA Assay Kit, 5–100 ng на флуориметре Qubit 3.0 (Invitrogen, USA).

3.5.3 Очистка поли-А фракции из тотальной РНК

Очистка поли(А) фракции РНК проводилась на магнитных шариках с помощью набора Thermo Fisher Scientific Poly(A)Purist™-MAG Purification Kit Invitrogen согласно инструкции из 100 мкг тотальной РНК. Все этапы очистки проводили с использованием наконечников с фильтрами и пробирок с низкой сорбцией нуклеиновых кислот к поверхности.

Образцы тотальной РНК количеством 100 мкг доводили водой, очищенной от нуклеаз, до объема 200 мкл, после чего добавляли равный объем двукратного связывающего раствора (2X Binding Solution) и аккуратно перемешивали.

Для подготовки магнитных частиц с поли(Т) хвостами (Oligo(dT) MagBeads) за час до выделения ставили ёмкость с частицами на ротационный перемешиватель для обеспечения равномерного распределения частиц по буферу.

Отбирали в пробирки по 10 мкл магнитных частиц, затем устанавливали в магнитный штатив и после того, как все частицы сконцентрировались, отбирали буфер. Затем вынимали пробирку из штатива и добавляли 100 мкл промывочного раствора 1. Аккуратно постукивая по пробирке, ресуспендировали смесь до полного растворения конгломерата частиц и ставили пробирку обратно в магнитный штатив, после чего повторяли отмывку частиц еще раз.

После двух отмывок к частицам добавляли образец подготовленной РНК, аккуратно перемешивали и инкубировали сначала в течение 5 минут при 70°C, затем 45 минут при комнатной температуре в режиме покачивания (21 rpm) для связывания РНК с частицами. После инкубации ставили пробирки в магнитный штатив, концентрировали частицы и отбирали раствор. Далее дважды промывали частицы равным объемом (200 мкл) промывочного раствора 1, и дважды – промывочным раствором 2. Затем к частицам добавляли 200 мкл горячего (70°C) раствора для хранения и аккуратно перемешивали. Пробирку с раствором ставили в магнитный штатив на 1-1,5 мин, после чего переносили супернатант с поли(А) в

новые пробирки. Затем ещё раз повторяли промывку горячим раствором для хранения, после чего получали 400 мкл поли(А) фракции РНК. Далее добавляли к образцам 40 мкл 5 М ацетата аммония, 1 мкл гликогена и 1 мл 96% этанола и инкубировали в течение 1 часа при -70°C . После инкубации центрифугировали образцы на скорости 13000 g в течение 50 минут при 4°C . Затем аккуратно отбирали раствор, промывали осадок 500 мл 70% этанола и снова центрифугировали образцы на скорости 13000 g в течение 5 минут при 4°C . Далее тщательно отбирали спирт из пробирок и подсушивали образцы на льду в течение 3 минут. Затем растворяли осадок в 15 мкл воды, очищенной от нуклеаз, и хранили при -70°C .

3.5.4 Электрофорез ДНК в агарозном геле

Электрофорез проводили в 1% агарозном геле с буфером ТАЕ. Для приготовления геля использовался 1 г агарозы, 2 мл 50x ТАЕ буфера, 2 мкл бромистого этидия для визуализации нуклеиновых кислот и дистиллированная вода до объема 100 мл. Гель нагревался до полного растворения агарозы и переливался в заливочную камеру. Затем остужался до комнатной температуры и полностью застывал.

Для электрофореза использовалась форезная камера Baryene subMIDI с источником питания PowerPack Basic (Bio-Rad Laboratories). Форезную камеру заполняли 500 мл раствора ТАЕ.

В качестве маркера использовали 5 мкл ДНК маркера Step50 plus (Биолабмикс, Россия). Для исследования брали 4 мкл образца с добавлением 1 мкл загрузочного буфера (Gel loading Dye, Evrogen, Россия). Электрофорез проводили при 70 V в течение 45 минут. Визуализация результатов электрофореза проводилась с помощью УФ-трансиллюминатора Квант-С (Helicon, Россия) и программы IC Measure.

3.5.5. Выделение плазмидной ДНК

Плазмидную ДНК выделяли с помощью QIAGEN Plasmid Mini Kit (Qiagen, Germany). Разбавляли стартовую культуру в 500 раз средой LB с антибиотиком и инкубировали на шейкере (300 об/мин) при 37°C 12-16 часов. Центрифугировали

бактериальную суспензию на 6000 g в течение 15 минут при 4°C и удаляли супернатант. Полученный бактериальный осадок ресуспендировали в 4 мл буфера P1. Добавляли 4 мл буфера P2, тщательно перемешивали и оставляли при комнатной температуре на 5 мин. В полученный раствор вносили 4 мл предварительно охлажденного буфера P3, сразу же перемешивали и помещали на лед на 15 мин. Далее центрифугировали суспензию на $\geq 20,000$ g в течении 30 минут при 4°C и сразу же отбирали супернатант. Снова центрифугировали супернатант при тех же условиях в течении 15 минут и переносили супернатант в QIAGEN-tip 100 предварительно уравновешенным буфером QBT. Далее дважды промывали QIAGEN-tip раствором Буфер QC. Элюировали плазмидную ДНК 5 мл буфера QF. Преципитацию ДНК осуществляли с использованием 3,5 мл изопропанола комнатной температуры. После перемешивания центрифугировали в течение 30 мин при 4°C на $\geq 15,000$ g. Удаляли супернатант и добавляли 2 мл 70% этанола. Снова центрифугировали при тех же условиях в течение 10 минут и удаляли супернатант. Высушивали осадок плазмидной ДНК в течение 5-10 мин и растворяли TE-буфером (pH 8.0).

3.5.6 ПЦР с реакцией обратной транскрипции

кДНК синтезировали с использованием наборов SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen, Россия) или MMLV RT kit (Evrogen) в соответствии с рекомендациями производителя, используя олиго(dT) праймеры, из 2 мкг тотальной РНК после обработки ДНКазой. Праймеры были разработаны с помощью Primer-BLAST [223]. Отрицательный контроль транскриптазы содержал РНК, не обработанную обратной транскриптазой, для подтверждения отсутствия ДНК в образцах. Реакции проводились в ДНК-амплификаторе T100 Thermal Cycler с термоблоком (Bio-Rad, USA). Продукты ОТ-ПЦР разделяли в 1,5% агарозном геле и визуализировали с помощью окрашивания бромистым этидием.

3.6 Методики транскрипционного анализа

3.6.1 Прямое секвенирование РНК с помощью нанопорового секвенирования
Полученная очищенная поли(А) фракция РНК в количестве 1 мкг использовалась для приготовления библиотек с использованием набора реактивов SQK-RNA002 по рекомендуемому протоколу.

На первом этапе подготовки к секвенированию к образцу РНК лигировали RTA адаптеры с использованием T4 DNA лигазы (NEB M0202). Для этого к образцам поли(А)РНК добавляли 3 мкл буфера для лигирования, 0,5 мкл контрольной РНК (RCS), 1 мкл RTA адаптера и 1,5 мкл T4 DNA лигазы, всё аккуратно перемешивали и инкубировали в течение 15 минут при комнатной температуре.

Далее достраивали вторую цепь ДНК с помощью ревертазы SuperScript III (Thermo Fisher Scientific). Для обратной транскрипции приготовили смесь из 9 мкл воды, очищенной от нуклеаз, 2 мкл 10мМ смеси нуклеотидов dNTP, 8 мкл буфера 5x First Strand Buffer и 4 мкл 0,1М ДТТ. К подготовленному образцу с пришитым адаптером добавили смесь для обратной транскрипции и 2 мкл ревертазы SuperScript III, всё аккуратно перемешивали. Далее инкубировали смесь в течение 50 минут при 50°C, затем терминировали реакцию при 70°C в течение 10 минут, после чего держали пробирку на льду.

Для очистки образца от ферментов использовали магнитные шарики RNA Clean XP beads (Beckman Coulter Life Sciences), которые предварительно в течение часа инкубировали при комнатной температуре на ротамиксе. В новую пробирку (1,5 мл LoBind) переносили образец и 80 мкл магнитных шариков, хорошо перемешивали и в течение 5 минут инкубировали на ротамиксе. Далее ставили пробирку с образцом в магнитный штатив для концентрации шариков. Отбирали раствор и добавляли к осадку 150 мкл 70% этанола, при этом не вынимая пробирку из штатива и не разбивая осадок. Далее проворачивали пробирку в штативе вокруг своей оси на 180°, перемещая пул частичек по стенке пробирки, добивались того, чтобы частицы сконцентрировались на стенке и затем еще на 180° до изначальной позиции. Затем отбирали этанол досуха, добавляли к частицам 20 мкл воды, очищенной от нуклеаз и в течение 5 минут перемешивали на ротамиксе. На этой

стадии происходила элюция гибридных молекул РНК-кДНК. Далее ставили пробирку в магнитный штатив, концентрировали шарики и переносили очищенный образец в новую пробирку.

Следующим этапом пробоподготовки было лигирование второго адаптера RMX с моторным белком. К образцу добавляли 8 мкл буфера для лигирования, 6 мкл RMX адаптера, 3 мкл воды и 3 мкл T4 DNA лигазы, всё аккуратно перемешивали и инкубировали в течение 15 минут при комнатной температуре на ротамиксе. После лигирования к образцу добавляли 45 мкл магнитных шариков и инкубировали 5 минут на ротамиксе, затем ставили пробирку в штатив и концентрировали частицы. Затем отбирали супернатант и дважды промывали шарики в 150 мкл промывочного буфера WSB. После каждой отмывки концентрировали частицы в магнитном штативе. После отмывок добавляли к осадку частиц 23 мкл элюирующего буфера, аккуратно перемешивали и инкубировали образец в течение 15 минут при комнатной температуре на ротамиксе. Затем концентрировали шарики в магнитном штативе и переносили супернатант с библиотекой в новую пробирку. Далее брали 1 мкл образца для измерения концентрации. Концентрацию библиотеки измеряли с помощью Qubit 3.0 и набора реактивов Qubit DNA HS Assay Kit.

Для загрузки образца в ячейку готовили загрузочную смесь: к 1 мл FB (Flush Buffer) добавляли 30 мкл FLT (Flush Tether) и хорошо перемешивали на вортексе.

Секвенирование проводилось на платформе MinION Oxford Nanopore. Для секвенирования использовались проточные ячейки FLO-MIN106 (ONT R9.4) и программное обеспечение MinKNOW (Oxford Nanopore Technologies). Ячейку предварительно выдерживали на комнатной температуре и устанавливали в секвенатор. Далее открывали в ячейке отверстие для загрузки и, используя дозатор с наконечником на 1000 мкл, аккуратно, не допуская появления пузырей, отбирали из отверстия 20-30 мкл жидкости, выкручивая на дозаторе необходимый объем. После этого аккуратно, не допуская появления пузырей, добавляли 800 мкл загрузочного микса, закрывали отверстие для загрузки и инкубировали 5 минут.

В это время готовили геномную библиотеку к загрузке в ячейку: добавляли к 22 мкл образца 15,5 мкл воды, очищенной от нуклеаз, и 37,5 мкл буфера RRB (RNA

Running Buffer). После инкубации открывали на ячейке два отверстия – отверстие для загрузки и SpotON - добавляли в отверстие для загрузки оставшийся объём загрузочного микса. Далее загружали 75 мкл подготовленной библиотеки в отверстие SpotON. Ячейку подключали при помощи USB провода к персональному компьютеру с установленным на нём программным обеспечением MinKnow и запускали процесс секвенирования продолжительностью 48 часов.

Секвенирование было проведено в трех повторах для протонем и четырех повторах для гаметофоров мха.

Анализ сырых данных секвенирования проводился в программе Guppy 4.0.15 (Oxford Nanopore Technologies). Для контроля качества полученных ридов использовался скрипт MinIONQC [224]. Для дальнейшего анализа использовались прочтения со средним качеством нуклеотида (Pherd quality score) больше 7.

Риды были выровнены на геном мха *Physcomitrium patens* V3.3 [225] с добавлением пластидного (NC_037465.1) и митохондриального (NC_007945.1) геномов *Physcomitrium patens*, а также последовательности гена енолазы *Saccharomyces cerevisiae*, которая включена в набор реактивов в качестве контрольной РНК (NC_001140.6). Для выравнивания использовалась программа Minimap 2.17 [226] с параметрами -ax splice -uf -k14 f -G2k --secondary=no. Программы Samtools 1.9 [227] и Bedtools 2.27.1 [228] использовались для дальнейшей работы с выровненными ридами: архивации, сортировки, визуализации, анализа пересечений.

3.6.2 Идентификация и анализ транскриптов

Для сборки транскриптома с использованием референсной последовательности использовали StringTie2 [229] в режиме сборки длинных ридов. Прочтения каждого образца использовались для сборки своего транскриптома, а затем все сборки объединялись командой merge. Для команды merge использовали параметры -i -g5. Для сравнения полученного транскриптома с различными имеющимися базами данных использовали программу GffCompare, а для конвертации между различными форматами аннотации генома использовали GffRead [230]. В качестве референсного транскриптома использовалась аннотация Phytozome V12. Для

анализа дифференциальной экспрессии длинных некодирующих РНК использовалась сборка с помощью программы Flair [231]. Однако Flair не подходит для квантификации моно-экзонных транскриптов. Поэтому для их анализа использовалась квантификация с помощью featureCounts 2.0.0 [232] с параметрами -s 1 -L -O --fracOverlapFeature 0.75 --fracOverlap 0.75. Как Flair, так и featureCounts адаптированы для работы с длинными прочтениями и квантифицируют риды методом простого подсчета, что позволяет сравнивать результаты этих двух методов.

3.7 Масс-спектрометрический анализ

3.7.1 Тандемный хромато-масс-спектрометрический анализ

Масс-спектрометрический анализ эндогенных и триптических пептидов проводили в трех биологических повторах. Анализ проводили на системе ВЭЖХ Ultimate 3000 RSLCnano, соединенной с масс-спектрометром Q Exactive Plus (ThermoFisher Scientific). Образцы загружались в самодельную колонку-ловушку $20 \times 0,1$ мм, заполненную сорбентом Inertsil ODS3 3 мкм (GLSciences), в загрузочном буфере (2 % ацетонитрила, 98 % воды, 0,1 % трифторуксусной кислоты (TFA)) при скорости потока 10 мкл/мин и разделялись при комнатной температуре в самодельной колонке из плавленого кварца $500 \times 0,1$ мм, заполненной Reprosil PUR C18AQ 1.9 (Dr. Maisch), в эмиттер, подготовленный с помощью P2000 Laser Puller (Sutter, США). Образцы элюировали линейным градиентом 80 % ацетонитрила, 19,9 % воды, 0,1 % муравьиной кислоты (буфер В) в 99,9 % воды, 0,1 % муравьиной кислоты (растворитель А) от 4 до 36 % растворителя В за 1 час при скорости потока 0,44 мкл/мин при комнатной температуре.

Данные МС были собраны в режиме DDA (информационно-зависимый сбор данных; Data Dependent Acquisition). Параметры MS1 были следующими: разрешение 140К, диапазон сканирования 350–2000, максимальное время инъекции 50 мс, AGC (автоматический контроль времени захвата; Automatic Gain Control) цели 3×10^6 . Ионы были изолированы в окне в 1,4 m/z и смещением 0,2 m/z, ориентируясь на 10 пиков максимальной интенсивности заряда от + 1 до + 6

зарядов, минимальное AGC 8×10^3 , с предпочтительным совпадением пептидов и исключением изотопа. Динамическое исключение было установлено на 40 с. Фрагментация MS2 проводилась в режиме HCD с разрешением 17,5К и 27 % NCE (оптимальная энергия фрагментации). Ионы накапливались в течение не более 50 мс с AGC цели 1×10^5 .

3.7.2 Мониторинг множественных реакций

Для мониторинга множественных реакций 6 пмоль контрольной смеси пептидов iRT (Biognosys, Швейцария) добавляли к 200 мл культуральной среды или к 5 мл буфера для экстракции для выделения внутриклеточных пептидов. Масс-спектрометрический анализ проводили на тройном квадрупольном масс-спектрометре QTRAP 4500 (Sciex, США), оснащенный источником ионов NanoSpray III (Sciex, США), соединенным с системой нано-ВЭЖХ nanoLC Ultra 2D+ (Eksigent, США). Система ВЭЖХ работала в режиме “trap-elute”. Смесь состоящую из 98,9 % воды, 1 % метанола и 0,1 % муравьиной кислоты (об./об.) использовали в качестве раствора для загрузки образца и подвижной фазы А. Подвижная фаза В представляла собой 99,9 % ацетонитрила, 0,1 % муравьиной кислоты (об/об). Образцы наносили на колонку-ловушку (10 мм, 100 мкм id, Aeris Peptide XB-C18, 2,6 мкм, 100 А) при скорости потока 2,5 мкл/мин в течение 10 минут и элюировали через разделительную колонку (20 см, 75 мкм id, Aeris Peptide XB-C18, 2,6 мкм, 100 А) при скорости потока 300 нл/мин. Общее время работы включало начальные 10 мин уравнивания колонки подвижной фазой А (0,1% муравьиной кислоты), затем градиент от 10–40% подвижной фазы В в течение 30 мин, 2 мин до достижения 95% подвижной фазы В, 10-минутная промывка раствором 95% подвижной фазы В и 5-минутное повторное уравнивание с подвижной фазой А. Хроматографические колонки термостатировали при 50°C. Целевые пептиды измеряли дважды, затем усредняли полученные значения, а общую площадь пика фрагмента, нормализованную к среднему сигналу каждого пептида, использовали в качестве показателя представленности пептидов. Список переходов был создан после оценки всех теоретических фрагментов “b” и “y” с

зарядом +1 для ионов-предшественников с зарядами +2 или +3. Пять транзиций для наиболее интенсивных ионов-предшественников были использованы для дальнейшего анализа синтетических пептидов. Данные были получены в режиме положительных ионов с общим временем сканирования 1,49 с и временем задержки 10 мс (разрешение Q1 и Q3 установлено на «UNIT» - 0,7 Да FWHH, пауза между диапазонами масс составляла 5 мс). Обработка данных включала выбор пиков (с ручной проверкой помех и отсутствующих сигналов), интеграцию пиков и экспорт с использованием программного обеспечения Skyline. Дальнейший анализ производился самодельным скриптом на R.

3.7.3 Анализ данных масс-спектрометрического анализа

Данные масс-спектрометрического анализа пептидомных образцов обчитывали с помощью PEAKS Studio версии 8.0 (Bioinform Inc., Калифорния, США) с использованием базы данных, содержащей белковые последовательности из Phytozome V12, объединенные с хлоропластными и митохондриальными белками (33,053 записи). Поиск проводился со следующими параметрами: допуск по массе прекурсора 10 ppm и допуск по массе фрагмента 0,05 Да; «режим ферментативного расщепления» был установлен как «неспецифический»; фиксированные модификации - карбамидометилирование (+ 57,02); переменные модификации - окисление (M) - +15,99 и ацетилирование (N-конец) - +42,01. Настройки прибора были установлены на «орби-орби» (орбитальная ловушка использовалась как для определения m/z предшественника, так и фрагментированного иона). Используемый коэффициент ложного обнаружения был равен 1 %.

Для поиска продуктов трансляции коротких рамок считывания тандемные масс-спектры пептидомных образцов исследовали по отдельности с помощью программного обеспечения PEAKS Studio версии 8.0 (Bioinform Inc., Калифорния, США) и MaxQuant v1.6.14 [233]. Для анализа использовали базу полипептидных последовательностей, содержащую аннотированные белки (33,053 записей) и предсказанные пептиды, кодируемые kORC.

Фильтр белкового FDR (false discovery rate) был отключен для поисковой машины MaxQuant, в то время как отсечка в 1% FDR использовалась для фильтрации совпадений спектров с пептидами, при этом неоднозначные совпадения отфильтровывались. Параметр «Digestion Mode» был установлен как «unspecific», а модификации не разрешались. Все остальные параметры были оставлены со значениями по умолчанию. После поиска пептидов с помощью алгоритма MaxQuant была использована более строгая стратегия фильтрации FDR для спектров совпадающих с пептидами от кОПС [234]. Специфичный FDR для этого класса пептидов рассчитывали как количество идентификаций в наборе “decoy” последовательностей, деленное на количество идентификаций предсказанных кОПС. Далее отсекали идентификации используя класс-специфичный FDR в 1%. Параметры поиска PEAKS 8.0: допуск по массе продуктов фрагментации 0,05 Да; допуск по массе родительского иона 10 ppm; без модификаций. Результаты были отфильтрованы по 1% FDR, но с порогом значимости не менее 20 (эквивалентно $P < 0,01$).

3.7.4 Количественный анализ протеомных данных

Для сравнительного количественного протеомного анализа на основе интенсивностей по MS1 без метки использовали программный пакет Progenesis LC-MS (Nonlinear Dynamics, Дарем, Северная Каролина, США). Для этого файлы необработанных данных были преобразованы в файлы .mzML с помощью конвертера данных AB SCIEX MS Data Converter (версия 1.3, ABSciex) и загружены в Progenesis. Файлы в формате .mgf были сгенерированы Progenesis LC-MS на основе поиска совпадений с белковой базой UniProt (35,414 белков для *P. patens*) с использованием программы Mascot версии 2.4.1 (Matrix Science, Бостон, Массачусетс, США). Поиск с помощью Mascot проводился по следующим параметрам: триптические пептиды, максимум одно пропущенное расщепление; состояние заряда пептида ограничено 1+, 2+ и 3+; допуск по массе иона-предшественника 20 ppm; допуск по массе фрагментов MS/MS 50 ppm; возможные модификации были окислением метионина (M) и карбамидометилированием. Для

расчета FDR была использована база обратных последовательностей (decoy). Пороговое значение FDR было установлено на уровне 5%. Двусторонний непарный t-критерий Стьюдента был использован для оценки достоверности результатов количественного определения.

Для сравнительного количественного анализа с использованием меток iTRAQ (Isobaric tag for relative and absolute quantitation) данные масс-спектрометрического анализа анализировали с помощью программного обеспечения PEAKS Studio версии 8.0, алгоритм PEAKS Q (Bioinform Inc., Калифорния, США). Для идентификаций использовали базу полипептидных последовательностей, содержащую аннотированные белки (33053 записей). Поиск в базе данных проводился по следующим параметрами: допуск по массе продуктов фрагментации 0,05 Да; допуск по массе родительского иона 10 ppm; фиксированная модификация – карбамидометилирование; переменные модификации — окисление метионина (M), дезамидирование (NQ) и ацетилирование (белок N-термин). Результаты были отфильтрованы по 1% FDR. Нормировку проводили путем усреднения интенсивности всех пептидов. Для усреднения использовались медианные значения. Белки считали дифференциально регулируемые, если кратность их изменений была выше 1,2 и порог значимости $-10\log P=20$, $P<0.05$ (дисперсионный анализ с коррекцией Benjamini и Hochberg FDR), критерием однородности дисперсии ($P> 0,05$) и нормальным распределением ($P> 0,05$).

3.8 Методы анализа интерактома

3.8.1 Иммунопреципитация антителами к FLAG эпитопу

Культивируемые в течение 6-7 дней клетки протонемы гомогенизировали растиранием в жидком азоте при помощи охлажденных ступки и пестика. 50-100 мг полученного гомогената лизировали в 450 мкл буфера (50 mM фосфат pH 7,2, 150 mM NaCl, 0,1% Tween-20) с добавлением 1:100 (v/v) ингибиторов протеаз (Sigma, USA) в течение 30 минут на льду при перемешивании. Далее лизаты центрифугировали при 13 000 g, 4 °C в течение 15 минут. К супернантам добавляли кросс-сшивающий реагент DSP (dithiobis(succinimidyl propionate); Thermo Fisher

Scientific Pierce) из 200 мМ стокового раствора в ДМСО до конечной концентрации 5 мМ и инкубировали в течение 2 часов на льду при перемешивании. Реакцию останавливали добавлением Tris-HCl pH 8 до концентрации 40 мМ в течение 15 мин и центрифугировали при 14 000 g, 4°C 10 мин. Полученный супернатант добавляли к 25 мкл уравновешенных лизирующим буфером магнитных частичек с иммобилизованными на их поверхности антителами против FLAG-эпитопа (антитела M2, Sigma). После двухчасовой инкубации на льду при помешивании сорбент отмывали дважды 700 мкл лизирующего буфера и один раз 700 мкл PBS (50 мМ фосфат pH 7,2, 150 мМ NaCl). Элюцию производили 125 мкл 100 нг/мкл раствором 3xFLAG-пептида (Sigma, USA) в PBS в течение 30 мин на льду при перемешивании. Сшивки белков, образованные DSP, разрушали добавлением в элюаты 20 мМ DTT и нагреванием до 50 °C. Элюаты подвергали трипсинолизу в растворе для последующего масс-спектрометрического определения белкового состава.

Образцы растирали в жидком азоте и гомогенизировали в 450 мкл литического буфера (50 мМ Tris-HCl pH 8, 150 мМ NaCl, 0,1% Tween-20) с добавлением 10 мкл/мл смеси ингибиторов протеаз (Sigma) и инкубировали на качалке на льду 30 мин. Остатки клеток осаждали центрифугированием (13 000g, 15 мин, 4°C, Eppendorf Centrifuge 5418 R), после чего к супернатанту добавляли сшивающий агент - DSP (Thermo Fisher Scientific Pierce), - в концентрации 5 мМ и инкубировали на качалке в течение 2 ч. Реакцию останавливали добавлением 40 мМ Tris-HCl pH 8 в течение 15 мин и центрифугировали 14 000 g в течении 10 мин. К супернатанту добавляли 50 мкл предотмытых литическим буфером магнитных частиц с иммобилизованными антителами против FLAG-эпитопа (клон M2, Sigma), инкубировали 2 ч на качалке на льду, после чего частицы дважды отмывали литическим буфером и один раз TBS (50 мМ Tris-HCl pH 8, 150 мМ NaCl). Белковые комплексы элюировали 125 мкл раствора 3xFLAG-пептида (100 нг/мкл, Sigma) в TBS в течение 30 мин на качалке на льду. Образованные DSP сшивки разрушали нагреванием образцов до 50°C в присутствии 20 мМ DTT.

3.8.2 Получение конъюгата стрептавидин-пептид в бактериальной системе экспрессии

Плазмиды pET-mSAV и pET-mSAV-pep4 сконструировали с использованием системы pET-плазмид (Novagen, USA). Они содержали кодирующую последовательность стрептавидина, слитую с последовательностью сайта TEV-протеазы (pET-mSAV) или те же последовательности, слитые с кодирующей последовательностью пептида FAMOSS (pET-mSAV-pep4). Для получения рекомбинантных белков использовали штамм *E. coli* Rosetta 2 (DE3). Экспрессию индуцировали добавлением IPTG до конечной концентрации 0,5 мМ. Клетки разрушали с помощью ультразвуковой обработки на установке Branson Sonifier 250 (VWR Scientific, USA). Затем центрифугировали лизат при 15,000 g, 25 мин и дважды промывали осадок 1% (v/v) Triton X-100. Полученный осадок растворяли в 6М гуанидин хлориде, pH 1,5, и диализовали против 200 мМ NaHCO₃ pH 8. Затем очищали белки на иминобиотин-агарозе (Thermo Fisher Scientific, USA) в соответствии с протоколом производителя.

3.8.3 Проведение аффинной хроматографии

Протонема дикого типа была измельчена в жидком азоте с использованием керамических ступки и пестика. К измельченной ткани был добавлен охлажденный (4°C) лизирующий буфер (PBS, 50 мМ фосфат pH 7,2, 150 мМ NaCl, 0,1% Triton X-100, 0,2% NP-40, 1:100 (об/об) раствор ингибиторов протеаз (Sigma, США) в пропорции 15 мг ткани - 100 мкл буфера. Лизат инкубировали на льду 30 минут при перемешивании и центрифугировали при 14000 g 15 мин., 4°C. К супернатантам добавляли 15 мкг рекомбинантного белка FAMOSS-стрептавидин (эксперимент) или стрептавидин (контроль) и инкубировали 1 час на льду при перемешивании. Затем стабилизировали белковые комплексы с помощью кросс-сшивающего реагента DSP (Thermo Fisher Scientific Pierce) в конечной концентрации 3 мМ и инкубировали в течение 2 часов на льду при перемешивании. Реакцию останавливали добавлением Tris-HCl pH 8 до концентрации 40 мМ в течение 15 мин на льду. Затем добавляли бикарбонат аммония до конечной

концентрации 50 мМ (из 2 М стокового раствора) и титровали 5 М NaOH до pH 11. Затем центрифугировали образцы при 14 000 g 10 мин., 4 °С. К полученным супернатантам добавляли 20 мкл иминобиотин-агарозы (Thermo Fisher Scientific Pierce), уравновешенной буфером для отмывок (50 мМ аммоний бикарбонат-NaOH pH 11, 500 мМ NaCl, 0,1% Triton X-100, 0,2% NP-40). Полученную смесь инкубировали на льду в течение 1 часа, после чего сорбент промывали 4 раза буфером для отмывок. Затем белковые комплексы элюировали 50 мМ ацетатом аммония pH 3 500 мМ NaCl, после чего нейтрализовали 1М раствором Tris до концентрации 20 мМ. Затем элюаты концентрировали и очищали на центрифужных картриджах Amicon Ultra-0.5 Centrifugal Filter Unit Ultracel-3K (Millipore). Сшивки DSP расщепляли нагреванием растворов элюатов до 50⁰С с 20 мМ ДТТ.

3.9 Флуоресцентная микроскопия

Для анализа везикулярного транспорта образцы протонемы окрашивали 5 мкМ SynaptoGreen C4 (эквивалент FM1-43, Biotium, USA) в течении 5 мин и отмывали 1 раз. Для анализа соотношения хлоронемных и каулонемных клеток клеточные стенки окрашивали 5 мин 10 мкг/мл йодида пропидия (Sigma-Aldrich, USA). Образование активных форм кислорода оценивали по интенсивности окрашивания 10 мкМ DCFH-DA (2',7'-Dichlorofluorescein Diacetate, Sigma-Aldrich, USA) в течение 10 мин. Жизнеспособность рассчитывали как долю клеток, окрашенных 10 мкМ флуоресцеиндиацетатом (ФДА, Sigma-Aldrich, USA) в течение 5 мин. Накопление АФК и жизнеспособность оценивали в суспензионной культуре протонемы с индуцибельной экспрессией микробелка *PSEP3* после индукции экспрессии 1 мкМ β-эстрадиола (Sigma, USA) в течение 24 ч (накопление АФК) или 48 ч (изменение жизнеспособности).

Флуоресценцию регистрировали при помощи флуоресцентного микроскопа ZEISS Imager.M2 (Zeiss, Germany), оборудованного цифровой камерой AxioCam 506 mono и программой ZEN 2.6 pro. Для регистрации флуоресценции красителей использовали следующие наборы фильтров: № 44 FITC (λex BP 475 nm/40 nm; λem

BP 530 nm/50 nm) для красителей SynaptoGreen C4, DCFH-DA и ФДА; № 20 Rhodamin (λ_{ex} BP 546/12 nm; λ_{em} BP 575–640 nm) - для йодида пропидия.

3.10 Методы биоинформатического анализа

3.10.1 Предсказание длинных некодирующих РНК на основе предсказанных транскриптов

Для создания полного набора транскриптов длинных некодирующих РНК *P. patens* были объединены 1498 длнкРНК из базы данных SANTATdb 2.0 [235], 9416 длнкРНК из базы GreeNC [236], 3018 длнкРНК из аннотации NCBI, 1512 длнкРНК из статьи Lung et al. [225] и 4648 из статьи Simopoulos et al. [237]. Используя программу bedtools транскрипты длнкРНК были объединены в локусы, а дублирующие транскрипты были исключены. Используя алгоритм TBLASTN, дополнительно отфильтровали длнкРНК, которые имели сходство с аннотированными белками viridiplantae из базы данных Uniprot ($E < 10^{-5}$; перекрытие $> 30\%$). Кроме того, дополнительно были отброшены длнкРНК со сходством с известными некодирующими РНК из базы данных Rfam ($E < 10^{-5}$).

3.10.2 Предсказание коротких рамок считывания с высоким кодирующим потенциалом

Для идентификации кОРС с высоким кодирующим потенциалом в геноме *P. patens* был использован инструмент sORFfinder [238] с разными настройками. Параметр -d был установлен на «b» для поиска последовательностей кОРС как в прямой, так и в обратной ориентации. Интронные последовательности и белок-кодирующие последовательности, которые использовали для обучения модели sORFfinder, были экстрагированы из аннотации Phytozome V12 с помощью программы gffread, а затем .fasta файлы, содержащие нуклеотидные последовательности, были получены с помощью команды getfasta из программы bedtools [228]. Для отбора кОРС, которые были расположены на экзонах транскриптов разного типа, были получены их координаты в геноме и пересечены с файлом аннотации генома *P.*

patens. Идентичные кОПС были отфильтрованы. Кроме того, кОПС, которые перекрывались с повторяющимися регионами, идентифицированными с помощью программы RepeatMasker [239], а также кОПС, содержащие части аннотированных основных и альтернативных изоформ белка *P. patens*, были удалены из набора данных.

Классификация кОПС с высоким кодирующим потенциалом осуществлялась следующим образом. На первом этапе были идентифицированы “длнкРНК-кОПС” путем пересечения с кОПС, локализованных на транскриптах из известных баз данных длнкРНК, таких как SNTATAdb [235], GreeNC [236], а также ранее опубликованном нами наборе данных [240]. Оставшиеся кОПС были пересечены с геномной аннотацией *P. patens* (Lang et al. 2018) и были определены как “генные” кОПС, т.е. “upstream” кОПС, которые расположены в 5'-нетранслируемых регионах мРНК; “downstream” кОПС, которые расположены в 3'-нетранслируемых регионах мРНК; “CDS”-кОПС, которые перекрываются с основной белок-кодирующей последовательностью мРНК, но со сдвигом рамки считывания +1 или +2; “interlaced” кОПС, которые перекрываются как с белок-кодирующей последовательностью, так и с нетранслируемыми регионами мРНК. “Upstream” кОПС и “downstream” кОПС были полностью отделены от более длинной рамки считывания, кодирующей белок, как описано ранее [241,242]. Все кОПС, расположенные на транскриптах, которые не были аннотированы в геноме *P. patens* V3.3, но были идентифицированы с использованием наших данных транскрипционного профилирования, были классифицированы как межгенные кОПС.

Альтернативный сплайсинг влияет на локализацию “генных” кОПС, приводя к трудностям при их классификации. Для обнаружения кОПС, которые подвергаются альтернативному сплайсингу, мы пересекли координаты коротких рамок считывания и координаты интронов всех изоформ транскриптов, используя программу bedtools. Короткие рамки считывания, которые перекрывались как по экзонам, так и по интронам с изоформами, были классифицированы как AC-кОПС.

3.10.3 Предсказание всех коротких рамок считывания на длинных некодирующих РНК

Короткие рамки считывания (кОРС) были предсказаны на транскриптах длнкРНК с помощью двух программных инструментов: MiPepid [243] и NCBI finder [244]. Короткие рамки считывания, которые начинались со стартового кодона “AUG” были предсказаны с помощью MiPepid. С помощью NCBI finder были предсказаны кОРС, начинающиеся с альтернативных старт кодонов - “CUG” и “UUG”. Короткие рамки считывания, предсказанные с помощью NCBI finder и находящиеся внутри рамок, начинающихся с “AUG”, были отброшены.

3.10.4 Анализ консервативности длнкРНК и коротких открытых рамок считывания

Транскриптомы зеленых водорослей, печеночников, мхов, антоцеротовых мхов, плауновидных, папоротников, хвойных и цветковых растений из проекта 100 транскриптомов растений OneKP [245] были загружены из репозитория https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/oneKP_capstone_2019. Для поиска схожих последовательностей между локусами длнкРНК *P. patens* и транскриптами из разных линий растений был использован алгоритм BLASTN ($E < 10^{-5}$).

Для анализа консервативности “генных” кОРС были использованы транскриптомы девяти видов растений, которые были загружены из Phytozome V12: *Sphagnum fallax* (версия 0.5), *Marchantia polymorpha* (версия 3.1), *Selaginella moellendorffii* (версия 1.0), *Spirodela polyrhiza* (версия 2), *Arabidopsis thaliana* (TAIR 10), *Zea mays* (Ensembl-18), *Oryza sativa* (версия 7), *Volvox carteri* (версия 2.1) и *Chlamydomonas reinhardtii* (версия 5.5). Транскриптом *Ceratodon purpureus* был собран *de novo* с использованием программы Trinity [246]. Для идентификации транскрибируемых гомологичных последовательностей использовали алгоритм TBLASTN (words=3). Для фильтрации только надежных выравниваний использовали следующие параметры: $E < 10^{-6}$ и покрытие $> 60\%$.

Чтобы идентифицировать ортологи “длнкРНК-кОРС”, в транскриптомах разных групп растений из проекта 1000 транскриптомов был использован поиск с

помощью алгоритма TBLASTN. Полученные аминокислотные выравнивания были отфильтрованы по значению $E < 10^{-3}$. Количество консервативных кОПС увеличилось только на 1,9% при менее жестких порогах фильтрации ($E < 10^{-2}$) и уменьшилось на 1,64% при более строгом пороге ($E < 10^{-4}$).

Для поиска ортологов FAMOSS в отдаленных видах растений был создан hmm профиль с помощью hmmbuild. Для создания hmm профиля гомологичные последовательности были выровнены с помощью программы Muscle v3.8.31 [247]. Дальнейший поиск был проведен с помощью программы hmmsearch [248].

3.10.5 Анализ эволюции коротких открытых рамок считывания

Выравнивание нуклеотидных последовательностей ортологичных транскриптов и аминокислотных последовательностей кодирующих последовательностей было выполнено с помощью программы Muscle [247]. Отобранные для дальнейшего расчета скоростей эволюции выравнивания не содержали внутренних стоп-кодонов, прерывающих рамку считывания, а стартовые кодоны ортологичных последовательностей совпадали. Кроме того, длина выравнивания ортологичных кодирующих последовательностей составляла $> 80\%$.

Для расчета показателя dN/dS были получены выравнивания нуклеотидных последовательностей на основе аминокислотных выравниваний ортологичных последовательностей с помощью программы PAL2NAL [249].

Для нуклеотидных последовательностей эволюционную дистанцию рассчитывали с использованием двухпараметрической модели Кимуры [250]. Скорость эволюции синонимичных и несинонимичных позиций (dS и dN , соответственно) в кодирующих последовательностях оценивали с использованием программного обеспечения PAML и метода максимального правдоподобия [251]. Значения dN и dS приводящие к ненадежной оценке скорости эволюции, были исключены из анализа.

Для анализа кодирующих последовательностей с признаками положительного отбора в дополнение к PAML был использован алгоритм HyPhy BUSTED [252]. BUSTED (Неограниченный статистический тест эпизодической диверсификации)

обеспечивает анализ на положительный отбор, по крайней мере, в одном сайте кодирующей последовательности, по крайней мере, в одной филогенетической ветви. Филогенетические деревья для анализа положительного отбора были построены для каждого выравнивания с помощью программы IQ-TREE 1.6.12 [253]. Мы проанализировали все результаты PAML и HyPhy с помощью собственных скриптов Python. Для расчета статистической значимости использовали подход Бенджамини-Хохберга [254] для корректировки множественного тестирования. Филогенетические деревья были визуализированы с помощью пакета Toytree [255].

3.10.6 Статистический анализ

Статистический анализ, обработка данных и визуализация были выполнены в Python v. 3.7.5 [256] с использованием модулей scipy 1.5.2 [257], seaborn 0.11.1 [258], numpy 1.20.1 [259], pandas 1.2.3 и upsetplot 0.5 [260].

3.10.7 Предсказание доменов и мотивов в полипептидных последовательностях

Белковые и пептидные последовательности сканировали на наличие функциональных доменов с помощью программы InterProScan 4.36 [261] с настройками по умолчанию. Для предсказания доменов, относящихся к малым секретиремым пептидам (SSPs) был использован веб-ресурс “SSP-Prediction”, доступный по адресу <http://mtsspdb.noble.org/prediction>. Регионы низкой сложности (LCR) были идентифицированы с помощью инструмента SEG с настройками по умолчанию. Для предсказания сигналов экспорта использовалась программа SignalP-5.0 [262], которая запускалась с параметрами по умолчанию. Для предсказания сигналов экспорта и трансмембранных доменов была использована программа TMHMM 2.0 [263], которая запускалась с параметрами по умолчанию. Все полученные результаты были отфильтрованы и проанализированы с помощью собственных скриптов python.

3.10.8 Предсказание антимикробной активности

Антимикробный потенциал эндогенных пептидов был предсказан на основе пептидных последовательностей с использованием трех серверов: CAMP (<http://www.camp3.bicnirrh.res.in>), iAMPpred (<http://cabgrid.res.in:8080/amppred/>) and ADAM (<http://bioinformatics.cs.ntou.edu.tw/ADAM/links.html>), использующих алгоритмы машинного обучения (метод опорных векторов; [SVM] support vector machines) для предсказания антимикробной активности в пептидах. В CAMP и iAMPpred для определения антимикробной активности устанавливалось пороговое значение 0,5; в ADAM - 0. Для дальнейшей работы использовались только пептиды с вероятностью ошибочной классификации меньше 5%.

3.10.9 Предсказание 3D-структуры пептидов

Чтобы предсказать трехмерную структуру пептида FAMOSS, мы использовали алгоритм AlphaFold2 [264], который был запущен как проект Google Colaboratory на GitHub с опцией пользовательского множественного выравнивания последовательностей (MSA). Для построения MSA для конкретного пептида мы сначала получили список из 100 наиболее похожих последовательностей. Затем мы выполнили множественное выравнивание белков с помощью онлайн-инструмента M-COFFEE с параметрами по умолчанию [265]. Полученный файл MSA использовали для анализа с помощью алгоритма AlphaFold2 с параметрами по умолчанию. Структуры визуализировались с помощью программы PyMol v.2.3.0 (<https://github.com/schrodinger/pymol-open-source>).

4. РЕЗУЛЬТАТЫ и ОБСУЖДЕНИЕ

4.1 Анализ нативных пептидомов модельного растения *P. patens*

4.1.1 Идентификация и анализ внутриклеточных пептидных пулов

На сегодняшний день основным методом идентификации эндогенных пептидов в тканях и биологических жидкостях является тандемный масс-спектрометрический

анализ. Несмотря на то, что к настоящему времени пептидные пулы тканей и жидкостей животных проанализированы достаточно подробно, крайне мало известно о растительных пептидомах.

4.1.1.1 Идентификация продуктов деградации белков-предшественников в протонеме и гаметофорах

На первом этапе исследования мы использовали масс-спектрометрический анализ для того, чтобы проанализировать внутриклеточные пептиды двух жизненных форм модельного растения - мха *Physcomytrium patens* (Рис. 4.1). Для идентификации полученных спектров была использована база аннотированных белков *P. patens* из Phytozome V12. Данный подход позволил нам идентифицировать пептиды размером от 5 до 78 а.о., являющиеся продуктами деградации известных белков. В гаметофорах был идентифицирован в общей сложности 4361 пептид, который образовался из 1859 белков-предшественников. Поскольку большинство идентифицированных предшественников были аннотированы в базе данных как "предсказанные" белки, для определения их возможных функций и локализации мы использовали анализ гомологии BLAST в базе данных белков зеленых растений (*Viridiplantae*).

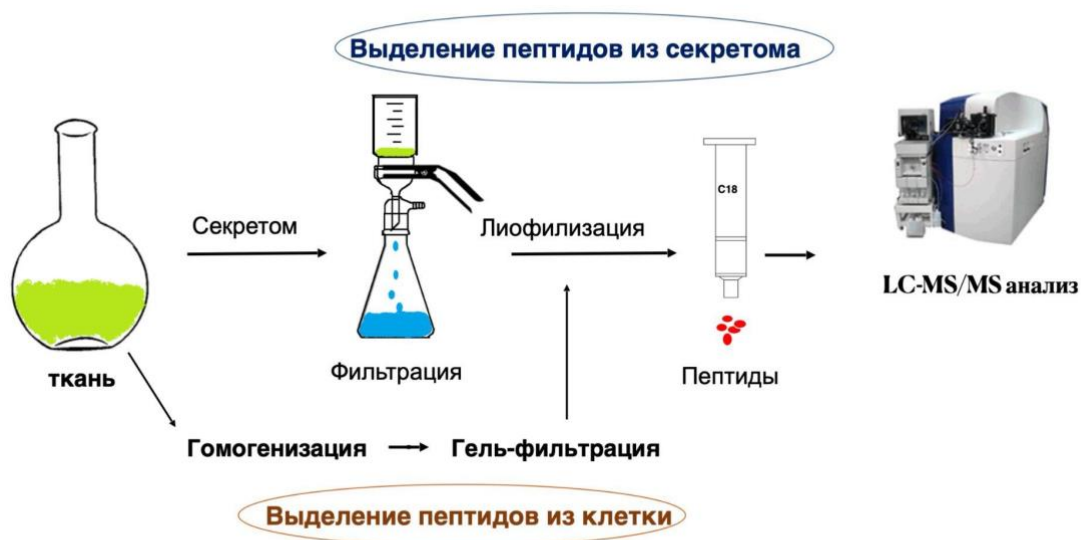


Рисунок 4.1 Схема выделения внутриклеточных и секретируемых эндогенных пептидов из протонемы мха *Physcomitrium patens*.

Однако функции наиболее пептидогенных белков-предшественников в гаметофорах, таких как A9U4I0, A9RXW5 и A9RXR3, из которых было идентифицировано соответственно 144, 115 и 111 уникальных пептидов, остались неизвестными. Кроме того, среди наиболее пептидогенных белков-предшественников 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).

В клетках протонемы мы идентифицировали 4332 пептида, которые являются фрагментами 1554 белков-предшественников. Основные пептидогенные белки-предшественники в протонеме - это большая субъединица рибулозо-1,5-бисфосфаткарбоксилазы/оксигеназы (RuBisCO; 111 нативных пептидов), фактор элонгации 1-альфа (A9RGA5; 80 нативных пептидов), хлоропластная фруктозо-бисфосфат альдолаза (A9RX76; 48 нативных пептидов) и субъединица реакционного центра фотосистемы I (A9SPD7, 47 нативных пептидов). Таким образом, внутриклеточные пептидные пулы растений включают в себя несколько тысяч пептидов - продуктов деградации аннотированных белков (**Рис. 4.2**).

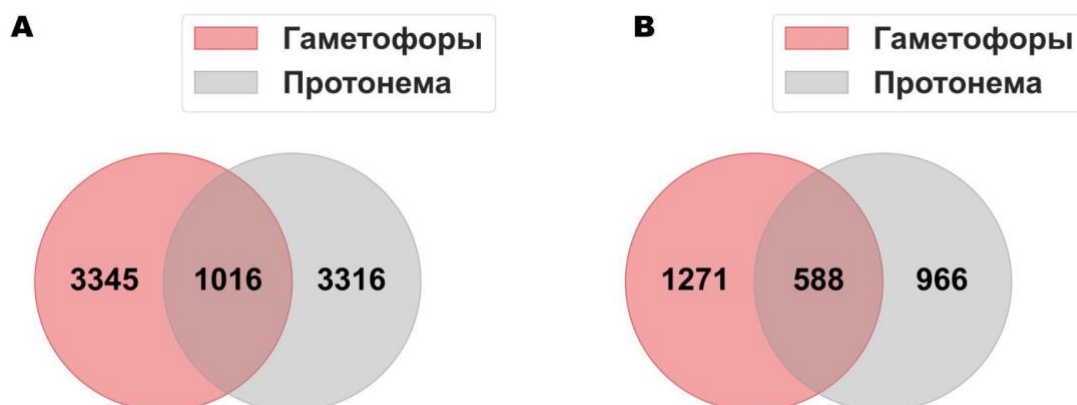


Рисунок 4.2 Диаграммы Венна, показывающие пересечение идентифицированных в протонеме и гаметофорах **A** - пептидов, **B** - белков-предшественников.

4.1.1.2 Биоинформатический анализ физико-химических свойств внутриклеточных пептидов

Далее мы сравнили предсказанные уровни гидрофобности (показатели “gravy” и “aromaticity”), нестабильности (“instability”) и значение изоэлектрической точки (“pI” при pH=7) идентифицированных пептидов в гаметофорах и протонеме мха (**Рис. 4.3**). Данный анализ позволил нам сравнить физико-химические свойства пептидных пулов, выделенных из двух разных типов ткани. 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 состав пептидных пулов оказывает влияние как состав белков в соответствующей ткани, так и процесс пробоподготовки, который, например, мог привести к снижению количества гидрофобных пептидов.

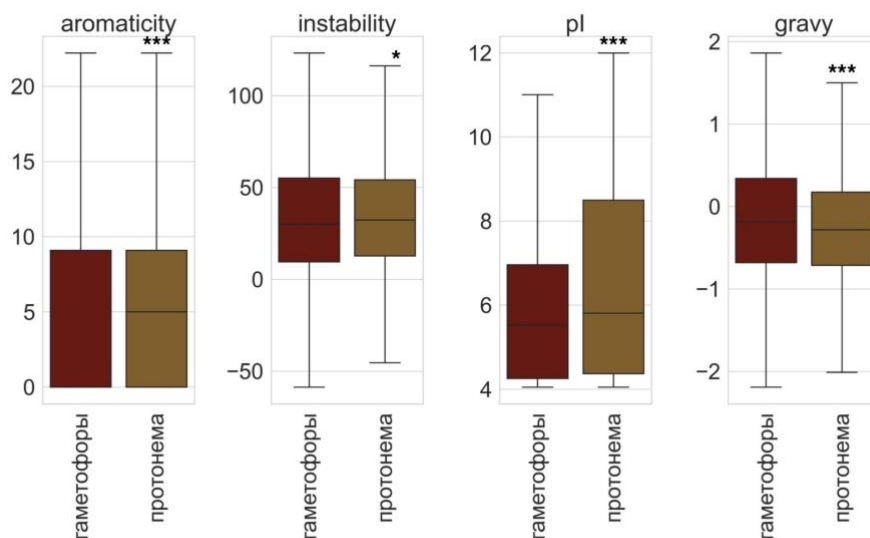


Рисунок 4.3 Сравнение предсказанных на основе аминокислотных последовательностей процента ароматических аминокислот ("aromaticity"), уровня гидрофобности ("gravy"),

стабильности ('instability') и значения изоэлектрической точки ("pI" при pH=7) пептидов, выделенных из протонемы и гаметофоров мха. *** $P < 0.0001$; * $P < 0.05$.

4.1.1.3 Анализ шаблонов деградации белков-предшественников

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 при помощи протеасом. После того, как первичные фрагменты высвобождаются из протеасомы, они далее деградируют на более мелкие пептиды с помощью амино- и карбоксипептидаз. Примеры покрытия некоторых белков-предшественников пептидами, идентифицированными в протонеме, показаны на Рисунке 4.4.

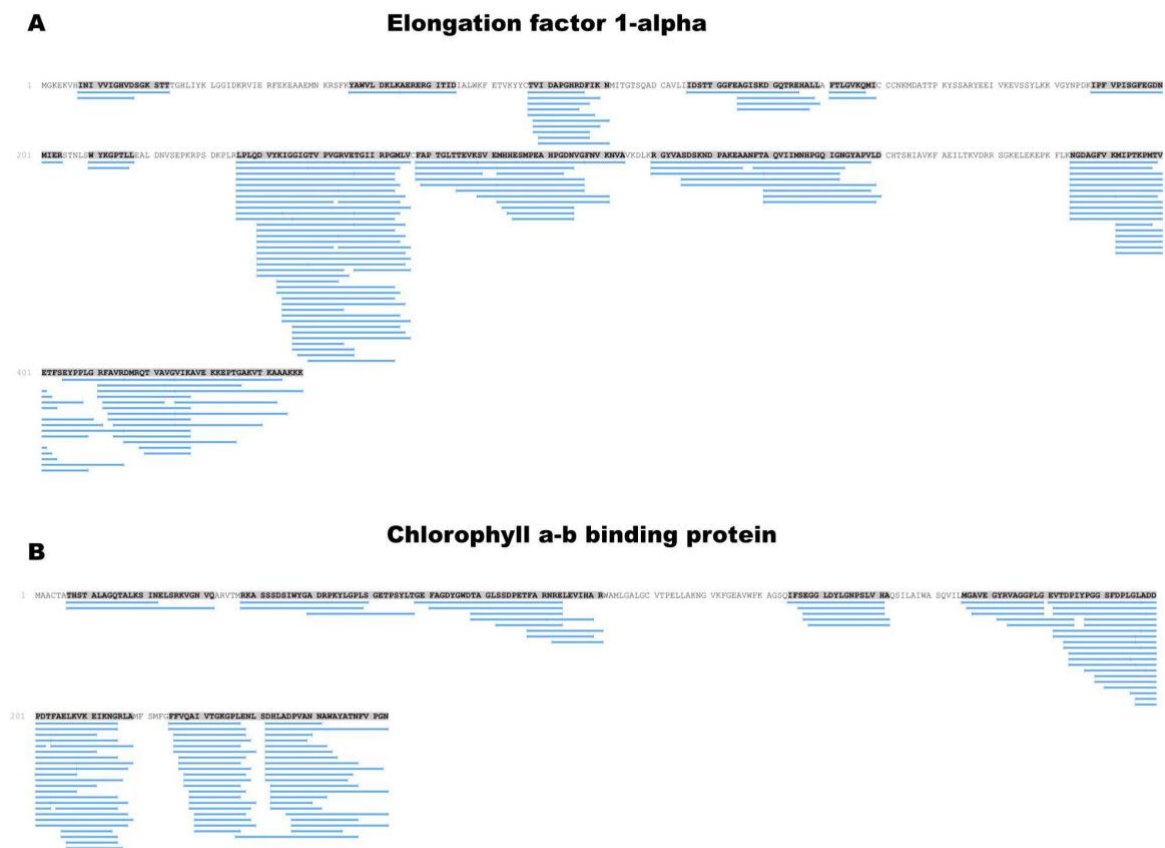


Рисунок 4.4 Анализ распределения по последовательности белка идентифицированных пептидов для **A** - фактора элонгации 1a и **B** - хлорофилл-связывающего белка. Пептидные последовательности схематично обозначены как синие линии.

Затем мы сравнили шаблоны деградации отдельных белков в двух жизненных формах - протонеме и гаметофорах. Множество белков-предшественников, например, таких как фруктозо-бисфосфат альдолаза (A9RX76), фактор элонгации 1-альфа (A9RGA5), липоксигеназа (A9U1H4), имели сходные профили покрытия эндогенными пептидами (**Рис. 4.5A и B**). В то время как для некоторых белков, в частности, A9RD61 (основной аллерген Mald 1), шаблон деградации был различным в двух жизненных формах - протонеме и гаметофорах (**Рис. 4.5C**).

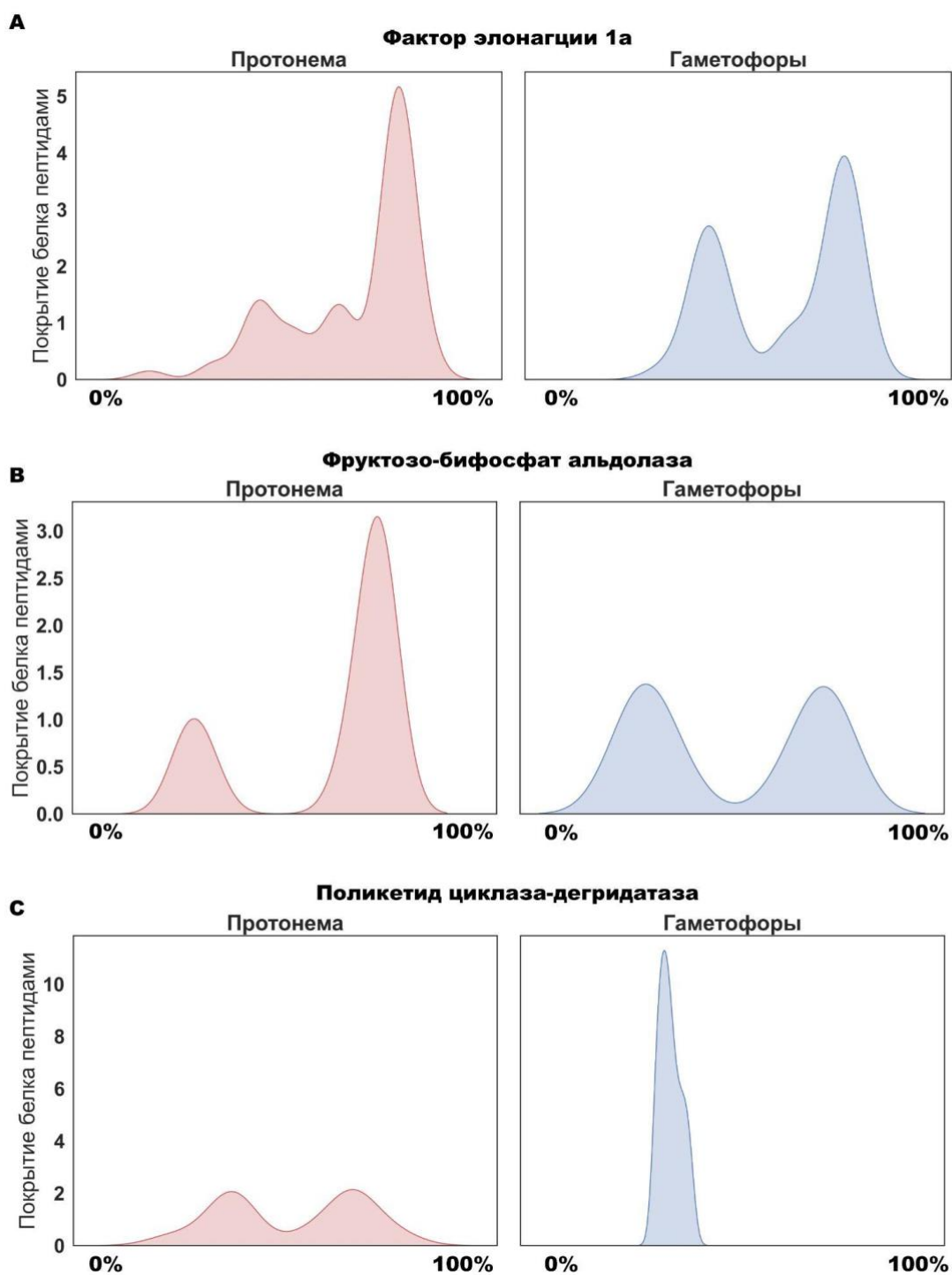


Рисунок 4.5 Анализ распределения идентифицированных пептидов по длине белка с помощью графика плотности. **А** - фактор элонгации (A9RGA5); **В** - фруктозо-бифосфат альдолаза (A9RX76); **С** - основной аллерген Mald 1 (A9RD61).

4.1.1.4 Анализ белков-предшественников пептидов из протонемы и гаметофоров

Анализ локализации белков-предшественников пептидов в гаметофорах и протонеме мха показал, что значительная доля пептидов (примерно 25% от общего количества) образуется при деградации хлоропластных белков. Доли пептидов от митохондриальных (3% для протонемы и 10% для гаметофоров), ядерных (4% для протонемы и 5% для гаметофоров) предшественников, а также белков, участвующих в трансляции (8% для протонемы и 3% для гаметофоров), также были довольно высокими. Белки Late Embryogenesis Abundant (LEA), которые участвуют в защите растений от абиотических стрессов [266], были одними из наиболее пептидогенных белков-предшественников в гаметофорах. Количество пептидов от LEA белков в гаметофорах было выше, чем в протонеме (164 нативных пептида в гаметофорах, 38 пептидов в протонеме), в то время как пептиды других шапероноподобных белков, таких как белки теплового шока, были более распространены в протонеме (125 пептидов в протонеме, 66 в гаметофорах). На основе полученных данных можно сделать вывод, что, также как и в животных клетках, значительную долю пептидома растений составляют фрагменты белков органелл.

Таким образом, в пептидоме молодой и быстрорастущей протонемы мы идентифицировали большое количество пептидов, которые являются фрагментами белков хлоропластов. Большое количество пептидов, полученных из хлоропластных белков, может быть связано с быстрым ростом клеток протонемы, что приводит к интенсивному метаболизму белков, связанных с клеточными энергетическими процессами. Например, Nelson et al. продемонстрировали, что некоторые белки фотосинтетического аппарата клетки обновляются со скоростью выше средней [267]. Также белки хлоропластов могут являться источником аминокислот, как это было показано ранее для рубилозо-бифосфат карбоксилазы [268]. Мы также обнаружили, что белки, участвующие в углеводном обмене, биосинтезе растительных гормонов, терпеноидов и стероидов, являются одними из наиболее пептидогенных белков-предшественников. Мы предполагаем, что этот факт коррелирует со скоростью оборота этих белков в клетках.

В дополнение к фрагментам хлоропластных белков, пептидом гаметофоров также содержит большое количество пептидов от белков-шаперонов и белков, связанных с ответом на недостаток влаги, такие как LEA, аквапорины, AWPМ-19-подобные белки и формиатдегидрогеназы. Возможным объяснением может быть тот факт, что гаметофоры, в отличие от протонемы, растут в воздушной среде, и регуляция водного баланса особенно важна. Можно предположить, что увеличение представленности таких белков-предшественников в протеоме приводит к увеличению количества эндогенных пептидов от этих предшественников. Однако для того, чтобы сделать окончательный вывод, нам необходимо знать скорость синтеза/деградации данных белков и механизмы их деградации. Белки LEA были впервые обнаружены в семенах растений [269,270], а у *P. patens* они экспрессируются на базальном уровне в гаметофорах, что также может быть связано с защитой от водного дефицита [271]. Неизвестно, имеют ли эндогенные пептиды от шапероноподобных белков, таких как LEA, какие-либо защитные функции в клетках мха, хотя предыдущие исследования показывают, что короткие фрагменты белка LEA могут независимо выполнять функции шаперонов [272].

Далее мы провели функциональный анализ аннотированных белков-предшественников пептидов из протонемы и гаметофоров с использованием базы данных DAVID (The Database for Annotation, Visualization and Integrated Discovery; <http://david.abcc.ncifcrf.gov>). 1691 белок-предшественник, который был сопоставлен с белками в базе данных DAVID, был разделен на 15 кластеров, содержащих 1030 белков. Наиболее представленными кластерами в протонеме и гаметофорах были белки-предшественники, локализованные в хлоропластах, и структурные компоненты рибосом. Другие белки-предшественники были вовлечены в процессы углеводного обмена, трансмембранной транспортной активности и биосинтеза растительных гормонов, терпеноидов и стероидов.

Можно предположить, что процесс деградации белков с образованием внутриклеточных пептидов происходит случайным образом и что увеличение представленности белка в протеоме коррелирует с увеличением количества эндогенных пептидов, продуктов гидролиза этого белка, в пептидоме. 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 (**Рис. 4.6**).

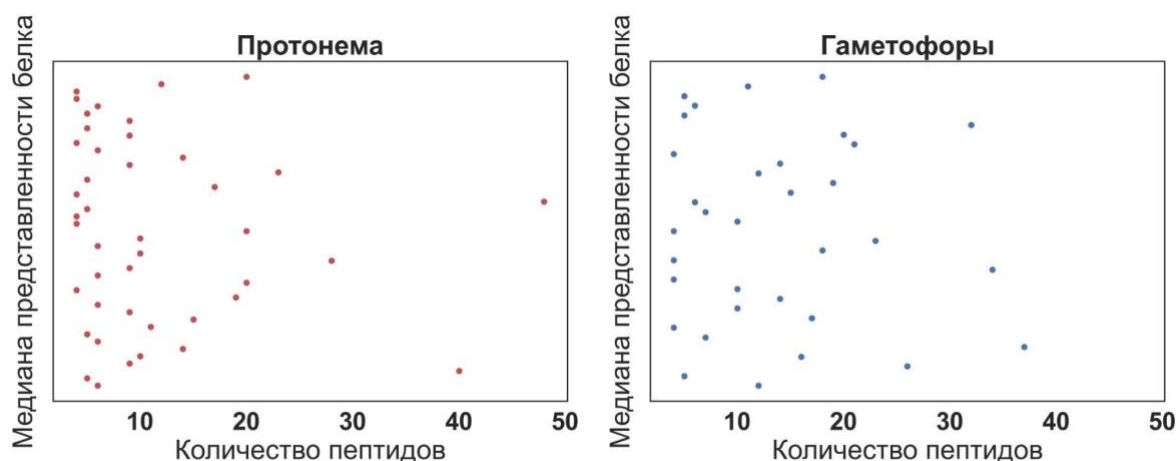


Рисунок 4.6. Диаграмма рассеивания, показывающая зависимость числа идентифицированных пептидов на белок-предшественник и представленность соответствующего белка в клетке (на основе анализа количественного анализа протеома протонемы и гаметофоров с использованием программного пакета Progenesis LC-MS (Nonlinear Dynamics, Durham, NC, USA).

Например, из таких белков-предшественников как A9U1H4 (линолеат 9S-липоксигеназа 4), A9RD61 (основной аллерген Mald 1) и A9U188 (глицератдегидрогеназа), выщеплялось больше эндогенных пептидов в протонеме в сравнении с гаметофорами, хотя в первой жизненной форме представленность белка была выше. В других случаях мы наблюдали увеличение количества эндогенных пептидов вместе с увеличением количества белка. Это такие белки-предшественники как A9RAS2 (белок, связывающий хлорофилл a-b, хлоропласт), A9SL09 (субъединица VI реакционного центра фотосистемы I), A9SX31 (супероксиддисмутаза [Cu-Zn]), A9T3W5 (малая цепь хлоропластной бифосфаткарбоксилазы) и ряд других. Полученные результаты могут быть объяснены несколькими причинами: 1) количество пептидов может быть связано

со скоростью синтеза/деградации белков (protein turnover), а не с концентрацией белка; 2) существуют различия в активности протеолитических ферментов в зависимости от стадии жизненного цикла; 3) деградация пептидов определенного состава зависит от стадии жизненного цикла; 4) особенности пробоподготовки пептидов для масс-спектрометрического анализа.

4.1.1.5 Возможные механизмы образования внутриклеточных пептидов

Можно предположить, что в растительной клетке значительная доля пептидов может продуцироваться не протеасомой, а различными органелльными протеазами. Чтобы проверить это предположение, мы далее сравнили аминокислоты в P1 позиции в нашем наборе пептидомных данных (**Рис. 4.7**) и в эндогенных пептидах, связанных с протеасомами в пептидомном датасете из клеток человека [39]. Известно, что протеазная активность протеасом нацелена на специфические аминокислотные мотивы, обогащенные отрицательно заряженными остатками (D, E; каспазоподобные), гидрофобными остатками (W, F, M, Y; химотрипсин) и положительно заряженными остатками (R, K; трипсиноподобные). Анализ C-концов пептидов, ассоциированных с протеасомой, указал на наличие каспазоподобной и химотрипсиноподобной, но не трипсиноподобной активности протеасом в наборе пептидома человека (**Рис. 4.7**). В нашем наборе данных аминокислоты в P1 позиции 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 органелльных предшественников, которые выщепляются специализированными протеазами; 3) особенностями выделения пептидов из растительной клетки.

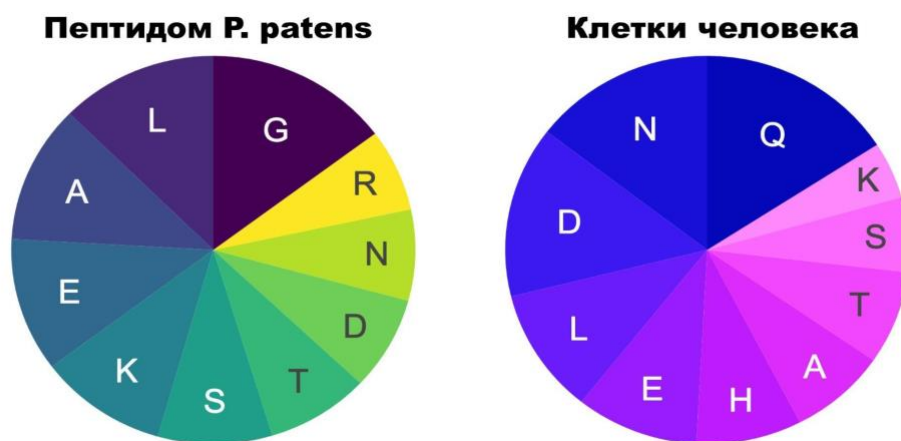


Рисунок 4.7 Частота встречаемости топ10 аминокислот на С-концах внутриклеточных пептидов мха в сравнении с пептидами человека, ассоциированными с протеасомами [39].

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

4.1.2 Идентификация и анализ внеклеточных пептидных пулов

4.1.2.1 Идентификация пептидов в секрете протонемы мха

Показано, что секретом растений содержит десятки белков [273], однако мало известно о составе секретируемых пептидных пулов и протеаз, ответственных за образование этих пептидных пулов. Исходя из этого, мы провели всесторонний анализ пептидома секрета растений, используя в качестве модели клеточную среду, в которой культивировали протонему *P. patens*. Используя масс-спектрометрический анализ, в секрете протонемы мы идентифицировали 802 эндогенных пептида, являющихся продуктами протеолиза 220 белков-предшественников. Из них 387 эндогенных пептидов были идентифицированы, по меньшей мере, в двух биологических повторах, что позволяет предположить существование стабильного пула пептидов в секрете мха. По сравнению с внутриклеточным пептидомом протонемы, наиболее пептидогенными белками-предшественниками в секрете были мембранные и секретируемые белки, такие как экспансин A0A2K1K5N6 (Pr3c8_870V3.1.p; 72 уникальных пептида), который

является консервативным белком наружной мембраны; пектинэстераза A0A2K1IDP0 (Pr3c25_3840V3.1.p; 21 уникальный пептид); субтилизиноподобная сериновая протеаза A0A2K1JI70 (Pr3c14_17710V3.1.p; 19 уникальных пептидов). Последний белок является наиболее распространенной субтилизиноподобной протеазой, идентифицированной в секрете мха [274]. Мы также обнаружили, что около 20% пептидов секрета были общими с пептидомами клеток (**Рис. 4.8**). 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. Эти выводы согласуются с тем фактом, что секретом мха содержит внутриклеточные белки, в том числе, хлоропластные [273,274].

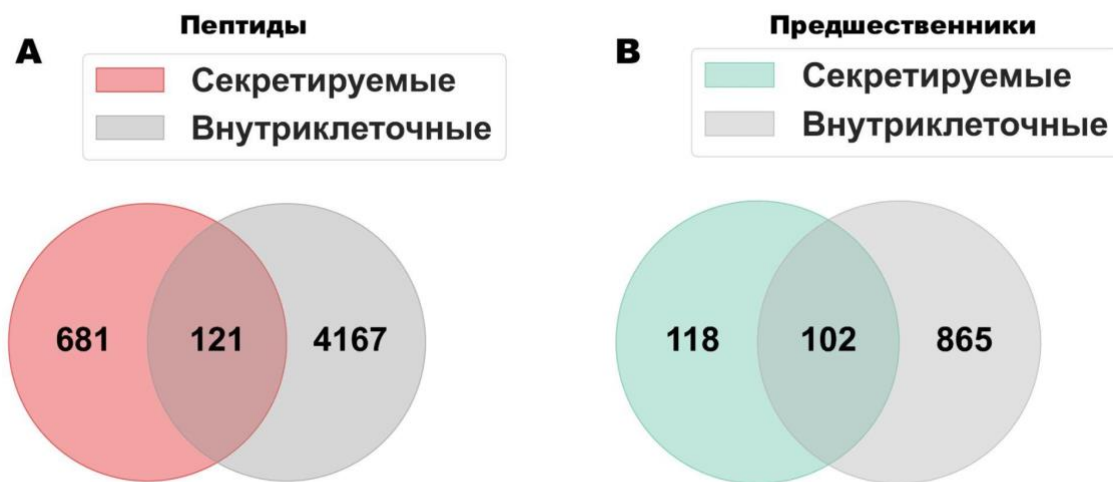


Рисунок 4.8 Диаграмма Венна, показывающая сравнение эндогенных пептидов (**А**) и их предшественников (**В**) между внутриклеточными и секретируемыми пептидными пулами ($n = 3$ независимых биологических повтора)

4.1.2.2 Сравнение физико-химических свойств внутриклеточных и внеклеточных пептидов

Используя биоинформатические подходы, далее мы оценили различия в физико-химических свойствах внеклеточных и внутриклеточных пептидов. На первом этапе были рассчитаны средние значения 544 индексов из базы AAindex для

каждого пептида из внутриклеточного и секретируемого пептидома. Сравнение внутриклеточных и внеклеточных пептидов выявило достоверные различия (критерий «Манна-Уитни», P -value < 0.00001) по 197 (36%) индексам. Среди шести индексов, показывающих наиболее выраженные различия между клеточными пептидами и пептидами секретома (критерий «Манна-Уитни», $P < 10^{-25}$), три индекса (KLEP840101, ZIMJ680104 и FAUJ880111) указывали на то, что пептиды секретома имели меньше положительно заряженных аминокислот, чем клеточные пептиды (Рис. 4.9). Затем мы выполнили иерархическую кластеризацию Уорда всех 544 индексов и выявили шесть различных кластеров. Кластер, содержащий AA-индексы, отражающие гидрофобность аминокислот, был обогащен секретомными пептидами (точный критерий Фишера, $P < 0.05$).

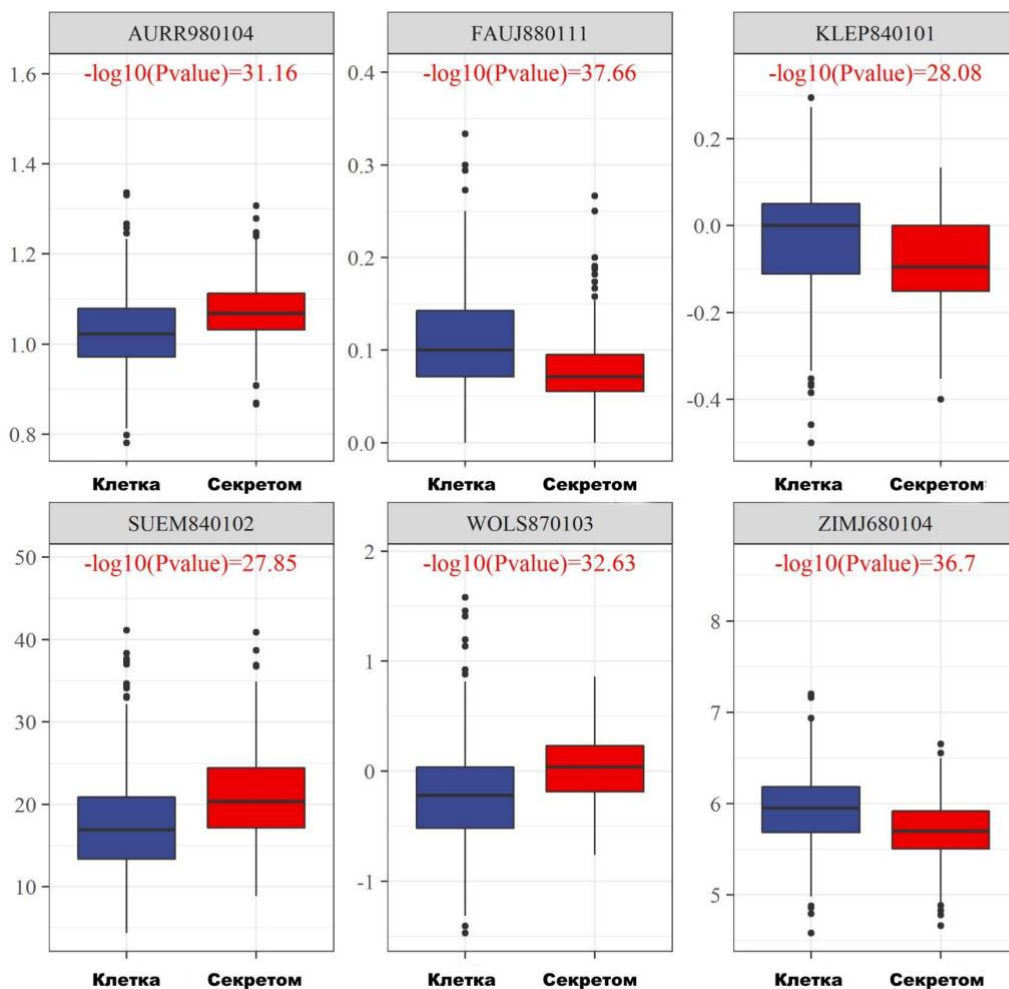


Рисунок 4.9 Различия в средних значениях шести наиболее значимых AAindex индексов ($P < 10^{-25}$). KLEP840101: заряд пептида, FAUJ880111: положительный заряд, ZIMJ680104:

изоэлектрическая точка, WOLS870103: значение Z3, AURR980104: нормализованная позиционная остаточная частота на концах спирали N' и SUEM840102: параметр Zimm-Bragg $\times 1.0E4$.

Чтобы получить более полное представление о механизмах формирования внеклеточного пула пептидов, мы предсказали в белках-предшественниках мембранные альфа-спирали. Этот анализ показал, что доля предсказанных трансмембранных доменов была схожа для предшественников внутриклеточных и секретируемых пептидов (27% и 33%, соответственно). Однако пептиды, происшедшие из областей белков, расположенных внутри клетки, были значительно больше представлены во внутриклеточном пептидоме клетки (точный критерий Фишера, $P < 10^{-6}$). Напротив, пептиды, происходящие из областей белков, которые, как предсказано, находятся за пределами клеточной мембраны, были значительно больше представлены в секрете (точный тест Фишера, $P = 0,006516$). Эти результаты дают дополнительные доказательства того, что идентифицированный пул пептидов действительно является частью секрета, а не артефактом, связанным с пробоподготовкой образца.

Используя биоинформатический анализ, мы дополнительно предсказали гидрофобность (показатели “gravity” и “aromaticity”), стабильность (“instability”) и значение изоэлектрической точки (“pI” при pH=7) у пептидов, идентифицированных в секрете, и внутриклеточных пептидов (**Рис. 4.10**). Мы обнаружили значительно более высокий процент ароматических аминокислот и значительно более низкие значения pI у секретируемых пептидов по сравнению с внутриклеточными пептидами. Таким образом, аминокислотный состав и физико-химические свойства пептидов значительно различаются у клеточных и секретируемых пептидомов.

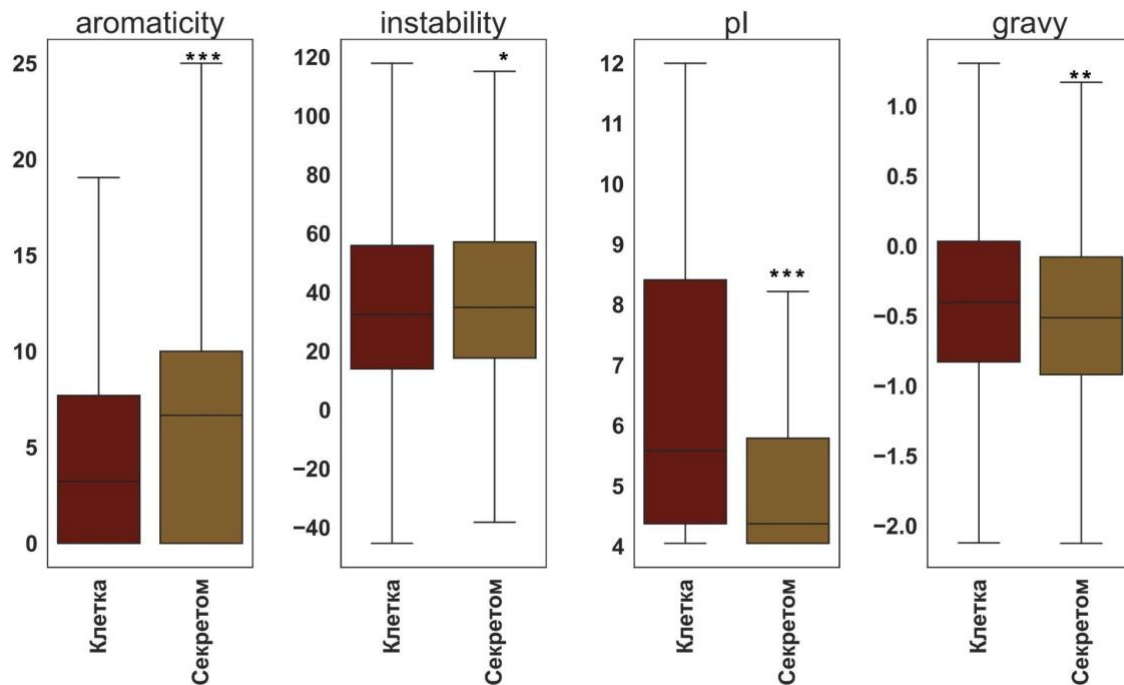


Рисунок 4.10 Сравнение предсказанных на основе аминокислотных последовательностей значений процента ароматических аминокислот ("aromaticity"), уровня гидрофобности ("gravy"), стабильности ('instability') и значения изоэлектрической точки ("pI" при pH=7) у секретируемых и внутриклеточных пептидов. *** $P < 0.0001$; * $P < 0.05$.

4.1.2.3 Анализ белков-предшественников внеклеточных пептидов

Предыдущие исследования внеклеточного протеома *P. patens* показали присутствие внутриклеточных белков в секрете [273,274]. Мы сравнили предшественников пептидов секрета с белками, ранее идентифицированными в секретах *P. patens* [273,274]. Перекрытие было достаточно низким - всего 13 белков были общими среди всех трех наборов данных (Рис. 4.11). Такое незначительное пересечение белков-предшественников внеклеточных пептидов из наших данных и секретируемых белков из других исследований может указывать на то, что внеклеточный пептидом не является побочным продуктом деградации внеклеточных белков.

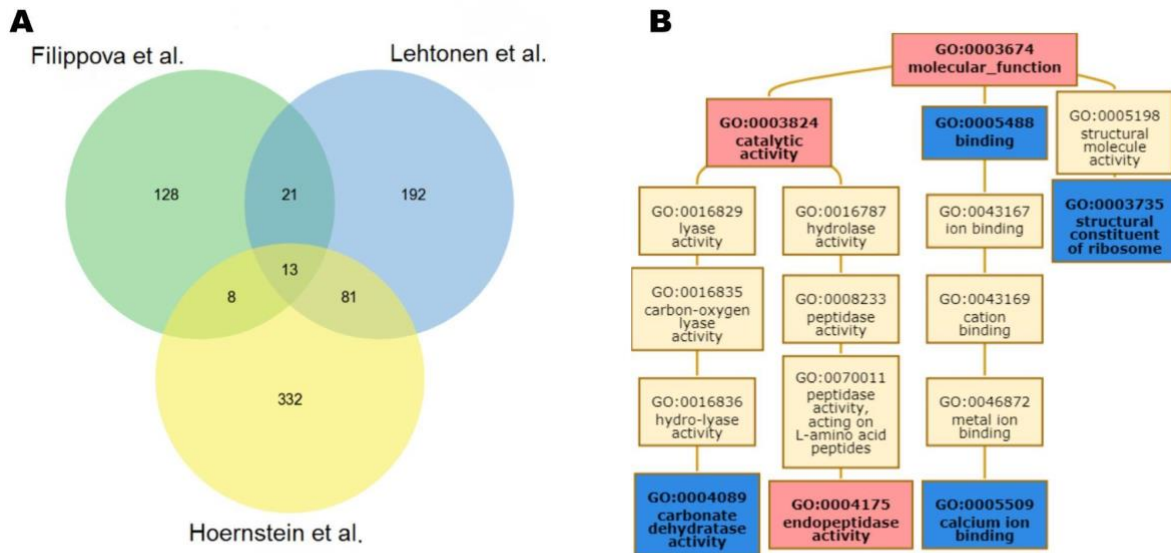


Рисунок 4.11 А - Диаграмма Венна, показывающая сравнение между идентифицированными белками-предшественниками пептидов в наших экспериментах [275] и секретируемыми белками, ранее идентифицированными в протеомных экспериментах у *P.patens* [273,274]; **В** - Анализ обогащения молекулярных функций белков-предшественников с использованием сервиса GO View (<https://github.com/foolin/goview>). Результаты обогащения выделены красным. Для GO: 0004175, $P = 0,008$.

Среди общих белков мы обнаружили субтилизин-подобную протеазу - кукумизин (Pr3c14_17710) и белки, участвующие в ремоделировании клеточной стенки растений (пектинэстеразы и другие).

Анализ терминов GO показал, что большая часть белков-предшественников пептидов в нашем наборе данных участвовала в фотосинтетических реакциях и модификациях клеточной стенки (деградация пектина). Мы идентифицировали 12 белков хлоропластной локализации, таких как psbA и psbB белки фотосистемы II, малая и большая субъединицы RUBISCO. Интересно, что в секрете не удалось идентифицировать пептиды от митохондриальных белков. Анализ обогащения терминов GO белков-предшественников пептидов показал, что наиболее существенно обогащены такие молекулярные функции как каталитическая ($P=0,000216$) и эндопептидазная активности (значение $P=0,008$).

Наши результаты подтверждают предыдущие данные, полученные при анализе пептидома млекопитающих. Например, показано, что мозг мыши продуцирует большое количество секретируемых нативных пептидов, образующихся из внутриклеточных белков, которые в основном расположены в митохондриях и цитозоле, а не из белков секреторного пути [36]. Кроме того, человеческая плазма содержит тысячи пептидов, включая пептидные гормоны и фрагменты внутриклеточных белков [40].

4.1.2.4 Механизмы образования внеклеточных пептидов

Недавние исследования показали, что определенные типы протеаз могут участвовать в образовании нативных пептидов в жидкостях человеческого тела [276]. Однако какие протеазы ответственны за формирование внеклеточного пептидома у растений, практически неизвестно. Поэтому мы *in silico* оценили типы протеаз, которые могут отвечать за формирование пептидных пулов в секрете растений. Такие программы как Proteasix (<http://www.proteasix.org>), Prosper (<https://prosper.erc.monash.edu.au>) и ExPASy PeptideCutter (https://web.expasy.org/peptide_cutter) были ранее использованы для предсказания протеаз, участвующих в генерации пептидных пулов у млекопитающих и бактерий, соответственно [277,278]. Однако такие подходы еще не были опробованы для растений. Сначала мы применили ранее упомянутый инструмент ExPASy PeptideCutter и 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 в секрете вносят сериновые протеазы. Далее мы использовали матрицы специфичности протеаз из базы данных MEROPS (<https://www.ebi.ac.uk/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. 4.12).

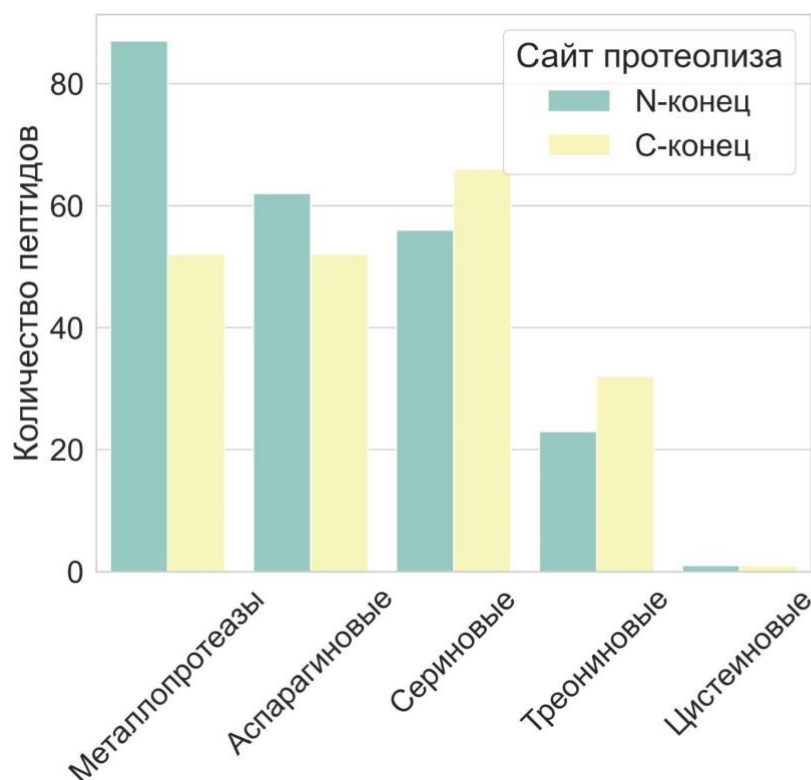


Рисунок 4.12 Гистограмма, иллюстрирующая сопоставление N- и C-концов идентифицированных пептидов с матрицами специфичности различных типов протеаз из базы MEROPS.

Ранее проведенный анализ внеклеточного протеома *P. patens* выявил наличие протеаз серинового типа, цистеиновых протеаз, аспартат- и металлопротеаз в секрете протонемы [274]. Согласно этой работе, сериновая протеаза типа S8 (субтилизин-подобная) была одним из наиболее широко представленных белков во внеклеточном протеоме *P. patens*. Нокаут этой протеазы приводит к резкому снижению внеклеточной протеолитической активности при культивировании протонемы *P. patens*. Кроме того, протеазы аспарагинового типа и металлопротеиназы также имели высокую представленность в секрете мха [274]. Это согласуется с нашим *in silico* анализом сайтов расщепления пептидов, согласно

которому металлопротеазы и протеазы серинового типа участвуют в образовании большинства уникальных пептидов в секрете мха.

Также наши результаты согласуются с недавними исследованиями пептидома млекопитающих, которые предполагают, что протеазы вносят значительный вклад в образование протеома и пептидома тканей и биологических жидкостей. Например, у человека самые распространенные классы протеаз, участвующих в формировании пептидомов жидкостей организма, это протеазы серинового типа и металлопротеиназы. Кроме того, сериновые и аспарагиновые протеазы, как известно, играют важную роль в различных физиологических процессах [279]. Более ранние исследования подтвердили участие аспарагиновых протеаз в ответе на стрессовые факторы и в процессах роста и развития растений [280]. Например, аспарагиновая протеаза из семейства пепсинов - фитепсин, играет важную роль в защите растений от стрессовых факторов. Показано, что у *Fagopyrum esculentum*, экспрессия фитепсина повышается при различных абиотических стрессах [281]. У арабидопсиса секретлируемая сериновая протеаза S8 участвует в распознавании патогенов и активации сигнальных процессов за счет расщепления мембранных белков [282].

Таким образом, наши данные по анализу секрета растительной клетки позволяют предположить, что внеклеточные пептидные пулы животной и растительной клетки представляют собой сложную смесь специализированных пептидных гормонов и фрагментов белков-предшественников, включая полноразмерные или более короткие внутриклеточные пептиды. 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 секретлируемых пептидных пулов в растениях. Внутриклеточные пептиды, образующиеся при протеасомальной деградации предшественников, могут быть экспортированы в секрет. Например, небольшая часть пептидома транспортируется транспортерами АТФ-связывающей кассеты (ABC) в

эндоплазматический ретикулум (ER) и лизосомы у животных [283,284]. Кроме того, белки хлоропластов могут подвергаться гидролизу в специальных везикулах или транспортироваться в вакуоль. Пептиды далее могут быть экспортированы в апопласт во время слияния вакуольных пузырьков с плазматической мембраной. Наконец, функциональные белки и/или специализированные предшественники биоактивных пептидов могут быть секретированы и подвергаться действию различных протеаз во внеклеточном пространстве.

4.1.3 Влияние стрессовых факторов на состав пептидных пулов растений

В ряде недавних исследований было показано, что стрессовые факторы влияют на механизмы деградации белков, и это может приводить к изменениям пептидных пулов и образованию биоактивных пептидов [62,285,286]. Мы использовали несколько моделей стрессовых воздействий - повреждение клеточной стенки; обработка стрессовыми гормонами - жасмоновой и салициловой кислотами, - для того, чтобы оценить их влияние на состав внутриклеточных и секретируемых пептидных пулов мха *P. patens*.

4.1.3.1 Анализ пептидома протопластов мха

Протопласты мха *P. patens* высвобождаются после обработки молодой ткани протонемы драйзелазой, смесью природных ферментов из *Basidiomycetes* sp. и проявляющей ламинариазную, ксиланазную и целлюлазную активности. Обработка драйзелазой моделирует повреждение клеточной стенки в результате атаки фитопатогенов и приводит к индукции иммунного ответа у растений [287]. Кроме того, протопласты являются хорошей моделью для изучения репрограммирования соматических растительных клеток [288].

Используя масс-спектрометрический анализ, мы проанализировали пептидом протопластов и обнаружили, что он претерпевает значительные изменения в сравнении с пептидомом протонемы и содержит 20,427 пептидов размером от 6 до 78 а.о., которые были фрагментами 3648 белков-предшественников (**Рис. 4.13**).

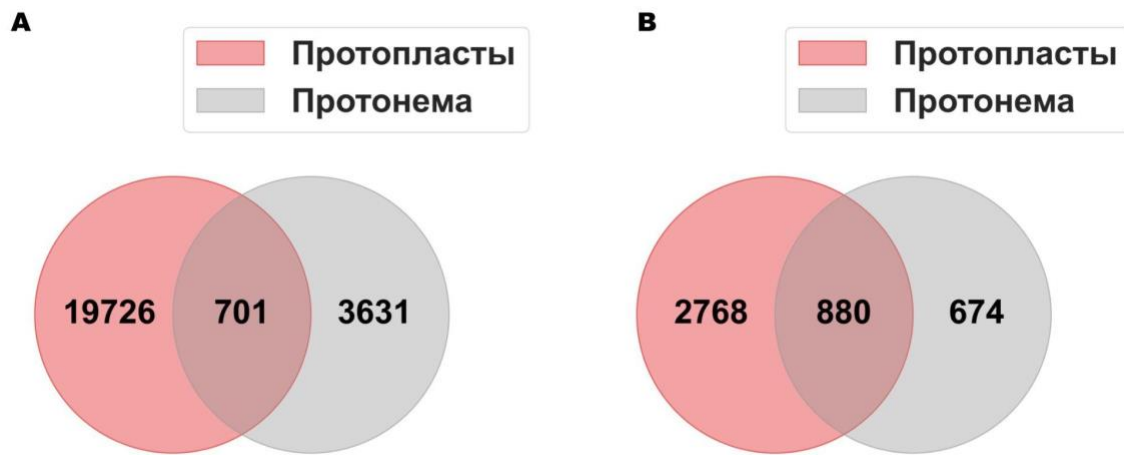


Рисунок 4.13 Диаграммы Венна, показывающие пересечение идентифицированных в протопластах и протонеме мха **А** - пептидов, **В** - белков-предшественников.

Примечательно, что значительная часть пептидома представляла собой "пептидные лестницы", в которых пептиды отличались друг от друга деградацией в С- или N-концах. Сравнение пептидов, идентифицированных в протонеме, гаметофорах и протопластах, показало, что пути деградации белков могут отличаться в зависимости от типа ткани (**Рис. 4.14**).

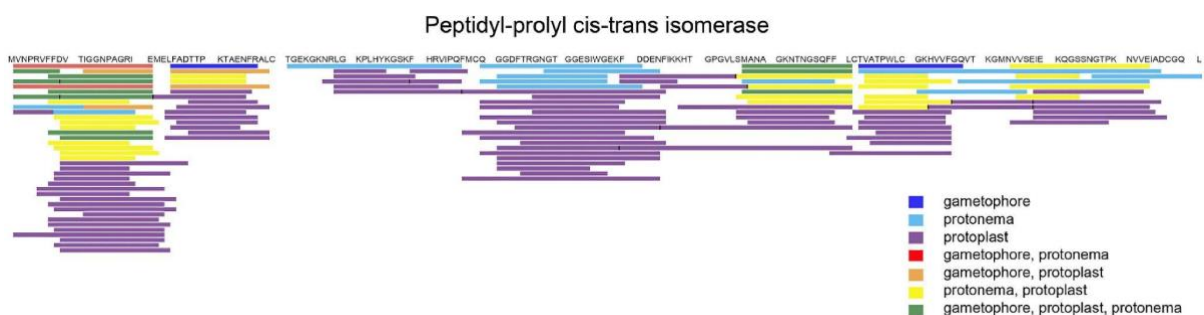


Рисунок 4.14 Пептидное выравнивание белка-предшественника A9TK88 (пептидил-пролил-цис-транс-изомеразы). Полосы указывают на положение идентифицированных пептидов.

Чтобы оценить долю пептидов, которые могли образоваться в результате гибели клеток и не являлись нативным пептидомом клеток, мы оценили жизнеспособность

протопластов, используя краситель трипановый синий, который окрашивает мертвые клетки и ткани. В ходе анализа нам не удалось идентифицировать значительное количество окрашенных клеток (данные не показаны). Такие результаты показывают, что мертвые или умирающие протопласты не оказали существенного влияния на наши результаты. Мы также не обнаружили значительного количества пептидов в промывном растворе, который используется в процессе получения протопластов (данные не показаны).

В пептидоме протопластов мы идентифицировали большое количество пептидов, полученных из хлоропластных белков. Наибольшее количество пептидов образовалось из большой субъединицы фермента рибулозо-1,5-бисфосфат карбоксилазы. Среди других наиболее пептидогенных белков-предшественников были субъединица II-2 реакционного центра фотосистемы I, активатора рибулозо-1,5-бисфосфат карбоксилазы, карбоангидраза, фактор элонгации трансляции 1-альфа, липоксигеназа и пластоцианин. Все эти данные указывают на интенсивную деградацию хлоропластных белков в протопластах. Имеются данные о том, что хлоропластные белки, такие как RuBisCO, являются основными мишенями для цистеиновой протеазы в вакуолях растительных клеток [289].

Для того, чтобы проверить, связано ли такое существенное увеличение количества пептидов от хлоропластных белков с их деградацией непосредственно в этой органелле, мы выполнили пептидомный анализ хлоропластов. В хлоропластах, выделенных из протопластов, мы идентифицировали 82 уникальных пептида, которые были фрагментами 21 белка-предшественника; только три пептида были фрагментами большой субъединицы RuBisCO, 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.

Увеличение количества эндогенных пептидов в ответ на повреждающее воздействие на клеточную стенку может указывать на функциональность пептидома клетки в условиях биотического стресса. Это может быть, например,

связано с активным выщеплением биоактивных пептидов из специализированных белков предшественников. Чтобы изучить факторы, влияющие на эти изменения в пептидных пулах протонемы мха, мы произвели обработку драйзелазой более низкой концентрации, чем та, которая использовалась при выделении протопластов. При обработке клеток протонемы 0,025% раствором драйзелазы протопласты не образовывались. Однако обработка протонемы драйзелазой в этой концентрации увеличила количество нативных пептидов во внутриклеточном пептидоме примерно в 2,5 раза по сравнению с контролем (**Рис. 4.15**). Этот результат свидетельствует о том, что повреждающее воздействие на клетку усиливает процессы протеолиза и генерацию большого количества эндогенных пептидов.

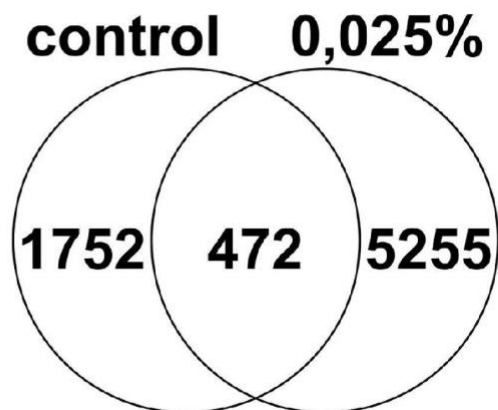


Рисунок 4.15 Диаграмма Венна, показывающая количество пептидов, идентифицированных протонеме, обработанной 0,025% раствором драйзелазы, по сравнению с необработанным контролем (control).

В связи с большим количеством пептидов от хлоропластных белков, мы также оценили воздействие усиления протеолиза на фотосинтетический аппарат хлоропластов. В клетках протонемы, подвергшихся воздействию света, рабочая квантовая эффективность фотохимической активности фотосистемы II ($\Phi\text{PSII}_{\text{light}}$) начала снижаться в начале мацерации, и к моменту выхода протопластов максимальное значение ΦPSII $\Phi\text{PSII}_{\text{max}}$ и, особенно, рабочие значения при

умеренной интенсивности света ($100 \text{ мкмоль} \cdot \text{м}^{-2} \cdot \text{с}^{-1}$), $\text{PSII}_{100\mu\text{E}}$ были значительно ниже, чем в интактных клетках протонемы. Уменьшение $\Phi\text{PSII}_{\text{light}}$ указывает на нарушение оттока электронов из фотосистемы II, что может быть связано с подавлением поглощения продуктов световой фазы (АТФ и НАДФН). В нормальных условиях основным потребителем АТФ и НАДФН в фотосинтетических клетках является система фиксации углекислого газа; таким образом, повреждение ферментов цикла Кальвина может снизить $\text{PSII}_{\text{light}}$ (как PSII_{max} , так и, особенно, PSII_{100}). Этот результат согласуется с наблюдаемой деградацией большой субъединицы RuBisCO. Примечательно, что ΦPSII протонемы также уменьшается при обработке низкими концентрациями драйзелазы (которые не приводят к выделению протопластов, но, тем не менее, разрушают некоторые клеточные белки). Таким образом, снижение ΦPSII не было связано с механическим повреждением клеток и хлоропластов, а указывает на активную деградацию фотосинтетических белков с образованием пула пептидов. Известно, что защитный ответ у растений связан со снижением экспрессии генов, участвующих в фотосинтезе, а также сопровождается замедлением роста растений [290]. Однако физиологическое значение и механизмы этого явления до сих пор мало изучены. Наши данные о фотосинтетической активности клеток мха при выделении протопластов показывают, что снижение фотосинтетической активности протонемы зависит от обработки драйзелазой и коррелирует с увеличением количества пептидов - продуктов деградации хлоропластных белков. Другими словами, наблюдаемое значительное увеличение количества нативных пептидов из белков хлоропластов в протопластах может быть связано с важными физиологическими эффектами. Этот результат указывает на сложную природу стресса, испытываемого клетками во время изоляции протопластов. Используя транскрипционное профилирование (RNA-seq), 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, абиотические стимулы (GO: 0009628), холод (GO: 0009409), температуру (GO: 0009266) и окислительный стресс (GO: 0006979) (Рис. 4.16).

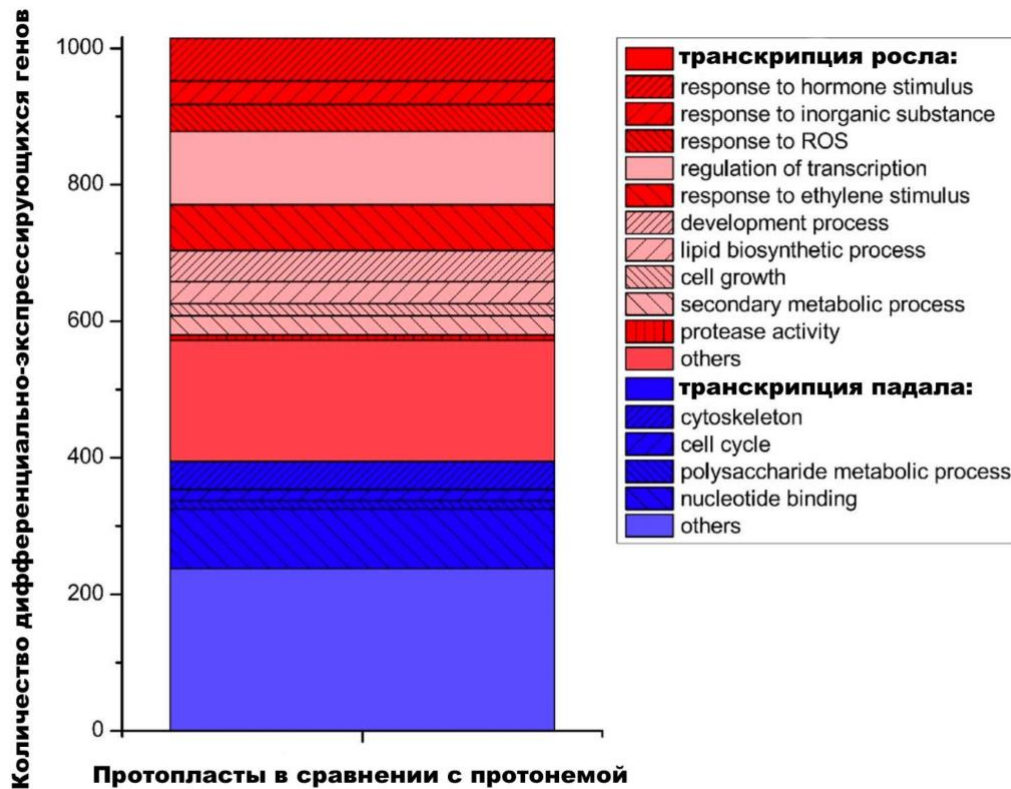


Рисунок 4.16 Распределение дифференциально экспрессируемых (DE) генов по кластерам, полученным с помощью онлайн сервиса DAVID (<https://david.ncifcrf.gov>).

Например, увеличилась экспрессия восьми транскрипционных факторов WRKY, участвующих в регуляции биотического и абиотического стресса, старения и процессов развития [291]. Также 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 [292]. Жасмоновая кислота — это гормон, относящийся к

классу оксипинов, который не был обнаружен у мха *P. patens*. Вместо жасмоновой кислоты для стрессового сигналинга у *P. patens* служит ее предшественник - 12-оксофитодиеновая кислота [293]. Поскольку синтез оксипинов 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 на обработку драйзелазой.

Мы также обнаружили дифференциально экспрессируемые гены, продукты которых могут участвовать в генерации пептидных пулов. Например, наблюдалось значительное увеличение транскрипции Pp1s166_98V6, кодирующего предполагаемый активирующий белок протеасомы 200 (PA200). Кроме того, ряд генов, кодирующих протеазы, например Pp1s78_186V6 (субтилизин-подобная сериновая протеаза 2), Pp1s112_240V6 (субтилизин-подобная сериновая протеаза 3), Pp1s39_149V6 (гомолог митохондриальной протеазы FtsH3), показали более высокий уровень транскрипции в протопластах. Интересно, что мы идентифицировали значительное увеличение уровней транскрипции генов пептидных транспортеров, таких как Pp1s114_7V6 (пептидный транспортер 1) и Pp1s72_96V6 (пептидный транспортер 5). 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 [294]. Этот факт может быть связан с увеличением пула свободных аминокислот и/или олигопептидов, а также с усилением прямого транспорта пептидов из клетки (Рис. 4.17).

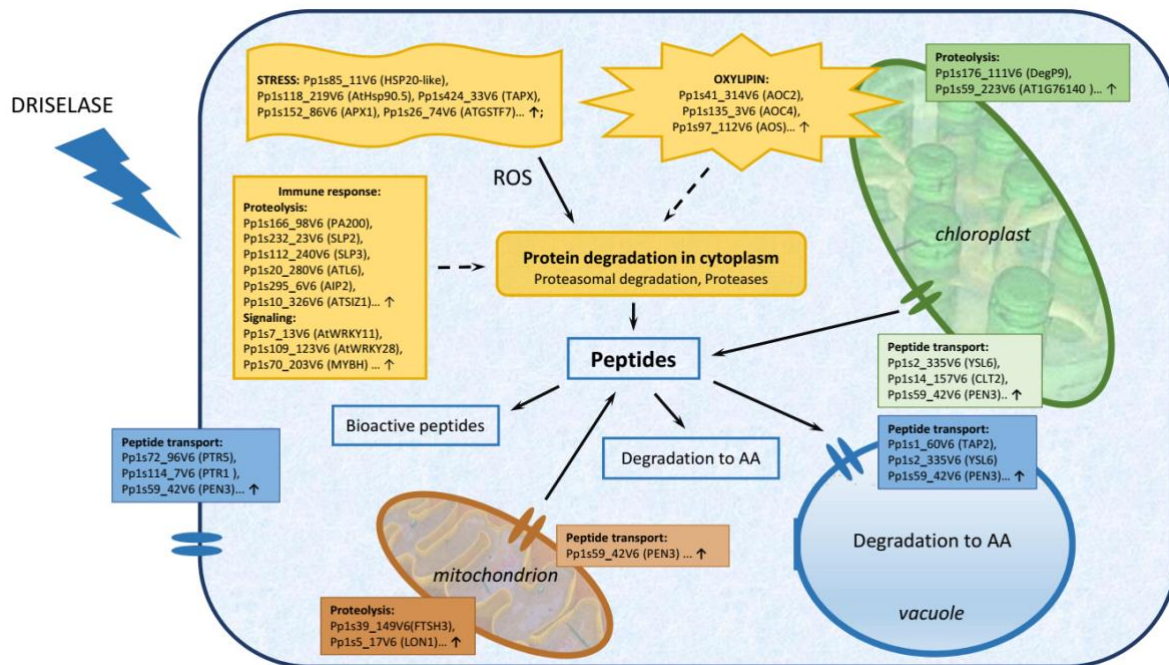


Рисунок 4.17 Модель образования пептидов в протопластах. Мы использовали данные транскрипционного профилирования для построения модели, описывающей пептидогенез в протопластах. Показаны гены, уровень транскрипции которых растет в протопластах, и связанные с факторами, влияющими на деградацию белка, такими как реакция на стрессы, приводящая к увеличению активных форм кислорода, активации иммунитета и синтезу гормонов стресса. Мы сделали вывод о том, что некоторые пептиды расщепляются до аминокислот, а другая фракция может быть биологически активной. Схема также указывает на активированные гены, кодирующие протеазы и переносчики пептидов, которые также могут вносить вклад в пептидом клетки. Стрелками показаны некоторые гены, экспрессия которых увеличилась в 2 раза.

Рост числа пептидов в растениях, подвергшихся биотическому стрессу, также может быть связан с защитной или сигнальной ролью таких пептидов [62,119]. Мы оценили антимикробный потенциал идентифицированных в протопластах пептидов (см. ниже) и выявили 81 эндогенный пептид с высокой потенциальной антимикробной активностью. Индукция образования потенциальных антимикробных пептидов в стрессовых условиях указывает на возможную функцию пептидных пулов, формирующихся в таких условиях у растений. Эта гипотеза согласуется с увеличением уровней транскрипции пептидных

транспортеров, которые могут играть важную роль в двунаправленном транспорте пептидов. Интересно, что ряд переносчиков пептидов, таких как PEN3 (Pp1s59_42V6) и CLT2 (Pp1s14_157V6), могут функционировать в транспортировке таких пептидов и индуцируются при биотическом стрессе.

Мы предложили несколько гипотез для объяснения резкого увеличения количества специфических пептидов, которые встречаются только в протопластном пептидоме: 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.

4.1.3.2 Анализ пептидома клеток, обработанных стрессовыми гормонами

Согласно нашим данным, процесс повреждения клеточных стенок драйзелазой индуцирует иммунный ответ и приводит к значительному изменению пула внутриклеточных пептидов. Кроме того, мы обнаружили индукцию генов, связанных с синтезом стрессового гормона - жасмоновой кислоты. Ранее было показано, что *P. patens* отвечает на обработку стрессовым гормоном метилжасмонатом ингибированием роста и индукцией экспрессии генов, связанных с патогенезом [293]. Чтобы определить, влияет ли этот гормон на процесс образования пептидов, мы использовали масс-спектрометрический анализа для идентификации внутриклеточных и внеклеточных эндогенных пептидов в контрольных и обработанных метилжасмонатом растениях *P. patens*. Для того, чтобы проверить одинакова ли эффективность выделения пептидов разных молекулярных масс при помощи выбранной нами методики, в процессе пробоподготовки (клеточный гомогенат или культуральная среда в случае секрета) мы добавили одиннадцать синтетических пептидов (iRT), имеющих широкий диапазон времени удерживания на хроматографической колонке. Затем синтетические пептиды были идентифицированы как в секрете, так и в образцах клеток с помощью эксперимента по мониторингу множественных реакций (MRM).

Данный эксперимент продемонстрировал, что в процессе пробоподготовки выделяется весь диапазон возможных пептидов (**Рис. 4.18**).

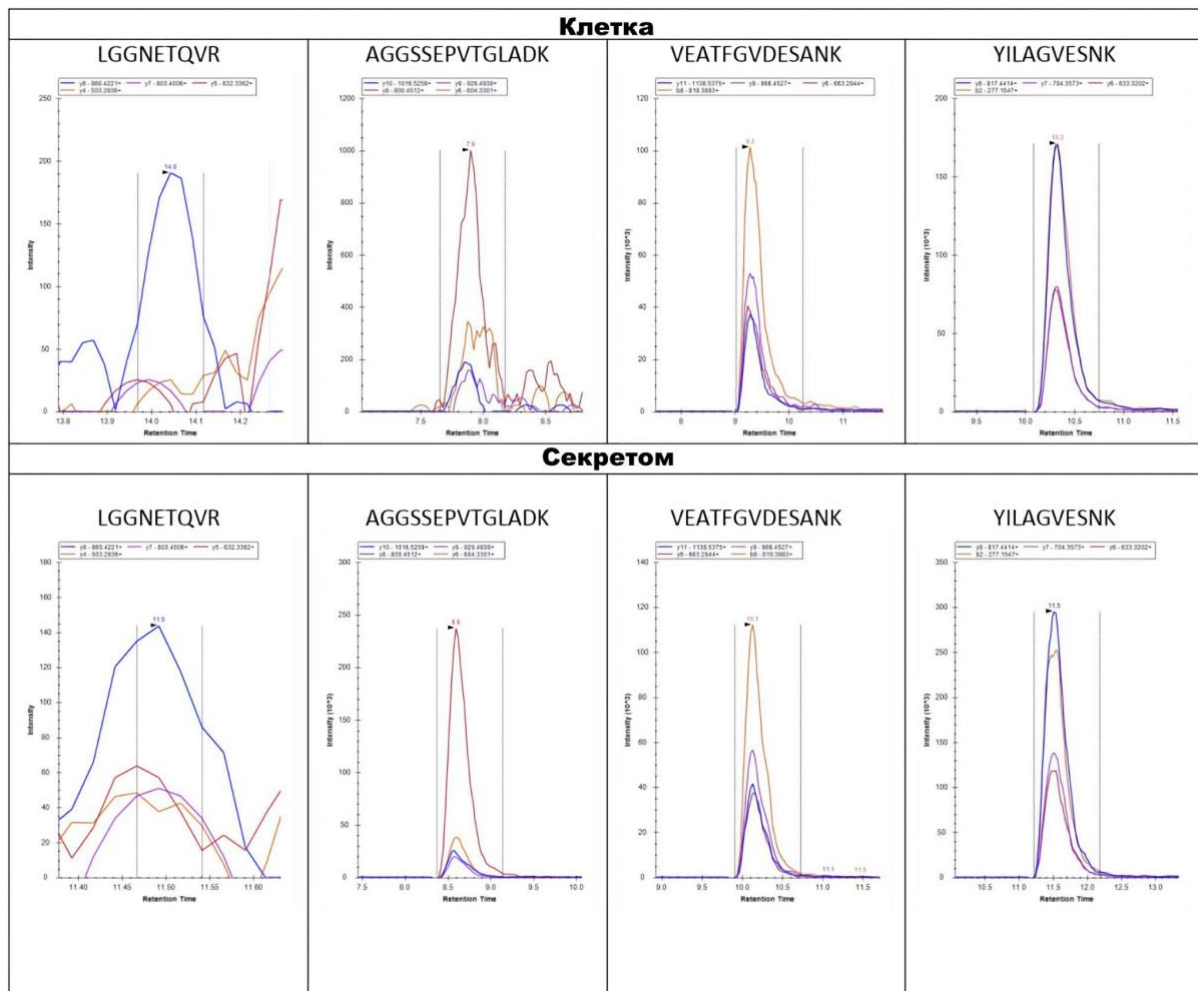


Рисунок 4.18 Результаты анализа с помощью мониторинга множественных реакций (MRM) стандартных пептидов iRT, добавленных при выделении внутриклеточных и секретируемых пептидов.

Далее мы исследовали эндогенные пептиды, экстрагированные из клеток протонемы *P. patens*, которая была обработана в течение часа 0,4 мМ раствором метилжасмоната (МЖ). 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. 4.19**).

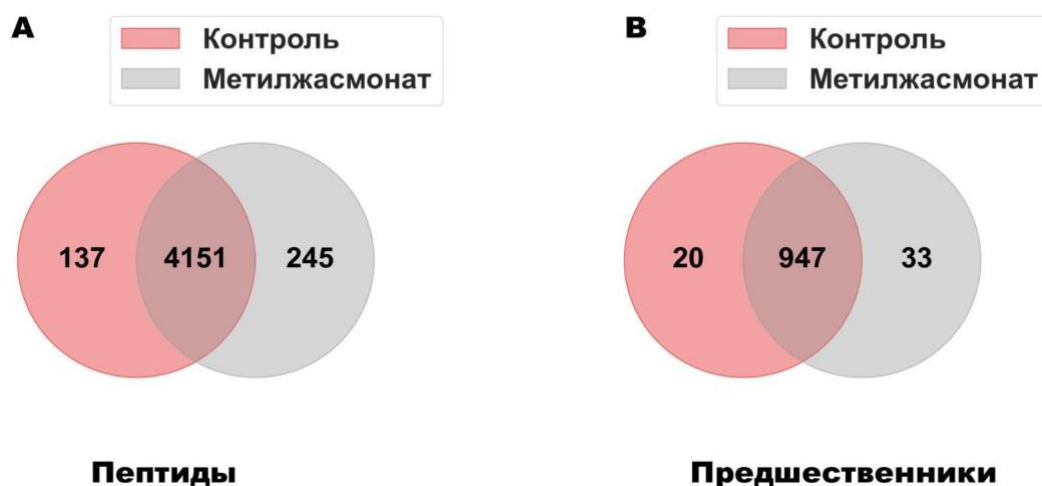


Рисунок 4.19 Сравнение идентифицированных внутриклеточных пептидов и их белков-предшественников между контрольными образцами и образцами, обработанными метилжасмонатом. Диаграмма Венна, показывающая сравнение эндогенных пептидов (А) и их предшественников (В) с контрольными клетками протонемы и клетками протонемы, обработанными метилжасмонатом ($n = 3$ независимых биологических повтора).

Хотя значительная часть внутриклеточных пептидных пулов была идентична с контрольными и обработанными клетками, 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. При этом мы обнаружили высокую степень корреляции между измеренной интенсивности пептидов между обработанными метилжасмонатом и контрольными образцами (коэффициент корреляции Пирсона $r = 0,68$, $P < 10^{-15}$) (**Рис. 4.20**).

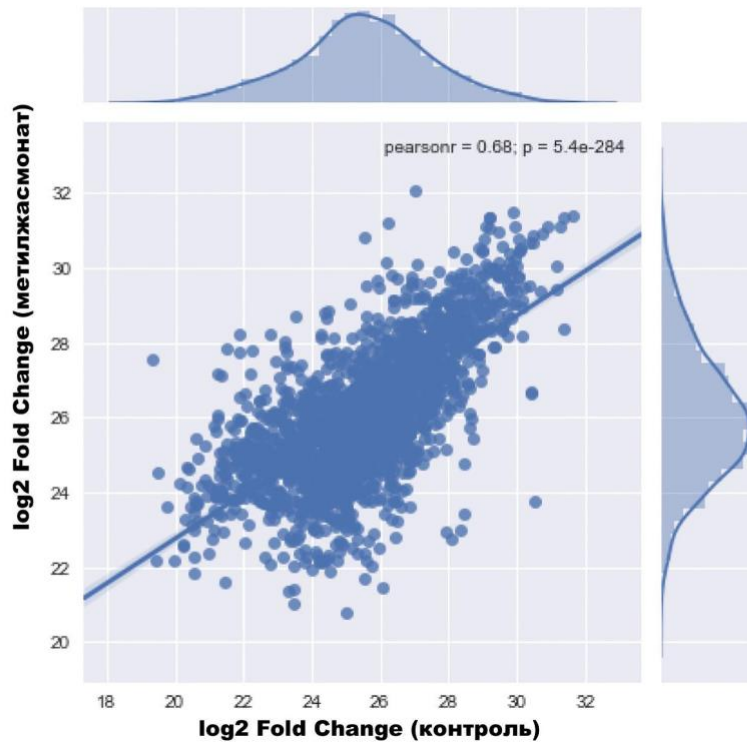


Рисунок 4.20 Диаграмма рассеяния значений интенсивностей пептидов (экстрагированная ионная хроматограмма), которые были выделены из обработанных метилжасмонатом и контрольных образцов.

Большинство пептидов в клеточном пептидоме происходит из белков, участвующих в фотосинтезе, цикле Кальвина, гликолизе и биосинтезе сахарозы. Образцы, обработанные МЖ, имели больше идентифицированных пептидов на белок по сравнению с контролем.

Затем мы сравнили шаблоны деградации белков-предшественников, идентифицированных как в контрольных образцах, так и в образцах, обработанных метилжасмонатом. Принимая во внимание возможность образования "пептидных лестниц", мы использовали 10% "окно" от длины белка-предшественника для оценки изменений в деградации белков. Мы обнаружили, что модели деградации белков были довольно схожи у обработанных метилжасмонатом и контрольными образцами (**Рис. 4.21**).

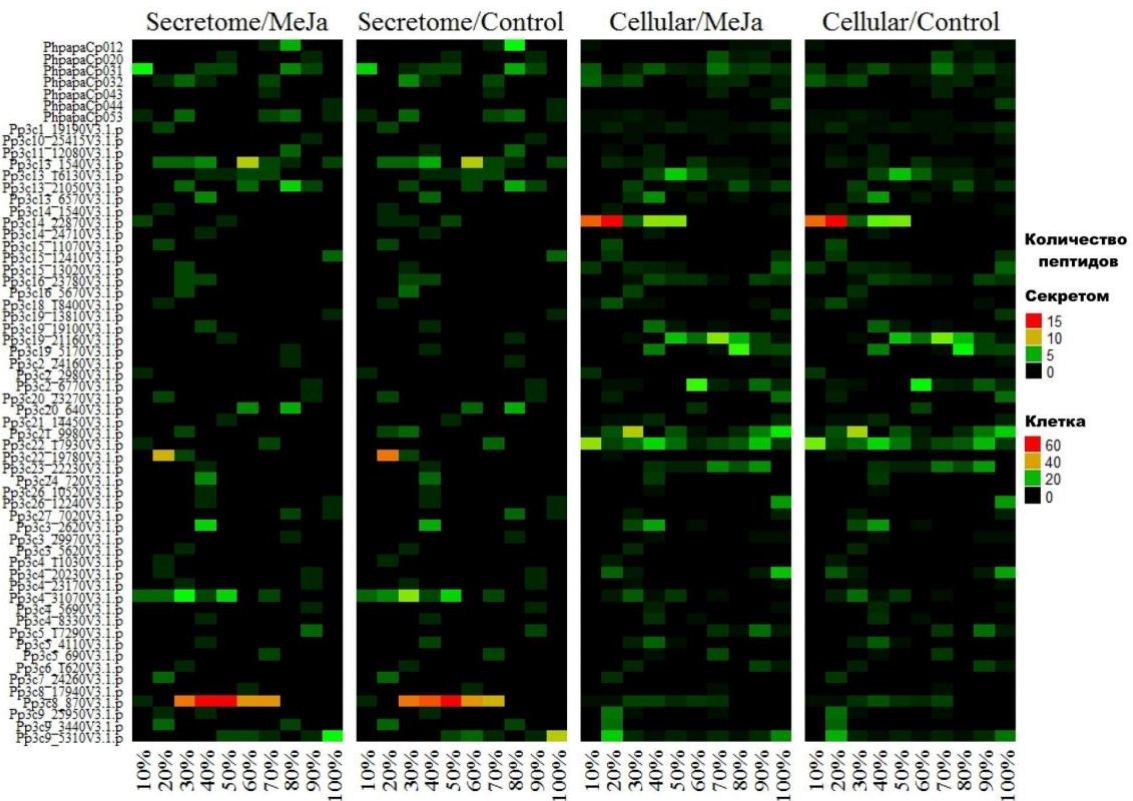


Рисунок 4.21 Тепловая карта, показывающая пример шаблонов деградации белков, обнаруженных в секретоме и клеточном пептидоме. Цвета соответствуют количеству идентифицированных эндогенных пептидов. Ось X показывает процентное соотношение, соответствующее длине белка.

Мы подсчитали, что доля пептидов, происходящих из белков-предшественников, уникальных для образцов, обработанных метилжасмонатом, составляет менее половины всех уникальных пептидов. Другими словами, вызванный обработкой метилжасмонатом стрессовый ответ увеличивал скорость протеолиза белков, а также приводил к деградации новых предшественников. В настоящее время неизвестно, является ли этот процесс результатом изменения активности протеасом/протеаз или пептиды образуются как целевой ответ на стресс.

Таким образом, обработка метилжасмонатом привела к увеличению количества уникальных пептидов в протоме 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 пептидов, схожих между контрольными клетками и клетками, обработанными метилжасмонатом.

Каким образом стрессовые условия могут влиять на деградацию белков и изменения пептидных пулов? Недавние исследования показали, что кислородный стресс может ингибировать активность протеасом, при этом увеличивая активность субъединицы 20S, которая разрушает окисленные белки [295]. Кроме того, было показано, что протеасомальная активность необходима для ответа растений на атаку фитопатогенов [296]. Например, *Pseudomonas syringae* (Pst) секретирует эффекторный белок для подавления активности протеасом и блокирования передачи сигналов салициловой кислоты [296,297]. Помимо протеасомальной активности, стрессовые условия или атака фитопатогенов активируют различные протеазы в апопласте, например, фитаспазу [298,299]. Наши результаты показывают, что обработка метилжасмонатом приводит к увеличению количества уникальных пептидов в протонеме мха. Таким образом, стрессовые фитогормоны не только вызывают деградацию специфических репрессоров факторов транскрипции, но также изменяют пулы пептидов в клетке и секрете. Ответ на биотический стресс, по-видимому, включает в себя индуцированный фитогормонами протеолиз функциональных белков, который может приводить к образованию биоактивных пептидов, например, обладающих антимикробной активностью.

4.1.3.3 Анализ пептидома секретомов, обработанных стрессовыми гормонами растений

Растения выделяют во внеклеточное пространство различные соединения, включая белки и пептиды, которые вместе составляют секретом клетки [300]. В стрессовых условиях в секретом могут выделяться биоактивные пептиды, участвующие в регуляции клеточного ответа на какое-либо воздействие. Например, стрессовый гормон растений метилжасмонат индуцирует высвобождение регуляторного иммунного пептида CAPE1 из белка-предшественника PR1 у пасленовых [119].

Также показано, что салициловая кислота может индуцировать высвобождение иммунных пептидов у растений [114]. Однако в целом, данные о секретируемом пептидоме растений в условиях стрессовых воздействий скудны и требуют дальнейшего изучения. Исходя из этого мы проанализировали пептиды в секрете мха *P. patens* в нормальных условиях и после обработки двумя стрессовыми гормонами - метилжасмонатом и салициловой кислотой.

Для анализа пептидома секрета растений мы провели масс-спектрометрический анализ пептидов, выделенных из жидкой питательной среды, в которой культивировали протонему *P. patens*, обработанную 0,4 мМ метилжасмонатом. Принимая во внимание сложность анализируемых выборок и стохастический характер выбора предшественников для МС анализа, для дальнейшего анализа использовали только пептиды, идентифицированные как минимум в двух биологических повторах. После часовой обработки 0,4 мМ метилжасмонатом было идентифицировано только 11 новых пептидов в секрете по сравнению с контрольными образцами. Таким образом, обработка метилжасмонатом не вызывала серьезных изменений в секретируемом пептидоме (**Рис. 4.22А**).

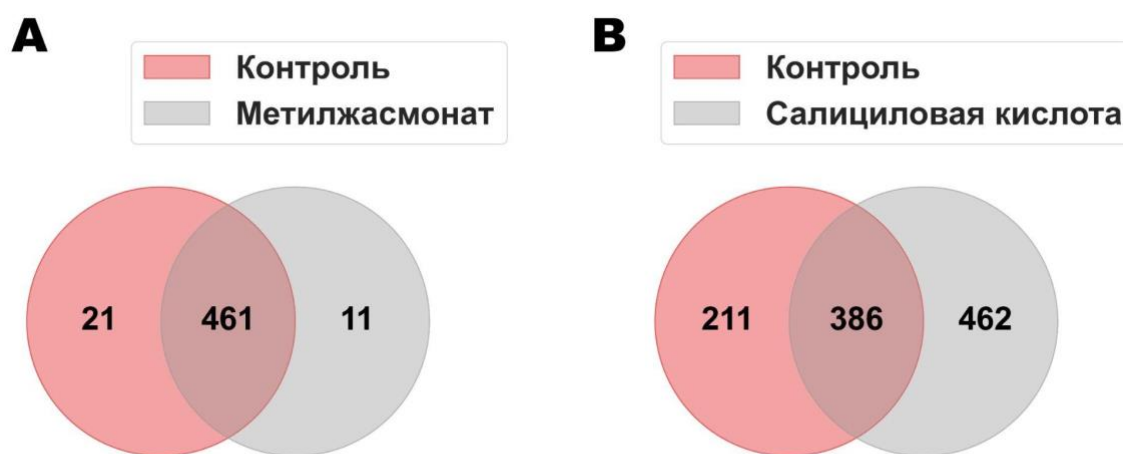


Рисунок 4.22 Диаграмма Венна, показывающая сравнение количества эндогенных пептидов в секрете протонемы контрольных и обработанных метилжасмонатом (**А**) и салициловой кислотой (**В**) образцов.

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 секрете 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 [21].

Протеомный анализ секрета *P. patens* показал, что обработка индуктором иммунного ответа - хитозаном - индуцировала секрецию таких белков, как α -экспансин, пектинметилестераза, фотосистема II, липоксигеназа и белки LEA [273]. Мы обнаружили пептиды от этих белков как в контрольных, так и в обработанных метилжасмонатом клетках. Также протеомный анализ секрета мха показал наличие сложной внеклеточной протеолитической сети, включающей различные типы протеаз и их ингибиторы [274]. Однако вклад этих протеаз в процесс образования внеклеточного пептидома пока неясен.

Известно, что предшественники биоактивных пептидов и гормонов растений часто представляют собой небольшие белки без известных доменов [2]. В секрете протонемы мха, обработанной метилжасмонатом, мы идентифицировали пептиды, происходящие из тридцати девяти белков размером менее 200 аминокислотных остатков. Три из них были предсказанными белками, которые являлись предшественниками пептидов, представленность которых увеличивалась после обработки метилжасмонатом. Возможные функции этих пептидов еще предстоит выяснить.

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 (**Рис. 4.22В**). Представленности пептидов (согласно значениям экстрагированных ионных хроматограмм, ХИС), идентифицированных в контроле и при обработке

салициловой кислотой, хорошо коррелировали друг с другом (коэффициент корреляции Пирсона = 0,7; $P < 10^{-15}$). Это указывает на то, что обработка салициловой кислотой, скорее, привела к протеолизу новых предшественников, чем к изменению представленности компонентов стабильного пептидного пула. Например, число пептидов, которые образовывались из наиболее пептидогенных предшественников в контрольных образцах, практически не менялось после обработки салициловой кислотой.

Ранее было показано, что биоактивные пептиды растений чаще образуется при протеолизе С-концевого региона белка-предшественника [301]. Согласно нашей оценке, приблизительно 23% и 26% всех идентифицированных пептидов образовались из 50 аминокислот на С-конце белков-предшественников в контрольном и обработанном СА секретах, соответственно. Чтобы понять, существуют ли какие-то отличия этих пептидов от остального пептидного пула, мы сравнили их физико-химические свойства в обработанных СА и контрольных образцах. Этот анализ показал значительные различия (ранговый критерий Вилкоксона, $P < 0.01$) только по одному индексу: WEBA780101 (“Значение времени удержания при хроматографии с высоким содержанием соли”). Значение этого индекса рассматривается как один из предикторов антимикробных пептидов [302]. Затем мы проанализировали, какой тип протеаз может быть ответственен за образование пептидов в контроле и образцах, обработанных СА. Используя подход, который мы ранее применили для анализа секретлируемого пептидома (см. выше), мы показали, что в контрольных и обработанных СА образцах десятки уникальных пептидов содержали сайты расщепления $\beta 1c$ субъединицей протеасомы (треониновой протеазой). 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 (**Рис. 4.23**).

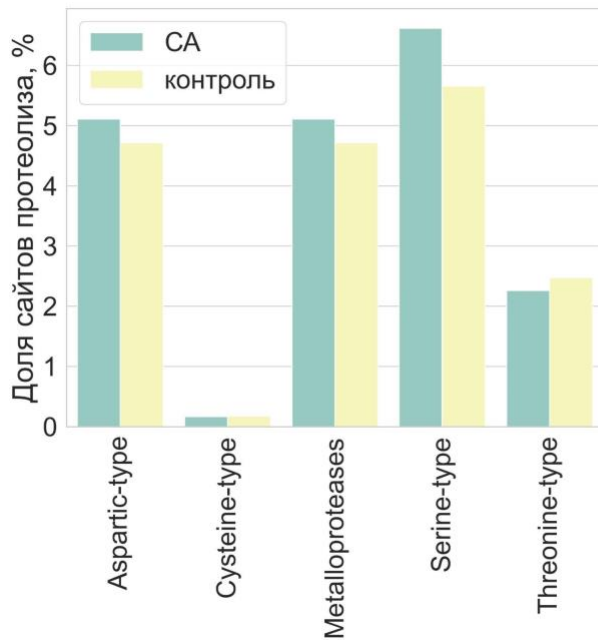


Рисунок 4.23 Гистограмма, показывающая доли предсказанных сайтов протеолиза (% от общего количества проанализированных N- и С-концов пептидов) для каждого типа протеаз в идентифицированных пептидных пулах

В пептидоме растений, обработанных СА, мы наблюдали деградацию 134 новых белков-предшественников. Согласно анализу терминов GO, эта группа предшественников была обогащена белками, связывающими кальций ($P = 1,32e-10$) и относящимися к структурным компонентам рибосомы ($P = 0,00004$).

Так же, как и в эксперименте с обработкой метилжасмонатом, мы проанализировали количество белков-предшественников пептидов длиной менее 200 а.о. Суммарно в контрольных образцах мы идентифицировали пятьдесят четыре эндогенных пептида (т.е. 8% от общего пептидома) от таких предшественников. Обработка СА индуцировала протеолиз небольших белков, в том числе, с неизвестными функциями. Такие белки составляли около 30% от всех идентифицированных предшественников. Мы идентифицировали 135 эндогенных пептидов из небольших предшественников в секретоме, обработанном СА. Среди них были обнаружены пептиды из небольших белков, связанных со стрессовым ответом, таких как реморин, белок теплового шока HSP20 и другие. Анализ обогащения терминов GO показал, что наиболее значительно обогащенной

молекулярной функцией небольших предшественников была структурная составляющая рибосомы ($P = 0,000196$).

4.1.3.4 Анализ посттрансляционных модификаций в пептидном пуле секрета протонемы, обработанной салициловой кислотой

Для того, чтобы оценить влияние обработки СА на модификации пептидов в секрете, а также идентифицировать известные пептидные гормоны растений, мы проанализировали данные масс-спектрометрии с учетом следующих модификаций: гидроксирование пролина, сульфатирование тирозина, O-гликозильрование (арабинозильрование) и N-концевое ацетильрование. N-концевое ацетильрование является наиболее распространенной и стабильной модификацией белков у растений, в то время как гидроксирование пролина, сульфатирование тирозина и арабинозильрование являются типичными модификациями растительных пептидных гормонов. Например, эти модификации были обнаружены в пептидах CLE, CEP и RALF, которые секретируются корнями сосудистых растений [303]. Мы обнаружили, что доля ацетильрованных по N-концу пептидов оставалась постоянной — 7,6% и 7,9% пептидов в контрольных и в обработанных СА образцах соответственно. При этом приблизительно 17% (108 пептидов) эндогенных пептидов в контроле содержали гидроксированный пролин (**Рис. 4.24**). Обработка СА привела к значительному снижению общего количества гидроксированных пептидов (8% в СА; тест Фишера, $P=7,29 \times 10^{-8}$). Однако мы идентифицировали 30 уникальных для обработки СА пептидов с модифицированным пролином. При этом мы не обнаружили арабинозильрованных пептидов, а идентификация пептидов с сульфатированием тирозина была недостаточно надежной. Следовательно, количество пептидов с сульфатированием тирозина и/или арабинозильрованных пептидов во внеклеточном пространстве растений подлежит дальнейшему изучению.

были идентифицированы четыре пептида от таких предшественников с модификациями пролина: FLLNELALP(hy)AG, HSSDP(hy)NFP(hy)FSIK, QP(hy)GSSACVTP(hy)LP(hy)QLS и VTENLEVMP(hy)STTIR. Два предшественника - это белки с неизвестными функциями, тогда как остальные содержали домены DUF674 и NIKESHI. Интересно, что размеры этих пептидов и количество модификаций были аналогичны известным пептидным гормонам, таким как CEP и CLE. Гены, кодирующие предшественники CLE и RALF, были предсказаны в геноме *P. patens*, однако пептиды из данных семейств не были обнаружены в нашем анализе. Это может говорить о том, что используемая нами методика анализа пептидома не является достаточно чувствительной для обнаружения пептидных гормонов растений, которые предположительно присутствуют в низких концентрациях. Например, идентификация пептидных гормонов CEP у люцерны стала возможной благодаря использованию мутантных линий, сверхэкспрессирующих MtCEP1 [304]. Еще одна возможная причина заключается в том, что мы анализировали секретом протонемы, тогда как в предыдущих исследованиях использовались интактные корни, культура бородатых корней ("hairy roots"), сок ксилемы или культура растительной ткани. Это согласуется с недавним исследованием Whitewoods с соавт., в котором экспрессия генов, кодирующих пептиды CLE у мха, была обнаружена в гаметофорах, но не в нитях протонемы [133].

Среди модифицированных пептидов, идентифицированных после обработки СА, мы обнаружили два пептида с гидроксированными пролинами — GP(hy)SLIFP(hy)ARSPV и VAVKYYSDDWP(hy), причём один из них выщеплялся из нефункционального предшественника, а другой из белка ZETA-COAT PROTEIN//SUBFAMILY NOT NAMED. Для проверки биологической активности данных пептидов необходим дополнительный анализ.

Таким образом, согласно нашим данным, гидроксирование пролина было наиболее распространенной из проверенных нами модификаций пептидов. Гидроксирование пролина - распространенная модификация растительных пептидных гормонов, которая связана с активностью ферментов пролил-4-

гидроксилаз. Недавно у *P. patens* были идентифицированы шесть гомологичных генов пролил-4-гидроксилаз (P4H), в том числе, P4H1, который необходим для гидроксилирования пролина у белков [305]. Интересно, что большинство хорошо изученных гидроксилированных биоактивных пептидов участвуют в регуляции процессов роста и развития, тогда как пептиды, связанные со стрессовым ответом, такие как AtPEPs, ZmPEPs, Zip1 и CAPE1, не подвергаются пост-трансляционным модификациям [120,306]. Можно предположить, что уменьшение количества гидроксилированных пептидов после обработки СА связано с образованием пептидов, которые могут играть важную роль в реакциях на стрессовые воздействия. Например, в отличие от гидроксилированных пептидов, количество ацетилированных пептидов не изменялось при обработке СА. Этот факт может указывать на важность гидроксилирования пролина в реакциях растений на стрессовые воздействия. Полученные нами данные могут также пролить свет на механизмы возникновения пептидных гормонов у растений.

4.1.4 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. Можно предположить, что структура протеома разных организмов, методы выделения пептидов из разных тканей влияют на идентифицируемый нами с помощью масс-спектрометрии пептидом. Эти различия могут затруднить сравнительный анализ наборов пептидных данных, полученных от разных организмов.

Мы 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 модельного организма *P. patens* с наборами эндогенных пептидов, идентифицированных при анализе бактерий [307]; дрожжей [45]; хлопка [308]; плазмы крови человека [40] и пептидов, связанных с протеасомой у человека [39]. Поскольку на сегодняшний

день наши представления о пептидомах тесно связаны с масс-спектрометрическим анализом выделенных пептидов из тканей и биологических жидкостей, это должно накладывать определенные ограничения на длину и аминокислотный состав детектируемых молекул. В отобранных нами для анализа наборах данных средняя длина идентифицированных пептидов колеблется от 11 до 18 аминокислотных остатков и остаётся практически одинаковой для внутри- и внеклеточных пептидомов разных организмов.

Для того, чтобы оценить схожесть механизмов генерации пептидных пулов у разных организмов, мы сравнили С-концевые аминокислоты (P1 позиция сайта протеолиза) в отобранных наборах данных. Наш анализ показал, что лизин (K) и аргинин (R) были одними из наиболее представленных аминокислот в разных наборах данных (**Рис. 4.25**). Это может указывать на то, что протеазы с трипсиноподобной активностью вносят значительный вклад в формирование нативных пептидомов у разных организмов. Интересно, что хотя протеасомальная деградация белков предполагается основным механизмом формирования внутриклеточных пептидных пулов, проанализированные нами внутриклеточные пептидомы отличались от специфического набора пептидов, связанных с протеасомой, которые были изучены у человека [39]. Известно, что протеолитическая активность протеасом включает каспазоподобную - мотивы, обогащенные отрицательно заряженными аминокислотами (D, E); химотрипсин-подобную - обогащение гидрофобными (W, F, M, Y) аминокислотами или трипсин-подобную - гидролиз после положительно заряженных (R и K) аминокислот [39]. В своем исследовании Wolf-Levy и др. предполагают, что низкая представленность положительно заряженных аминокислот на С-концах проанализированных пептидов обусловлена либо с биологическими, либо с техническими причинами. В целом, вариабельность С-концевых аминокислотах в разных наборах пептидомных данных может быть вызвана различными причинами, такими как неспецифический протеолиз, действием экзопептидаз, технические особенности метода выделения или особенностями масс-спектрометрического анализа.



Рисунок 4.25 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.

Согласно целому ряду исследований, большинство известных растительных пептидных гормонов закодировано в С-концевой части белка-предшественника [21,27]. При этом предшественники пептидных гормонов у животных зачастую имеют многодоменную структуру, позволяющую генерировать множество идентичных или совершенно различных функциональных пептидов, распределенных по длине одного препротейна [16,309].

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 [36]. 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. 4.26). Действительно, сравнение различных наборов данных показало, что пептиды, высвобождаемые с С- или N-конца предшественника, как правило, более представлены во внутриклеточных или внеклеточных пептидомах, чем мы могли бы ожидать при случайном расщеплении 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. Уникальные шаблоны распределения пептидов из плазмы крови человека и корней хлопка могут отражать техническую вариабельность или специфичность соответствующих пептидомов.

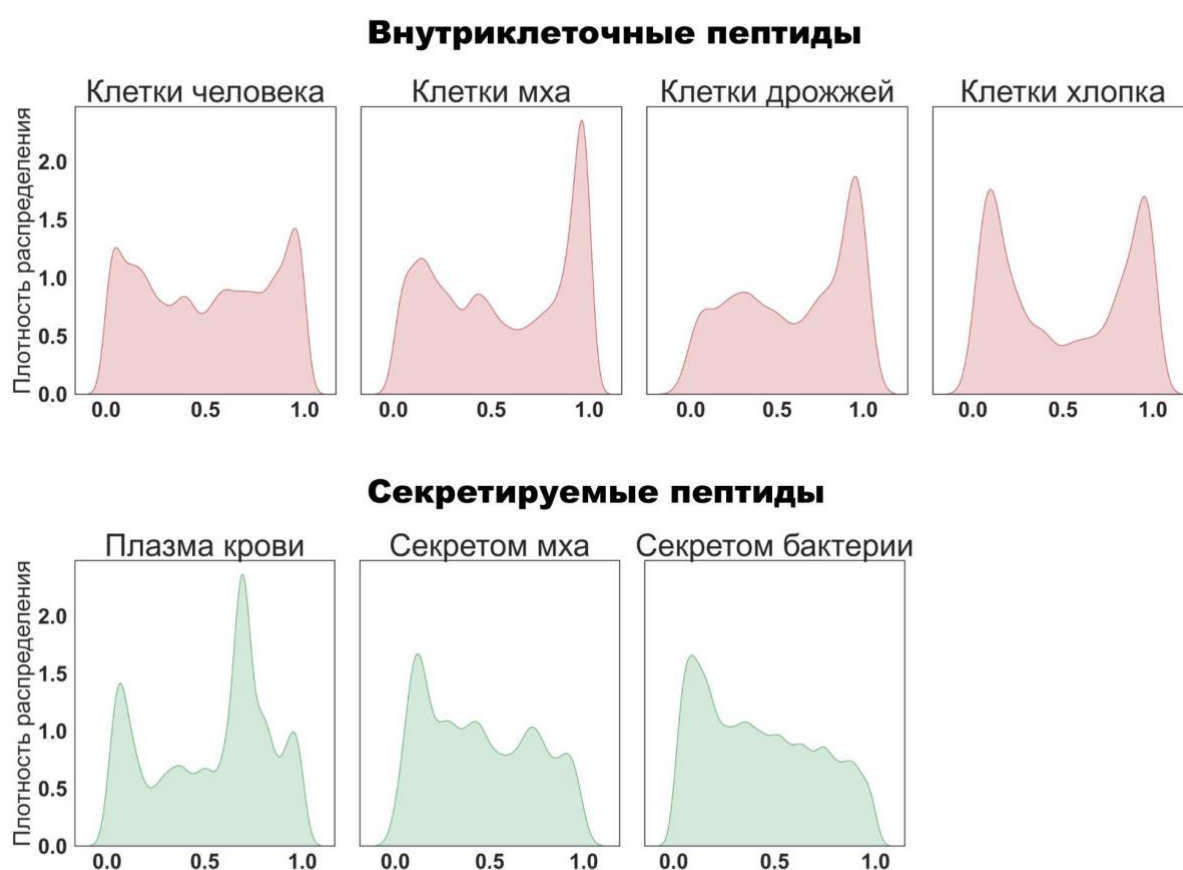


Рисунок 4.26 Распределение (density plot) идентифицированных с помощью масс-спектрометрии пептидов по длине предшественников в наборах пептидомных данных мха, человека, дрожжей, хлопка и бактерий. Положения каждого идентифицированного пептида были нормализованы к длине белка.

В недавнем исследовании пептидома клеток человека НК-2, обработанных TGF- β 1, было показано, что индексы "GRAVY", указывающие на гидрофобность пептидной последовательности эндогенных пептидов были, в основном, ниже нуля, и большинство из них, предположительно, были гидрофильными [203]. Далее мы сравнили индексы "GRAVY" пептидов из различных организмов, чтобы оценить степень их гидрофобности (Рис. 4.27А). Индексы "GRAVY" и доля ароматических аминокислот (Рис. 4.27В) были значительно ниже почти во всех проанализированных пептидомных наборах данных по сравнению с пептидами, случайно сгенерированными из тех же белков-предшественников. На основе этого анализа можно предположить, что идентификация в масс-спектрометрических экспериментах менее гидрофобных пептидов, чем ожидалось бы при случайном протеолизе белков, может быть общей тенденцией пептидомных исследований. Это может отражать как биологические тенденции в составе клеточных и секретируемых пептидомов, так и особенности подготовки образцов и анализа LC-MS/MS. Например, очень гидрофильные короткие пептиды могут быть потеряны при разделении на колонках C18 [201].

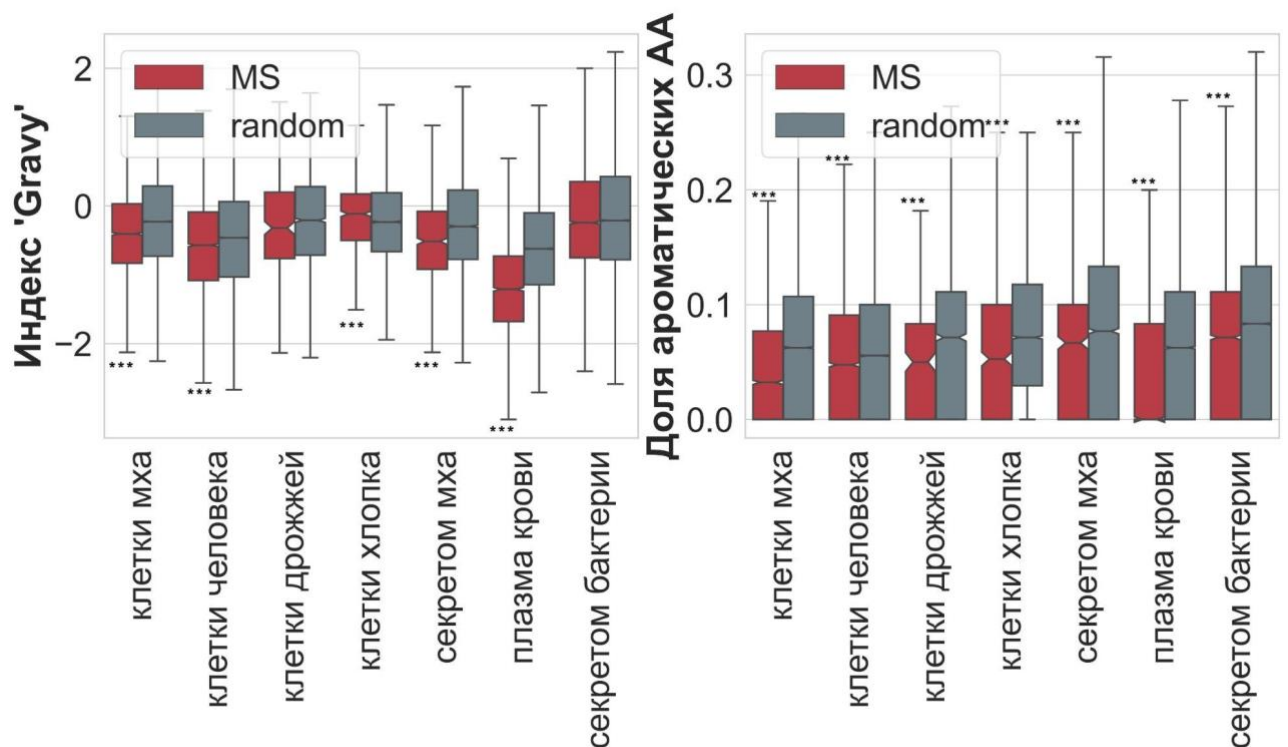


Рисунок 4.27 Сравнение рассчитанных индексов "Gravy" и доли ароматических кислот между

пептидами, идентифицированными с помощью масс-спектрометрии и наборами случайных пептидов, которые были сгенерированы из соответствующих белков-предшественников для каждого набора данных. Все расчеты были выполнены с помощью Biopython [310]. *** - $P < 10^{-15}$ критерий Манна-Уитни.

Таким образом, недавний прогресс в масс-спектрометрическом анализе существенно расширил наши знания/представления о составе внутри- и внеклеточных пептидных пулов. Тем не менее, современные пептидомные подходы на основе масс-спектрометрического анализа имеют некоторые недостатки, такие как (1) проблемы с обнаружением низкопредставленных пептидов; (2) масс-спектрометрическая идентификация пептидов с определенными физико-химическими свойствами; (3) неполные аннотации протеома для некоторых организмов, которые требуют дальнейших улучшений; и (4) трудности в корректной идентификации модифицированных нативных пептидов. Поэтому необходим дальнейший прогресс в области пептидомного анализа, связанный с улучшением детекции пептидов и уменьшением количества артефактов во время пробоподготовки. Еще более важным аспектом является разработка подходов к идентификации и функциональному анализу ранее неаннотированных компонентов клеточных и секретируемых пептидомов.

4.2 Идентификация продуктов трансляции коротких открытых рамок считывания в пептидоме *P. patens*

Помимо специфического и неспецифического протеолиза белков-предшественников, еще одним малоизученным источником нативных пептидов могут являться 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 [69]. Однако данный источник пептидов ранее системно не анализировали у растений, и вклад кОРС в формирование пептидома растений неизвестен. Исходя из этого, мы провели

всесторонний анализ кОПС и продуктов их трансляции у растений, используя *P. patens* как модель, обладающую целым рядом преимуществ для такого рода исследований.

Поскольку масс-спектрометрический анализ является единственным прямым методом детекции белков или пептидов в протеоме клетки, мы выбрали данный подход для идентификации транслирующихся кОПС в клетках растений. Следует отметить, что анализ данных масс-спектрометрического анализа основан на сопоставлении *in silico* полученных на приборе масс-спектров с теоретическими спектрами пептидов из базы белковых последовательностей данного организма [234]. Таким образом, как и в случае протеомных исследований, идентификация продуктов трансляции кОПС зависит от полноты используемой базы последовательностей. Наиболее подходящая для поиска база должна содержать пептиды, кодируемые кОПС, которые действительно могут присутствовать в пептидоме организма с небольшим количеством нерелевантных последовательностей. Очевидно, что транслирующиеся кОПС отсутствуют в широко используемых базах данных, таких как Ensembl (<https://www.ensembl.org>), RefSeq (<https://www.ncbi.nlm.nih.gov/refseq>) и UniProtKB (<https://www.uniprot.org>). Поэтому для поиска пептидов или микробелков, кодируемых кОПС в пептидоме, необходимо создавать специальные базы последовательностей, содержащие как можно большее количество потенциально транслирующихся кОПС.

4.2.1 Поиск коротких открытых рамок считывания (кОПС) с высоким кодирующим потенциалом в геноме *P. patens*

Существует несколько стратегий аннотации кодирующих кОПС в геноме исследуемого организма и создания совмещенной (белки+кОПС) 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 [311].

Такие совмещенные базы данных содержат большое количество в действительности нетранслирующихся последовательностей и, следовательно, настолько большие, что это приводит к большому количеству ложных идентификаций и экспоненциальному увеличению времени поиска [234]. Также несуществующие пептидные последовательности, присутствующие в таких базах данных, делают ненадежной оценку достоверности совпадений спектров пептидов, что еще больше усугубляет трудности с обнаружением низкопредставленных пептидов.

По этим причинам для поиска потенциально транслирующихся коротких рамок считывания в геноме *P. patens* мы выбрали подход с предсказанием всех кОПС 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 [238], который ищет в геноме безинтронные кОПС, начинающиеся со стартового кодона AUG и длиной от 30 до 300 нуклеотидов. Данный алгоритм основан на поиске различий в составе к-меров между кодирующими и некодирующими последовательностями и последующим вычислением кодирующего потенциала кОПС и классификацией с использованием теоремы Байеса [238]. Таким образом, предсказание кодирующих кОПС зависит от состава используемых для обучения модели кодирующих (CDSs) и некодирующих (NCDSs) последовательностей. Еще одним настраиваемым параметром для идентификации кодирующих кОПС являются априорные вероятности $P(\text{CDS})$ и $P(\text{NCDSs})$. По умолчанию значения для $P(\text{CDS})$ и $P(\text{NCDSs})$ установлены равными 0,5 и 0,5 соответственно. Однако этот параметр также можно настраивать.

На первом этапе для подбора оптимальных параметров предсказания кодирующих кОПС в геноме *P. patens* мы использовали несколько вариантов сочетаний кодирующих и некодирующих последовательностей: в качестве некодирующих - интроны, а также различные вариации кодирующих - белок-кодирующие последовательности (CDS), экзоны, нетранслирующиеся последовательности мРНК (UTRs). Для ускорения процесса кодирующие кОПС на этом этапе были

предсказаны только на одной цепи, и различные комбинации параметров были проанализированы. Например, существенная разница в предсказанных кОРС наблюдалась при использовании для обучения модели в качестве кодирующих последовательностей нетранслируемых регионов (UTR) и экзонов аннотированных генов из Phytozome V12 (**Рис. 4.28**).

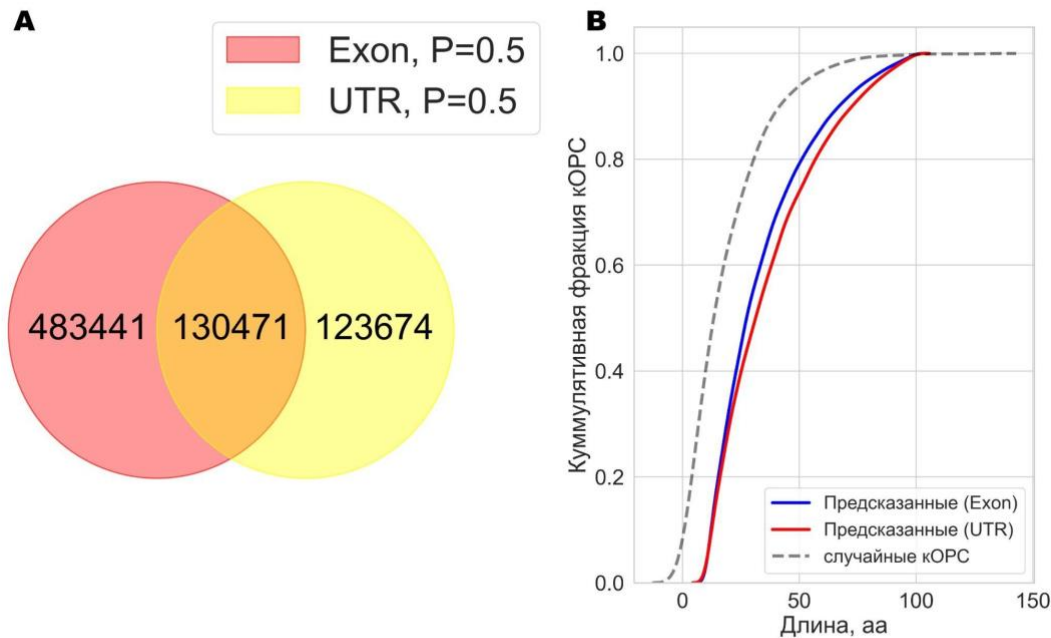


Рисунок 4.28 А - диаграмма Венна, показывающая сравнение количества предсказанных кодирующих кОРС в геноме *P. patens* с помощью sORF finder с разными параметрами: “Exon, P=0.5” - кодирующие последовательности - экзоны, не кодирующие - интроны, P(CDS)=0.5; “UTR, P=0.5” - кодирующие последовательности - нетранслирующие регионы мРНК, не кодирующие - интроны, P(CDS)=0.5. **Б** - График иллюстрирует кумулятивное распределение длин предсказанных кОРС с разными параметрами кодирующих кОРС; к категории "случайные кОРС" относятся короткие рамки, которые могут случайно образовываться в геноме *P. patens* (GC=45.9%, медиана равна 13 кодонам).

Проанализировав разные наборы кодирующих кОРС, предсказанных с разными параметрами, мы обнаружили, что наиболее оптимальным вариантом для обучения модели является использование белок-кодирующих последовательностей и интронов из аннотации Phytozome V12 для *P. patens*. Кроме того, мы сравнили

предсказание кодирующих кОПС с разными значениями априорных вероятностей (Рис. 4.29).

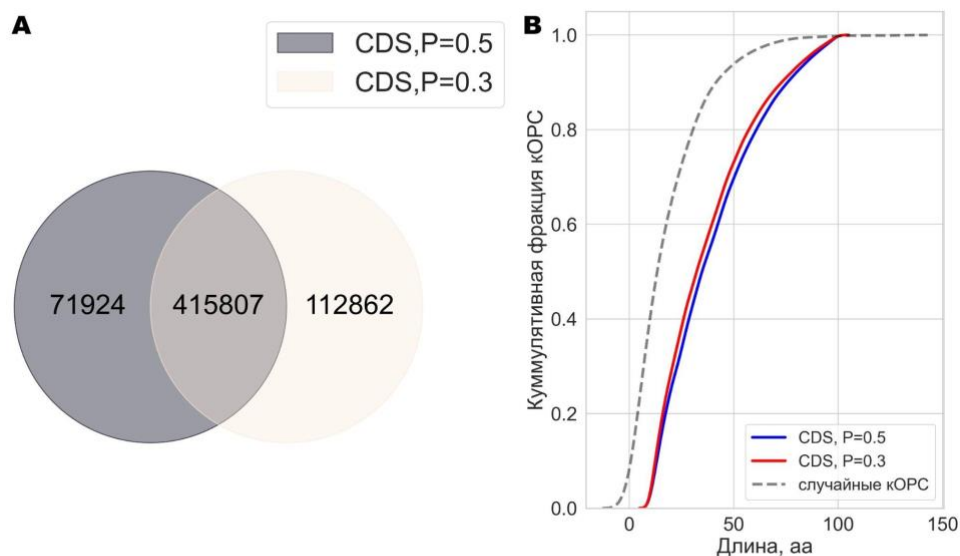


Рисунок 4.29 А - диаграмма Венна, показывающая сравнение количества предсказанных на обеих цепях кодирующих кОПС в геноме *P. patens* с помощью программы sORF finder с разными параметрами: “CDS, P=0.5”, белок-кодирующие последовательности, некодирующие - интроны, P(CDS)=0.5; “CDS, P=0.3”, белок-кодирующие последовательности, некодирующие - интроны, P(CDS)=0.3. **Б** - График иллюстрирует куммулятивное распределение длин предсказанных кОПС с разными параметрами кодирующих кОПС; к категории "случайные кОПС" относятся короткие рамки, которые могут случайно образовываться в геноме *P. patens* (GC=45.9%, медиана равна 13 кодонам).

Таким образом, для дальнейшего анализа были использованы кОПС, кодирующий потенциал которых был предсказан с помощью следующих параметров: кодирующие последовательности - белок-кодирующие, не пересекающиеся с интронами, некодирующие - интроны из аннотации Phytozome V12, P(CDS)=0.3. Этот подход привел к идентификации 638,439 кОПС 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*. Длина кОПС с высоким кодирующим потенциалом значительно отличалась от тех,

которые мы могли бы ожидать при случайной генерации кОРС в геноме *P. patens* (критерий «Манна-Уитни», $P < 10^{-15}$).

4.2.2 Классификация кОРС с высоким кодирующим потенциалом

В результате предсказания потенциальных пептид-кодирующих коротких рамок считывания с помощью программы sORF finder в геноме *P. patens* были идентифицированы несколько сотен тысяч кОРС. Далее мы провели фильтрацию идентифицированных кОРС, используя следующие параметры: 1) является частью аннотированных белков; 2) кОРС не находится на транскрипте; 3) встречается в геноме несколько раз. В результате мы отобрали 70,095 уникальных кОРС, расположенных на транскриптах различного типа, включая мРНК и длнкРНК из двух баз данных - SANTATAdb [235] и GreeNC [236]. Таким образом, мы получили набор кОРС длиной от 30 до 300 нуклеотидов, которые были расположены на 33,981 транскрипте (22 969 генов).

Мы классифицировали кОРС на основе их положения на транскриптах: 63,109 “генных кОРС” (расположены на транскриптах мРНК); 1241 “межгенных” кОРС (расположены на транскриптах из нашего набора данных [312] и не аннотированных в текущей версии Phytozome V12); и 5745 “длнкРНК-кОРС” (расположены на длнкРНК из баз SANTATAdb [235], GreeNC [236], или предсказанных в нашем наборе транскриптов [240]). “Генные кОРС” были классифицированы как 11,998 “upstream” кОРС, которые расположены в 5'-нетранслируемых регионах мРНК; 9443 “downstream” кОРС, которые расположены в 3'-нетранслируемых регионах мРНК; 36,732 “CDS”-кОРС, которые перекрываются с основной белок-кодирующей последовательностью мРНК, но со сдвигом рамки считывания +1 или +2; 3485 “interlaced” кОРС, которые перекрываются как с белок-кодирующей последовательностью, так и с нетранслируемыми регионами мРНК (Рис. 4.30).

Исходя из используемых параметров поиска программы sORF finder [238], ожидаемо, что наш набор кОРС с высоким кодирующим потенциалом был обогащен “CDS”-кОРС (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 использовали в качестве отрицательного контроля.

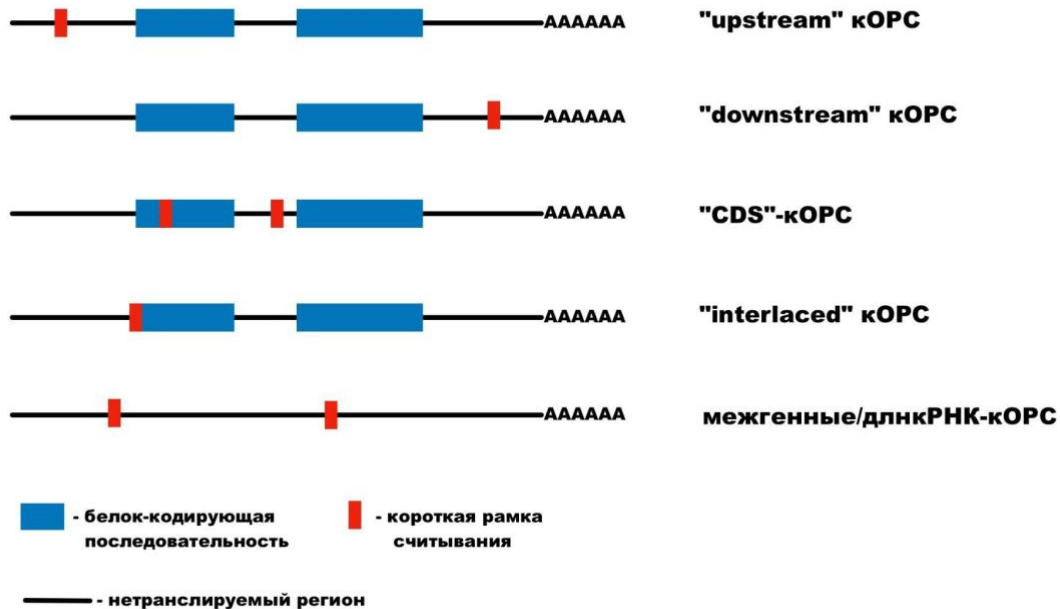


Рисунок 4.30 Предлагаемая классификация кОПС в соответствии с типом транскрипта и положением короткой рамки считывания: “upstream” кОПС (uORFs) и “downstream” кОПС (dORFs) из нетранслируемых областей (UTR) мРНК; “CDS”-кОПС, которые перекрываются с белок-кодирующими последовательностями, но расположены в альтернативных рамках считывания или являются короткими изоформами, образующимися в результате альтернативного сплайсинга; “interlaced” кОПС, которые перекрывают как кодирующую последовательность, так и UTR на одном и том же транскрипте; длинкРНК-кОПС и межгенные кОПС, которые расположены на транскриптах аннотированных как длинные некодирующие РНК или не содержащие белок-кодирующих последовательностей.

Длина кОПС в нашем наборе данных варьировала от 30 до 300 нуклеотидов (**Рис. 4.31**). 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 $P = 0,0021$).

Далее мы проанализировали консенсусные последовательности Козака [313], окружающие стартовые кодоны транскрибирующихся кОПС. 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) [313]. Согласно этой оценке, 41,816 ($\approx 60\%$) предсказанных нами кОПС 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 $P < 2,2 \times 10^{-16}$).

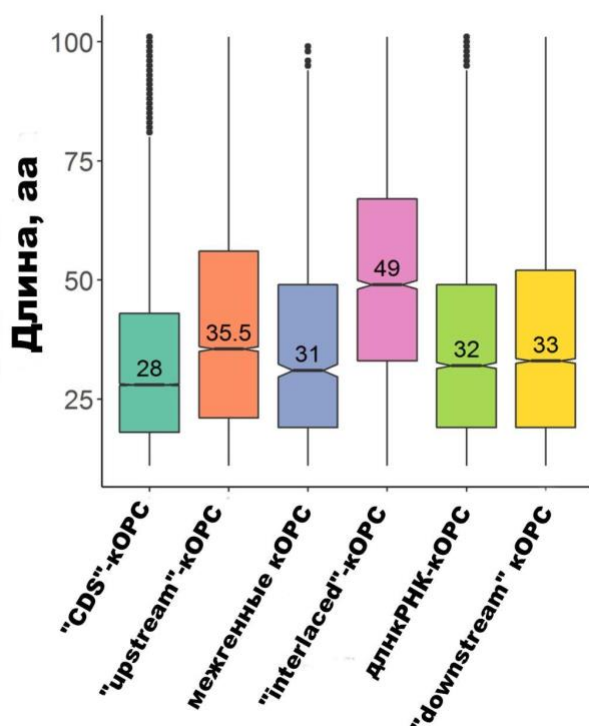


Рисунок 4.31 Сравнение длины различных типов кОПС.

Далее мы провели сравнительный анализ обогащения терминов GO в 4 группах “генных” кОПС, которые были локализованы на мРНК транскриптах. Чтобы исключить возможность того, что различия между группами обусловлены только структурными различиями в генах, несущих кОПС (например, гены с более длинным 5'-UTR имеют больше шансов иметь “upstream” кОПС), мы также оценили обогащение терминов GO в наборе генов, из которых извлекали случайные фрагменты экзонов (см. выше). Термины GO, которые были обогащены в обоих наборах данных, были исключены из дальнейшего рассмотрения. Анализ показал значительное (скорректированное значение $P < 0,01$) обогащение терминов GO в генах, несущих “CDS”-кОПС и “upstream” кОПС. Набор генов, обладающих “CDS”-кОПС, был обогащен терминами GO, связанными со связыванием белка и трансферазной активностью, в то время как гены, обладающие “upstream” кОПС, участвуют в передаче сигналов и регуляции транскрипции (Рис. 4.32). Такие контрастные различия в предполагаемых функциях между генами с различными местоположениями кОПС могут указывать на их роль и/или их пептидов в различных уровнях клеточной регуляции.

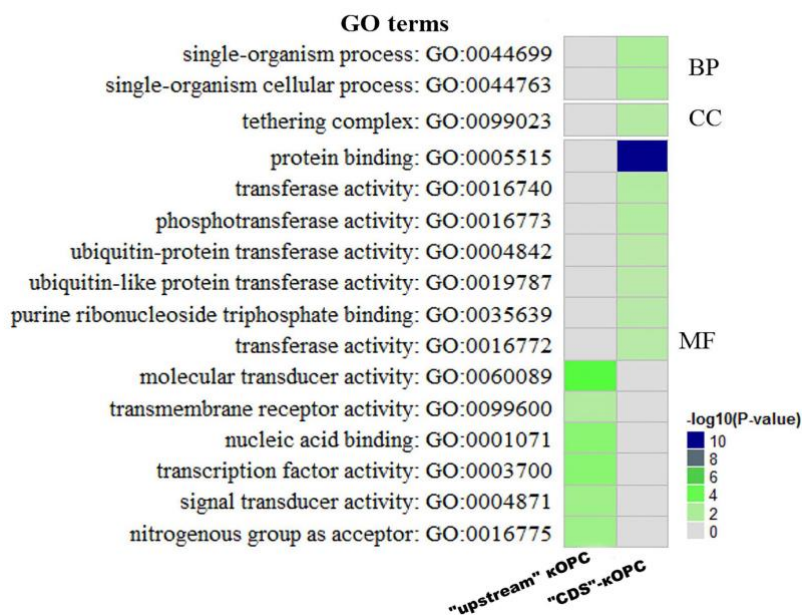


Рисунок 4.32 Результаты анализа обогащения терминами GO для генов, обладающих “upstream” кОПС и “CDS”-кОПС. BP, CC и MF - это “Биологические процессы”, “Локализация в клетке” и “Молекулярные функции” соответственно.

4.2.3 Анализ консервативности “генных” коротких открытых рамок считывания

Считается, что консервативность кодирующей последовательности является маркером ее функциональности. Исходя из этого, мы оценили консервативность и степень эволюционного давления на “генные” кОПС 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, а именно: *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. Ожидаемо, 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 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. 4.33А).

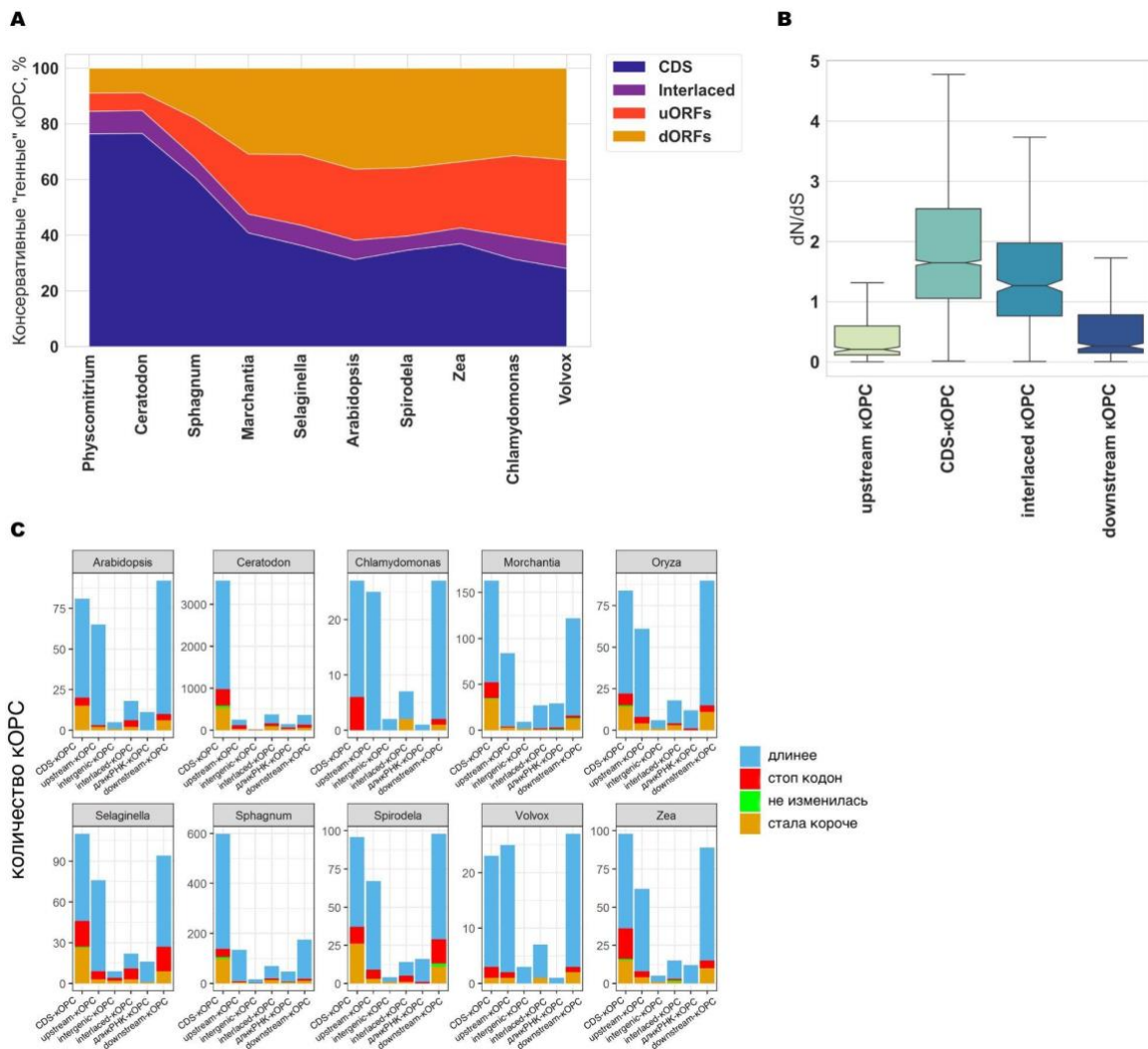


Рисунок 4.33 А - Доля каждого типа "генных" кОРС, имеющих возможных ортологов у десяти видов растений. В - Распределение соотношений dN/dS в разных типах "генных" кОРС, посчитанных на основе выравнивания с *S. purpureus*; $P < 0,0001$ по критерию суммы рангов Крускала–Уоллиса. С - Сравнение длины кОРС, обнаруженных у *P. patens* с длиной предполагаемых ортологов у десяти отобранных видов растений.

Мы также сравнили консервативность "генных" кОРС и 158 маленьких (< 100 aa) белков без интронов из аннотации генома *P. patens*. Доли консервативных кОРС и этих белков значительно различались (7,2% кОРС vs 86% маленьких белков), указывая на низкую консервативность последовательностей кОРС, 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. 4.33В). 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. Поскольку “CDS”-кОПС могут располагаться в области белковых доменов и влиять на структуру и функцию основного белка, то процесс отбора транскрибирующихся рамок этого типа должен иметь определенную специфику. Можно предположить, что транскрибируемые “CDS”-кОПС могут формироваться в быстро меняющихся областях белков (например, в неупорядоченных регионах).

Возможная эволюция некодирующих частей генома в белок-кодирующие гены активно обсуждается в последнее время [314,315], и даже показаны примеры возникновения таких генов [316,317]. Однако обнаружение таких случаев появления новых белок-кодирующих генов из длинных некодирующих РНК,

несущих короткие рамки считывания, является нетривиальной задачей [317,318]. В нашем наборе данных мы оценили, имеют ли возможные ортологи наших кОРС ту же длину, что и исходная кОРС. Согласно нашей оценке, предполагаемые ортологи предсказанных у *P. patens* кОРС в большинстве случаев имеют тенденцию к изменению своей длины (Рис. 4.33С). Большинство таких случаев связаны с увеличением кодирующей последовательности, тем самым показывая, что они могут играть роль исходного материала для отбора. Вероятно, эволюция коротких открытых рамок считывания может включать этап их элонгации [68] и впоследствии приводить к появлению белок-кодирующих последовательностей, однако для того, чтобы оценить масштаб этого явления в растительных геномах, требуются дополнительные исследования.

4.2.4 Идентификация пептидов, кодируемых короткими открытыми рамками считывания

Согласно данным рибосомального профилирования, разные типы кОРС могут подвергаться трансляции [86,319,320]. Однако на сегодняшний момент наиболее весомым доказательством наличия продуктов трансляции кОРС в пептидоме клетки нужно считать их идентификацию при помощи масс-спектрометрического анализа. Сравнение полученных ранее результатов рибосомального профилирования и масс-спектрометрического анализа позволяет предположить, что МС обнаруживает только пептиды от интенсивно транслирующихся рамок [86,189]. Тем не менее, к настоящему времени масс-спектрометрический анализ подтвердил наличие нескольких десятков продуктов трансляции кОРС в пептидомах клеток животных [7,190,321]. С другой стороны, в недавнем методическом исследовании не было обнаружено никаких технических препятствий для детекции пептидов от кОРС с помощью методов масс-спектрометрии [322].

Получение доказательств трансляции кОРС является важным шагом на пути к идентификации функциональных коротких открытых рамок считывания. Мы решили выяснить, какое количество предсказанных нами кОРС можно обнаружить

в пептидоме клеток *P. patens* при помощи масс-спектрометрического анализа. Для этого был проведен поиск продуктов трансляции кОПС в полученных нами данных масс-спектрометрического анализа пептидомов и протеомов протонемы, протопластов и гаметофоров *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 поиска мы использовали поисковую программу MaxQuant [233] и использовали базу данных, содержащую 32 926 аннотированных белков из последней версии генома *P. patens* (V3.3) [225], 85 хлоропластных белков, 42 митохондриальных белка и 70 095 кОПС. Совпадения пар теоретический пептид — настоящий спектр (peptide spectrum matches (PSMs)) были идентифицированы с уровнем ложноположительных результатов (FDR), равным 1%. Это привело к идентификации 1177 PSMs, 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 [234]. 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. Далее мы провели оценку полученных спектров и отбросили недостаточно хорошие идентификации. В результате мы получили набор “достоверно” транслирующихся 46 кОПС: 17 в гаметофорах, 29 в протонеме и 14 в протопластах (Рис. 4.34А). Достаточно небольшое количество “общих” для разных типов клеток *P. patens* транслирующихся кОПС может указывать, скорее, на их тканеспецифичную трансляцию, а не на технические ограничения в их детекции. 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. 4.34В). Поскольку считается, что “upstream” кОПС в основном играют регуляторную роль, влияя на трансляцию белок-кодирующей последовательности, такая большая длина идентифицированных нами транслирующихся “upstream” кОПС, скорее, может указывать на идентификацию продуктов трансляции полицистронных транскриптов, как было недавно показано у растений [188,323]. Тем не менее, “генные” кОПС были наиболее представленной группой “достоверно” транслирующихся кОПС. Например, “CDS”-кОПС составляли около 41,3% (19 кОПС) всех рамок (Рис. 4.34С), “upstream” кОПС (6 кОПС, 13%) и “downstream” кОПС (9 кОПС, 19,6%). В целом, мы обнаружили, что только пять транслирующихся кОПС являются консервативными и четыре из них находились под действием негативного отбора ($dN/dS < 0.5$). Таким образом, анализ консервативности транслирующихся кОПС показал, что только 11% из них являются консервативными.

Кроме того, была обнаружена “достоверная” трансляция девяти кОПС, расположенных на длнкРНК. Уровень транскрипции некоторых днРНК (согласно нашим предыдущим данным [240] и экспрессионному атласу Phytozome V12) и признаки трансляции соответствующих длнкРНК-кОПС представлены на Рис. 4.34D.

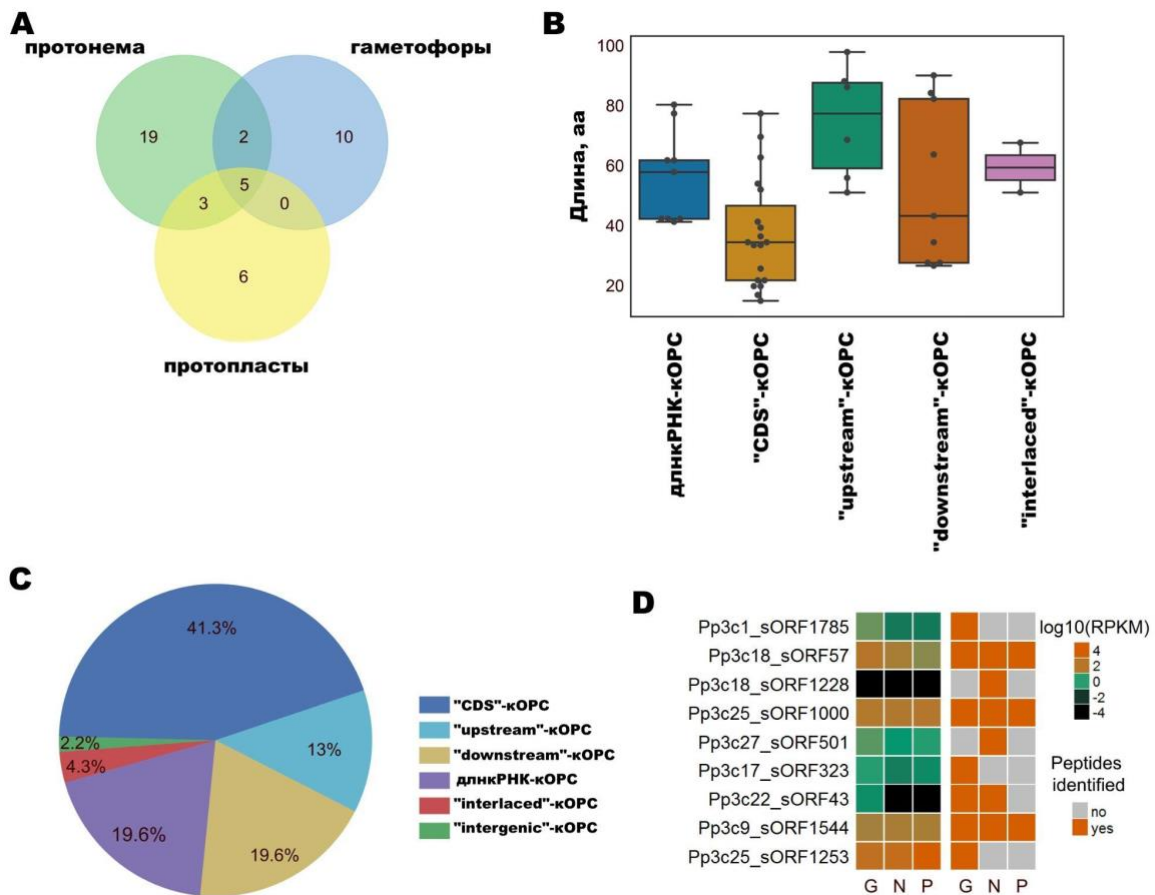


Рисунок 4.34 **A** - Диаграмма Венна, показывающая распределение идентифицированных транслируемых “достоверных” кОПС среди трех типов клеток мха. **B** - Распределение длины различных групп транслирующихся “достоверных” кОПС. **C** - Диаграмма, показывающая процентное соотношение разных типов транслирующихся кОПС. **D** - Тепловая карта, показывающая уровни экспрессии (\log_{10} [RPKM]) для днкРНК (слева), несущих кОПС (днкРНК-кОПС), и бинарная тепловая карта, показывающая доказательства трансляции (определяется по тому, был ли пептид идентифицирован [коричневый] или нет [серый] в МС данных) для соответствующих днкРНК-кОПС (справа) в трех тканях мха: гаметофоры (G), протонема (N) и протопласты (P).

Три из этих “днкРНК-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. Эти данные могут указывать на биологическую значимость пептидов, транслируемых с этих кОПС (см. ниже).

Мы обнаружили, что спектры, подтверждающие идентификацию пептидов, кодируемых короткими рамками считывания, в целом имеют более низкое качество, чем те, которые сопоставлены с последовательностями белков (Рис. 4.35). Это согласуется с данными, полученными ранее для коротких рамок у животных [7,190]. Интересно, что качество спектров, поддерживающих идентификацию продуктов трансляции кОПС, было выше в «пептидном» наборе данных.

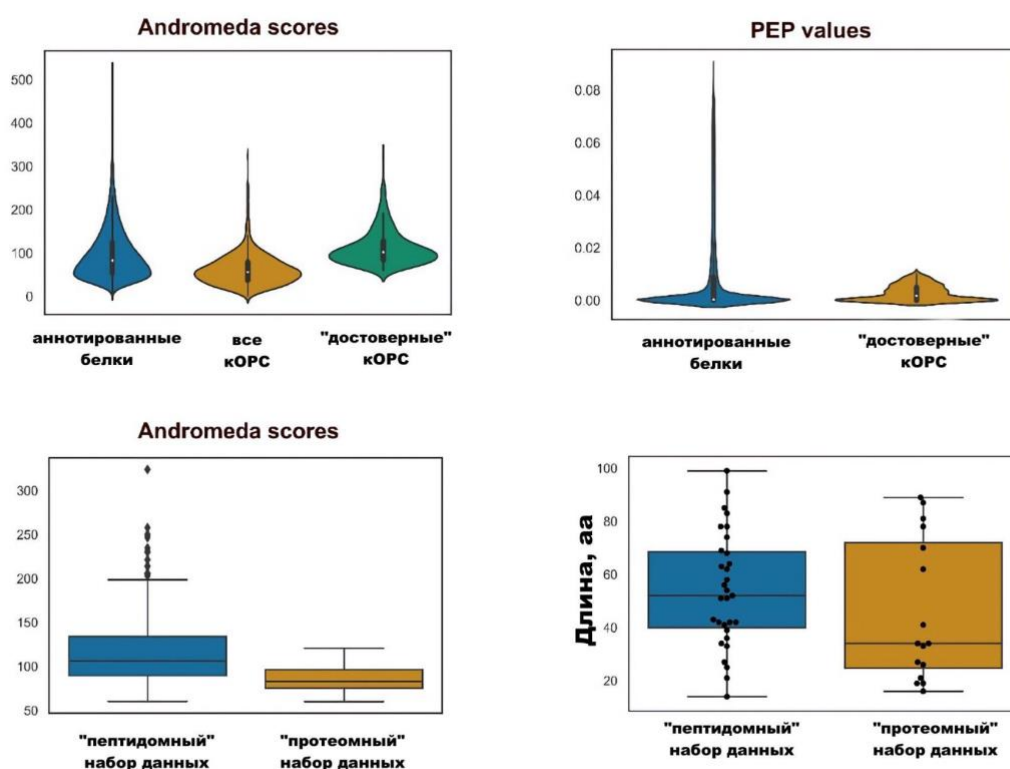


Рисунок 4.35 Сравнение значений апостериорной вероятности (PEP) и Score для PSMs, поддерживающих идентификацию пептидов и белковых последовательностей, кодируемых кОПС: **A, B** – Распределение значений Score/PEP для поискового алгоритма Andromeda для пептидов, относящихся к аннотированным последовательностям белков, идентифицированных кОПС («все кОПС», PSM FDR 1%), кОПС с высокой «достоверностью» («достоверные кОПС», PSM FDR = 1%, PEP \leq 0,01, Score \geq 60). **C** - Сравнение значений Score для PSMs, полученных из поискового алгоритма Andromeda для кОПС высокой достоверности в «пептидном» и «протеомном» наборе масс-спектрометрических данных; **D** – Распределение длин кОПС с высокой достоверностью в «пептидных» и «протеомных» данных.

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 пептидных молекул; и (2) пептиды от кОПС 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 кОПС в клетке. Интересно отметить, что мы не наблюдаем какого-либо значительного перекрытия между кОПС, идентифицированными с помощью протеомного и пептидомного подходов. Таким образом, эти результаты демонстрируют преимущества использования взаимодополняющих подходов для более полной идентификации продуктов трансляции кОПС.

4.2.5 Совместная экспрессия коротких рамок считывания и белков, которые расположены на одном транскрипте

Ранее было показано, что у эукариот альтернативная рамка считывания может транслироваться наряду с основной белок-кодирующей последовательностью [324–326]. Для того, чтобы обнаружить совместную трансляцию основной белок-кодирующей рамки и кОПС из одного и того же локуса (предположительно на полицистронном транскрипте), мы далее проанализировали результаты идентификаций данных масс-спектрометрического анализа. Используя нефильтрованный набор транслирующихся кОПС (PSMs FDR=1%), мы идентифицировали 144 гена, для которых, по нашим данным МС, транслируются не менее двух рамок считывания (одна основная рамка считывания и одна кОПС). Восемьдесят два из этих случаев были связаны с совместной трансляцией “CDS”-кОПС и белок-кодирующей рамки. Используя набор “достоверно” транслирующихся кОПС, мы идентифицировали семь локусов, по крайней мере, с двумя транслируемыми рамками, включая пять “CDS”-кОПС, 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 (Рис. 4.36).

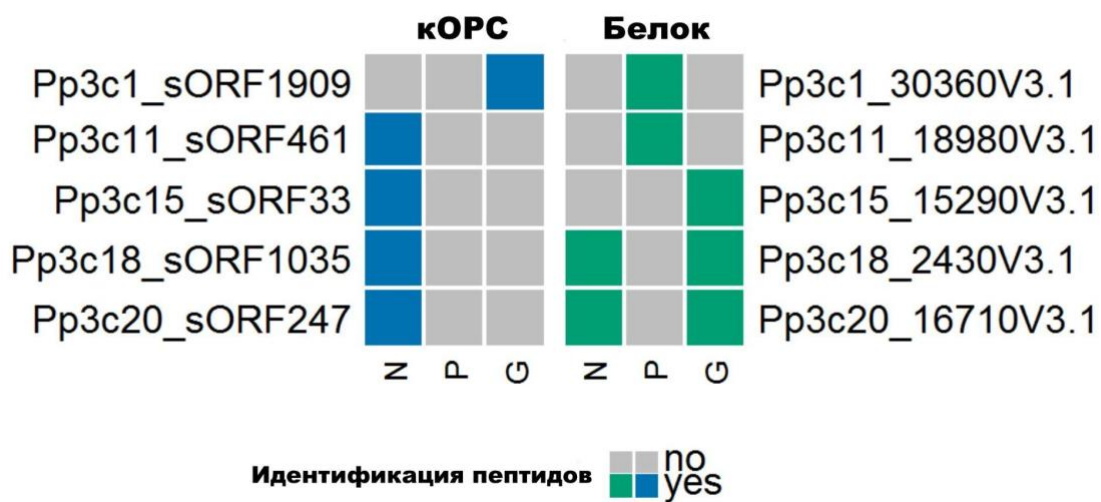


Рисунок 4.36 Тепловая карта, показывающая тканеспецифическую экспрессию "достоверных" кОПС и белков, расположенных в одном локусе в трех типах тканей мха. G, N и P соответствуют гаметофорам, протонеме и протопластам соответственно. Показаны только белки, подтвержденные более чем тремя уникальными триптическими пептидами в данных МС.

4.2.6 Альтернативный сплайсинг регулирует количество кОПС на транскриптах

Альтернативный сплайсинг (АС) — это процесс вырезания интронов и объединения экзонов в различных комбинациях, что приводит к формированию разных изоформ одного и того же гена. Альтернативный сплайсинг является универсальным для эукариотических организмов, и показано, что более 50% генов *P. patens* подвергаются АС [327]. Можно предположить, что АС может приводить к изменению количества, длины и положения на транскрипте различных групп

кОРС. Мы провели анализ, чтобы определить, какое количество кОРС может меняться при АС. Суммарно мы обнаружили 6092 альтернативно-сплайсированных кОРС (АС-кОРС), принадлежащих 4389 транскриптам. Из них “CDS”-кОРС были значительно представлены, в то время как остальные типы “генных” кОРС были значительно недопредставлены среди АС-кОРС по сравнению с контрольным набором случайных экзонных фрагментов (Рис. 4.37А). Мы обнаружили, что примерно половина всего набора АС-кОРС подвергалась полному вырезанию из соответствующей изоформы (Рис. 4.37В).

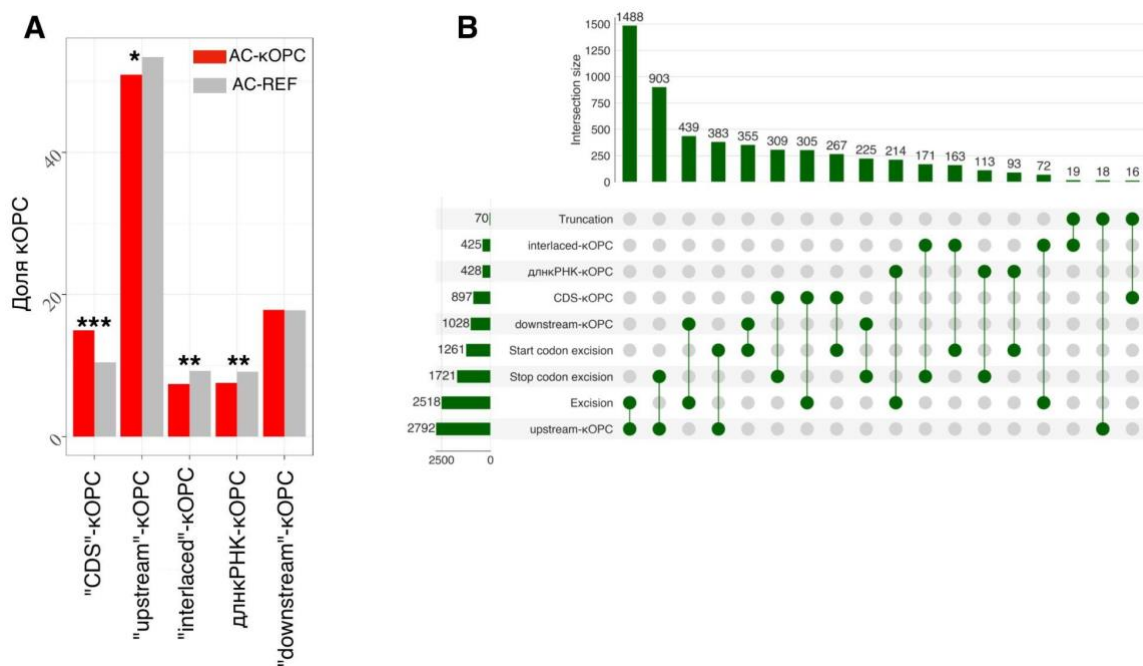


Рисунок 4.37 А - Анализ обогащения различных типов коротких рамок считывания среди альтернативно-сплайсирующихся кОРС (АС-кОРС) в сравнении с контрольным набором случайных экзонных фрагментов (АС-REF); точный тест Фишера. $***P < 10^{-9}$; $**P < 0,001$; $*P < 0,05$. **В** - Диаграмма UpSet, показывающая количество кОРС, которые подвергаются изменениям при альтернативном сплайсинге: “Truncation” - изменение длины кОРС в результате вырезания участка последовательности между старт и стоп-кодонами короткой рамки считывания; “Stop codon excision” - вырезание стоп-кодона кОРС при альтернативном сплайсинге; “Start codon excision” - вырезание стартового кодона кОРС при альтернативном сплайсинге; “Excision” - полное удаление кОРС из изоформы при альтернативном сплайсинге.

Наиболее подвержены действию АС оказались “upstream” кОРС. Наибольшее количество случаев полного удаления кОРС из изоформы происходило именно в этом случае, в сравнении с другими группами кОРС (57% против 20-44%, точный критерий Фишера $P < 0,00001$). Кроме того, эволюционно консервативные кОРС (консервированные у более чем 1 вида) были слабо представлены в наборе АС-кОРС, которые подверглись полному удалению из изоформ (точный критерий Фишера $P = 6,76 \times 10^{-42}$) по сравнению с другими наборами АС-кОРС. Среди транслирующихся “достоверных” АС-кОРС мы обнаружили три, затронутые вырезанием стоп-кодона в определенных изоформах. Еще две из транслируемых АС-кОРС были связаны с вырезанием старт-кодона в изоформах.

Затем мы случайным образом отобрали 13 различных АС-кОРС с/без признаков трансляции и провели поиск соответствующих изоформ в транскриптомах протонемы, гаметофоров и протопластов *P. patens* используя ПЦР, совмещенный с обратной транскрипцией. Этот анализ выявил транскрипцию соответствующих изоформ, подтвердив, что они действительно могут быть транскрибированы и транслированы (пример на **Рис. 4.38**). Более того, некоторые кОРС содержали изоформы, демонстрирующие тканеспецифическую транскрипцию. Эти результаты согласуются с данными масс-спектрометрического анализа и могут указывать на то, что трансляция некоторых кОРС в действительности является тканеспецифичной и может регулироваться АС.

Затем мы провели анализ терминов GO для генов, несущих АС-кОРС, и обнаружили значительное обогащение (точный критерий Фишера, значение $P < 0,01$) для тринадцати молекулярных функций. Десять из них относились к связыванию нуклеиновых кислот (GO:0001071, GO:0003700), активности передачи сигналов (GO:0004871), активности аминопептидаз (GO:0004177), активности трансфераз (GO:0003950, GO :0016772, GO:0016775) и активности киназ (GO:0004672, GO:0004673, GO:0000155). Эти результаты демонстрируют, что альтернативный сплайсинг действительно влияет на ландшафт кОРС в регуляторных генах и что короткие рамки могут играть важную роль в регуляции определенных изоформ.

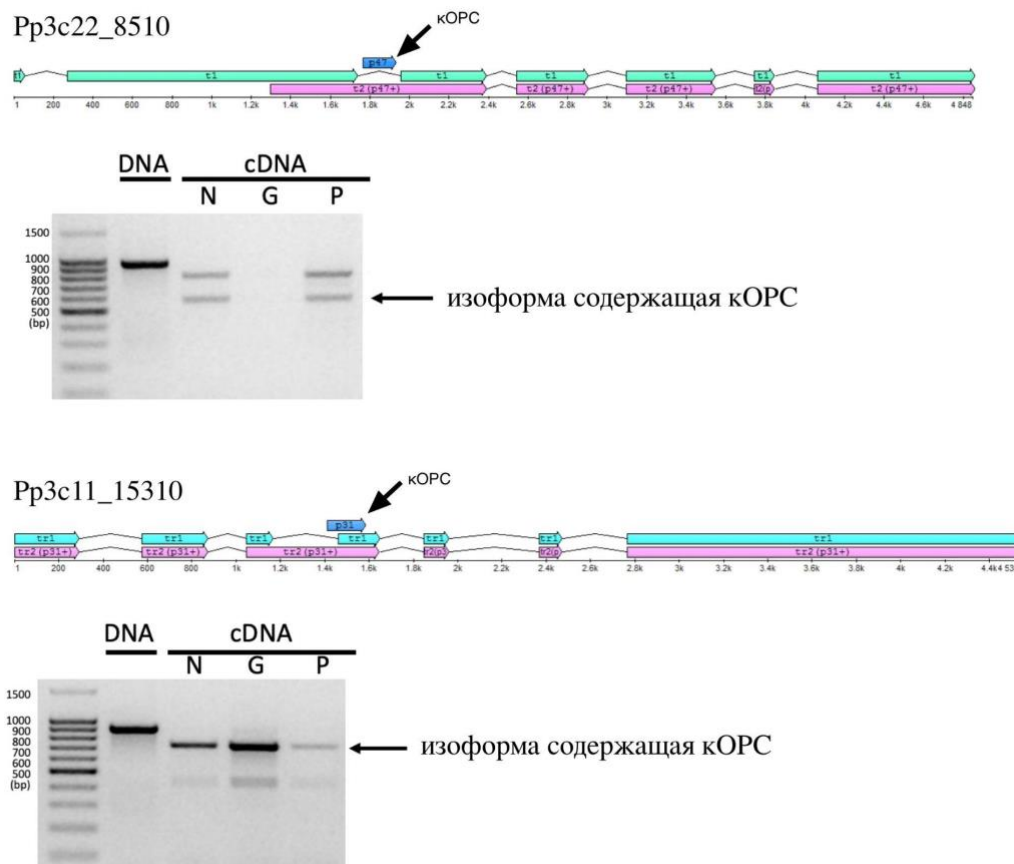


Рисунок 4.38 Пример анализа транскрипции изоформ, содержащих кОРС, с помощью реакции обратной транскрипции, совмещенной с полимеразной цепной реакцией (ОТ-ПЦР). Показана экзон-интронная структура изоформ мРНК и результат электрофоретического анализа продуктов ОТ-ПЦР. G, N и P соответствуют гаметофитам, протонеме и протопластам соответственно.

Ранее было показано, что наличие коротких рамок считывания в 5'-лидерной последовательности мРНК может приводить к запуску системы нонсенс-опосредованного распада РНК (NMD) [328]. Согласно нашим результатам, наиболее частым событием альтернативного сплайсинга являлось полное исключение “upstream” кОРС из изоформы. Мы предположили, что такие “события” могут широко использоваться для регуляции деградации определенных транскриптов. В недавних исследованиях было показано, что появление преждевременного стоп-кодона в результате событий альтернативного сплайсинга может привести к распаду мРНК [329,330]. Используя недавно опубликованные

данные транскрипционного анализа мутантов *P. patens* с выключенной системой нонсенс-опосредованного распада [331], мы исследовали, присутствуют ли наши транслируемые кОРС в транскриптах, предсказанных как мишени данного механизма. Однако мы идентифицировали только одну транслируемую “CDS”-кОРС, которая потенциально присутствовала на таком транскрипте. Для более детального анализа необходимо было бы проанализировать пептидомы соответствующих мутантных линий. Следовательно, трудно судить, могут ли АС-кОРС из нашего набора коротких рамок с высоким кодирующим потенциалом запускать нонсенс-независимую деградацию транскриптов. Таким образом, для подтверждения этой гипотезы требуются дальнейшие исследования.

4.2.7 Анализ кОРС - потенциальных регуляторов белок-белковых взаимодействий

Конкурентные ингибиторы белок-белковых взаимодействий (protein-protein interactions, PPI) называются микропротеинами (miPs) или малыми интерферирующими пептидами (siPEPs). Они могут образовываться при альтернативном сплайсинге или возникать в процессе эволюции при дупликации генов [332–334]. Используя поиск с помощью BLASTP (значение E -value $<10^{-6}$), мы идентифицировали 363 АС-кОРС, которые частично перекрывались с основной белок-кодирующей рамкой считывания в локусе и, следовательно, содержали часть белковой последовательности (цис-кОРС). Мы обнаружили, что 60 цис-sORF содержат внутренне неупорядоченные области [335], а еще 30 содержат части функциональных доменов. Среди них были рамки с протеинкиназным доменом (PS50011) и *tub*-подобным ДНК-связывающим доменом (TIGR01557). Пептиды от соответствующих кОРС можно рассматривать как потенциальных кандидатов в микропротеины [334]. Интересно, что гены, содержащие цис-кОРС, были обогащены киназными и киназоподобными доменами. Анализ обогащения терминов GO также выявил значительное обогащение функциями, связанными с модификациями белка, такими как GO:0006468 (фосфорилирование) или GO:0036211 (процесс модификации белков). Известно, что малые интерферирующие пептиды могут действовать как доминантно-негативные

репрессоры транскрипционных факторов. Среди генов, содержащих цис-кОПС, мы обнаружили транскрипционные факторы, такие как GROWTH-REGULATING FACTOR (Pr3c20_10590), белки, содержащие домен “цинковые пальцы” C2H2 (например, Pr3c1_16920)), домен ВТВ/POZ (например, Pr3c16_9230), ДНК-связывающий домен В3 (например, Pr3c7_7990) и факторы транскрипции типа MYB-CC (например, Pr3c21_2850).

Мы нашли доказательства возможной трансляции цис-кОПС в наших МС данных, и на **Рис. 4.39А** показан один из примеров такой рамки, которая транслируется и при этом перекрывается с основной белок-кодирующей последовательностью.

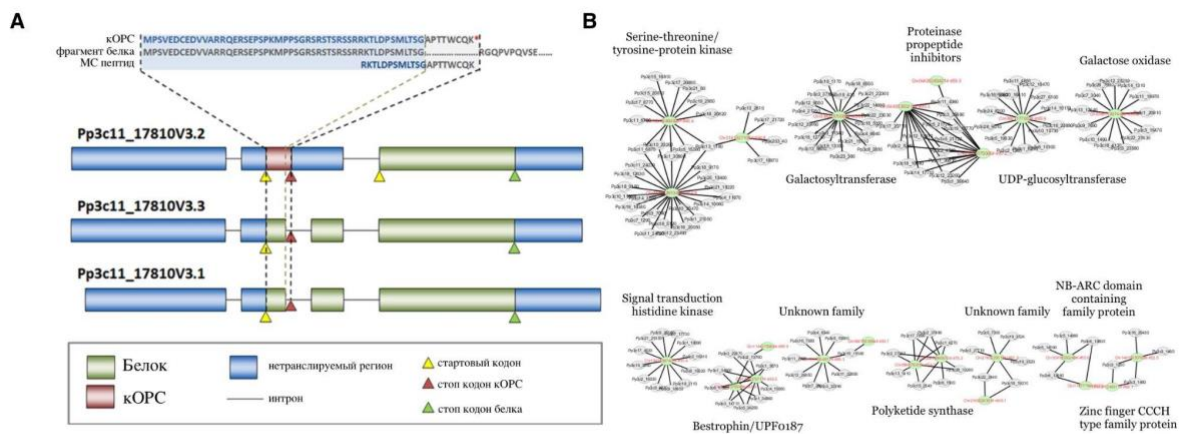


Рисунок 4.39 А - Пример транслируемой “CDS”-цис-кОПС, которая образуется при альтернативном сплайсинге и частично перекрывается с основной белок-кодирующей последовательностью гена Pr3c11_17810. “Удержание” интрона приводит к образованию изоформы с кОПС, а сплайсинг этого интрона приводит к вырезанию стоп-кодона этой кОПС и ее разрушению. На верхней панели показана аминокислотная последовательность пептида, кодируемого кОПС, пептида, обнаруженного в МС данных, и частично белка, транслируемого из основной рамки считывания. Черные и серые пунктирные линии отмечают границы кОПС и начала канонического интрона соответственно. Интрон-экзонная структура трех изоформ гена Pr3c11_17810 была получена из Phytozome v12. **В** - Кластеры генов (более 4 белков), у которых было обнаружено сходство с кОПС; Серые и зеленые узлы соответствуют кОПС и белкам соответственно. Семейства белков, показанные для каждого кластера, были предсказаны на основе данных аннотации генов из Phytozome v12.

Однако мы не идентифицировали цис-кОРС в наборе “достоверно” транслирующихся рамок. Это можно было бы объяснить значительным перекрытием последовательностей белков, тогда как мы отфильтровали «неоднозначные» масс-спектры. Таким образом, идентифицированные в нашем наборе данных потенциальные кандидаты требуют дальнейшей валидации.

Далее мы идентифицировали 272 кОРС, которые имели сходство с аннотированными белками *P. patens*, но были расположены на других транскриптах (транс-кОРС). Такие кОРС могли, например, возникнуть в результате дивергенции древних паралогичных генов, которая произошла после палеодупликации генома мхов [225]. Мы проанализировали консервативность этого класса кОРС и идентифицировали 159 (58,5%) транс-кОРС, которые имели возможных ортологов, по крайней мере, в одном из 10 видов растений (см. выше). Мы также обнаружили, что все эти транс-кОРС подвергаются сильному отбору ($dN/dS \ll 1$).

Мы далее идентифицировали транс-кОРС, которые имели сходство к большим белковым семействам, и сформировали несколько отдельных кластеров как минимум с четырьмя белками из одного семейства, но из разных локусов (**Рис. 4.39В**). Мы обнаружили кластеры генов, которые участвуют в белок-белковых и белок-нуклеиновых взаимодействиях (**Рис. 4.39В**).

Затем мы попытались ответить на вопрос, существует ли связь между экспрессией кОРС и схожих с ними белков? Для этого мы изучили данные по экспрессии кОРС и их потенциальным мишеням и сравнили распределение значений коэффициента корреляции их уровней транскрипции с таковыми для случайно выбранных пар генов (10 итераций). В среднем, пары кОРС-белок имели более высокие коэффициенты корреляции, чем случайно выбранные пары генов (ранговая сумма Уилкоксона и критерий Колмогорова-Смирнова $P < 0,05$), подразумевая, что кОРС и гены-мишени часто ко-экспрессируются.

4.2.8 Предсказание и классификация коротких рамок считывания на длинных некодирующих РНК

Используя программу sORF finder [238], мы обнаружили 5745 транскрибируемых “длнкРНК-кОПС”, которые обладали высоким кодирующим потенциалом, а 9 из них относились к “достоверно” транслирующим коротким рамкам (см. выше). Однако наш подход, основанный только на анализе кОПС с предсказанным высоким кодирующим потенциалом (по данным программы sORF finder) 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 [69,156]. Следовательно, поиск транслирующихся “длнкРНК-кОПС” и изучение их функций у растений представляет значительный интерес. Кроме того, изучение эволюции длнкРНК может пролить свет на появление *de novo* новых белок-кодирующих генов [316,336]. Поскольку случайная трансляция кОПС, 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 [315,337]. Поскольку системный анализ кОПС, кодируемых длинными некодирующими РНК, ранее не проводился у растений, и для того, чтобы получить представление об их функциях и эволюции. мы дополнительно провели всесторонний анализ этого типа кОПС, используя длнкРНК *P. patens*.

Для этого мы объединили в один набор аннотированные длнкРНК *P. patens*, которые были опубликованы ранее: 1498 длнкРНК из базы данных SANTATdb 2.0 [235], 9416 днРНК из GreeNC [236], 3018 длнкРНК из аннотации NCBI (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Physcomitrella_patens/100), 1512 длнкРНК из статьи Lang et al. [225] и 4648 длнкРНК из статьи Simopoulos et

al. [237]. 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 длнкРНК значительно отличались от мРНК (критерий Манна-Уитни, $P < 10^{-15}$). Кроме того, длнкРНК содержали значительно меньше экзонов (медиана=1), чем мРНК (медиана=6; критерий Манна-Уитни, $P < 0,000001$).

Далее, используя транскрипты длнкРНК, мы предсказали на них короткие открытые рамки считывания, сформировав два набора кОРС: начинающиеся с кодона “AUG” с помощью программы MiPePid [243] и начинающиеся с альтернативных кодонов “UUG” и “CUG” с помощью инструмента NCBI ORF finder [244]. Наиболее распространенным альтернативным сайтом инициации трансляции, который был идентифицирован с помощью метода рибосомального профайлинга как у растений, так и у млекопитающих, является “CUG” [176,338]. Однако примерно 65% предсказанных нами с помощью NCBI ORF finder коротких рамок считывания начинались с кодона “UUG”. 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 v12.1, 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. 4.40А**).

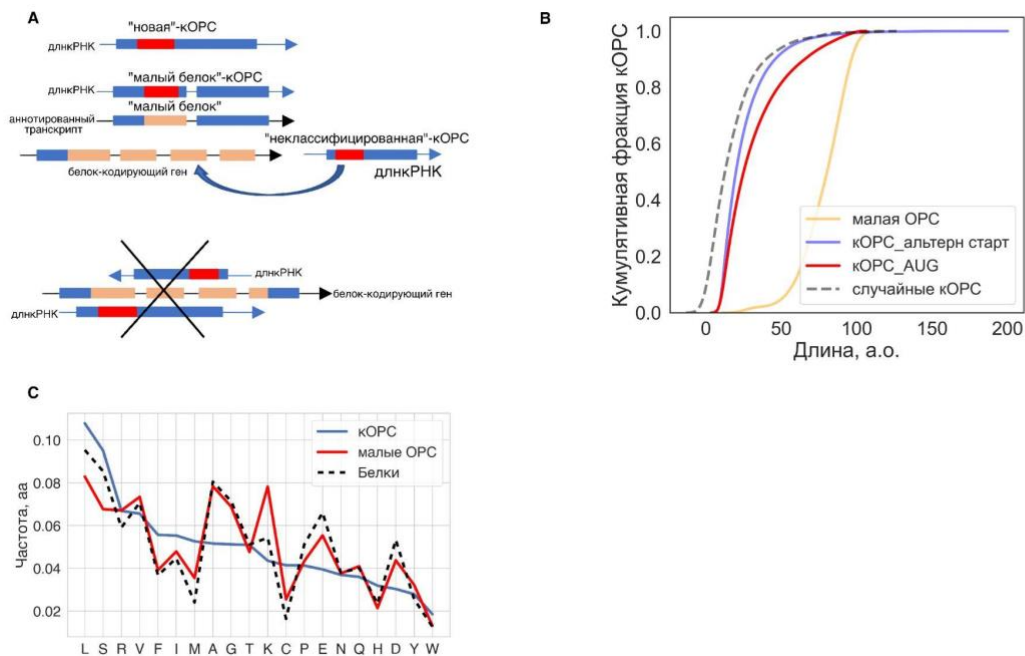


Рисунок 4.40 А- Предсказание и классификация кОРС; В - Кумулятивное распределение длин различных рамок считывания: «случайные кОРС» - короткие рамки считывания, которые теоретически могут появляться в геноме в зависимости от частоты появления стоп-кодонов; медиана длин кОРС, которые начинались со стартового кодона "AUG" составляли 25 а.о. и 20 а.о. для кОРС, начинающихся с альтернативных старт 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.

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 дlnкРНК, согласно аннотации Phytozome v12, имел предсказанный «малый белок», соответствующая ему кОРС была обозначена как "малый белок"-кОРС (Рис. 4.40А,

Табл. 4). Короткие рамки считывания, которые не соответствовали категориям «малый белок» и «неклассифицированные», были обозначены как «новые».

Таблица 4. Классификация предсказанных кОРС

| тип кОРС | программа MiPEPID | программа NCBI finder |
|-----------------------------|-------------------|-----------------------|
| «новые»-кОРС | 80972 (89,65%) | 84409 (99%) |
| «малый белок»-кОРС | 6974 (8%) | 101 (0,11%) |
| «неклассифицированный»-кОРС | 2111 (2,35) | 705 (0,89%) |
| Всего | 90057 | 85215 |

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 Муда $P < 10^{-15}$; **Рис. 4.40В**). Мы также обнаружили, что, в среднем, предсказанные кОРС имели менее богатый ГЦ-состав, чем кОРС, кодирующие белки (критерий Колмогорова-Смирнова, $P < 10^{-20}$).

Мы также провели поиск схожих последовательностей для кОРС из нашего набора с использованием алгоритма BLASTP (значение $E < 0,001$, $\geq 50\%$ идентичности) в известных для животных базах данных коротких рамок считывания. Такой анализ выявил 16 кОРС, общих для *P. patens* и *Arabidopsis* [339,340], и четыре возможных ортолога в базе данных smProt [83].

Ранее было показано, что разные типы кОРС у животных существенно отличаются друг от друга по аминокислотному составу [68]. Аминокислотный состав

предсказанных нами пептидов, кодируемых короткими открытыми рамками считывания, отличался от функциональных белков, особенно содержанием некоторых гидрофобных аминокислот (**Рис. 4.40С**). Повышенные частоты метионина и цистеина, а также пониженные частоты аланина, глутамата и аспартата схожи с составом пептидов, кодируемых кОРС из длнкРНК млекопитающих [68].

4.2.9 Анализ консервативности кОРС, локализованных на длинных некодирующих РНК

4.2.9.1 Анализ консервативности транскриптов длнкРНК

Ранее было показано, что более 70% длнкРНК животных не имеют гомологов у видов, которые разделились более 50 миллионов лет назад [341]. Таким образом, на нуклеотидном уровне длнкРНК являются намного менее консервативными, чем мРНК. На первом этапе мы решили изучить консервативность нашего набора длнкРНК у растений. Для этого с использованием алгоритма BLASTN ($E < 0,00001$) мы выполнили поиск гомологов длнкРНК из *P. patens* в транскриптомах растений из проекта “1000 транскриптомов растений” [245]. 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 отличие от длнкРНК, транскрипты мРНК были гораздо более консервативны (**Рис. 4.41А**). Как и ожидалось, консервативные области в транскриптах мРНК были длиннее, чем в длнкРНК, со средней длиной 753 и 168 нуклеотидов соответственно (**Рис. 4.41В**). У мРНК, которые кодируют белки меньше 100 а.о. (малые ОРС), медиана длины консервативных участков была лишь немного больше, чем в длинных некодирующих РНК (226 vs 168 нуклеотидов соответственно; критерий Манна-Уитни, $P < 10^{-15}$), однако мРНК демонстрировали гораздо более сильную эволюционную консервативность (**Рис. 4.41А и В**). Таким образом, наши результаты показали, что у растений эволюционная

консервативность длнкРНК на нуклеотидном уровне существенно ниже, чем у мРНК. Этот результат совпадает с тем, который получен при сравнении консервативности мРНК и длнкРНК млекопитающих [78,342].

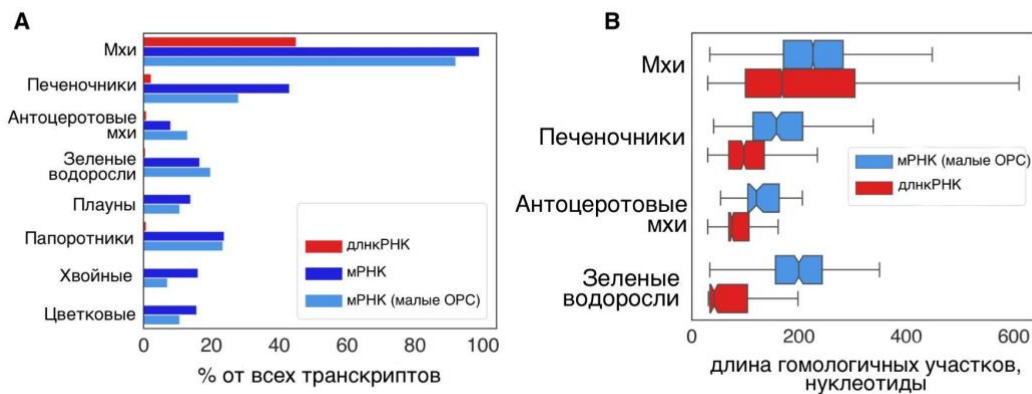


Рисунок 4.41 А - Сравнение консервативности длнкРНК и мРНК, выполненное с использованием транскриптомов из разных растительных таксонов; **Б** - Распределение длин консервативных областей длнкРНК и мРНК, кодирующих белки меньше 100 а.о.; Показаны медиана, квантили, 5-й и 95-й процентиля.

4.2.9.2 Анализ консервативности пептидов, кодируемых кОРС

Затем мы проанализировали консервативность пептидов/микробелков, кодируемых предсказанными нами кОРС. Для поиска гомологов в разных растительных отделах был использован алгоритм TBLASTN ($E < 0.001$) и 479 транскриптомов из проекта “1000 транскриптомов растений” [245]. 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. Ожидаемо, что наибольшее количество совпадений было обнаружено в транскриптоме ближайшего вида мха *Physcomitrium* sp. (“YERO”). У более отдаленных видов мхов количество предполагаемых ортологов колебалось от 1130 до 2887. 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. 4.42).

Таким образом, почти половина локусов нашего набора длнкРНК (4078 из 9168) содержала, по крайней мере, одну консервативную кОРС.

| | 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 4.42 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.

Доля парных выравниваний между последовательностями *P. patens* и “YERO”, содержащими у возможных ортологов стоп-кодона, прерывающие открытую рамку считывания, была значительно выше в кОРС, чем в малых ОРС (11,6% против 1,6% соответственно, хи-квадрат $P < 10^{-10}$), что согласуется с более низкой эволюционной консервативностью кОРС. Как и ожидалось, процент таких выравниваний был выше в случае «новых» кОРС по сравнению с другими типами (хи-квадрат, $P < 0,000001$). Однако доля высококонсервативных выравниваний (>80% идентичности и начинающихся с метионина), содержащих внутренние стоп-кодона, была сходна в кОРС и белках - 2% и 2,8% соответственно.

Мы далее 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). 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. 4.43**).

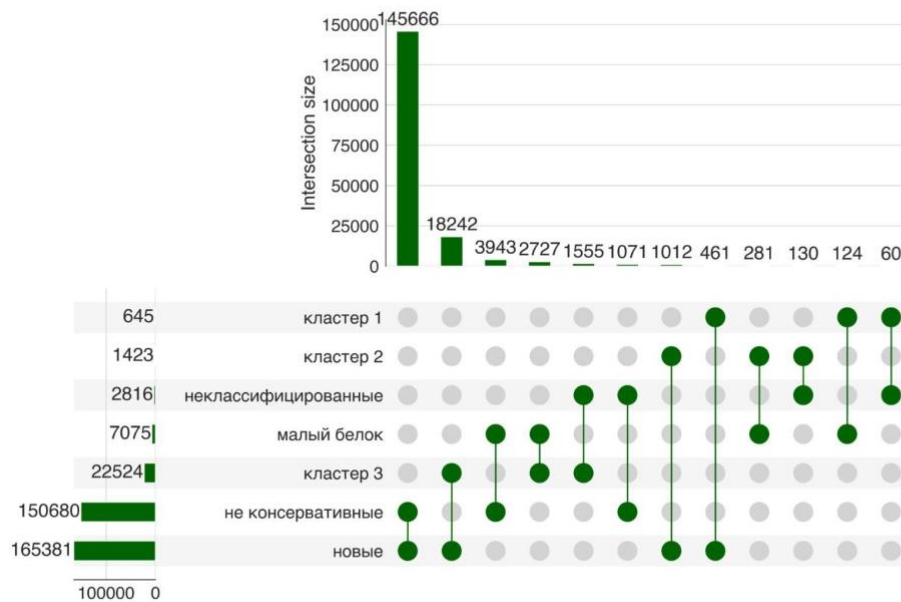


Рисунок 4.43 График UpSet показывающий пересечение трех консервативных кластеров и разных типов кОРС.

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. Короткие рамки считывания, которые имели сходство с аннотированными белками («неклассифицированные»-кОРС) или пересекались с предсказанными малыми белками («малый белок»-кОРС) были более консервативны. Тем не менее, достаточно большое количество этих типов кОРС были или неконсервативны или относились к кластеру 3. Поскольку эти кОРС могут быть обогащены известными белковыми доменами и связаны с уже

существующей аннотацией генома *P. patens*, можно предположить, что соответствующие длнкРНК могут быть остатками функциональных белков. Используя алгоритм BLASTX ($E < 0,00001$), мы провели поиск гомологов пептидов, кодируемых короткими рамками в базе данных Viridiplantae uniprot и примерно оценили долю таких остатков предковых белок-кодирующих генов, в нашем наборе длнкРНК примерно в 3%. Это соответствует результатам, полученным ранее на млекопитающих [318]. Эти результаты предполагают, 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 возникновение генов *de novo* из некодирующих областей. Образующиеся *de novo* “орфанные” гены принадлежат только одному таксону и составляют до 30% генов у некоторых эукариот [316,336,343,344]. Полученные нами результаты согласуются с предыдущими данными, описывающими эволюционную консервативность кОРС в геномах животных [68,190].

4.2.10 Анализ скоростей эволюции длинных некодирующих РНК и коротких открытых рамок считывания

Поскольку особенности эволюции длинных некодирующих РНК и расположенных на них кОРС являются малоизученными, мы проанализировали силу действующего на транскрипты длнкРНК и кодирующие последовательности отбора.

Сравнительный анализ нуклеотидных выравниваний между *P. patens* и *Physcomitrium* sp. показал, что распределения скоростей эволюции (Kd , двухпараметрическая модель Кимуры, K2P) существенно различались между длнкРНК и мРНК (критерий суммы рангов Краскела-Уоллиса, $P < 10^{-15}$), при этом длнкРНК эволюционировали значительно быстрее мРНК (Рис. 4.44А). При этом медианы распределения Kd у длнкРНК и мРНК, кодирующих малые ОРС, были близки (медиана=0,046 vs 0,048 соответственно) и значительно отличалась от другой подгруппы мРНК с более длинными ОРС (медиана=0,035; критерий Манна-Уитни, $P < 10^{-15}$ для обоих сравнений). Таким образом, мы выяснили, 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 [78,345,346].

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. 4.44В**).

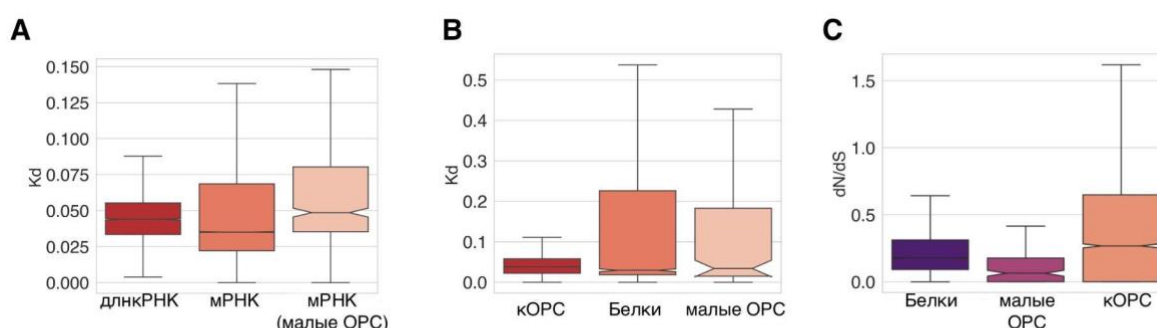


Рисунок 4.44 А - распределение значений Kd , рассчитанных на основе нуклеотидных выравниваний *P. patens* и *Physcomitrium* sp. длнкРНК ($n=4078$) и транскриптов мРНК ($n=15,926$ и $n=252$ для мРНК, кодирующих белки больше и меньше 100 аминокислот соответственно); $P < 10^{-15}$ по критерию суммы рангов Краскела-Уоллиса. **В** - распределение значений Kd , рассчитанное на основании аминокислотных выравниваний между *P. patens* и *Physcomitrium* sp. у кОРС и белковых ОРС; $P=0,2$ по критерию суммы рангов Краскела-Уоллиса. **С** - распределение отношений dN/dS у кОРС ($n=4022$) и функциональных белков ($n=8203$). Малые ОРС представляют собой белок-кодирующую последовательность размером менее 100 а.о.; $P < 0,0001$ по критерию суммы рангов Краскела-Уоллиса.

Ранее было показано, что консервативные области в межгенных длнкРНК животных были обогащены транслируемыми кОРС [342]. Следовательно, 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. Для сравнения были использованы белки из нашего контрольного набора. Как и ожидалось, последовательности, кодирующие белок (включая малые ОРС), имели более сильные признаки отрицательного отбора в сравнении с кОРС (критерий суммы рангов Крускала-Уоллиса, $P < 0,0001$; **Рис. 4.44С**). В целом, ~76% малых ОРС продемонстрировали признаки сильного негативного отбора ($dN/dS < 0,20$) по сравнению с ~45% кОРС (1771 коротких рамок, $dN/dS < 0,20$). Около 30% как кОРС, так и малых ОРС были обозначены нами как «высококонсервативные», т.е. ортологи без мутаций/замен или с единичными заменами в выравниваниях между *P. patens* и *Physcomitrium* sp. Однако в отличие от малых ОРС, ~83% этих «высококонсервативных» кОРС имели гомологи только у небольшого числа видов мхов (кластер 3, см. выше). Эти результаты показывают, 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.

Количество предполагаемых ортологов кОРС у более отдаленных видов мхов резко снижалось, что сопровождалось снижением значений dN/dS (**Рис. 4.45**). У более отдаленных видов мхов медианные значения dS у кОРС были аналогичны таковым в функциональных белках, тогда как значения dN были в два раза выше в кОРС (**Рис. 4.45**). Таким образом, достаточно большое количество кОРС имели признаки положительного отбора или могли подвергаться отбору на нуклеотидном уровне.

| A | Kd | dN/dS | dN | dS | ортологи |
|------------------------|-------|-------|-------|------|----------|
| | | | | | кОПС |
| Physcomitrium sp. | 0.038 | 0.27 | 0.034 | 0.05 | 4022 |
| Encalypta streptocarpa | 0.28 | 0.21 | 0.19 | 0.73 | 384 |
| Racomitrium elongatum | 0.32 | 0.19 | 0.18 | 0.84 | 360 |
| Racomitrium varium | 0.32 | 0.18 | 0.23 | 0.81 | 433 |
| Diphyscium foliosum | 0.35 | 0.18 | 0.25 | 0.94 | 357 |
| Buxbaumia aphylla | 0.42 | 0.14 | 0.28 | 1.0 | 266 |

| B | Kd | dN/dS | dN | dS | ортологи |
|------------------------|-------|-------|-------|-------|----------|
| | | | | | белков |
| Physcomitrium sp. | 0.029 | 0.18 | 0.018 | 0.069 | 8038 |
| Encalypta streptocarpa | 0.22 | 0.13 | 0.1 | 0.74 | 5448 |
| Racomitrium elongatum | 0.24 | 0.11 | 0.11 | 0.87 | 4198 |
| Racomitrium varium | 0.25 | 0.13 | 0.17 | 0.9 | 7038 |
| Diphyscium foliosum | 0.29 | 0.12 | 0.19 | 1.1 | 6317 |
| Buxbaumia aphylla | 0.33 | 0.095 | 0.21 | 1.3 | 4634 |

| C | Kd | dN/dS | dN | dS | ортологи |
|------------------------|-------|-------|-------|-------|-----------|
| | | | | | малых ОПС |
| Physcomitrium sp. | 0.034 | 0.063 | 0.014 | 0.097 | 165 |
| Encalypta streptocarpa | 0.17 | 0.059 | 0.046 | 0.73 | 147 |
| Racomitrium elongatum | 0.23 | 0.069 | 0.089 | 0.86 | 138 |
| Racomitrium varium | 0.21 | 0.058 | 0.14 | 0.91 | 147 |
| Diphyscium foliosum | 0.22 | 0.058 | 0.11 | 1.0 | 144 |
| Buxbaumia aphylla | 0.25 | 0.065 | 0.13 | 1.0 | 122 |

Рисунок 4.45 Медианы значений Kd и dN/dS , рассчитанные на основе выравнивания гомологичных последовательностей для шести видов мхов. **A** - таблица описывает рассчитанные показатели скоростей эволюции для потенциальных ортологов кОПС; **B** - таблица описывает рассчитанные показатели скоростей эволюции для потенциальных ортологов белков; **C** - таблица описывает рассчитанные показатели скоростей эволюции для потенциальных ортологов малых ОПС.

Можно предположить, что продукты трансляции кОПС могут быть материалом для дальнейшего отбора и играть важную роль в адаптации организма к изменяющимся условиям окружающей среды [68]. Исходя из этого, мы далее провели поиск кОПС и белков из наших наборов, которые имели признаки положительного отбора. Около 12% (507 кОПС) проанализированных выравниваний между *P. patens* и *Physcomitrium* sp. имели значения $dN/dS > 1$. Напротив, в наборе малых ОПС не было обнаружено последовательностей с $dN/dS > 1$ (**Рис. 4.46**).

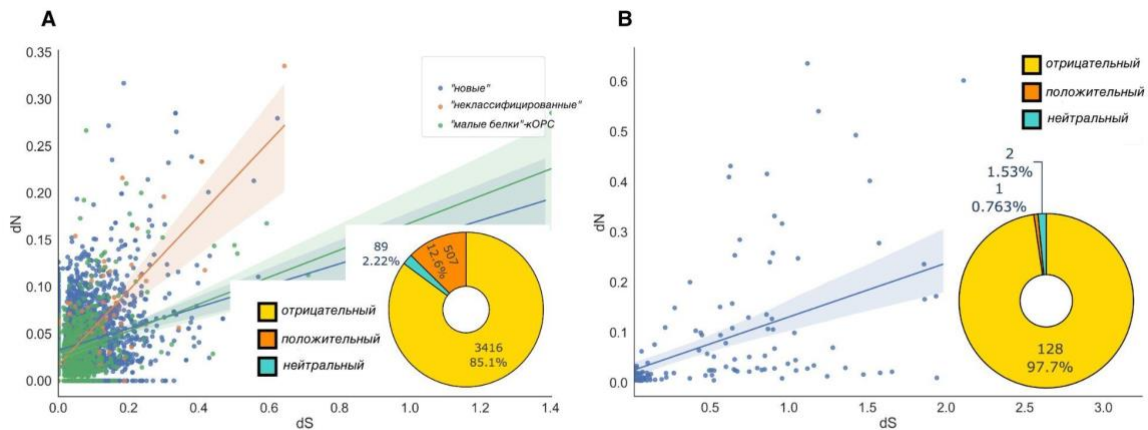


Рисунок 4.45 **А** - Распределение значений dN и dS из выравниваний кОПС; круговая диаграмма показывает количество кОПС, классифицированных на основе значений dN/dS ; “отрицательный” - $dN/dS < 1$, “положительный” - $dN/dS > 1$, “нейтральный” - $dN/dS \sim 1$. **В** - Распределение значений dN по сравнению со значениями dS для малых ОПС (белков менее 100 а.о.); круговая диаграмма показывает количество белков, классифицированных на основе значений dN/dS ; “отрицательный” - $dN/dS < 1$, “положительный” - $dN/dS > 1$, “нейтральный” - $dN/dS \sim 1$.

Для дальнейшего анализа признаков положительного отбора был использован алгоритм HyPhy BUSTED [252]. Алгоритм BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification) проводит генный (не сайт-специфичный) анализ на положительный отбор, отвечая на вопрос, испытал ли ген положительный отбор, по крайней мере, в одном сайте, по крайней мере, на одной филогенетической ветви. Используя данный подход, мы идентифицировали 125 кОПС, которые имели признаки положительного отбора (LRT, $P < 0,05$), включая ~ 16% (20/125) кОПС с $dN/dS > 1$. Таким образом, только ~ 4% из 507 кОПС с $dN/dS > 1$ были также идентифицированы с помощью алгоритма HyPhy-BUSTED как короткие рамки, на которые действует положительный отбор. Затем мы запустили алгоритм HyPhy-BUSTED для проверки положительного отбора у 398 кОПС и 146 малых ОПС, которые имели гомологов у близкородственных видов мха (от 3 до 6

видов). Около 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 длнкРНК, которые подвергаются давлению отбора на нуклеотидном уровне и могут быть функционально важны для РНК-РНК и/или РНК-белковых взаимодействий [342]. Тем не менее, поскольку кодирующие последовательности могут существовать также и на функциональных РНК [347], мы не можем отбросить возможность того, что кОРС в этих областях также кодируют пептиды. Учитывая, что доля синонимичных замен dS обычно ниже в альтернативных и новых экзонах по сравнению с конститутивными экзонами [348,349] и что части белков, кодируемых такими экзонами, часто содержат неупорядоченные области, мы предполагаем, что многие, если не большинство областей длнкРНК с низким значением dS , выполняют двойную функцию [350]. Хорошо известными примерами РНК, сочетающих кодирующие и некодирующие функции, являются растительные при-миРНК, которые участвуют в регуляции генов, а также кодируют функциональные пептиды [150,151]. Кроме того, некоторые кОРС могут участвовать в регуляции экспрессии длнкРНК, связывая соответствующие транскрипты с рибосомами [351] и запуская механизм нонсенс-опосредованного распада РНК [352]. Возможно, такая функция кОРС приводит к консервативности их положения на транскрипте длнкРНК. Такие позиционно-специфичные кОРС могут играть регуляторные роли, сходные с функциями многих “upstream” кОРС [68] и “downstream” кОРС [91] в мРНК.

4.2.11 Структурные особенности пептидов, кодируемых длнкРНК

4.2.11.1 Идентификация пептидов/микробелков содержащих регионы низкой сложности

Большое количество белков, особенно у эукариот, содержат регионы низкой сложности (РНС) различной длины. Несмотря на ряд гипотез о том, что эти регионы тесно связаны с эволюцией белков, их функции остаются слабо изученными [353,354]. 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 [355]. Мы предположили, что когда речь идет о коротких последовательностях кОРС, исключение регионов низкой сложности при поиске гомологичных последовательностей может привести к недооценке количества консервативных и потенциально функциональных пептидов, кодируемых длнкРНК. Для того, чтобы изучить этот вопрос, мы проанализировали рамки, начинающиеся с кодона “AUG”, на наличие РНС с использованием инструмента SEG [356]. Мы обнаружили, что примерно 10% коротких рамок (7831 кОРС), содержали предсказанные РНС (средняя длина=14 а.о.; **Рис. 4.47А**). В целом ~4% всех аминокислот, входящих в состав кОРС, были частью предсказанных РНС. Около 39% РНС, входящих в состав коротких рамок считывания, были либо обогащены одной аминокислотой (частота наиболее распространенной аминокислоты более чем в два раза превышает частоту любой другой аминокислоты), либо представляли собой идеальные тандемные повторы одной аминокислоты более 4 раз подряд. Наиболее часто такими аминокислотами служили фенилаланин (F), лейцин (L), серин (S) и лизин (K). Примечательно, что три из этих аминокислот кодируются AT-богатыми кодонами - F (UUU, UUC), L (CUU, CUC, CUA, CUG), K (AAA, AAG). Для сравнения аминокислотного состава РНС у белков и кОРС мы использовали регионы низкой сложности (средняя длина=16 а.о.), предсказанные в 12,524 функциональном белке, включая 79 малых ОРС.

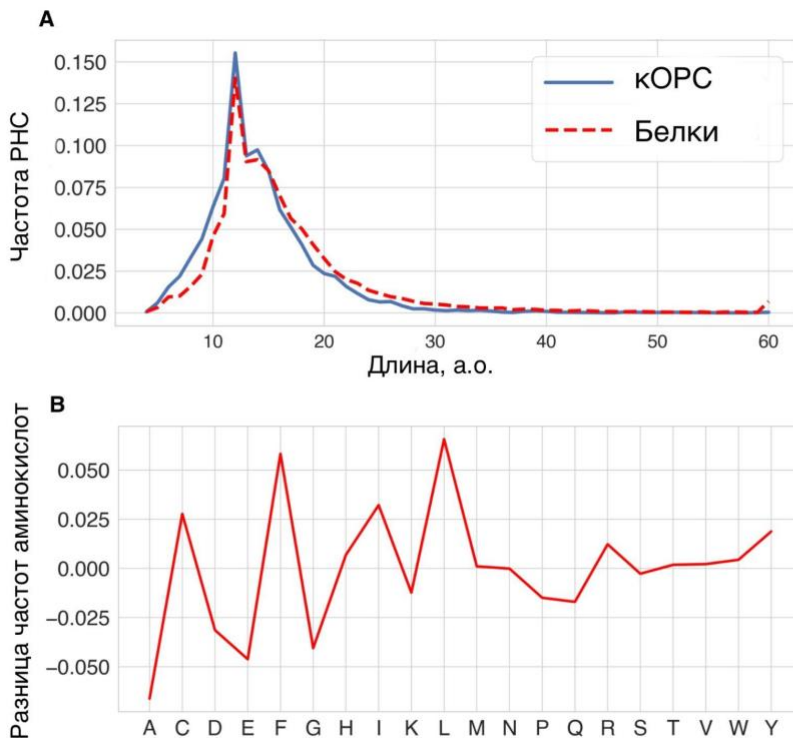


Рисунок 4.47 А - Распределение длин предсказанных РНС в кОРС и функциональных белках. **В** - частоты аминокислот входящих в состав регионов низкой сложности у кОРС; показано отличие от РНС, входящих в состав малых ОРС. КОРС, содержащие РНС, были значительно обогащены лейцином, фенилаланином и изолейцином в сравнении с малыми ОРС (точный критерий Фишера $P < 0,00001$).

В отличие от кОРС, около 69% регионов низкой сложности в белках были обогащены одной аминокислотой. Частоты аланина, глицина, глутаминовой и аспарагиновой аминокислот были значительно выше в РНС из белков, чем для идентифицированных у кОРС (точный критерий Фишера $P < 0,00001$). Это согласуется с полученными ранее данными по регионам низкой сложности у белков [353]. Мы отдельно сравнили состав РНС у кОРС и малых ОРС. Было обнаружено, что РНС в 79 малых ОРС были значительно обогащены лизином и аланином по сравнению с кОРС (хи-квадрат, значение $P < 0,00001$; **Рис. 4.47В**). В свою очередь, регионы низкой сложности у кОРС были значительно обогащены неполярными аминокислотами, такими как лейцин, фенилаланин и изолейцин (точное значение критерия Фишера $P < 0,00001$).

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

Поскольку нуклеотидный состав областей ДНК, в которых “появляются” транскрибирующиеся кОПС, может определять их дальнейшую эволюцию, мы далее проанализировали, отличаются ли показатели Kd (скорость эволюции) и dN/dS (направление отбора) у кОПС, которые находятся в регионах низкой сложности, от других коротких рамок. Интересно, что хотя распределение скоростей эволюции Kd существенно не различалось, отношения dN/dS значительно различались между кОПС, содержащими РНС, и остальными кОПС (критерий Манна-Уитни, $P = 4 \times 10^{-10}$). Кроме того, доля кОПС с $dN/dS > 1$ была значительно выше у кОПС, содержащих РНС, чем у остальных коротких рамок (25% vs 12% соответственно; точный критерий Фишера $P < 0,00001$; Табл. 5). Мы обнаружили, что скорости эволюции (значения Kd и dN/dS) существенно не отличались между белками размером менее 100 а.о. (малые ОПС), которые содержали РНС и теми, которые не содержали регионов низкой сложности.

Таблица 5. Скорость эволюции кОПС с регионами низкой сложности и трансмембранными доменами в сравнении с малыми ОПС.

| | dN/dS , медиана | Kd , медиана |
|----------------------|-------------------|----------------|
| кОПС (РНС) | 0.44 | 0.04 |
| кОПС (без РНС) | 0.26 | 0.04 |
| кОПС (ТМ домен) | 0.41 | 0.04 |
| кОПС (без ТМ домена) | 0.26 | 0.04 |
| малые ОПС | 0.06 | 0.03 |

Таким образом, кОПС, которые содержат РНС, по-видимому, эволюционируют в условиях слабого очищающего отбора, а некоторые могут даже подвергаться

положительному отбору, тогда как малые ОРС, в том числе, содержащие РНС, подвергаются значительно более сильному очищающему отбору.

4.2.11.2 Идентификация пептидов/микробелков содержащих трансмембранные домены и сигналы экспорта

Мало известно об эволюции генов, которые возникли *de novo* из некодирующих областей генома. Такие новые рамки считывания, особенно АТ-богатые, могут быть склонны к образованию трансмембранных (ТМ) доменов [316], которые могут также служить сигналами сортировки белков [357]. Для того, что оценить, как много предсказанных нами кОРС содержат трансмембранные и сигнальные последовательности, мы использовали два алгоритма: ТМНМ 2.0 [263] и SignalP-5.0 [262]. В результате предсказания двумя алгоритмами было идентифицировано 9472 потенциально секретируемых кОРС. Кроме того, 4978 кОРС содержали предсказанные трансмембранные домены (ТМ-кОРС), т.е. потенциально могли являться маленькими трансмембранными белками. Короткие рамки считывания, которые содержали предсказанные ТМ домены, были значительно длиннее в сравнении с другими кОРС (критерий Манна-Уитни, $P < 10^{-15}$; **Рис. 4.49А**), включая консервативные кОРС (критерий Манна-Уитни, $P < 10^{-10}$; **Рис. 4.49В**). При этом длины «новых» кОРС с ТМ доменом были значительно меньше, чем у других типов ТМ-кОРС («малый белок» и «неклассифицированные», критерий Манна-Уитни, $P < 10^{-10}$, **Рис. 4.49С**).

Мы обнаружили, что 1182 кОРС с трансмембранным доменом были консервативны (ортолог, по крайней мере, у одного вида), и около 70% (821/1182) из них принадлежали к типу «новых» кОРС. Процент кОРС с ТМ доменом был немного выше, по сравнению с набором всех коротких рамок считывания (~ 5% vs ~ 3% соответственно; хи-квадрат, $P < 10^{-15}$).

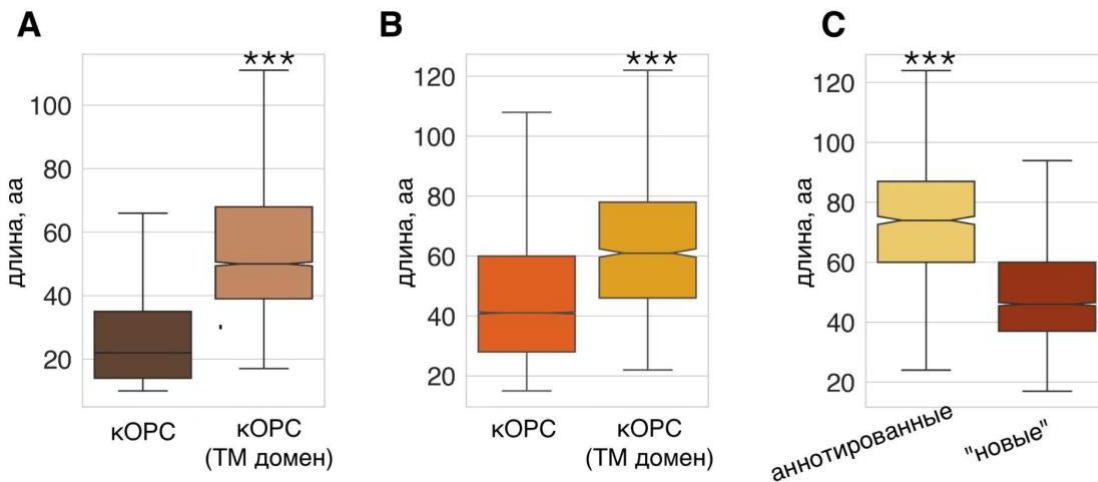


Рисунок 4.49 А - Длина всех кОРС с/без предсказанного трансмембранного домена. Графики показывают медиану, квартили и 5-й и 95-й процентиля. В - Длина консервативных кОРС с/без предсказанного трансмембранного домена. Графики показывают медиану, квартили и 5-й и 95-й процентиля. С - длина «новых» и аннотированных (типы «малый белок» и «неклассифицированные») консервативных кОРС, содержащих трансмембранный домен. Графики показывают медиану, квартили и 5-й и 95-й процентиля. *** $P < 10^{-10}$

Предполагается, что возникающие *de novo* новые, адаптивные гены образуются в АТ-богатых областях генома [316]. Согласно нашим данным, те локусы длнкРНК, которые кодировали ТМ-кОРС имели более низкий ГЦ состав (критерий Колмогорова-Смирнова, $P < 10^{-20}$; **Рис. 4.50А**). 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. 4.50В**). Также, мы обнаружили, что неконсервативные ТМ-кОРС значительно менее богаты ГЦ, чем консервативные (**Рис. 4.50С**). Этот результат может указывать на отбор против сильно гидрофобных кОРС, возможно, из-за их способности к агрегации внутри клетки.

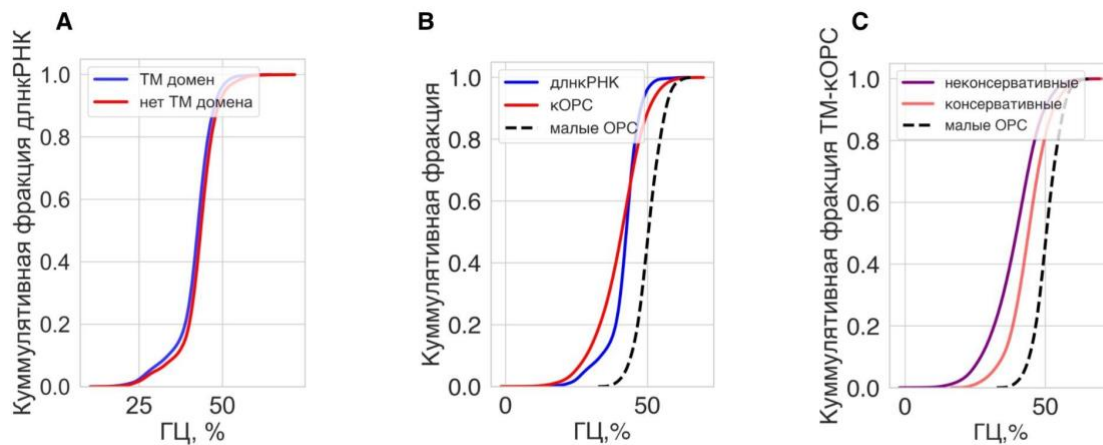


Рисунок 4.50 А - Куммулятивное распределение ГЦ-состава локусов длнкРНК с/без ТМ-кОРС; В - Куммулятивное распределение ГЦ-состава локусов длнкРНК с ТМ-кОРС, без ТМ-кОРС и малыми ОРС; С - Куммулятивное распределение ГЦ-состава консервативных и неконсервативных ТМ-кОРС и малых ОРС.

Затем мы проверили, отличаются ли скорости эволюции кОРС, содержащих ТМ домен, от других кОРС, и обнаружили, что значения dN/dS были значительно выше в наборе ТМ-кОРС (критерий Манна-Уитни, $P=1,3 \times 10^{-5}$; Табл. 2). Это свидетельствует о том, что действие отбора на ТМ-кОРС ниже, и они эволюционируют быстрее, чем остальные кОРС.

Ранее мы обнаружили, что предполагаемые ортологи предсказанных у *P. patens* “генных” кОРС (см. выше) в большинстве случаев имеют тенденцию к изменению своей длины. Кроме того, процесс элонгации кодирующей рамки был предложен в качестве модели эволюции генов, возникающих *de novo* [68]. Таким примером из полученных нами данных является “новая” консервативная ТМ-кОРС, кодирующая 51 а.о. пептид, содержащий предсказанную N-концевую сигнальную последовательность (4–26 а.о.). Эта кОРС содержит специфический консервативный мотив [P***R*R***LR] на С-конце, который является общим со множеством неохарактеризованных небольших белков из базы данных RefSeq (Рис. 4.51).



Рисунок 4.51 Множественное выравнивание последовательностей выбранных ортологов «новой» 50 а.о. ТМ-кОРС.

Другим примером такой кОРС, элонгация которой у ортологов происходила на С-конце, является потенциально секретируемый микробелок из 66 аминокислот (Рис. 4.52).

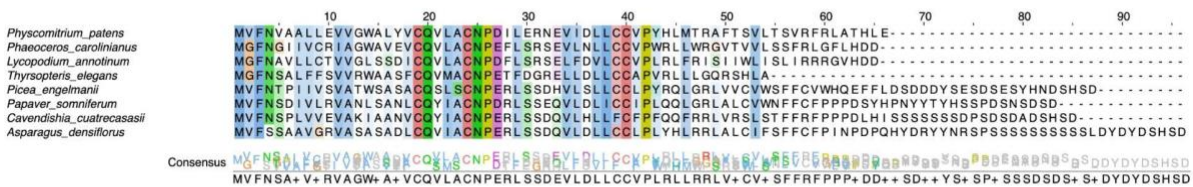


Рисунок 4.52 Множественное выравнивание последовательностей ортологов секретируемого 66-а.о. микробелка, кодируемого длиной некодирующей РНК.

С-конец более длинных ортологов этой кОРС содержал участки идентичных аминокислот и был похож на регионы низкой сложности. Эти примеры указывают на возможную эволюцию кОРС путем мутаций в стоп-кодоне, что приводит к элонгации транслируемой части, которая при этом является неконсервативной.

В целом, роль регионов низкой сложности в происхождении и эволюции секретируемых и трансмембранных пептидов/микробелков, кодируемых различными транскриптами, плохо изучена [354]. Поскольку трансмембранные домены состоят преимущественно из неполярных аминокислотных остатков, мы проанализировали перекрытие между регионами низкой сложности и трансмембранными доменами в нашем наборе коротких рамок. Приблизительно 35% кОРС, которые начинались с кодона “AUG”, также содержали предсказанные

PHC и в 90% случаев они перекрывались как минимум на 50%. Учитывая, что кОРС, которые содержат регионы низкой сложности, были обогащены неполярными аминокислотами, было ожидаемо, что ~35% из них содержат сигналы экспорта или трансмембранные домены (Рис. 4.53).

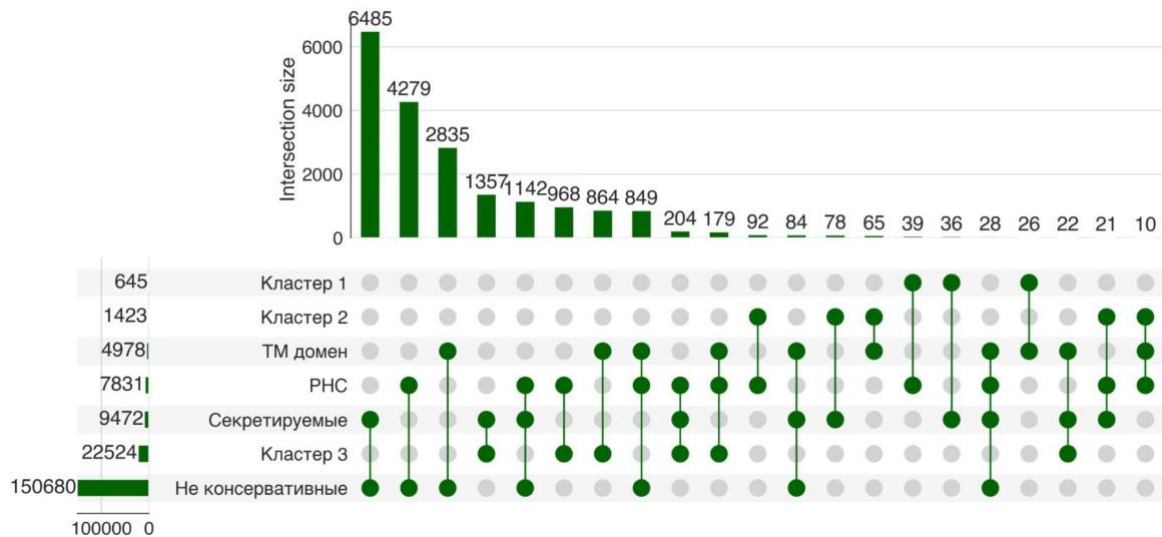


Рисунок 4.53 График UpSet, показывающий пересечение кОРС, кластеризованных по уровню консервативности, с короткими рамками, которые содержат предсказанные регионы низкой сложности (PHC), трансмембранные (TM) домены и сигналы экспорта из клетки. Кластер 1 - высокий уровень консервативности кОРС в разных растительных отделах; Кластер 2 - гомологичные последовательности найдены у многих видов мхов, а также идентифицируются в отделах печеночников и антоцеротовых мхов; Кластер 3 - кОРС имеют возможных ортологов только в ближайших видах мхов.

Кроме того, 937 кОРС с регионами низкой сложности (80% из которых «новые» кОРС) были обогащены такими аминокислотами как фенилаланин, лейцин, валин и изолейцин, но не перекрывались с предсказанными трансмембранными регионами. Можно предположить, что некоторые из них также могут локализоваться в мембранах. Учитывая особенности поиска ортологов для коротких рамок с PHC, мы не можем исключить возможность того, что уровень консервативности кОРС с трансмембранными спиралями недооценивается, и большое количество функциональных пептидов/микробелков пропускаются при аннотации геномов растений.

Примером такой рамки является 89 а.о. микробелок, который содержит один регион низкой комплексности, перекрывающийся с сигналом экспорта (предсказано инструментами TMHMM 2.0 и SignalP-5.0), и другой РНС, богатый пролином на С-конце (Рис. 4.54). Поиск с использованием алгоритма TBLASTN без фильтрации таких регионов (параметр `seg='No'`) значительно расширил количество идентифицированных ортологов в ряде видов мхов, предполагая широкое распространение данной кОРС.

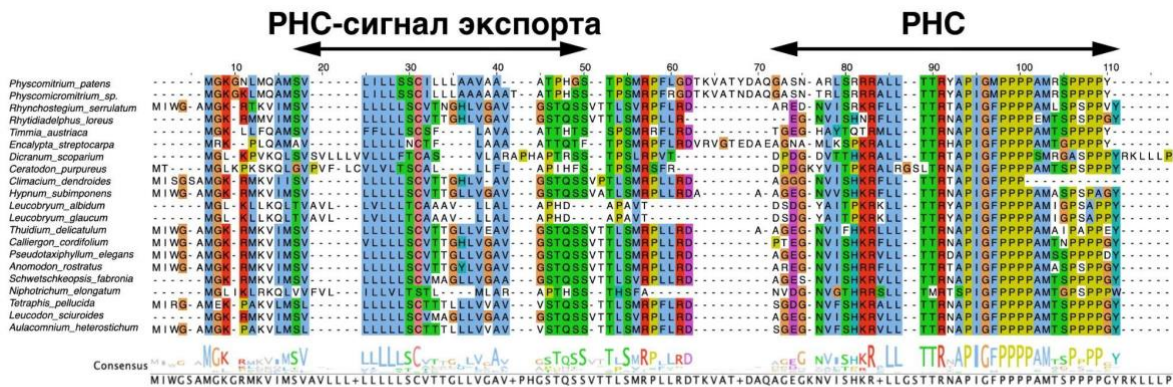


Рисунок 4.54 Множественное выравнивание аминокислотных последовательностей 89 а.о. микробелка, кодируемого кОРС, и ортологов из отдельных видов мхов.

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 [358], длнкРНК могут преимущественно продуцировать пептиды, содержащие трансмембранные домены [316]. Новые пептиды, содержащие ТМ домены, могут быть защищены в мембранах от деградации или взаимодействий с цитоплазматическими белками, которые токсичны для клетки.

4.2.11.3 Идентификация пептидов/микробелков, содержащих функциональные домены

Из предыдущих исследований пептидов/микробелков, кодируемых короткими рамками считывания, известно, что они могут быть обогащены внутренне

неупорядоченными областями в сравнении с аннотированными функциональными малыми белками [190]. Для того, чтобы проанализировать наличие возможных доменов и мотивов в нашем наборе кОПС, мы использовали программу InterProscan [261]. Согласно этому анализу, 95% кОПС не содержали какие-либо предсказанные мотивы и домены. Наиболее распространенным мотивом в оставшихся 5% кОПС являлся “consensus disorder prediction” (внутренне неупорядоченные области), который были обнаружены у 93% (8595/9189) таких коротких рамок. Ожидается, что консервативные рамки содержали большое количество предсказанных доменов. Так, 3357 кОПС, гомологи которых были обнаружены как минимум у 5 видов мхов, значительно обогащены известными доменами и мотивами по сравнению с остальными консервативными кОПС (точный критерий Фишера, $P < 10^{-5}$). Затем мы сравнили наш набор кОПС с функциональными белками длиной менее 100 а.о. (малыми ОПС). Наиболее распространенные функции были связаны с различными рибосомными белками (около 16% малых ОПС).

4.2.11.4 Идентификация кОПС, кодирующих предшественники биоактивных пептидов

Поскольку предшественники биоактивных пептидов часто представляют собой небольшие белки, большая часть последовательности которых неконсервативна, соответствующие гены могут быть ошибочно аннотированы как длнкРНК. Для того, чтобы обнаружить потенциальные предшественники биоактивных пептидов в нашем наборе кОПС, мы использовали специальную программу “SSP prediction tool” [359], которая использует скрытые марковские модели известных семейств секретируемых биоактивных пептидов для их поиска в отобранном геноме. Благодаря такому подходу, мы идентифицировали 45 кОПС с «известными» последовательностями пептидных гормонов, фитоцитокинов и антимикробных пептидов. Наиболее консервативные кОПС содержали мотивы различных семейств растительных антимикробных пептидов, богатых цистеином (CRP), таких как CRP5310 (дефенсин-подобные белки) или CRP5660 — белки, богатые глицином (GRP; **Рис. 4.55**). Другим распространенным консервативным семейством пептидов в нашем наборе кОПС является TAXIMIN (TAX), которые у сосудистых

растений участвуют в регуляции дифференциации латеральных органов [360]. Детальный анализ двух кОПС, содержащих предсказанный домен CLE10 (CLAVATA3/ESR-related), показал, что они являются предшественниками CLV3-подобных пептидов PpCLE5 и PpCLE7, идентифицированных ранее у *P. patens* [133]. Мы также идентифицировали три консервативных кОПС, сходных с пептидами DEVIL/ROTUNDIFOLIA (DVL/ROT), которые, как известно, кодируются короткими рамками считывания у растений [2]. Ортологи кОПС с доменом DVL/ROT были ранее идентифицированы во всех линиях растений, кроме зеленых водорослей.

Некоторые кОПС и малые ОПС содержали предсказанный домен «Probable lipid transfer» (неспецифические белки-переносчики липидов (ns-LTP)). Маленькие белки с этим доменом образуют большие семейства генов у всех наземных растений, но ранее они не были идентифицированы у водорослей [361]. Эти белки считаются одними из ключевых для адаптации растений к наземным условиям существования. Используя более мягкие пороги фильтрации, мы провели поиск кОПС с этим доменом в транскриптомах зеленых водорослей. Из 15 кОПС, 5 давали выравнивание не менее чем на 8 различных транскриптов в водорослях рода *Dunaliella* - *D. salina* и *D. tertiolecta*.

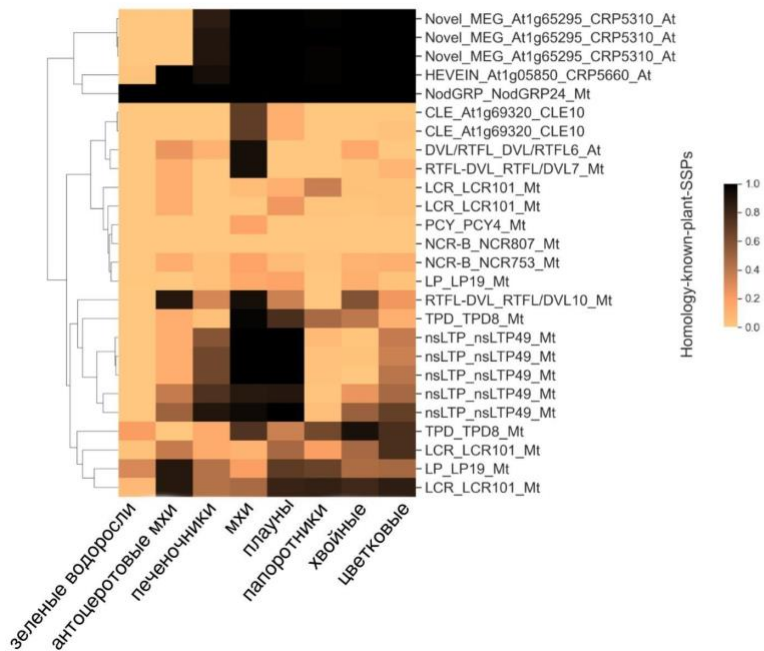


Рисунок 4.55 Тепловая карта, показывающая консервативность кОРС, содержащих наиболее представленные мотивы биоактивных пептидов. Количество ортологов было нормализовано на количество видов.

Таким образом, мы показали, что зеленые водоросли обладают предшественниками микробелков, неспецифических переносчиков липидов. Этот пример показывает, что некоторые из предсказанных длнкРНК действительно могут кодировать предшественники биоактивных пептидов и, следовательно, количество таких семейств у растений может быть в действительности выше, чем предполагается на данный момент.

4.2.12 Анализ экспрессии длнкРНК и трансляции коротких рамок считывания

Для того, чтобы проанализировать экспрессию нашего набора длнкРНК и сравнить ее с мРНК, мы провели прямое нанопоровое секвенирование полиА(+) фракций РНК, выделенных из протонемы ($n = 3$ биологических повтора) и гаметофоров ($n = 4$ биологических повтора) *P. patens*. Нанопоровое секвенирование позволяет охарактеризовать полноразмерные транскрипты РНК в транскриптомах [362,363], являясь незаменимым инструментом для анализа длнкРНК [364].

Этот анализ подтвердил транскрипцию примерно 57% (5249/9168) локусов, кодирующих длнкРНК. Относительно низкая доля выявленных длнкРНК может быть объяснена следующими причинами: (i) экспрессия длнкРНК тканеспецифична, так что многие из транскриптов длнкРНК из нашего набора не экспрессируются в использованных нами условиях; (ii) длнкРНК с низкой экспрессией не были детектированы при нанопоровом секвенировании; (iii) различия в технологиях секвенирования и сборки транскриптов между данными нанопорового секвенирования и секвенирования с короткими ридами.

Около 30% транскриптов, которые были идентифицированы нами с помощью нанопорового секвенирования, подтверждали экзон-интронную структуру длнкРНК из нашего набора данных. Для дальнейшего анализа мы использовали данные по уровню транскрипции этих 1678 локусов длнкРНК.

Ранее было показано, что частота инициации трансляции в альтернативных стартовых кодонах была значительно ниже в том случае, если соответствующая рамка считывания была расположена после кодонов “AUG” [365]. Это может повлиять на распределение по длине транскрипта консервативных и потенциально транслируемых кОПС разного типа. Чтобы оценить степень вероятности этой тенденции, мы проанализировали распределение разных типов кОПС по длине идентифицированных с помощью нанопорового секвенирования длнкРНК. Было обнаружено, что распределение обоих типов (начинающихся с “AUG” и альтернативных старт-кодонов) неконсервативных кОПС является бимодальным и значительно отличается от распределения консервативных кОПС (критерий Колмогорова-Смирнова, $P < 10^{-15}$; Рис. 4.56).

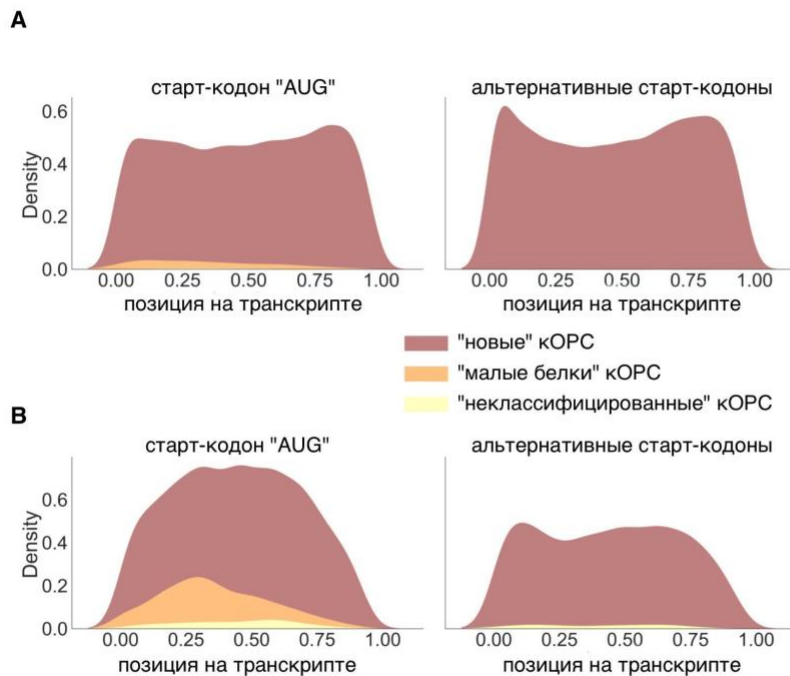


Рисунок 4.56 Распределение позиций стартовых кодонов неконсервативных (А) и консервативных (В) КОРС. Начальное положение каждой КОРС было нормализовано к длине транскрипта.

В частности, консервативные «малые белки»-КОРС, которые начинались с кодона «AUG», как правило, были значительно ближе к 5'-концу транскриптов, чем «новые» и «неклассифицированные» (критерий Колмогорова-Смирнова, $P < 10^{-20}$). Но так как точное предсказание начала КОРС с альтернативными стартовыми кодонами может быть затруднено, эти наблюдения требуют дальнейшей проверки, например с использованием рибосомального профайлинга [365].

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 [366,367]. Согласно нашим результатам нанопорового секвенирования, уровень транскрипции мРНК был значительно выше в протонеме (критерий суммы рангов Краскела-Уоллиса, $P < 10^{-15}$) и гаметофорах (критерий суммы рангов Краскела-Уоллиса, $P < 10^{-15}$) по сравнению с длнкРНК (**Рис. 4.57А**).

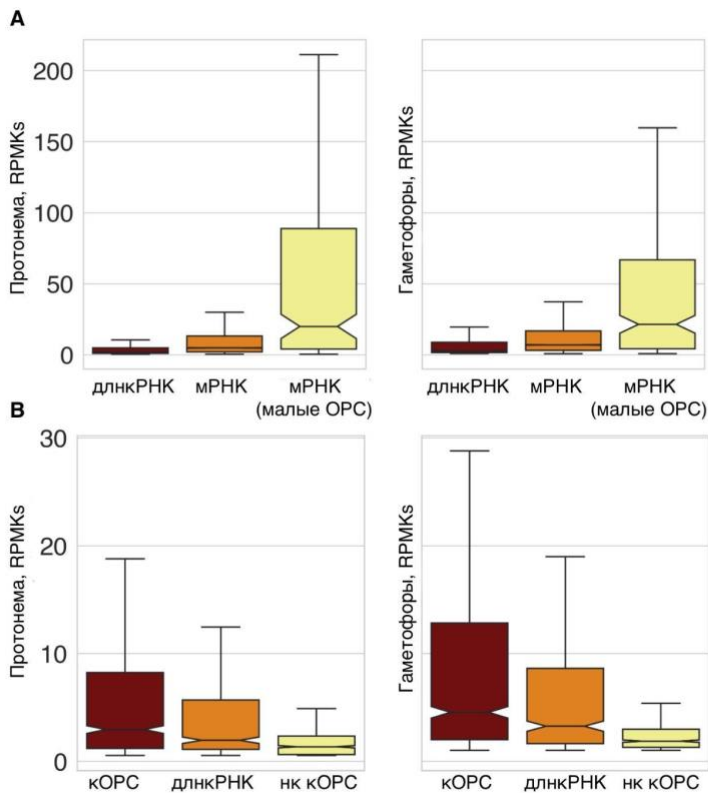


Рисунок 4.57 А - сравнение уровня транскрипции длнкРНК и мРНК в гаметофорах и протонеме; мРНК (малые ОРС) — подмножество мРНК, кодирующих белки размером менее 100 а.о.; В - уровень транскрипции консервативных кОРС, консервативных длнкРНК и неконсервативных кОРС (нк длнкРНК) в протонеме и гаметофорах.

Уровень транскрипции длнкРНК, содержащих консервативные кОРС ($n=629$), был значительно выше, чем уровень экспрессии длнкРНК содержащие отличные от кОРС консервативные на нуклеотидном уровне области ($n=451$), а также неконсервативные кОРС ($n=598$) в протонеме (критерий суммы рангов Краскела-Уоллиса, $P < 10^{-15}$) и гаметофоры (критерий суммы рангов Краскела-Уоллиса, $P < 10^{-15}$; **Рис. 4.57В**). Однако мы не обнаружили существенных различий между уровнями транскрипции длнкРНК, содержащих кОРС разного уровня консервативности (кластеры 1,2 и 3). Таким образом, полученные нами результаты показали, что 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 [368]. Таким образом, консервативность кодирующих участков положительно коррелирует с

уровнем транскрипции мРНК и длнкРНК и, по-видимому, является универсальным правилом. Следовательно, длнкРНК с консервативными кОРС можно рассматривать как мРНК с предсказанным низким кодирующим потенциалом.

Затем мы использовали масс-спектрометрический анализ для идентификации транслируемых кОРС. Для этого были использованы результаты МС анализа пептидомов гаметофоров и протонемы *P. patens*. пептидомы. Для поиска использовали базу последовательностей, содержащую аннотированные белки и все предсказанные кОРС. Выявление транслирующихся кОРС с использованием масс-спектрометрии является сложной задачей из-за низкой экспрессии и быстрой деградации пептидов, кодируемых длнкРНК [69,311]. Используя два поисковых алгоритма - PEAKS 8.0 и MaxQuant, мы получили доказательства трансляции для 195 кОРС, в том числе 56 были идентифицированы обеими поисковыми системами (Рис. 4.58А). Такое количество пептидов, кодируемых длнкРНК, идентифицированное с помощью МС анализа, согласуется с результатами, полученными на клетках человека [69,369].

Ожидается, что приблизительно 44% транслируемых кОРС принадлежали к классу «новых» кОРС (Рис. 4.58В). Интересно, что около 31% «новых» транслируемых кОРС были неконсервативны, что свидетельствует о сильной изменчивости пептидов/микробелков, кодируемых длнкРНК. Как и ожидалось, «новые» транслируемые кОРС значительно преобладали среди неконсервативных коротких рамок (критерий хи-квадрат, $P < 10^{-15}$).

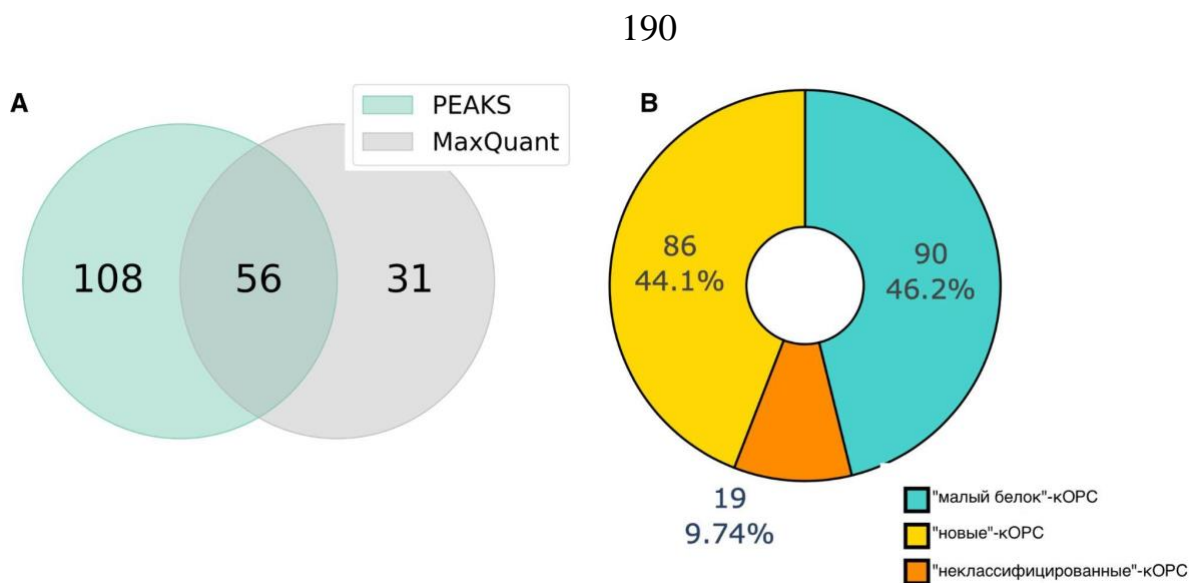


Рисунок 4.58 **А** - Диаграмма Венна, показывающая сравнение количества идентификаций транслирующихся кОРС двумя поисковыми системами — PEAKS 8.0 и MaxQuant; **Б** - Круговая диаграмма, показывающая количество транслирующихся кОРС разного типа.

Предполагается, что использование для продуктов трансляции кОРС протеомных стандартов идентификации белков (≥ 2 уникальных пептидов) будет достаточно для подтверждения их существования в пептидоме, но такие отсечки могут привести к потере части действительно транслирующихся пептидов [370]. Согласно нашим данным, трансляция 73 кОРС была подтверждена двумя и более уникальными пептидами. Среди них было 13 «новых» (8 неконсервативных) кОРС, в том числе, выявленные ранее «длнкРНК-кОРС». Также ранее было показано, что масс-спектрометрия может надежно подтвердить трансляцию кОРС из транскриптов с высокой экспрессией [86]. Мы проверили это предположение и оказалось, что как в протонеме, так и в гаметофорах уровень транскрипции длнкРНК с транслируемыми кОРС был значительно выше, чем без признаков трансляции (критерий Манна-Уитни, $P < 0,00001$), что означает, что в нашем анализе были обнаружены только кОРС из высокоэкспрессируемых длнкРНК. Поэтому количество действительно транслируемых кОРС в нашем исследовании, вероятно, существенно занижено.

4.3 Биологически активные компоненты пептидомов растений

В тканях и биологических жидкостях позвоночных, насекомых, грибов, растений и бактерий идентифицированы тысячи пептидов, различающихся по своему происхождению, количеству и возможным функциям. Эти пулы пептидов называются внутриклеточными или внеклеточными пептидомами, и помимо небольшой доли хорошо охарактеризованных пептидных гормонов и антимикробных пептидов, основная часть пептидома плохо изучена. Однако все больше данных свидетельствует о том, что множество биологически активных пептидов присутствуют в пептидомах различных организмов. Исходя из этого, мы поставили перед собой задачу оценить биологическую активность пептидов - продуктов гидролиза белков предшественников, а также пептидов/микробелков - продуктов трансляции коротких открытых рамок считывания.

4.3.1 Антимикробная активность компонентов пептидного пула

4.3.1.1 Антимикробные пептиды протопластов

Согласно нескольким исследованиям, функциональные белки растений и животных содержат аминокислотные последовательности, потенциально обладающие антимикробной активностью [106,371]. Мы предположили, что резкое увеличение количества эндогенных пептидов при обработке протонемы драйзелазой моделирует биотический стресс и может приводить к выщеплению из функциональных белков биоактивных пептидов, например, обладающих антимикробной активностью. Для того, чтобы проверить эту гипотезу, мы далее оценили антимикробный потенциал идентифицированных в протопластах пептидов с помощью специализированной программы поиска антимикробных последовательностей АМРА (Antimicrobial Sequence Scanning System, <http://tcoffee.crg.cat/apps/ampa/do>) [372]. Всего мы идентифицировали 81 пептид, уникальный для протопластов и обладающий предсказанной антимикробной активностью. Пептиды с наиболее высокой предсказанной антимикробной

активностью (индекс PV) являлись фрагментами различных функциональных, главным образом рибосомных белков (около 30% от всех пептидов).

Для оценки действия идентифицированных пептидов на микроорганизмы мы синтезировали четыре пептида (Табл. 6) и оценили их активность *in vitro*. Из этих пептидов наиболее высоким антимикробным потенциалом обладали три: VKLNKHIWSR (пеп 1, фрагмент рибосомального белка L31), FKTGKNRWF (пеп 2, фрагмент 60S рибосомального белка L27) и VISVKGVKGR (пеп 3, фрагмент рибосомального белка L14). Кроме того, согласно данным масс-спектрометрического анализа, пеп1 и пеп2 имели высокий уровень представленности в протопластах (Рис. 4.59).

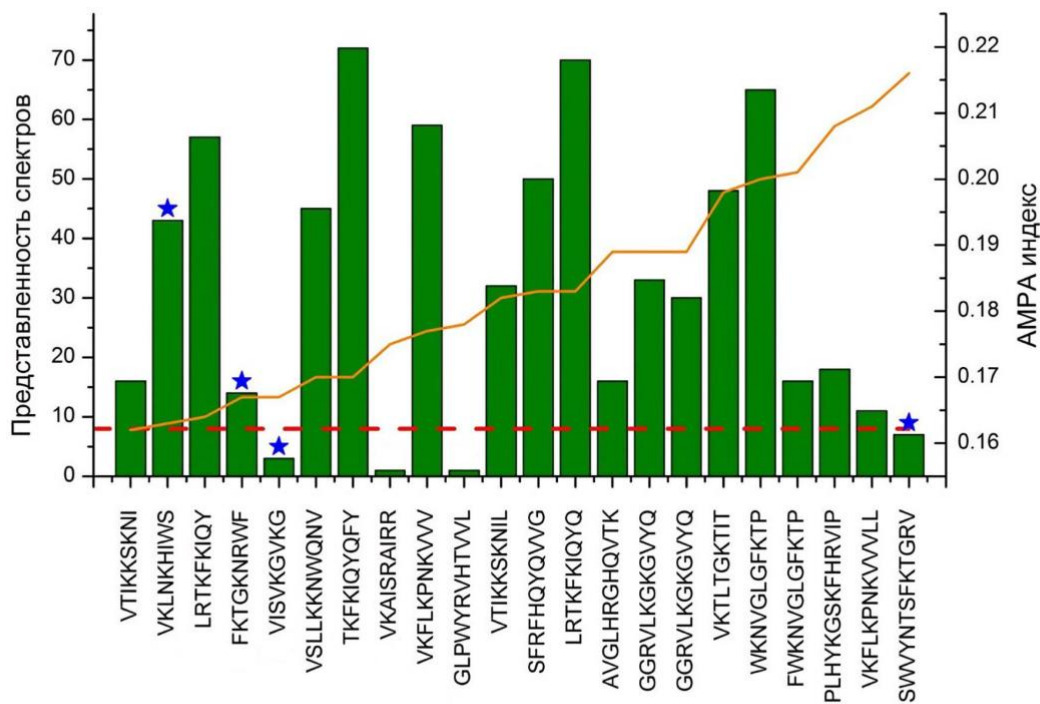


Рисунок 4.59 Диаграмма, показывающая сравнение индекса АМРА и представленности эндогенных пептидов (подсчет количества спектров). Звездочкой отмечены пептиды, отобранные для анализа антимикробной активности *in vitro*. Чем ниже индекс АМРА, тем выше предсказанная антимикробная активность.

В качестве контроля был использован четвертый синтезированный пептид - SWVYNTSFKTGRVF (пеп 4), являющийся фрагментом белка липоксигеназы, который, по данным сервиса AMPA, обладал наименьшей антимикробной активностью и был слабо представлен в образцах (Табл. 6, Рис. 4.59).

Таблица 6. Пептиды из протопластов, отобранные для изучения антимикробной активности *in vitro*

| ИД | Последовательность | Предшественник | Индекс AMPA | Количество спектров |
|------|--------------------|---|-------------|---------------------|
| пеп1 | VKLNKHIWSR | Pp3c10_8090 (рибосомальный белок L31) | 0,163 | 43 |
| пеп2 | FKTGKNRWWF | Pp3c24_20290 (рибосомальный белок L27) | 0,167 | 14 |
| пеп3 | VISVKGVKGR | Pp3c23_4520 (рибосомальный белок L14) | 0,167 | 3 |
| пеп4 | SWVYNTSFKTGRVF | Pp3c1_29300V3.1 (липоксигеназа) | 0,216 | 7 |

Для анализа антимикробной активности было решено использовать фитопатогенные бактерии родов *Clavibacter* и *Xanthomonas*. При этом *Clavibacter michiganensis* pv. *michiganensis* является одним из широко распространенных фитопатогенов и вызывает бактериальный рак у томата [373]. Интересно, что проведенный нами анализ показал, что при инокуляции гаметофоров мха *S.m.* pv. *michiganensis* не наблюдается изменений фенотипа, а симптомы инфекции отсутствовали в течение двух месяцев после инокуляции. В то же время различные штаммы *Xanthomonas* вызывали явно выраженные симптомы инфекции на гаметофорах *P. patens* (Рис. 4.60). Таким образом, дальнейший анализ

антимикробной активности отобранных пептидов проводили *in vitro* методом серийных разведений с использованием фитопатогенных бактерий *Bacillus subtilis* SHgw, *Clavibacter michiganensis* pv. *michiganensis*, *Xanthomonas arboricola* 3004, а также *Escherichia coli* K12. Для анализа использовали синтезированные пептиды в концентрациях 7 или 100 мкМ, оптическую плотность бактериальной суспензии измеряли каждые 24 часа в течение 3 суток.

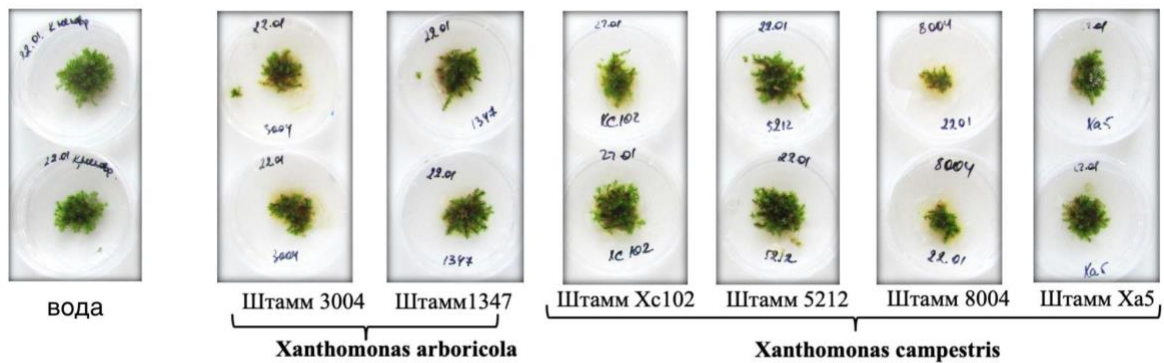


Рисунок 4.60 Фенотип гаметофоров мха, инокулированных бактериями рода *Xanthomonas*

Мы обнаружили, что пептиды пеп 1 и пеп 2 ингибируют рост фитопатогенных бактерий *X. arboricola* 3004 и *S.m.* pv. *michiganensis*, а также грам (-) бактерии *E.coli* K12. Наиболее сильный ингибирующий эффект пеп 1 в концентрации 100 мкМ был отмечен по отношению к бактерии *S.m.* pv. *michiganensis* (Рис. 4.61). При добавлении пеп 1 в концентрации 100 мкМ к *E.coli* K12 снижение роста наблюдали со вторых суток после посева. Пептид 2 в концентрации 7 и 100 мкМ наингибировал рост *X. arboricola* 3004 по сравнению с контролем на вторые, а *S.m.* pv. *michiganensis* – на первые-третьи сутки после посева. Ингибирование роста *E.coli* K12 при добавлении пептида 2 в концентрации 100 мкМ наблюдали на вторые сутки; через трое суток инкубации рост бактерии снизился на 15,2 %. Пептид 3 в наших экспериментах проявлял ингибирующий эффект только в отношении грамположительной бактерии *B. subtilis* SHgw; ингибирование роста

бактерии при добавлении пептида в концентрации 100 мкМ наблюдали уже через сутки после инокуляции питательной среды (Рис. 4.61).

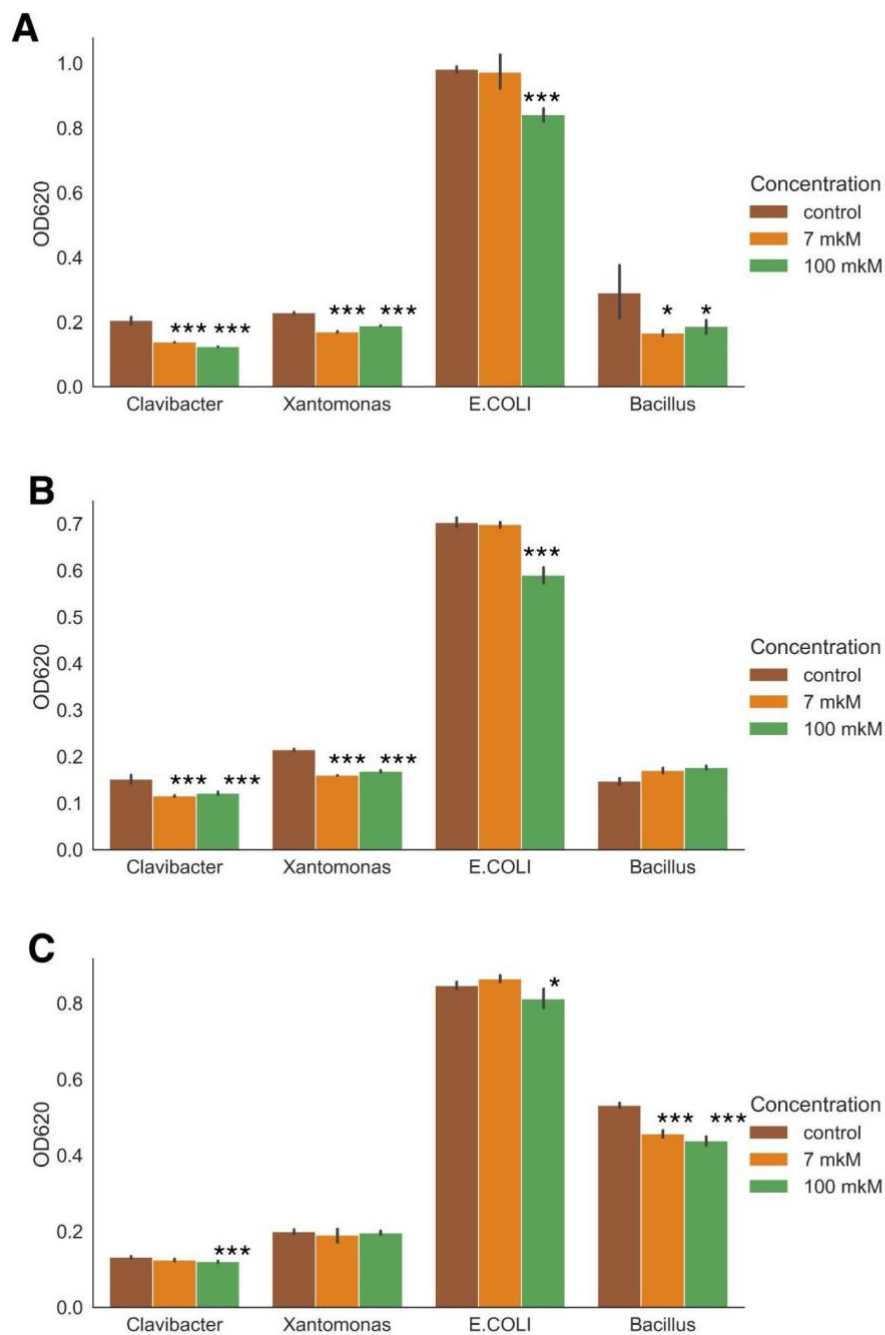


Рисунок 4.61 Анализ антибактериальной активности **А** – пеп 1, **В** – пеп 2, **С** – пеп 3 пептидов. Показана оптическая плотность культуры на 3 сутки инкубации после инокуляции питательной среды содержащей 1 мкл бактерии. Р-значение рассчитывали с помощью непарного t-критерия Стьюдента. (***) $P < 0,001$, (**) $P < 0,01$, (*) $P < 0,05$.

Ожидается, что пептид 4 не оказывал ингибирующего действия на рост изученных бактерий.

Таким образом, в результате анализа мы обнаружили, что три выбранных нами и синтезированных пептида - VKLNKHIWSR (пеп 1), и FKTGKNRFFF (пеп 2) и VISVKGVKGR (пеп 3) - оказывали ингибирующее действие на рост как грамположительных, так и грамотрицательных бактерий. Интересно, что все эти пептиды являлись фрагментами рибосомальных белков и были идентифицированы только в протопластах. Проведенные нами ранее эксперименты показали, что при образовании протопластов мха наблюдается существенное изменение пула эндогенных пептидов. Мы предполагаем, что стрессовые условия могут влиять на пептидом клетки, вызывая деградацию белков. В этом случае повышенный уровень протеолиза функциональных белков может создавать дополнительный барьер для распространения инфекции благодаря антимикробному действию образующихся пептидов.

Интересно, что данные об антимикробной активности рибосомальных белков и их фрагментов впервые были получены у животных. Так, было показано, что рибосомальные белки L35, L3 и L40, выделенные из эпидермальной слизи *Gadus morhua*, обладают антибактериальной активностью в отношении бактерии кишечной палочки *E. coli* [374]. Кроме того, N-концевой фрагмент рибосомального белка S30, выделенный из тканей радужной форели, подавлял рост целого ряда патогенных бактерий [375].

4.3.1.2 Пептиды с антимикробной активностью в секрете *P. patens*

Поскольку внеклеточная среда является первой границей взаимодействия фитопатогенов и растений, мы протестировали антимикробную активность секретомов, обработанных метилжасмонатом, 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 0,4 мМ МЖ (Рис. 4.62). Это позволяет предположить возможную роль эндогенных пептидов как быстро высвобождаемых противомикробных агентов. Известно, что некоторые оксипирины обладают антимикробной активностью, поэтому мы дополнительно протестировали секретомы, обработанные различными концентрациями метилжасмоната (0,05 мМ, 0,4 мМ и 1 мМ; Рис. 4.62). Наиболее выраженный бактериостатический эффект наблюдался при использовании секретома, обработанного 0,05 и 0,4 мМ МЖ, что, скорее, позволяет говорить о регуляторной роли МЖ, а не о действии в качестве антимикробного соединения.

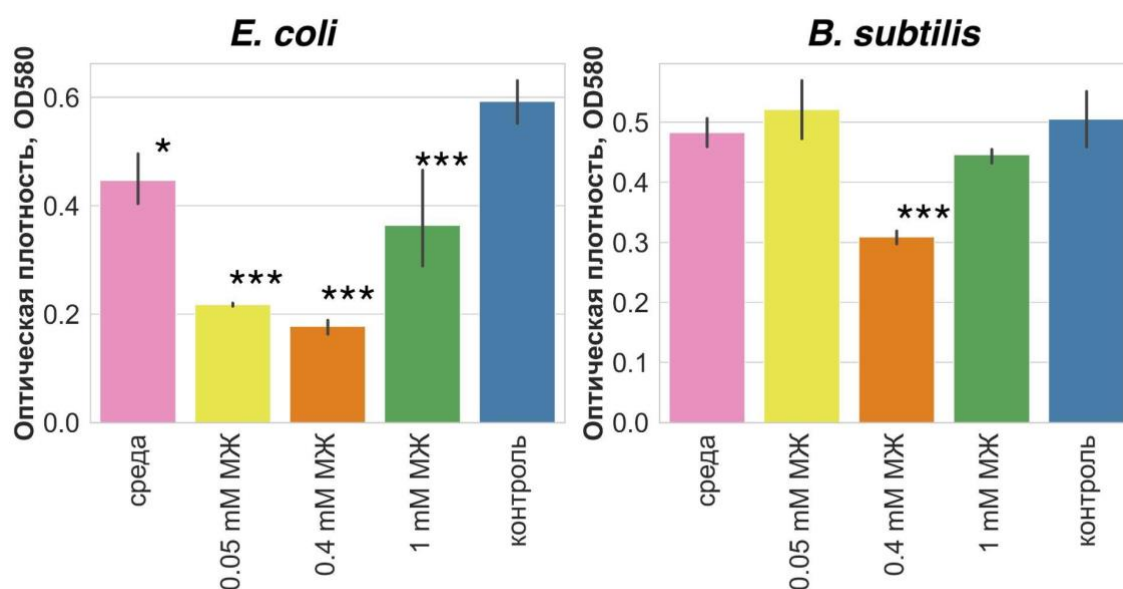


Рисунок 4.62 Анализ антибактериальной активности секретома *P. patens* обработанного разными концентрациями метилжасмоната. Гистограмма показывает оптическую плотность культур *E. coli* и *B. subtilis* после 24-часовой инкубации с контролем, обработанными метилжасмонатом образцами секретома и культуральной 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 -value рассчитывали с помощью теста Тьюки (показано сравнение образцов с контролем).

Затем мы провели дополнительные эксперименты, чтобы определить, могут ли полученные нами результаты быть связаны именно с высвобождением

антимикробных пептидов во время процессов деградации белка. Для этого в ходе обработке метилжасмонатом в культуральную среду добавляли коктейль ингибиторов протеаз, предотвращающий протеолитическую деградацию секретируемых белков. В результате, мы обнаружили значительное снижение бактериостатического эффекта (Рис. 4.63) в образцах, обработанных метилжасмонатом, но с добавлением ингибитора протеолитической активности. В зависимости от концентрации метилжасмоната, бактериостатический эффект полностью нивелировался. Это свидетельствует о возможной роли пептидов, образующихся при расщеплении белков, в качестве быстро высвобождаемых антимикробных агентов.

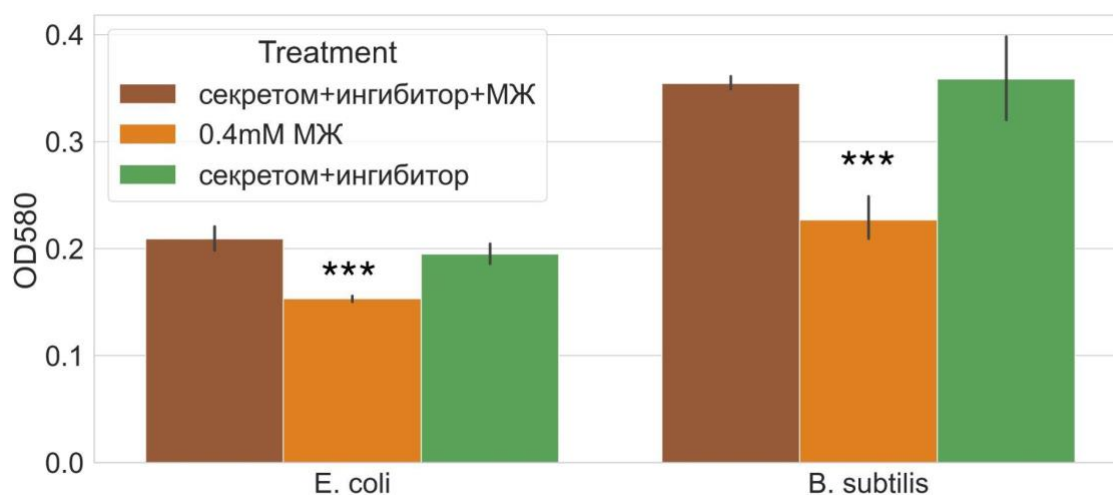


Рисунок 4.63 Анализ антибактериальной активности секрета *P. patens*. Гистограмма показывает оптическую плотность культур *E. coli* и *B. subtilis* после 24-часовой инкубации с секретомом *P. patens*, обработанного 0.4 мМ метилжасмонатом, ингибитором протеолитической активности и 0.4 мМ метилжасмонатом совместно с ингибитором. Столбцы ($M \pm SD$) представляют результаты трех независимых экспериментов, проведенных в трех повторностях. *** $P < 0.0001$. Значение P рассчитывали с помощью теста Тьюки (показано сравнение образцов, обработанных метилжасмонатом с образцами с добавлением ингибитора).

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 секретоме могут обладать антимикробными свойствами на основании предсказаний всех трех сервисов. Эти пептиды являются продуктами деградации различных групп белковых предшественников, включая Pp3c2_24160 (глицеральдегид-3-фосфатдегидрогеназа), Pp3c14_17710 (кукумизин) и другие. Мы также обнаружили, что некоторые пептиды, такие как IHNVGGAGDVVSVK и TDINLDLGDGKQG, образующиеся при деградации α -экспансина, также обладают антимикробной активностью. Представленность пептида с самым высоким антимикробным потенциалом INIINAPLQGFKIA (вырос в $\log_2 = 1,3$) увеличивалось при обработке МЖ. Мы также обнаружили увеличение количества потенциальных антимикробных пептидов в тканях протонемы, обработанных МЖ, по сравнению с контролем. Интересно, что представленность 37 (21%) пептидов с предсказанной антимикробной активностью в протонеме повышалось при обработке МЖ.

Для проверки биологической активности предсказанных антимикробных пептидов мы отобрали восемь (пять из клеточного пептидома и три из секретомы) на основании следующих критериев: физико-химические свойства, высокий предполагаемый антимикробный потенциал и изменение содержания при обработке МЖ по сравнению с контролем (Табл. 7).

Таблица 7. Список синтетических пептидов, используемых для проверки антимикробной активности

| № | Пептид | Предшественник | Ген | Индекс AMPpred | FC, log2 |
|------------------------|------------------------------------|--|--------------|----------------|----------|
| Внутриклеточные | | | | | |
| 1 | LVQIGTKIVGVGRNYAA H | Fumarylacetoacetate hydrolase domain- containing protein 1 | Pp3c9_26130 | 0.832 | 4.2 |
| 2 | AAQGQKIENTKLAGAAG DILSGLAAYGKLD | Predicted | Pp3c22_17930 | 0.884 | 0.9 |

| | | | | | |
|-----------------|-----------------------|---|--------------|-------|----------------------------------|
| 3 | VAAVAPKFATLKPLG | Chloroplast chaperonin 21 | Pp3c19_4270 | 0.811 | 0.7 |
| 4 | KIKVAINGFGRIG | Glyceraldehyde-3-phosphate dehydrogenase | Pp3c2_24160 | 0.998 | -1.9 |
| 5 | IVPTSTGAAKAVALVLPN LK | Glyceraldehyde-3-phosphate dehydrogenase | Pp3c2_24160 | 0.597 | 3.1 |
| Секретом | | | | | |
| 6 | INIINAPLQGFKIA | Predicted | Pp3c14_22870 | 0.941 | 1.3 (secretome) 4.1 (cell) |
| 7 | TDINLDLGDGKQG | Expa6 gene for putative alpha expansin protein family EXPA6 | Pp3c8_870 | 0.513 | 1.6 |
| 8 | VVDLLAPYRRGGKIG | Predicted | PhpapaCp032 | 0.507 | -3.2 |

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 секретома - составляла 64 и 16 мкг/мл соответственно (Табл. 8).

Таблица 8. Значения минимальной ингибирующей концентрации для двух пептидов

| Пептид | <i>E. coli</i> | | <i>B. subtilis</i> | |
|--------------------|----------------|-----------------|--------------------|-----------------|
| | МИК, мкг/мл | % ингибирования | МИК, мкг/мл | % ингибирования |
| LVQIGTKIVGVGRNYAAH | 64 | >90 | 64 | >90 |
| INIINAPLQGFKIA | 16 | >90 | 32 | >90 |

В качестве положительного контроля использовали известный противомикробный пептид мелиттин. Мы получили схожие МИК для INIINAPLQGFKIA и мелиттина (16 и 8 мкг/мл соответственно), что позволяет предположить возможную роль идентифицированного пептида в качестве антимикробного агента. Кроме того, клеточный пептид KIKVAINGFGRIG (фрагмент глицеральдегид-3-фосфатдегидрогеназы (ГАФД)) подавлял рост бактерий в концентрации 128 мкг/мл на 1-й день после инокуляции.

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.

4.3.2 Функции пептидов, кодируемых длнкРНК у растений

Согласно полученным нами результатам, пептидомы растений, помимо продуктов деградации белков, содержат сотни продуктов трансляции кОРС. Учитывая, что многие из транслирующихся кОРС были высококонсервативны, можно предположить, что соответствующие пептиды являются функциональными. Однако функции большинства консервативных и неконсервативных пептидов/микробелков, кодируемых короткими рамками считывания, у растений неизвестны. При этом, особый интерес представляют пептиды, кодируемые кОРС, расположенными на многочисленных транскриптах длнкРНК. В этом контексте нам было бы важно определить биологические функции таких пептидов.

Мы показали, что некоторые части “длнкРНК-кОРС” демонстрируют тканеспецифические паттерны транскрипции и трансляции, в то время как другие экспрессируются во всех типах клеток мхов. Можно предположить, что стабильно экспрессируемые консервативные кОРС могут продуцировать пептиды, играющие фундаментальную роль в различных клеточных процессах. Чтобы проверить эту

гипотезу, мы исследовали влияние сверхэкспрессии и нокаута четырех “днкРНК-кОРС”, которые были обозначены как: *PSEP25*, *PSEP18*, *PSEP1* и *PSEP3* (Табл. 9).

Таблица 9. Список “днкРНК-кОРС”, отобранных для функционального анализа

| ИД | Последовательность | Размер |
|---------------|--|---------|
| <i>PSEP25</i> | MVQSKQGLSLLKFIPKVIRPQTSDVSSAVLWGTAAACGL WLVQPFDWIKEQITGPKEESK | 61 а.о. |
| <i>PSEP18</i> | MQAFTDTQGYSSFNRPATTAATTPPEVVGEFGGKGWRP SS | 40 а.о. |
| <i>PSEP1</i> | MVQPLLARLASAAEFVALPGAILVAYFSTSRSTEPKRDH RK | 41 а.о. |
| <i>PSEP3</i> | MVHQDNSGSLRSFNHPNPPNPPNPPSNPPVVRNPSSG RTPHPYPPPHNYNGYPN | 57 а.о. |

Трансляция этих четырех “днкРНК-кОРС” была надежно подтверждена методом масс-спектрометрического анализа пептидомных наборов данных с помощью нескольких уникальных пептидов, а их стартовые кодоны находятся в «сильном» и «среднем» контексте Козак. Мы получили несколько независимых мутантных линий для каждой из этих “днкРНК-кОРС” и проанализировали изменения в фенотипе, а также влияние сверхэкспрессии/нокаут пептида на протеом клеток.

4.3.2.1 Пептид *PSEP25*

Для изучения влияния на фенотип нокаута/сверхэкспрессии данной кОРС были получены мутантные линии *P. patens*. Для сверхэкспрессии соответствующего микробелка, обозначенного *PSEP25*, была создана генетическая конструкция, в которой кодирующая последовательность регулировалась 35S промотором (Рис. 4.64А). Для получения мутантных линий *P. patens*, нокаутных по данной короткой рамке считывания, использовали технологию CRISPR/Cas9 (Рис. 4.64В).

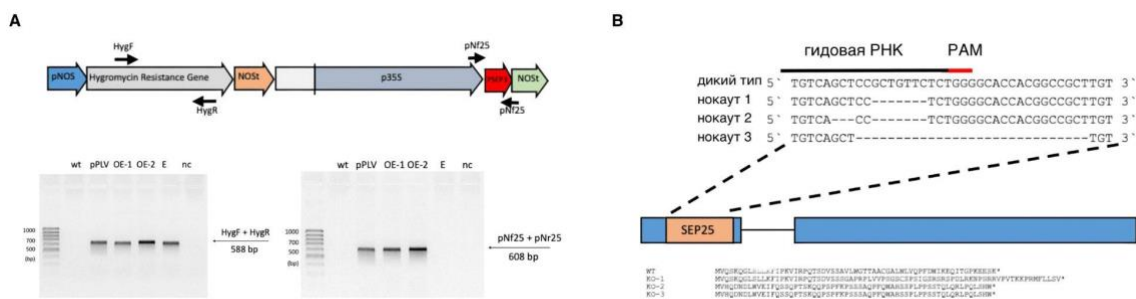


Рисунок 4.64 **A** - схема кассеты экспрессии, используемой для получения мутантных линий со сверхэкспрессией микробелка *PSEP25*. Указаны промотор и терминатор NOS (pNOS и NOSr), промотор для микробелка (p35S), специфические праймеры (pNf25, pNr25, HygF и HygR), последовательность кОРС (показана красным). Показаны результаты ПЦР-анализа интеграции экспрессионной кассеты. Длина фрагментов: 588 п.н. для пары HygF и HygR. Wt - дикий тип, pPLV - плаزمида pPLV-Hpa-25FR-SacI, OE1, OE2, - линии со сверхэкспрессией микробелка, E - контроль, трансформированный пустой плазмидой, nc - отрицательный контроль. **B** - Схема нокаута короткой рамки считывания с помощью технологии CRISPR/Cas9. Показана последовательность гидовой РНК и сайт PAM. Показаны последовательности нокаутов в сравнении с диким типом. Секвенирование по Сэнгеру подтвердило делецию семи, десяти и двадцати девяти нуклеотидов у линий дикого типа и линий с нокаутом микробелка.

Для дальнейшего анализа фенотипов были использованы 3 нокаутных и 2 линии *P. patens* со сверхэкспрессией *PSEP25*. Нокаут данного микробелка, кодируемого консервативной “длинкРНК-кОРС”, приводил к снижению скорости роста и изменению архитектуры нитей протонемы на среде без глюкозы, но с добавлением тартрата аммония (**Рис. 4.65A**). 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.

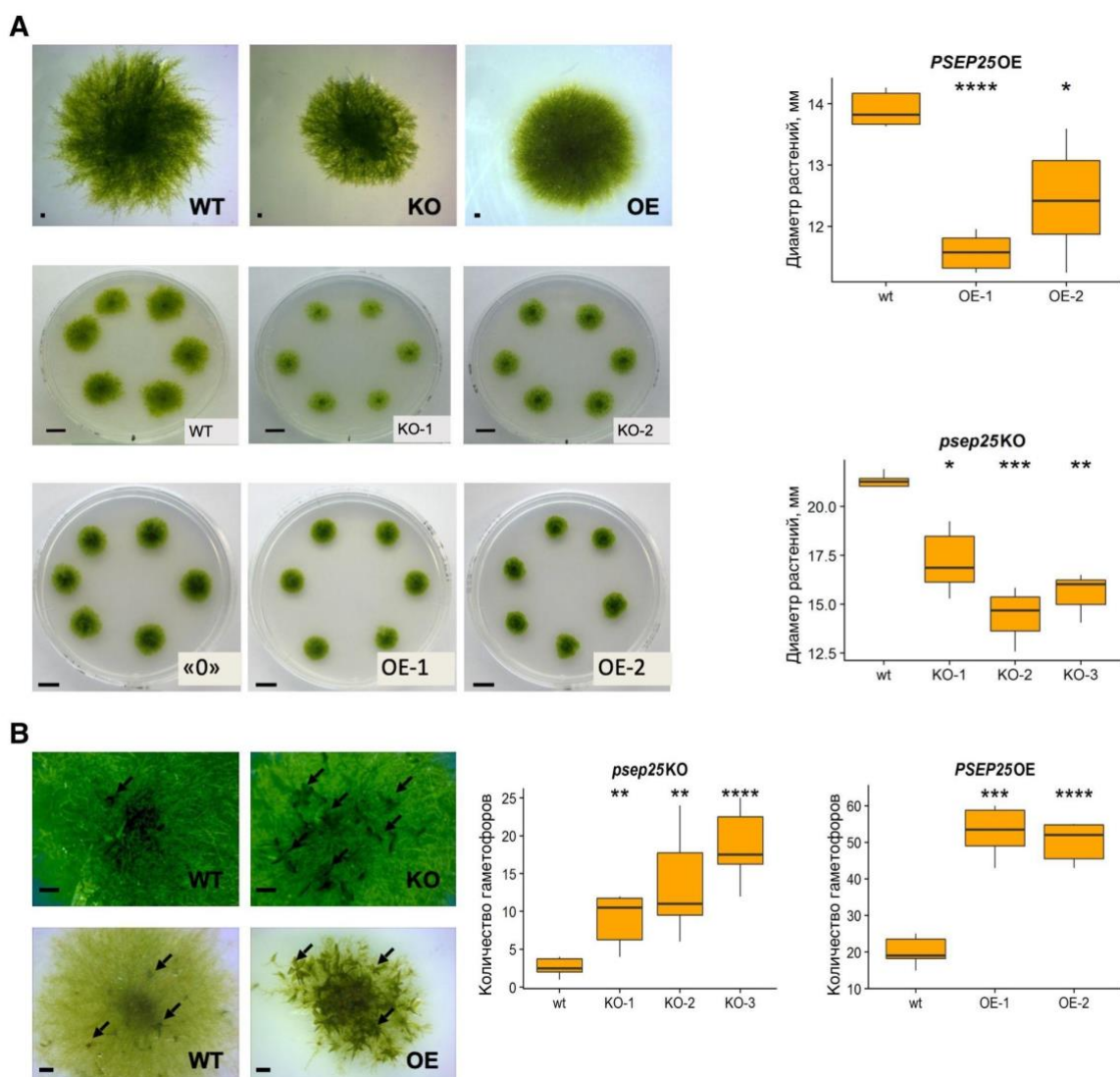


Рисунок 4.65 Морфология растений *P. patens* дикого типа (WT) и мутантных линий по микробелку *PSEP25*. **A** - Фенотипы нокаутных линий (*psep25* KO) и линий со сверхэкспрессией (*PSEP25* OE) микробелка, выращенные на среде BCDAT. Графики показывают различия в диаметре растений мха дикого типа и мутантных линий; «0» - линий со сверхэкспрессией пустой плазмиды. **B** - Количество гаметофоров у дикого типа и мутантных линий. Стрелками показаны гаметофоров. 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. (***) $P < 0,0001$, (***) $P < 0,001$, (**) $P < 0,01$, (*) $P < 0,05$.

В отличие от линий *P. patens* со сверхэкспрессией *PSEP25*, у нокаутов по данному микробелку (*psep25* KO) наблюдалось значительное увеличение числа листовых побегов на среде без глюкозы, но с добавлением тартрата аммония, который

обычно снижает количество гаметофоров (**Рис. 4.65В**). В свою очередь, мутантные линии со сверхэкспрессией *PSEP25* также показали увеличение количества гаметофоров по сравнению с диким типом, но на твердой среде без тартрата аммония (**Рис. 4.65В**). Таким образом, любые изменения в экспрессии *PSEP25* (как сверхэкспрессия, так и нокаут) стимулировали образование зачатков гаметофоров. Следует отметить, что наблюдаемые эффекты зависели от стимуляции/ингибирования роста каулонемных нитей.

Мы далее оценили влияние данного микробелка на протеом клетки, проведя сравнительный количественный анализ с изобарными метками iTRAQ растений дикого типа и мутантных линий. У нокаутов по *PSEP25* мы идентифицировали всего 14 дифференциально экспрессирующихся белков, представленность 9 из которых была выше у мутантов, чем у дикого типа. Интересно, что 4 из 9 белков, представленность которых росла в нокаутных линиях, относились к рибосомальным белкам (A9TRK1 - рибосомальный белок L21, A9RZI8 - 50S рибосомальный белок L30e-like, A9RUW5 - 60S рибосомальный белок L6E, A9TVZ2 - дифитинсинтаза), и по данным базы String DB, объединялись в один функциональный кластер. При этом, мы не обнаружили существенных изменений в протеомах линий со сверхэкспрессией *PSEP25*. 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*.

4.3.2.2 Пептид *PSEP18*

Далее мы провели анализ еще одного пептида, размером 40 а.о., который кодируется неконсервативной “длнкРНК-кОРС”. Данный пептид относится к группе неконсервативных и, по нашим данным, является примером транслирующейся кОРС, которая не находится под стабилизирующим отбором. Тем не менее, пептид имеет предсказанный мотив “consensus disorder prediction” (внутренне неупорядоченные области), транскрипция соответствующей длнкРНК

подтверждена нашими данными нанопорового секвенирования, а масс-спектрометрический анализ надежно показал трансляцию соответствующей кОПС. Данный пептид был обозначен *PSEP18* и для дальнейшего анализа были получены 2 мутантные линии со сверхэкспрессией данной кОПС и 2 нокаутные линии (**Рис. 4.66**).

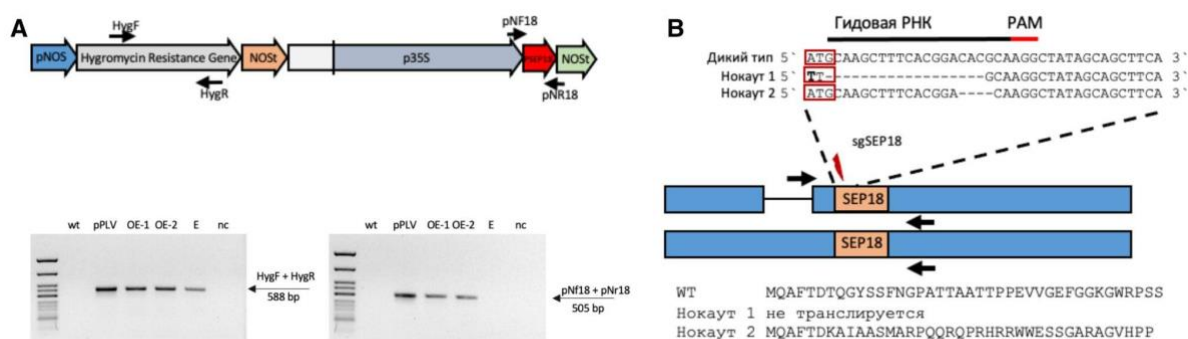


Рисунок 4.66 **А** - схема кассеты экспрессии, используемой для получения мутантных линий со сверхэкспрессией микробелка *PSEP18*. Указаны промотор и терминатор NOS (pNOS и NOS), промотор для микробелка (p35S), специфические праймеры (pNf25, pNr25, HygF и HygR), последовательность кОПС (показана красным). Показаны результаты ПЦР-анализа интеграции экспрессионной кассеты. Длина фрагментов: 588 п.н. для пары HygF и HygR. Wt - дикий тип, pPLV - плаزمида pPLV-Нра-25FR-SacI, OE1, OE2, - линии со сверхэкспрессией микробелка, E - контроль, трансформированный пустой плазмидой, nc - отрицательный контроль. **В** - Схема нокаута короткой рамки считывания с помощью технологии CRISPR/Cas9. Показана последовательность гидовой РНК и сайт PAM. Показаны нуклеотидные и аминокислотные последовательности нокаутов в сравнении с диким типом.

Нокаут данного пептида (*psep18* KO) приводил к небольшому уменьшению диаметра растений мха на среде с глюкозой и без тартрата аммония (**Рис. 4.67**). Однако только нокаутная линия (КО-1) с делецией стартового кодона продемонстрировала значительное снижение скорости роста по сравнению с диким типом (**Рис. 4.67**). Принимая во внимание, что кОПС могут запускать механизм нонсенс-опосредованной деградации РНК [328], этот факт требует более детального изучения. В этом случае, функциональной может оказаться длнкРНК, а трансляция *PSEP18* будет, скорее, играть регуляторную роль.

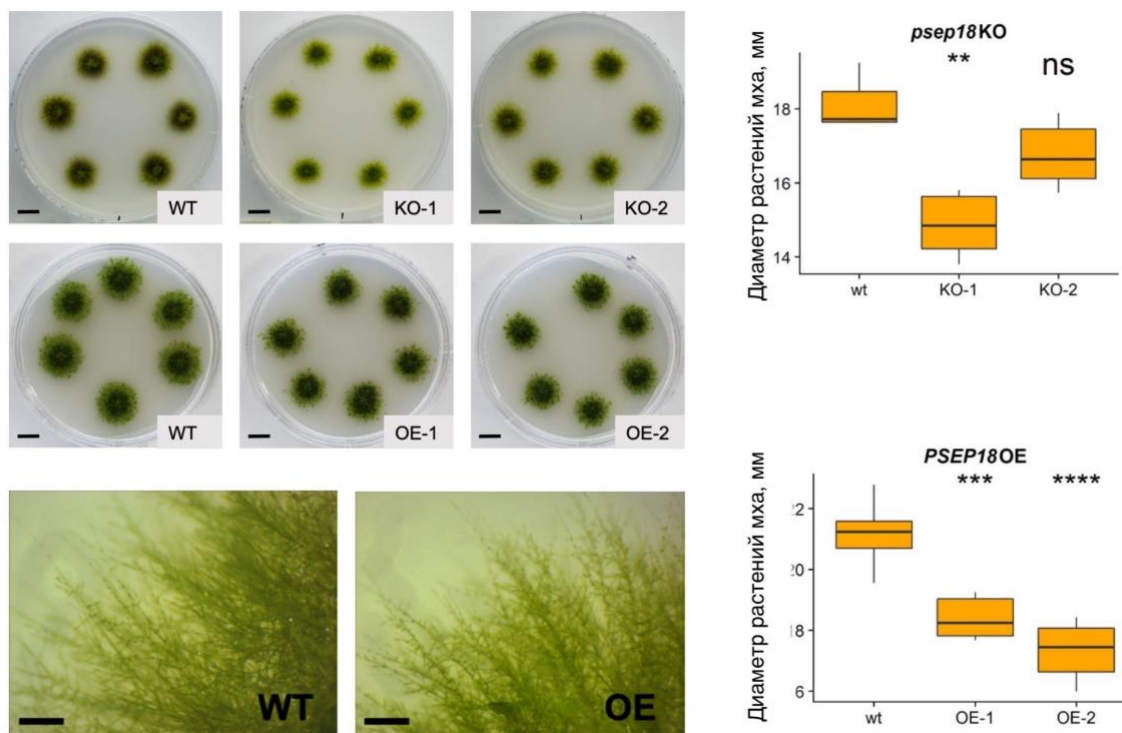


Рисунок 4.67 Морфология растений *P. patens* дикого типа (WT) и мутантных линий по микробелку PSEP18. Показаны фенотипы нокаутных линий (KO-1 и KO-2) и линий со сверхэкспрессией (OE-1 и OE-2) микробелка, выращенные на среде BCD с добавлением 0,5% глюкозы. Графики показывают различия в диаметре растений мха дикого типа и мутантных линий. Р-значение рассчитывали с помощью непарного t-критерия Стьюдента. (***) $P < 0,0001$, (***) $P < 0,001$, (**) $P < 0,01$, (*) $P < 0,05$.

Линии со сверхэкспрессией *PSEP18* OE продемонстрировали значительное снижение скорости роста по сравнению с диким типом на среде с глюкозой (**Рис. 4.67**). Однако как у мутантных линий со сверхэкспрессией пептида, так и у нокаутов мы не наблюдали каких-либо изменений в архитектуре протонемы, числе листовых побегов или ветвлении филаментов.

Тем не менее, поскольку мы наблюдали изменения фенотипа мутантных по *PSEP18* линий *P. patens*, мы провели количественный сравнительный протеомный анализ с использованием меток iTRAQ. У нокаутной линии KO-1 значимо изменилось 8 белковых групп. Представленность четырех белковых групп возросла у нокаутных растений (A0A2K1IJ08 - белок хлоридного канала, A0A2K1IE5 -

рибосомальный белок L15, A0A2K1J0Z5 - неохарактеризованный белок, A0A2K1L2D7 - SAM-зависимая метилтрансфераза), тогда как представленность остальных белков падала (A0A2K1JC42 - белок фотосистемы I psaG/psaK, A9SF97 - белок с цитохром с оксидазной активностью, A9SGK7 - Nucleoid-associated protein YbaB-like domain superfamily, A0A2K1JS08 - Nucleotide-binding alpha-beta plait domain superfamily). SAM-зависимые метилтрансферазы - это большая группа ферментов, осуществляющая перенос метильных групп на различные субстраты, которые, в свою очередь, участвуют в целом ряде процессов, включая, например, регуляцию экспрессии генов, защиту от вирусов и синтез этилена. Однако мы не обнаружили существенных изменений в протеомах линий со сверхэкспрессией. Таким образом, полученные результаты позволяют предположить, что эволюционная консервативность может быть маркером функциональности “длнкРНК-кОПС”.

4.3.2.3 Пептид PSEPI

Данный 41 а.о. пептид кодируется транскриптом CNT2064811, который был предсказан как длинная некодирующая РНК в базе данных SANGATAdb [235], но пропущен в аннотации Phytozome V12 [225]. Мы получили 2 нокаутные линии и 3 линии со сверхэкспрессией для 41 а.о. пептида, кодируемого длинной некодирующей РНК для функционального анализа (**Рис. 4.68**).

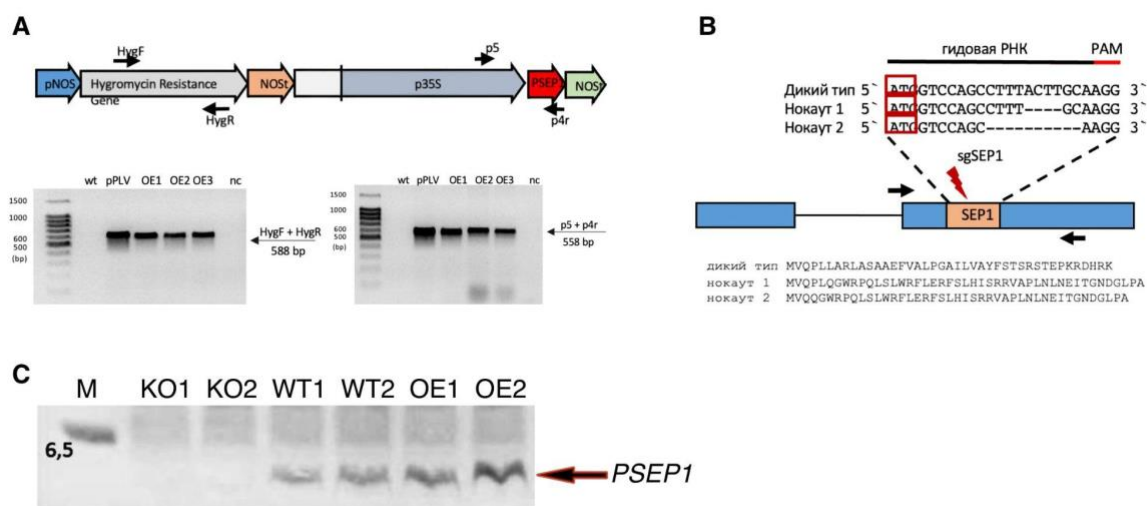


Рисунок 4.68 **A** - схема кассеты экспрессии, используемой для получения мутантных линий со сверхэкспрессией микробелка *PSEPI*. Указаны промотор и терминатор NOS (pNOS и NOS_t), промотор для микробелка (p35S), специфические праймеры (pNf25, pNr25, HygF и HygR), последовательность кОПС (показана красным). Показаны результаты ПЦР-анализа интеграции экспрессионной кассеты. Длина фрагментов: 588 п.н. для пары HygF и HygR. wt- дикий тип, pPLV - плаزمида pPLV-Нра-25FR-SacI, OE1, OE2, OE3 - линии со сверхэкспрессией микробелка, nc - отрицательный контроль. **B** - Схема нокаута короткой рамки считывания с помощью технологии CRISPR/Cas9. Показана последовательность гидовой РНК и сайт PAM. Показаны нуклеотидные и аминокислотные последовательности нокаутов в сравнении с диким типом. **C** - Детекция пептида в растения дикого типа (WT1, WT2), нокаутных линиях (KO1, KO2) и линиях со сверхэкспрессией (OE1, OE2), используя иммуноблот.

Известно, что протонема мхов состоит из двух морфологически различимых типов клеток: хлоронемы и каулонемы. Нити хлоронемы образуются первыми при прорастании спор, а каулонемные клетки развиваются позже [376]. Клетки каулонемы характеризуются более высокой скоростью роста, меньшим количеством хлоропластов и косыми клеточными перегородками [377]. 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. 4.69A**). Также наблюдалась небольшая, но значительная разница в скорости роста между линиями дикого типа и мутантными линиями *psep1*KO, выращенными на твердой среде и в жидкой

культуре без глюкозы (**Рис. 4.69В**). Соотношение каулонемных и хлоронемных клеток было значительно выше в линии со сверхэкспрессией *PSEP1* 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*. Мы также обнаружили, что быстрый рост линий со сверхэкспрессией *PSEP1* сопровождался более быстрой гибелью клеток протонемы.

Учитывая консервативность данного пептида, наличие предсказанного трансмембранного домена, а также разнообразные фенотипические эффекты, которые демонстрировали мутантные линии, мы назвали данный пептид FAMOSS (“FAst-growing MOSS”) и более подробно изучили его функции (см. ниже главу “Пептид FAMOSS - регулятор полярного роста у растений”).

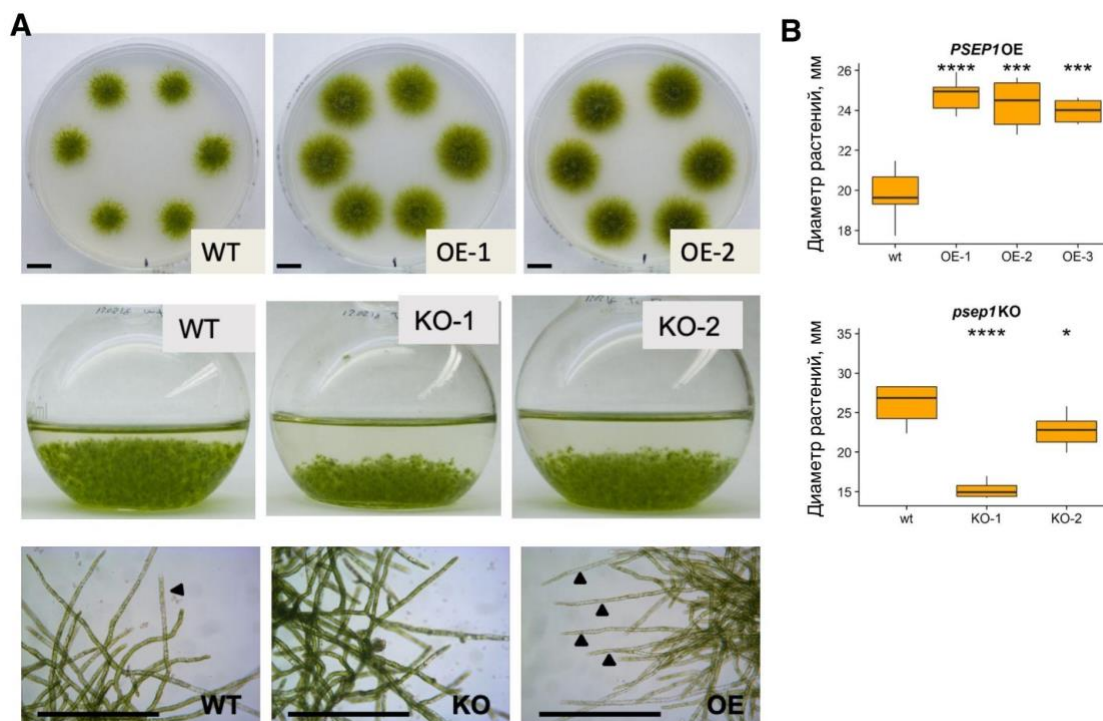


Рисунок 4.69 Морфология растений *P. patens* дикого типа (WT) и мутантных линий по микробелку *PSEP1*. Показаны фенотипы нокаутных линий (KO-1 и KO-2), выращенных в жидкой питательной среде, и линий со сверхэкспрессией (OE-1 и OE-2) микробелка, выращенные на среде BCD. Показано увеличение количества 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. (***) $P < 0,0001$, (***) $P < 0,001$, (**) $P < 0,01$, (*) $P < 0,05$.

4.3.2.4 Пенмуд PSEP3

Пептиды и белки, богатые пролином, играют важную роль в каскадах передачи сигналов, развитии растений и устойчивости к стрессу [378,379]. Наш MS-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. 4.70А). Вследствие наличия региона низкой сложности, только поиск с использованием алгоритма TBLASTN с отключенной по таким участкам фильтрацией (SEG="no") позволил выявить ортологи *PSEP3* у 27 видов мхов и одного вида печеночника (Рис. 4.70В).

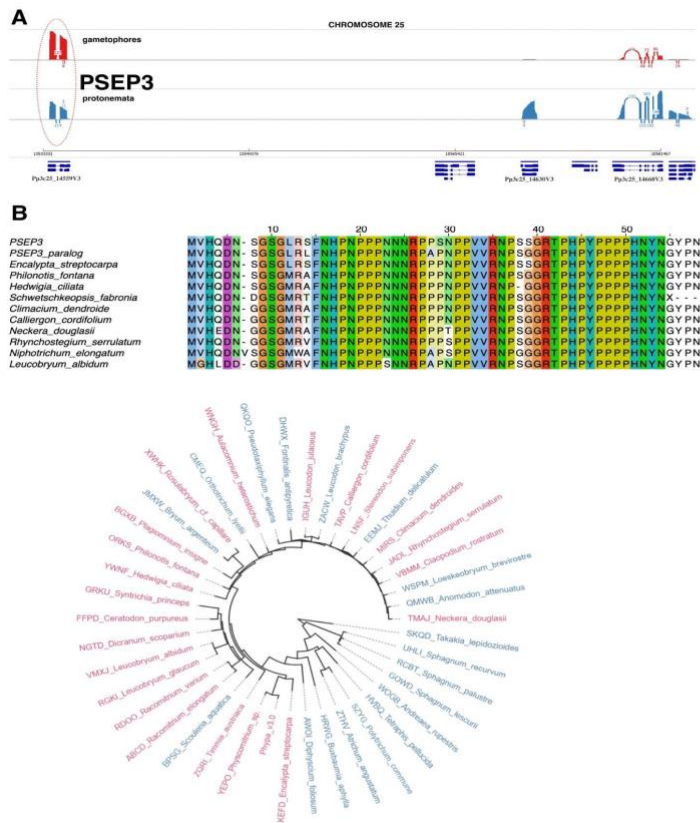


Рисунок 4.70 А - график Сашими, показывающий уровень транскрипции *PSEP3* и окружающего региона на хромосоме 25; данные получены на основе нанопорового секвенирования. В -

множественное попарное выравнивание выбранных ортологов *PSEP3*, а также его паралога. Филогенетическое дерево видов мхов, построено по данным проекта OneKP. Красным цветом показаны виды с идентифицированными ортологами *PSEP3*.

Для дальнейшего функционального анализа мы получили две линии с нокаутом данного пептида (**Рис. 4.71А**), однако получение линий со сверхэкспрессией было затруднено из-за сильного ингибирования роста и гибели клеток в мутантных линиях. Для того, чтобы проверить влияние сверхэкспрессии микробелка на фенотип *P. patens*, мы создали мутантные линии, в которых кодирующая кОПС находилась под регуляцией индуцибельного промотора (система индукции β -эстрадиолом [218]; **Рис. 4.71В**). Индукция экспрессии с помощью добавления эстрадиола обеспечивается за счет белка-активатора транскрипции - XVE, который связывается с LexA оператором и индуцирует транскрипцию с Pm35S промотора нижележащей рамки.

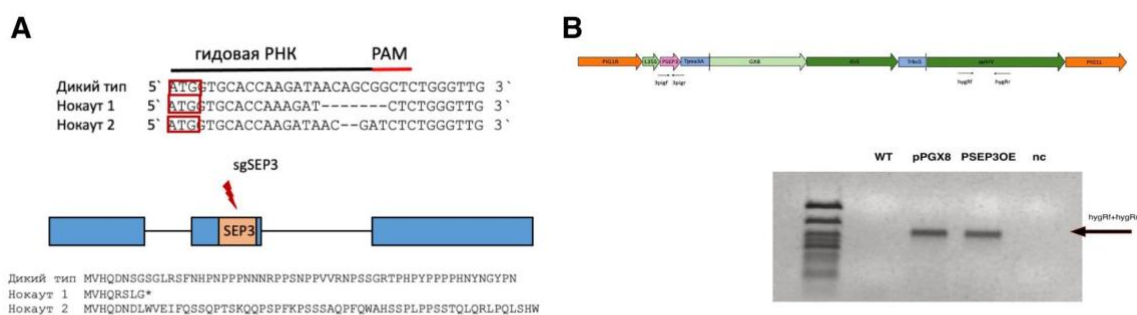


Рисунок 4.71 А - Схема нокаута короткой рамки считывания с помощью технологии CRISPR/Cas9. Показана последовательность гидовой РНК и сайт PAM. Показаны нуклеотидные и аминокислотные последовательности нокаутов в сравнении с диким типом. **В** - Схема генетической конструкции (pPGX8_PSEP3), использованной для получения мутантных линий со сверхэкспрессией *PSEP3*. PIG1R и PIG1L - последовательности для гомологичной рекомбинации с нейтральным локусом PIG1 у *P. patens*; L35S - восемь копий LexA слитых с промотором 35S вируса мозаики цветной капусты; PSEP3 - последовательность пептида *PSEP3*; Треа3А — терминатор; GX8 - промотор GX; показаны активатор VP16 и домен рецептора эстрогена (XVE); терминатор TrbcS, кассета экспрессии гигромицинофосфотрансферазы (aphIV). ПЦР-анализ

интеграции экспрессионной кассеты. Длина фрагментов - 744 п.н. для пары праймеров *hygRf* и *hygRr*; нк-отрицательный контроль.

У линий с нокаутом по *PSEP3* (*psep3* KO) наблюдалось снижение скорости роста и изменение ветвления протонемных филаментов (**Рис. 4.72А**). 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. 4.72В**), которые мы детектировали с помощью красителя флуоресцеин диацетат через 48 часов после добавления эстрадиола (**Рис. 4.72С**). Таким образом, невозможность получения линий с неконтролируемой сверхэкспрессией данного микробелка связана с повышенной гибелью клеток *P. patens*. Для того, чтобы разобраться в причинах такой реакции клеток на повышенную экспрессию *PSEP3*, мы далее измерили уровень активных форм кислорода (АФК) при индукции микробелка с помощью красителя DCFH-DA, который не проявляет специфичности к какому то виду АФК, но позволяет обнаружить изменение их количества. В результате проведенного анализа, мы обнаружили значительное увеличение уровня АФК в клетках протонемы мутантной линии со сверхэкспрессией *PSEP3* через 24 часа после индукции (ANOVA, $P < 0,001$; **Рис. 4.72D**).

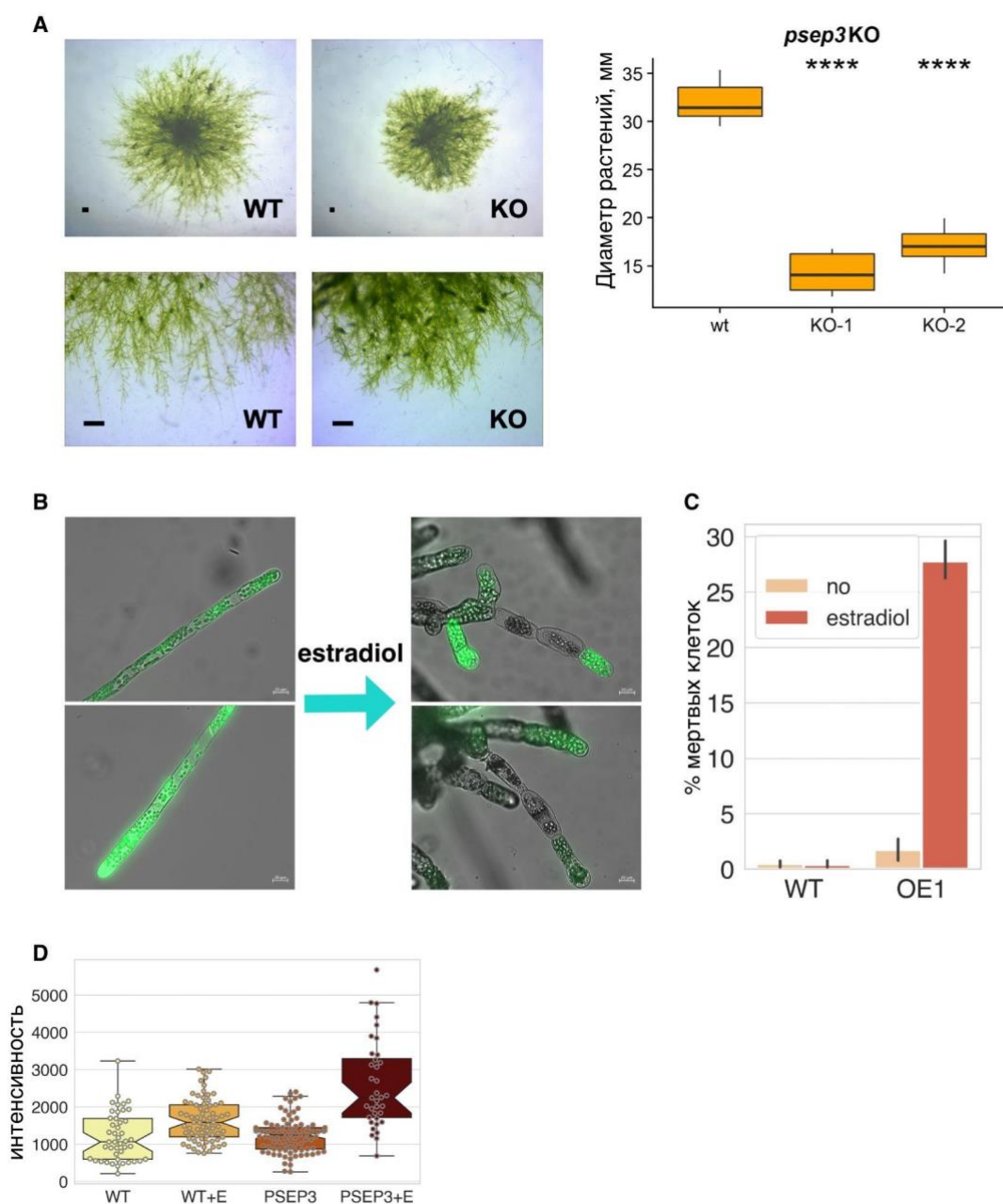


Рисунок 4.72 **A** - Фенотипы линий *psep3* KO и *PSEP3* OE, выращенных на среде BCD: “WT” - дикий тип; “KO” - нокаут *PSEP3*. Масштабная линейка: 0,5 мм. На графике показан диаметр растений мха дикого типа и с нокаутом *PSEP3*. *P*-значение рассчитывали с помощью непарного *t*-критерия Стьюдента. (****) $P < 0,0001$, (***) $P < 0,001$, (*) $P < 0,05$. **B**, **C** - влияние индукции сверхэкспрессии *PSEP3* на жизнеспособность клеток, измеренное с помощью красителя флуоресцеин диацетат (FDA). Экспрессию *PSEP3* индуцировали обработкой эстрадиолом в жидкой культуре; WT - растения дикого типа, OE1 - линия со сверхэкспрессией *PSEP3*. **D** - 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).

Далее мы использовали количественный сравнительный протеомный анализ с метками iTRAQ, для оценки различий протеомов растений дикого типа и мутантных линий. После индукции экспрессии *PSEP3*, мы идентифицировали 167 белковых групп, представленность которых значимо менялась по сравнению с растениями дикого типа (Fold Change > 1,2, $P < 0,01$, **Рис. 4.73А**). Среди белков, представленность которых наиболее существенно выросла в мутантной линии, мы идентифицировали светособирающие хлорофилл а/б-связывающие белки *Lhcb1*, *Lhcb2* и *Lhcb5*. Также, анализ терминов GO показал, что белки, растущие при сверхэкспрессии *PSEP3*, были обогащены функциями «фотосинтез» (GO:0015979, $P < 10^{-15}$) и «генерация метаболитов-предшественников и энергии» (GO:0006091, $P < 10^{-5}$). В группу белков, представленность которых снижалась при сверхэкспрессии *PSEP3*, входили метакаспаза-4 (Pr3c2_20840V3) и ксилоглюкановая эндо-трансгликозилаза (Pr3c25_4050V3). Таким образом, сверхэкспрессия *PSEP3* приводила к существенным изменениям в протеоме клеток *P. patens* и индукции гибели клеток. Следует отметить возможную связь между индукцией АФК и увеличением экспрессии генов фотосинтетического аппарата хлоропластов.

Влияние нокаута *PSEP3* на протеом протонемы было менее выраженным. Мы идентифицировали 56 дифференциально экспрессируемых белковых групп (Fold Change > 1,2, $P < 0,01$; **Рис. 4.73В**). Анализ терминов GO показал обогащение термином «оксидоредуктазная активность» (GO:0016491, $P < 0,0001$). Среди наиболее меняющихся белков были обнаружены «тиоредоксин х» и «тиоредоксин m (митохондриального) типа». Среди белков, представленность которых значительно выросла, были рибосомный белок 60s (RP, Pr3c20_19190V3) и фосфоглюкомутаза (PGM, Pr3c16_20760V3).

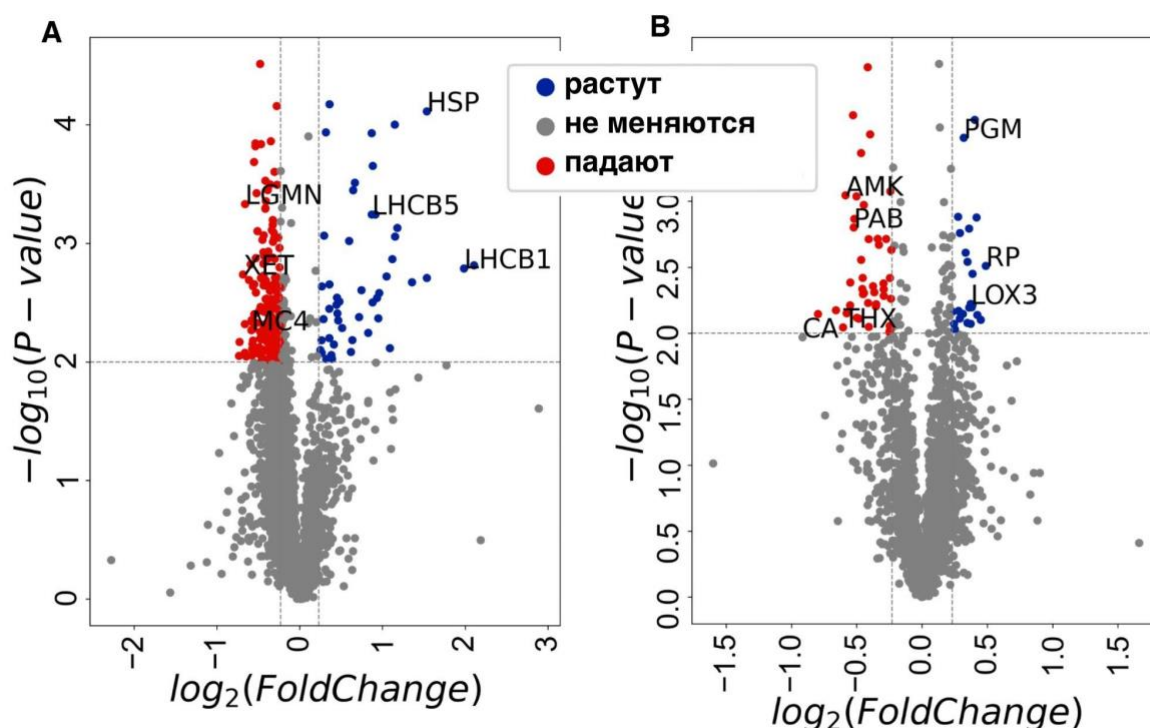


Рисунок 4.73 А - График “volcano plot” всех белков, количественно определенных во время протеомного анализа в линии со сверхэкспрессией *PSEP3*. В - График “volcano plot” всех белков, количественно определенных во время протеомного анализа мутантной линии, нокаутной по *PSEP3*. Белки, содержание которых значительно изменилось ($P < 0.01$), показаны цветом. Синие точки указывают на белки с повышенной экспрессией, а красные точки указывают на белки с пониженной экспрессией в мутантных линиях.

Таким образом, сверхэкспрессия *PSEP3* оказывала сильное негативное влияние на клетки *P. patens*, в отличие от нокаута по *PSEP3*. Ранее было показано, что сверхэкспрессия случайных пептидов может приводить к видимым изменениям фенотипа растений [143]. Следовательно, наблюдаемые нами эффекты при сверхэкспрессии *PSEP3* могут быть объяснены как биологическим эффектом, например, связанным со взаимодействием с протеомом клетки, так и быть артефактом не физиологической концентрации микробелка. По нашему мнению, в первом случае можно ожидать точной регуляции экспрессии *PSEP3* в тканях *P. patens*.

Поскольку проведенное нами транскрипционное профилирование не позволяет оценить пространственную экспрессию данного микробелка в тканях *P. patens*, мы

использовали систему CRISPR-Cas9 для слияния в одной рамке считывания репортерной последовательности гена *GUS* с последовательностью *PSEP3* непосредственно в геноме *P. patens*. Такой подход позволил нам оценить активность нативного промотора *PSEP3* в разных тканях. В результате анализа мутантных линий мы обнаружили четкое пространственное разделение экспрессии данного микробелка (Рис. 4.74).

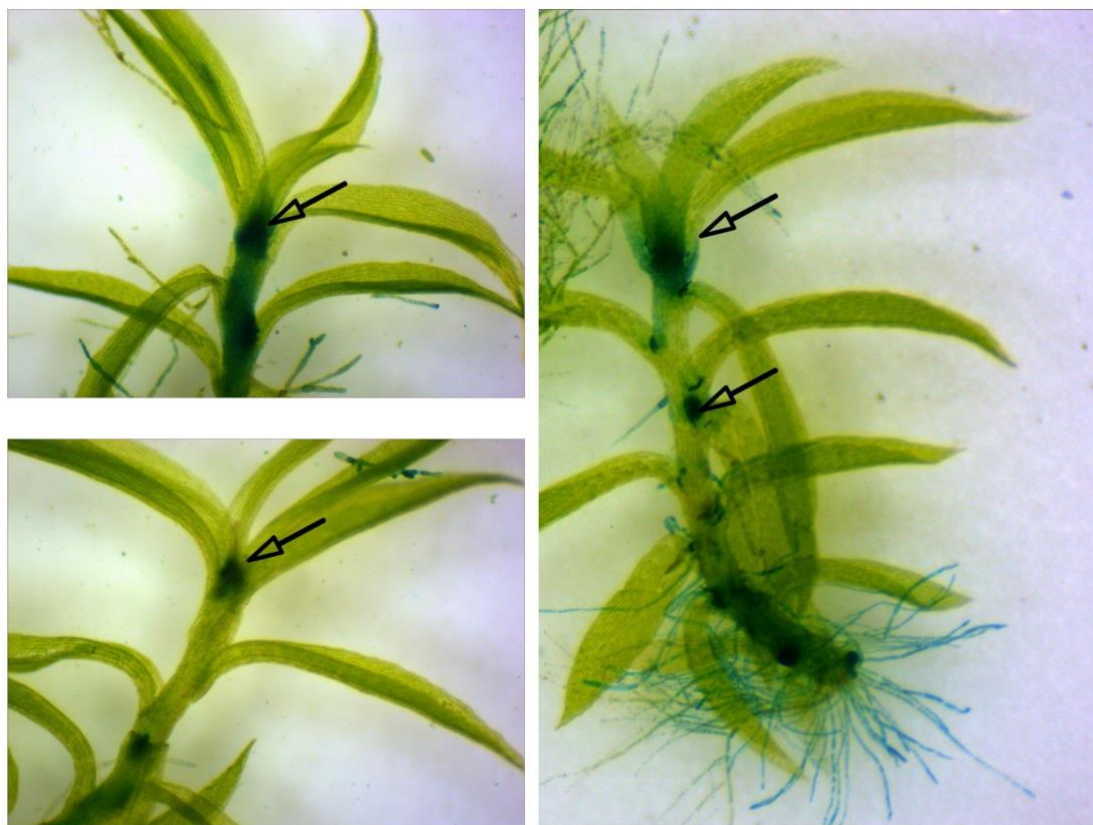


Рисунок 4.74 Анализ пространственной экспрессии микробелка *PSEP3-GUS* в тканях *P. patens*. Стрелками показаны некоторые области экспрессии *PSEP3* у гаметофоров, которые находятся в регионах, содержащих активно делящиеся клетки.

Особенно интересной нам представляется локальная экспрессия в активно делящихся “точках роста” гаметофоров, что может указывать на важную роль данного микробелка в регуляции процессов развития у *P. patens*.

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

растений. Учитывая проблемы в идентификации ортологов таких пептидов, кодируемых кОПС даже у близких видов, а следовательно, и проблемы с анализом их эволюции и оценки функциональной значимости, мы предполагаем, что количество функциональных генов данного класса у растений существенно недооценено.

Согласно нашим результатам, нокаут отобранных нами пептидов, кодируемых длнкРНК, не был летальным для *P. patens*, но влиял на рост и развитие растений. Все полученные нокауты показали снижение скорости роста по сравнению с растениями дикого типа. Напротив, мы обнаружили, что растения со сверхэкспрессией пептидов/микробелков, кодируемых длнкРНК, демонстрируют больше фенотипических различий по сравнению с растениями дикого типа и нокаутами. Мы наблюдали как значительное увеличение скорости роста (*PSEP1*) и числа облиственных побегов (*PSEP25*), так и снижение скорости роста у линий со сверхэкспрессией *PSEP3*, *PSEP18* и *PSEP25*. Различия между линиями дикого типа и мутантными линиями часто проявлялись только в специфических условиях, например, при выращивании растений на твердых или жидких средах с/без глюкозы или тартрата аммония. Эти данные могут указывать на адаптивную функцию пептидов/микробелков, кодируемых длнкРНК, в определенных условиях роста и развития растений.

4.3.3 Пептид FAMOSS - регулятор полярного роста у растений

Ряд недавних исследований показал, что пептиды, кодируемые кОПС расположенных на длнкРНК и содержащие трансмембранные домены, адаптированы к тому, чтобы быть предшественниками новых функциональных генов [315,316]. Действительно, большое количество функционально аннотированных пептидов, транслируемых с длинных некодирующих РНК у животных, локализованы в мембранах различных органелл [68]. Однако кодирующие функциональные трансмембранные растительные пептиды/микробелки, кодируемые “длнкРНК-кОПС”, еще не описаны. Для того,

чтобы понять возможные функции таких пептидов у растений, мы сконцентрировались на изучении функций 41-аа пептида *PSEPI* (см. выше), который мы, основываясь на характерном фенотипе, переименовали в *FAMOSS* («FAst-growing MOSS").

Анализ транскрипционных единиц на участке генома вокруг CNT2064811 (днкРНК, кодирующая *FAMOSS*) показал, что белок-кодирующий ген Pr3c9_20610 расположен в непосредственной близости или перекрывается с соответствующей кОПС. Чтобы исключить возможность того, что пептид *FAMOSS* в действительности является “upstream” или “downstream” кОПС, локализованной на аннотированной мРНК, мы использовали данные прямого нанопорового секвенирования РНК. В этом случае технология нанопорового секвенирования обеспечила точную сборку и идентификацию экзон-интронной структуры транскриптов в данном регионе 9 хромосомы. Этот анализ подтвердил, что CNT2064811 является независимой транскрипционной единицей с более высоким уровнем экспрессии, чем окружающие белок-кодирующие гены (**Рис. 4.75А**). Чтобы дополнительно изучить активность промотора CNT2064811 в разных тканях и условиях, мы использовали технологию CRISPR-Cas9 для слияния репортерной последовательности гена GUS с последовательностью *FAMOSS* в геноме мха (**Рис. 4.75В**). Анализ полученных мутантов на разных средах показал, что экспрессия *FAMOSS* как в протонеме, так и в гаметофорах не носит тканеспецифичный характер (**Рис. 4.75С**).

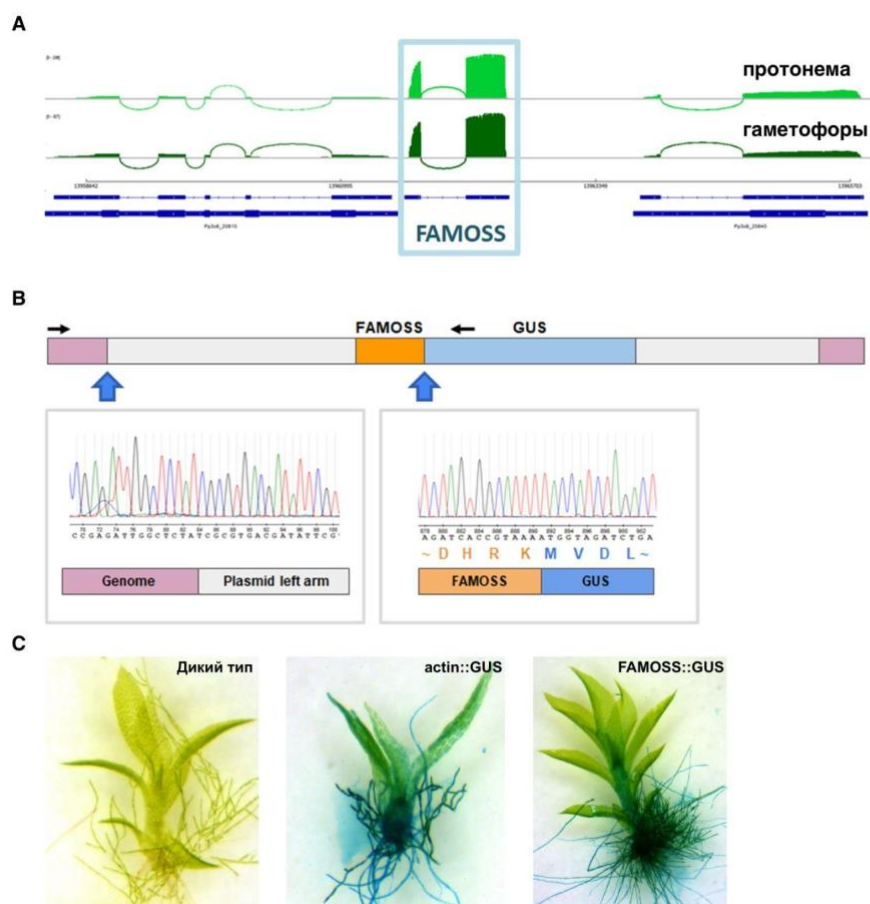


Рисунок 4.75 **A** - График Сашими, построенный на основе данных прямого нанопорового секвенирования образцов протонемы и гаметофоров. **B** - Схема слитого белка FAMOSS+GUS под контролем промотора FAMOSS. Черные стрелки указывают на праймеры, использованные для идентификации. Показаны результаты секвенирования геномной ДНК. **C** - Анализ пространственной экспрессии пептида FAMOSS в тканях *P. patens*. Показаны растения *P. patens* после окрашивания GUS: “дикий тип” - растения без вставки гена GUS; “actin::GUS” - растения со вставкой гена GUS под актиновым промотором; “FAMOSS::GUS” - растения со слитым белком - FAMOSS+GUS.

Затем мы провели поиск ортологов *FAMOSS* в различных растительных таксонах, используя транскриптомы из проекта OneKP [245]). На первом этапе мы использовали алгоритм TBLASTN ($E < 0,001$) и вручную проверили полученные результаты. Этот поиск выявил возможные ортологи у печеночников, мхов и папоротников, давая основания предположить, что этот пептид высококонсервативен у растений. Кроме того, ортолог пептида *FAMOSS* был

обнаружен в транскриптоме водоросли *Zygnemopsis* sp.. Хотя С-конец пептида FAMOSS был плохо консервативен, анализ выравниваний между возможными ортологами в зеленых водорослях, мхах, печеночниках и папоротниках обнаружил ряд высококонсервативных остатков (**Рис. 4.76А**). Тем не менее, ортологи FAMOSS были не найдены у роголистников, плаунов и цветковых растений.

Учитывая, что идентификация гомологов коротких последовательностей у отдаленных видов может быть затруднена, мы затем выполнили поиск с использованием программы HMMER ($E < 0,01$), используя построенные на полученных при предыдущем поиске выравнивания для построения профайлов HMM. Для анализа были использованы транскриптомы голосеменных и покрытосеменных растений из проекта OneKP [245]. Такой подход позволил нам идентифицировать возможные ортологи (**Рис. 4.76В**) у восьми видов Gymnospermae (12 транскриптов, средний размер рамки = 44 аминокислотных остатка), у 108 видов покрытосеменных из группы Розиды (Eudicots/Rosids; 124 транскрипта, средний размер рамки = 47 а.о.), у 41 вида из группы Астериды (Eudicots/Asterids; 53 транскрипта, средний размер рамки = 53 а.о.) и у 25 растений из клады Basal Eudicots (38 транскриптов, средний размер рамки = 46 а.о.); однако необходимы дальнейшие исследования, чтобы подтвердить, что эти пептиды являются истинными ортологами FAMOSS. Мы обнаружили, что соответствующие транскрипты не кодируют полипептиды длиннее 100 а.о., а 50% (63/124) транскриптов из Eudicots/Rosids, 66% (35/53) из Eudicots/Asterids и 71% (27/38) из клады Basal Eudicots были аннотированы как длнкРНК в базе данных длинных некодирующих последовательностей у покрытосеменных, AlnС [380]. Эти результаты показывают, что транскрипты, кодирующие консервативные пептиды/микробелки, могут быть неправильно аннотированы как длнкРНК. Наши результаты показывают, что 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.

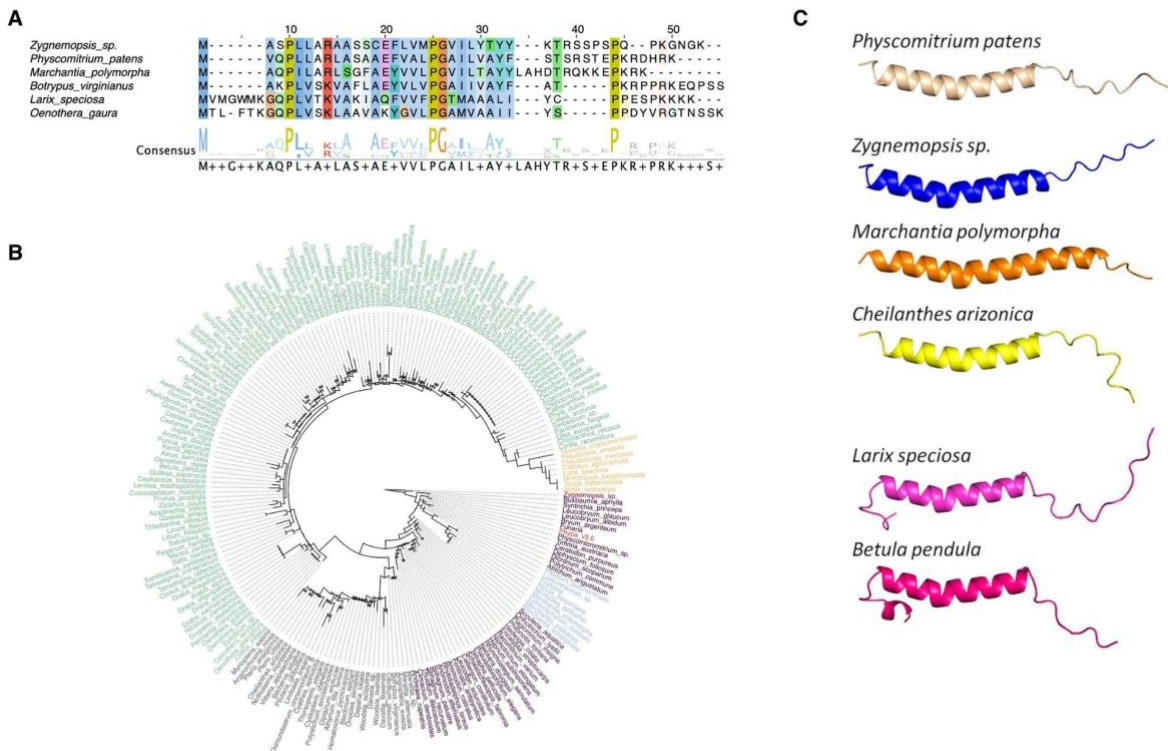


Рисунок 4.76 А - Выравнивание пептида FAMOSS и ортологических последовательностей из разных отделов растений. **В** - Филогенетическое дерево идентифицированных ортологов FAMOSS. Мхи - бордовый; печеночники - синий; водоросли - красный; папоротники - серый; голосеменные - оранжевый; покрытосеменные - зеленый. **С** - Структура пептида FAMOSS и ортологов из разных отделов растений, предсказанные с помощью алгоритма AlphaFold2.

Для дальнейшего изучения взаимосвязи между консервативными аминокислотными остатками и структурными свойствами *FAMOSS* мы предсказали его трехмерную структуру, используя недавно разработанный алгоритм машинного обучения AlphaFold2 [264]. Согласно полученной структурной модели, N-концевая и центральная части пептида представляют собой альфа-спираль с небольшим изломом в центре, вероятно, связанным с консервативным остатком пролина. В то же время, неконсервативный С-конец является неупорядоченным (Рис. 4.76С).

4.3.3.1 Анализ протеома мутантных линий

Затем мы использовали изобарные метки для относительного и абсолютного количественного анализа для выявления изменений в протеомах мутантных линий в сравнении с диким типом. Для фильтрации дифференциально регулируемых белковых групп (ДРБ) использовали следующие параметры: изменение представленности (Fold Change, $FC > 1,20$ или $< 0,83$; $P < 0,01$, односторонний ANOVA). Используя эти параметры, мы идентифицировали 118 ДРБ в линии со сверхэкспрессией и 75 ДРБ в нокаутной линии. Белки, экспрессия которых значительно выросла в линии со сверхэкспрессией *FAMOSS*, были рибосомные белки, таких как Pr3c2_29010 (RP-S30e), Pr3c24_17080 (RP-S20) и Pr3c12_16410 (rbfA) (**Рис. 4.77А**). Интересно, что представленность белков, которые участвуют в реакции на стресс, такие как связанные с патогенезом белки (PR-10) Pr3c2_27350 и Pr3c7_19850 (PR-10), Pr3c5_22560 (диригентный белок; DIR), Pr3c5_22560 (халконсинтаза; CHS) и Pr3c4_22490 (алленоксидциклаза; АОС), значительно снизились в мутантной линии. Еще одним белком, представленность которого снизилась в линии со сверхэкспрессией *FAMOSS* и связанном с регуляцией вторичного метаболизма и защитным ответом, является Pr3c9_1620 (полифенолоксидаза; PPO) [381]. Таким образом, полученные результаты позволяют предположить, что сверхэкспрессия *FAMOSS* будет приводить к уменьшению устойчивости к различным стрессовым условиям. Изменения протеома нокаутной линии были менее выражены (**Рис. 4.77В**). К наиболее меняющимся белкам относились большая и малая субъединицы RuBisCO и Pr3c11_15790 (глицеральдегид-3-фосфатдегидрогеназа (GAPDH)). Одним из возможных объяснений этого результата является преобладание у нокаутной линии богатых хлоропластами хлоронемных филаментов. Интересно, что представленность белков, связанных со стрессовым ответом, таких как Pr3c7_19850 (PR10), Pr3c6_6545 (диригентный белок, DIR) и Pr3c1_18940 (фенилаланин-аммиак-лиаза; PAL), также была снижена в нокаутной линии, как и в протеоме линии со сверхэкспрессией *FAMOSS*.

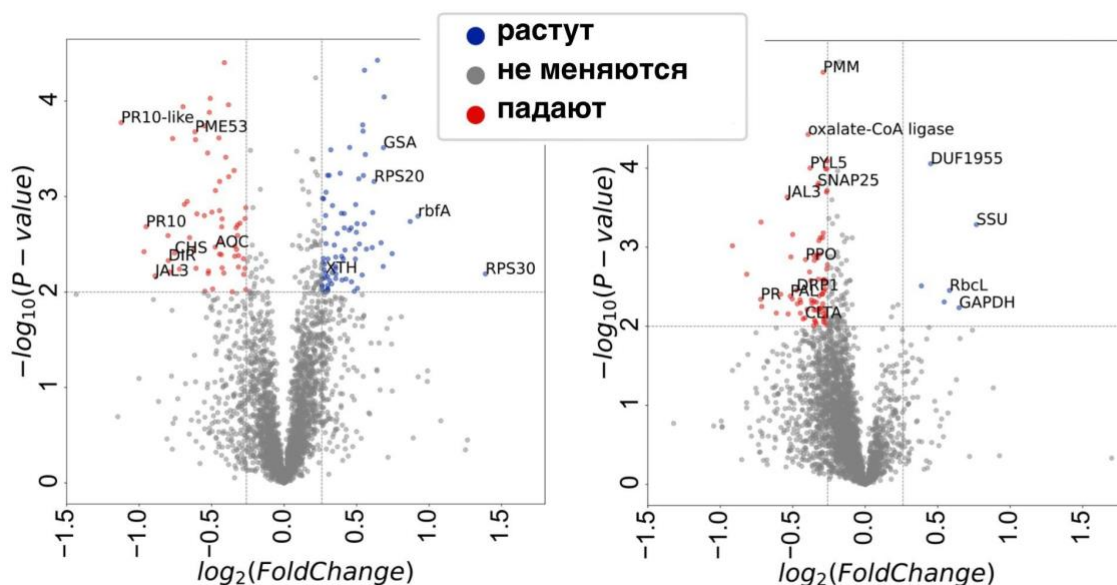


Рисунок 4.77 **А** - График “volcano plot” всех белков, количественно определенных во время протеомного анализа в линии со сверхэкспрессией *FAMOSS*. **В** - График “volcano plot” всех белков, количественно определенных во время протеомного анализа мутантной линии, нокаутной по *FAMOSS*. Белки, содержание которых значительно изменилось ($P < 0.01$), показаны цветом. Синие точки указывают на белки с повышенной экспрессией, а красные точки указывают на белки с пониженной экспрессией в мутантных линиях. GSA, глутамат-1-полуальдегид 2,1-аминомутаза; rbfA – рибосомо связывающий фактор А; RPS20, рибосомный белок S20; RPS30, рибосомный белок S30; AOC, алленоксидциклаза; GST, глутатион-S-трансфераза; CHS, халконсинтаза; XTH, белок ксилоглюкан-эндотрансглюкозилаза/гидролаза 23; PME, пектинметилэстераза; rbcL, большая субъединица рибулозо-1,5-бисфосфаткарбоксилазы/оксигеназы; GAPDH, глицеральдегид-3-фосфатдегидрогеназа А; JAL3, лектин 3; PPO, полифенолоксидаза; PAL, фенилаланин-аммиак-лиаза; PR10, белок, связанный с патогенезом; DIR, белок диригент; SNAP25, синаптосомально-ассоциированный белок 25; ASPL, белок содержащий домен UBX; DRP1, белок динамин 1E; CLTA, легкая цепь клатрина; PYL5, белок устойчивости к пиробактину 5; PMM, фосфоманномутаза; SSU, малые субъединицы рибулозо-1,5-бисфосфаткарбоксилазы/оксигеназы; DUF1955, домен с неизвестной функцией.

Стоит отметить, что представленность таких белков как Pr3c7_19850 (PR10), Pr3c18_10760 (глутаминсинтетаза), Pr3c9_1620 (полифенолоксидаза) и Pr3c1_1940 (белок с лектиновым доменом) снижалась у всех мутантных линий.

Кроме того, мы также идентифицировали группу ДРБ, связанных с везикулярным транспортом, представленность которых снижалась в нокаутной линии. Это был динамин-подобный белок 1С (DRP1С, Pp3c19_4870), который регулирует образование везикул при транспортировке после аппарата Гольджи [382,383]; легкая цепь клатрина (Pp3c7_10220; CТLА); и связанный с синапсосомами белок SNAP25 (Pp3c2_7850). SNAP25 представляет собой белок из группы SNARE (soluble NSF attachment receptor), который участвует не только в везикулярном транспорте в нормальных условиях, но также в регуляции связанных с клеточной стенкой и защитных белков во время биотического стресса [384]. Таким образом, количественный протеомный анализ показал, что экспрессия белков, участвующих в ответе на различные стрессовые условия, была снижена в обеих мутантных линиях.

4.3.3.2 Анализ интерактома пептида FAMOSS

Согласно последним исследованиям, подавляющее количество уже изученных пептидов/микробелков, кодируемых кОРС, у животных выполняли свои биологические функции при взаимодействии с белками [69]. Следовательно, одним из необходимых этапов в установлении функций пептидов, кодируемых кОРС, является анализ их интерактомов.

Для того, чтобы попытаться понять функции *FAMOSS*, мы идентифицировали возможных белковых партнеров, с которыми он взаимодействует, используя два подхода: метод белковой преципитации (“pull-down assay”) и ко-иммунопреципитацию (“co-IP”; **Рис. 4.78А**). Для метода белковой преципитации использовали рекомбинантный пептид *FAMOSS*, слитый со стрептавидином на N-конце (SAV-*FAMOSS*). Такой подход позволил идентифицировать 323 уникальные в сравнении с контролем (рекомбинантный стрептавидин) белковые группы в образцах SAV-*FAMOSS*. Помимо большого количества рибосомальных и хлоропластных белков, которые наиболее часто являются контаминантами

интерактивных экспериментов [215], мы обнаружили высококонсервативные белки - малые ГТФазы Rab-типа.

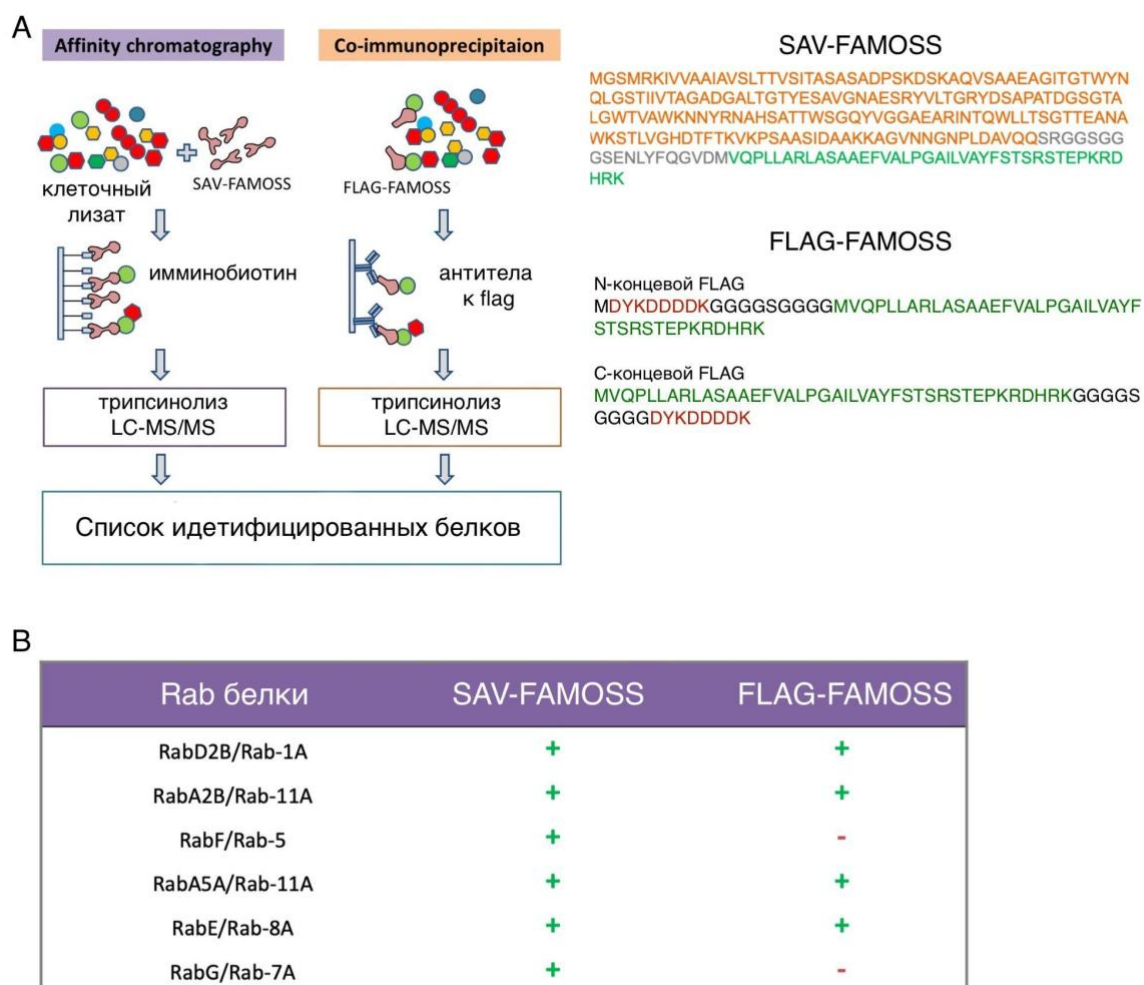


Рисунок 4.78 **A** - Схема эксперимента по анализу интерактома. Показаны последовательности слитых белков, использовавшихся для экспериментов. Цветом выделены последовательности FAMOSS, стрептавидина и FLAG. **B** - Rab-белки, идентифицированные с помощью масс-спектрометрического анализа в интерактивных экспериментах. Белки, обнаруженные в образцах, отмечены знаком плюс, а не обнаруженные – знаком минус.

Известно, что ГТФазы этого типа участвуют в регуляции разнообразных клеточных процессов, включая перенос везикул, регуляцию полярного и реакцию на стрессовые факторы [385–387]. Список ГТФаз Rab-типа, уникальных для образцов SAV-FAMOSS, включал несколько белков из этого семейства - RabD2B/Rab-1A, RabA2B/Rab-11A, RabF/Rab-5, RabA5A/Rab-11A, RabE/Rab-8A и RabG/Rab-7A

(Рис. 4.78В). Среди этих белков наибольшее покрытие пептидами (33%) и представленность наблюдались для белка Pp3c14_15250 (RabA2B/Rab-11A).

Чтобы подтвердить полученные результаты, мы создали мутантные линии, сверхэкспрессирующие *FAMOSS*, слитый с 8-аминокислотной пептидной меткой - flag (Рис. 4.78А). Пептидную метку помещали на С-конец (мутантная линия 1028) или на N-конец (мутантная линия 1032) пептида *FAMOSS* (Рис. 4.78А). Проведя иммунопреципитацию со специфическими антителами на flag, мы также идентифицировали несколько белков Rab-типа: RabD2B/Rab-1A, RabA2B/Rab-11A, RabA5A/Rab-11A и RabE/Rab-8A в интерактоме линии 1032 (n = 3 независимых эксперимента). Эти белки уникальны для интерактомов FLAG-*FAMOSS* и не обнаружены в контрольных образцах. Таким образом, мы не можем исключать, что пептид *FAMOSS* взаимодействует с несколькими белками Rab-типа. Таким образом, наши результаты позволяют предположить, что пептид *FAMOSS* может регулировать белки Rab-типа. Известно, что белки ГТФазы Rab-типа регулируют слияние внутриклеточных везикул с плазматическими мембранами, тем самым стимулируя интенсивность их транспорта к апексу клетки и увеличивая скорость полярного роста. Это согласуется с наблюдаемыми нами фенотипами линий со сверхэкспрессией *FAMOSS*.

4.3.3.3 Анализ везикулярного транспорта в мутантных линиях

Рост нитей протонемы *P. patens* осуществляется посредством полярного роста апикальных клеток, подобно пыльцевым трубкам или корневым волоскам цветковых растений [388,389]. Известно, что полярный рост у растений требует активного апикального везикулярного транспорта, регулируемого различными семействами малых ГТФаз, включая Rab-белки [390].

Чтобы проанализировать интенсивность везикулярного транспорта в апикальных клетках протонемы дикого типа и мутантных линий, мы использовали флуоресцентный краситель SynaptoGreen C4 (также известный как FM1-43) [391]. Краситель FM1-43 широко используется для окрашивания везикул в пыльцевых трубках, корневых волосках и апикальных клетках протонемы [392–394]. Мы

обнаружили, что интенсивность везикулярного транспорта была значительно выше в апикальных клетках протонемы линии со сверхэкспрессией *FAMOSS* в сравнении с диким типом и нокаутными линиями (ANOVA, $P < 0,001$). Это согласуется с быстрым ростом каулонемных нитей в сверхэкспрессирующей линии. Напротив, интенсивность флуоресценции была значительно снижена в апикальных клетках линии с нокаутом *FAMOSS* по сравнению с клетками линии дикого типа и линии со сверхэкспрессией (ANOVA, $P < 0,001$; **Рис. 4.79**).

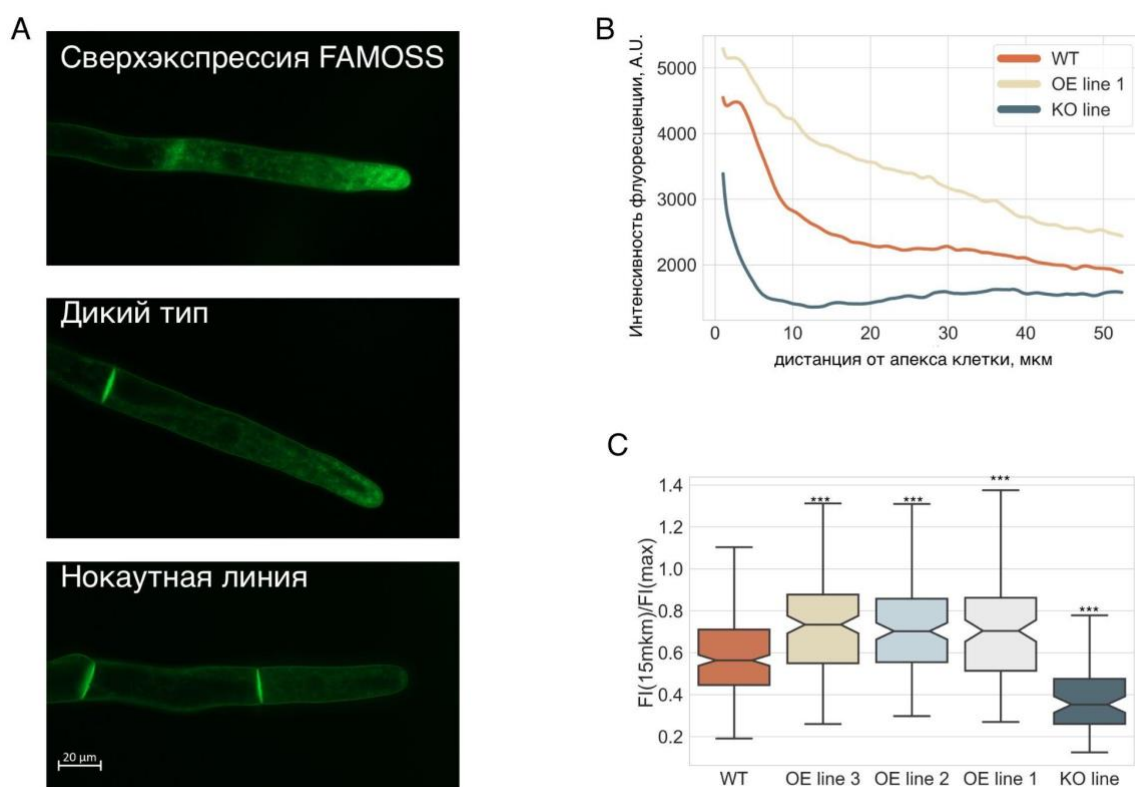


Рисунок 4.79 А - Анализ интенсивности везикулярного транспорта с помощью красителя SynaptoGreen C4 в апикальных клетках дикого типа и мутантных линий. В - Профиль флуоресценции SynaptoGreen C4 в апикальных клетках “WT” - дикого типа, “OE line 1” - линии со сверхэкспрессией, “KO line” - нокаутной линии. С - Диаграмма, показывающая значения отношения интенсивности окрашивания SynaptoGreen C4 на расстоянии 15 мкм от апекса клеток к максимальной интенсивности у дикого типа и мутантных линий. *** критерий Манна-Уитни, $P < 0,001$.

Эти результаты позволяют предположить, что пептид *FAMOSS* может функционировать как возможный модулятор везикулярного транспорта во время полярного роста посредством его взаимодействия с малыми ГТФазами Rab-типа. В этой связи роль ГТФазы Rab-типа как возможных партнеров пептида *FAMOSS* согласуется со скоростью роста протонемы, изменениями в протеоме и интенсивным ростом каулонемы в линиях со сверхэкспрессией пептида. Например, количественный протеомный анализ показал, что представленность некоторых белков, связанных с везикулярным транспортом, была снижена у нокаутных линий. Таким образом, мы предполагаем, что взаимодействие пептида *FAMOSS* с Rab-белками может влиять на интенсивность везикулярного транспорта в апикальных клетках *P. patens* и косвенно регулировать скорость роста протонемы. Таким образом, пептид *FAMOSS* можно считать важным компонентом регуляции везикулярного транспорта у растений и дальнейшие исследования помогут более точно прояснить роль данного пептида.

5. 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”. При этом очень важную роль в формировании клеточного пептидома растений играют белки хлоропластов. 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 деградама” 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 (“длнкРНК-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 “длнкРНК-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 плохо консервативные пептиды могут не иметь четко определенной функции, но служить предшественниками функциональных генов. В-третьих, существует подмножество “длнкРНК-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 деградома”, являются пептиды, транслирующиеся с длнкРНК.

Поскольку ранее функции пептидов, кодируемых длнкРНК, 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 длнкРНК, 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 длнкРНК. В свете этих находок мы предположили, что пептиды, кодируемые длнкРНК, могут быть важны для адаптации к определенным условиям внешней среды. Кроме того, полученные данные указывают на то, что “длнкРНК-кОРС” с регионами низкой сложности или содержащие трансмембранный домен могут быть адаптированы к тому, чтобы стать предшественниками новых функциональных генов. 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. 5.1).

Для ряда обнаруженных пептидов была показана биологическая значимость.

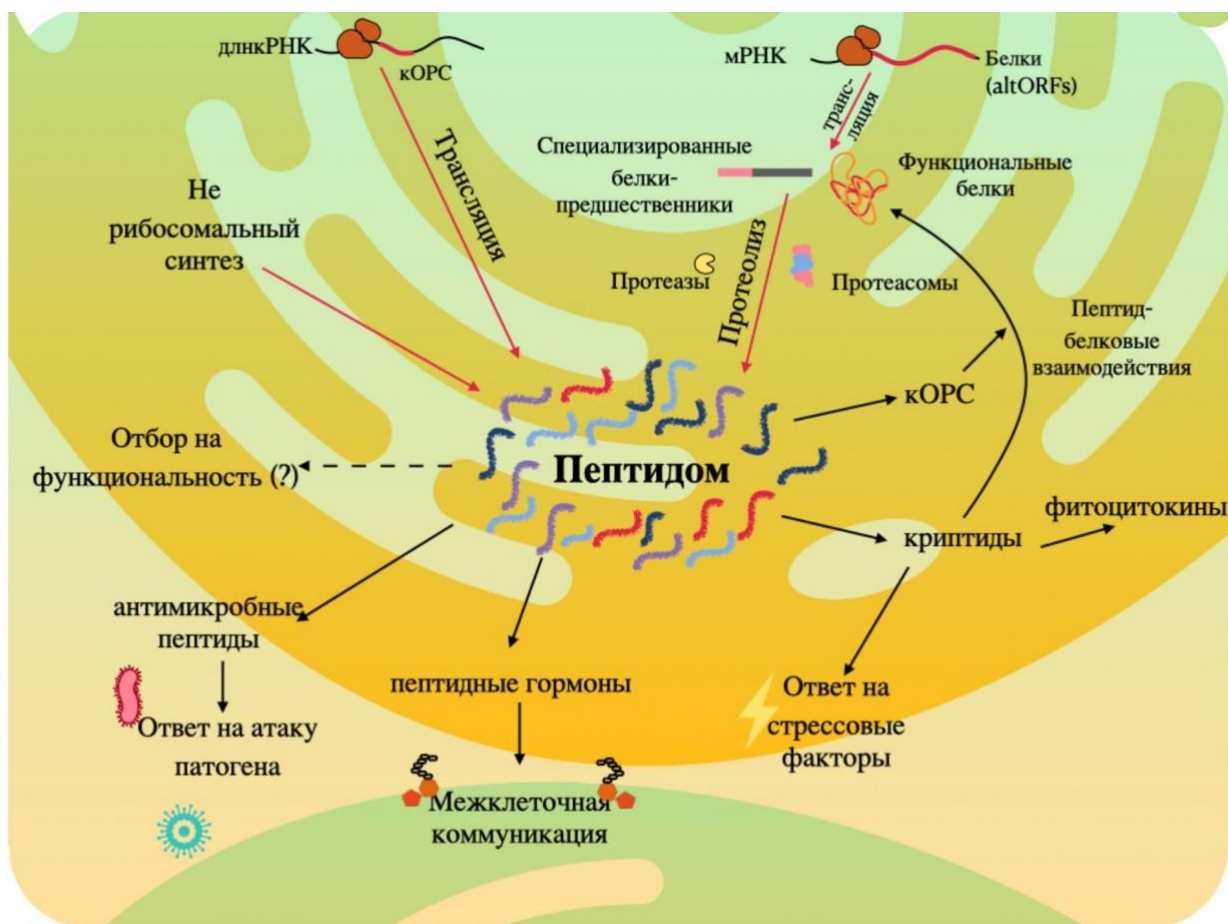


Рисунок 5.1 Общая схема путей образования и функций пептидов у растений.

6. 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 транслирующихся 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.

7. СПИСОК СОКРАЩЕНИЙ

КОРС - короткая открытая рамка считывания

длнкРНК - длинные некодирующие РНК

ЭПР - эндоплазматический ретикулум

МАРР - масс-спектрометрический анализ протеолитических пептидов (mass spectrometry analysis of proteolytic peptides)

АМП - Антимикробные пептиды

ЛТР - липид-переносящие белки (lipid transfer proteins)

РЕРs - элиситорные пептиды растений

МАРК - митоген-активируемые киназы (Mitogen Activated Protein Kinase)

DAMP - молекулы, связанные с повреждениями (damage-associated molecular patterns)

miPEP - микроРНК-кодируемые пептиды

МС - масс-спектрометрический

SCX - сильный катионный обменник (strong cation exchange)

HILIC - гидрофильная хроматография (Hydrophilic interaction chromatography)

CID - Диссоциация, активируемая соударением (Collision-Induced Dissociation)

HCD - Диссоциация, активируемая соударением при повышенной энергии (Higher Energy Collision Dissociation)

ETD - диссоциация с переносом электронов (Electron-transfer dissociation)

ЭДТА - Этилендиаминтетрауксусная кислота

MES - Морфолино-2-этансульфоновая кислота

САА - хлорацетамид

ТСЕР - трис(2-карбоксиэтил)фосфин

CRISPR/Cas 9 - Сгруппированные регулярно расположенные короткие палиндромные повторы (Clustered Regularly Interspaced Short Palindromic Repeats/Cas9)

LB - Luria broth (среда Лурия-Бертания)

МИК - минимальная ингибирующая концентрация

ПААГ – полиакриламидный гель

ДСН – додецилсульфат натрия

DDT - дитиотрейтол

iTRAQ - набор для относительной и абсолютной квантификации (Isobaric tag for relative and absolute quantitation)

SCX - катионно-обменная хроматография

NBT/BCIP - 5-бром-4-хлор-3-индолилфосфат/нитро-синий тетразолий

ПЦР – полимеразная цепная реакция

ВЭЖХ – высокоэффективная жидкостная хроматография

DDA - информационно-зависимый сбор данных (Data Dependent Acquisition)

AGC - автоматический контроль времени захвата (Automatic Gain Control)

Да – дальтон

FDR - уровень ложноположительных результатов (false discovery rate)

DSP - дитиобис (сукцинимидилпропионат) (dithiobis(succinimidyl propionate))

IPTG - изопропил- β -D-1-тиогалактопиранозид

BLAST - Базовый инструмент поиска местного выравнивания (Basic Local Alignment Search Tool)

SSPs - малые декретируемые пептиды

GO - онтология генов (gene ontology)

МЖ - метилжасмонат

СА - салициловая кислота

CDS - кодирующая последовательность (Coding sequence)

NCDS - декодирующая последовательность (non-coding sequence)

UTRs - ретранслируемый регион (untranslated region)

PSM - совпадения спектров пептидов (peptide spectrum matches)

АС - альтернативный сплайсинг

NMD - нонсенс-опосредованный распад (nonsense mediated decay)

siPEPs - малые интерферирующие пептиды

ТМ домен - трансмембранный домен

ОРС - открытая рамка считывания

BUSTED - генный анализ на положительный отбор (Branch-Site Unrestricted Statistical Test for Episodic Diversification)

РНС - регион низкой сложности

8. СПИСОК ЛИТЕРАТУРЫ

1. Olsson V. et al. Look Closely, the Beautiful May Be Small: Precursor-Derived Peptides in Plants // *Annu. Rev. Plant Biol.* 2019. Vol. 70. P. 153–186.
2. Tavormina P. et al. The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions // *Plant Cell.* 2015. Vol. 27, № 8. P. 2095–2118.
3. Pires-daSilva A., Sommer R.J. The evolution of signalling pathways in animal development // *Nat. Rev. Genet.* 2003. Vol. 4, № 1. P. 39–49.
4. Ivanov V.T. et al. Peptidomics: A modern approach to biodiversity of peptides // *J. Macromol. Sci. Part A Pure Appl. Chem. De Gruyter*, 2006. Vol. 78, № 5. P. 963–975.
5. Ivanov V.T., Yatskin O.N. Peptidomics: a logical sequel to proteomics // *Expert Rev. Proteomics.* 2005. Vol. 2, № 4. P. 463–473.
6. Ueki N. et al. Cryptides: functional cryptic peptides hidden in protein structures // *Biopolymers.* 2007. Vol. 88, № 2. P. 190–198.
7. Slavoff S.A. et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells // *Nat. Chem. Biol.* 2013. Vol. 9, № 1. P. 59–64.
8. de Araujo C.B. et al. Intracellular Peptides in Cell Biology and Pharmacology // *Biomolecules.* 2019. Vol. 9, № 4.
9. Yuan S. et al. Genome-wide survey of soybean papain-like cysteine proteases and their expression analysis in root nodule symbiosis // *BMC Plant Biol.* 2020. Vol. 20, № 1. P. 517.
10. van der Hoorn R.A.L. Plant proteases: from phenotypes to molecular mechanisms // *Annu. Rev. Plant Biol.* 2008. Vol. 59. P. 191–223.
11. Rautengarten C. et al. Inferring hypotheses on functional relationships of genes: Analysis of the *Arabidopsis thaliana* subtilase gene family // *PLoS Comput. Biol.* 2005. Vol. 1, № 4. P. e40.
12. Reichardt S. et al. The tomato subtilase family includes several cell death-related proteinases with caspase specificity // *Sci. Rep.* 2018. Vol. 8, № 1. P. 10531.
13. Figueiredo J. et al. Revisiting *Vitis vinifera* Subtilase Gene Family: A Possible Role in Grapevine Resistance against *Plasmopara viticola* // *Front. Plant Sci.* 2016. Vol. 7. P. 1783.
14. Srivastava R. et al. Regulation and processing of a plant peptide hormone, AtRALF23, in *Arabidopsis* // *Plant J.* 2009. Vol. 59, № 6. P. 930–939.
15. Stegmann M. et al. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling // *Science.* 2017. Vol. 355, № 6322. P. 287–289.
16. Schardon K. et al. Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases // *Science.* 2016. Vol. 354, № 6319. P. 1594–1597.
17. Ni J. et al. Characterization of a CLE processing activity // *Plant Mol. Biol.* 2011. Vol. 75, № 1-2. P. 67–75.
18. Fiers M., Ku K.L., Liu C.-M. CLE peptide ligands and their roles in establishing meristems // *Curr. Opin. Plant Biol.* 2007. Vol. 10, № 1. P. 39–43.
19. Stührwohldt N. et al. The biogenesis of CLEL peptides involves several processing events in consecutive compartments of the secretory pathway // *Elife.* 2020. Vol. 9.
20. Stührwohldt N. et al. Precursor processing by SBT3.8 and phytosulfokine signaling contribute to drought stress tolerance in *Arabidopsis* // *J. Exp. Bot.* 2021.
21. Chen Y.-L. et al. The role of peptides cleaved from protein precursors in eliciting plant stress reactions // *New Phytol.* 2020. Vol. 225, № 6. P. 2267–2282.
22. Hander T. et al. Damage on plants activates Ca²⁺-dependent metacaspases for release of immunomodulatory peptides // *Science.* 2019. Vol. 363, № 6433.
23. Hou S. et al. Damage-Associated Molecular Pattern-Triggered Immunity in Plants // *Front. Plant Sci.* 2019. Vol. 10. P. 646.
24. Wrzaczek M. et al. GRIM REAPER peptide binds to receptor kinase PRK5 to trigger cell death in *Arabidopsis* // *EMBO J.* 2015. Vol. 34, № 1. P. 55–66.
25. Tabata R. et al. Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling // *Science.* 2014. Vol. 346, № 6207. P. 343–346.

26. Matsubayashi Y. Posttranslationally modified small-peptide signals in plants // *Annu. Rev. Plant Biol.* 2014. Vol. 65. P. 385–413.
27. Stührwohldt N., Schaller A. Regulation of plant peptide hormones and growth factors by post-translational modification // *Plant Biol.* . 2019. Vol. 21 Suppl 1. P. 49–63.
28. Komori R. et al. Identification of tyrosylprotein sulfotransferase in Arabidopsis // *Proc. Natl. Acad. Sci. U. S. A.* 2009. Vol. 106, № 35. P. 15067–15072.
29. Velasquez S.M. et al. Complex regulation of prolyl-4-hydroxylases impacts root hair expansion // *Mol. Plant.* 2015. Vol. 8, № 5. P. 734–746.
30. Ogawa-Ohnishi M., Matsushita W., Matsubayashi Y. Identification of three hydroxyproline O-arabinosyltransferases in Arabidopsis thaliana // *Nat. Chem. Biol.* 2013. Vol. 9, № 11. P. 726–730.
31. MacAlister C.A. et al. Hydroxyproline O-arabinosyltransferase mutants oppositely alter tip growth in Arabidopsis thaliana and Physcomitrella patens // *Plant J. Wiley*, 2016. Vol. 85, № 2. P. 193–208.
32. McMANUS I.R. Synthesis of intracellular peptides in Torula utilis // *J. Biol. Chem.* 1958. Vol. 231, № 2. P. 777–785.
33. Connell G.E., Watson R.W. Intracellular peptides of Pseudomonas hydrophila // *Biochim. Biophys. Acta.* 1957. Vol. 24, № 1. P. 226–227.
34. Ferro E.S., Hyslop S., Camargo A.C.M. Intracellular peptides as putative natural regulators of protein interactions // *J. Neurochem.* 2004. Vol. 91, № 4. P. 769–777.
35. Rock K.L. et al. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules // *Cell.* 1994. Vol. 78, № 5. P. 761–771.
36. Fricker L.D. Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins // *Mol. Biosyst.* 2010. Vol. 6, № 8. P. 1355–1365.
37. Teixeira C.M.M. et al. Characterization of Intracellular Peptides from Zebrafish (Danio rerio) Brain // *Zebrafish.* 2019. Vol. 16, № 3. P. 240–251.
38. Gelman J.S. et al. Peptidomic analysis of human cell lines // *J. Proteome Res.* 2011. Vol. 10, № 4. P. 1583–1592.
39. Wolf-Levy H. et al. Revealing the cellular degradome by mass spectrometry analysis of proteasome-cleaved peptides // *Nat. Biotechnol.* 2018.
40. Parker B.L. et al. Multiplexed Temporal Quantification of the Exercise-regulated Plasma Peptidome // *Mol. Cell. Proteomics.* 2017. Vol. 16, № 12. P. 2055–2068.
41. Ziganshin R. et al. New method for peptide desorption from abundant blood proteins for plasma/serum peptidome analyses by mass spectrometry // *J. Proteomics.* 2011. Vol. 74, № 5. P. 595–606.
42. Ziganshin R.H. et al. The Pathogenesis of the Demyelinating Form of Guillain-Barre Syndrome (GBS): Proteo-peptidomic and Immunological Profiling of Physiological Fluids // *Mol. Cell. Proteomics.* 2016. Vol. 15, № 7. P. 2366–2378.
43. Kisselev A.F. et al. The Sizes of Peptides Generated from Protein by Mammalian 26 and 20 S Proteasomes: IMPLICATIONS FOR UNDERSTANDING THE DEGRADATIVE MECHANISM AND ANTIGEN PRESENTATION * // *J. Biol. Chem. Elsevier*, 1999. Vol. 274, № 6. P. 3363–3371.
44. Fricker L.D. et al. Peptidomic analysis of HEK293T cells: effect of the proteasome inhibitor epoxomicin on intracellular peptides // *J. Proteome Res.* 2012. Vol. 11, № 3. P. 1981–1990.
45. Dasgupta S. et al. Proteasome inhibitors alter levels of intracellular peptides in HEK293T and SH-SY5Y cells // *PLoS One.* 2014. Vol. 9, № 7. P. e103604.
46. Camargo A.C., Graeff F.G. Subcellular distribution and properties of the bradykinin inactivation system in rabbit brain homogenates // *Biochem. Pharmacol.* 1969. Vol. 18, № 2. P. 548–549.
47. Oliveira E.B., Martins A.R., Camargo A.C. Isolation of brain endopeptidases: influence of size and sequence of substrates structurally related to bradykinin // *Biochemistry.* 1976. Vol. 15, № 9. P. 1967–1974.
48. Checler F., Ferro E.S. Neurolysin: From Initial Detection to Latest Advances // *Neurochem. Res.* 2018. Vol. 43, № 11. P. 2017–2024.
49. Camargo A.C. et al. Brain endo-oligopeptidase A, a putative enkephalin converting enzyme // *J. Neurochem. Wiley*, 1987. Vol. 48, № 4. P. 1258–1263.
50. Villar-Cheda B. et al. Aging-related dysregulation of dopamine and angiotensin receptor interaction // *Neurobiol. Aging.* 2014. Vol. 35, № 7. P. 1726–1738.
51. Wu T.J., Pagano E., Mani S.K. A biological role for the gonadotrophin-releasing hormone (GnRH) metabolite, GnRH-(1-5) // *J. Neuroendocrinol. Wiley*, 2009. Vol. 21, № 4. P. 293–298.
52. Berti D.A. et al. Analysis of intracellular substrates and products of thimet oligopeptidase in human

- embryonic kidney 293 cells // *J. Biol. Chem.* 2009. Vol. 284, № 21. P. 14105–14116.
53. Russo L.C. et al. Inhibition of thimet oligopeptidase by siRNA alters specific intracellular peptides and potentiates isoproterenol signal transduction // *FEBS Lett.* 2012. Vol. 586, № 19. P. 3287–3292.
 54. Santos N. dos et al. Thimet Oligopeptidase (EC 3.4.24.15) Key Functions Suggested by Knockout Mice Phenotype Characterization // *Biomolecules.* 2019. Vol. 9, № 8. P. 382.
 55. Castro L.M. et al. Peptidomic analysis of the neurolysin-knockout mouse brain // *J. Proteomics.* 2014. Vol. 111. P. 238–248.
 56. Teixeira P.F. et al. Mechanism of Peptide Binding and Cleavage by the Human Mitochondrial Peptidase Neurolysin // *J. Mol. Biol.* 2018. Vol. 430, № 3. P. 348–362.
 57. Nishimura K., Kato Y., Sakamoto W. Essentials of Proteolytic Machineries in Chloroplasts // *Mol. Plant.* 2017. Vol. 10, № 1. P. 4–19.
 58. Ishida H. et al. Mobilization of rubisco and stroma-localized fluorescent proteins of chloroplasts to the vacuole by an ATG gene-dependent autophagic process // *Plant Physiol.* 2008. Vol. 148, № 1. P. 142–155.
 59. Izumi M. et al. Entire Photodamaged Chloroplasts Are Transported to the Central Vacuole by Autophagy // *Plant Cell.* 2017. Vol. 29, № 2. P. 377–394.
 60. Teixeira P.F. et al. A multi-step peptidolytic cascade for amino acid recovery in chloroplasts // *Nat. Chem. Biol.* 2017. Vol. 13, № 1. P. 15–17.
 61. Kmiec B., Glaser E. A novel mitochondrial and chloroplast peptidasome, PreP // *Physiol. Plant.* 2012. Vol. 145, № 1. P. 180–186.
 62. Kmiec B. et al. Accumulation of endogenous peptides triggers a pathogen stress response in *Arabidopsis thaliana* // *Plant J.* 2018. Vol. 96, № 4. P. 705–715.
 63. Guerra-Almeida D., Tschoeke D.A., Nunes-da-Fonseca R. Understanding small ORF diversity through a comprehensive transcription feature classification // *DNA Res.* 2021. Vol. 28, № 5.
 64. Yang X. et al. Discovery and annotation of small proteins using genomics, proteomics, and computational approaches // *Genome Res.* 2011. Vol. 21, № 4. P. 634–641.
 65. Andrews S.J., Rothnagel J.A. Emerging evidence for functional peptides encoded by short open reading frames // *Nat. Rev. Genet.* 2014. Vol. 15, № 3. P. 193–204.
 66. Storz G., Wolf Y.I., Ramamurthi K.S. Small proteins can no longer be ignored // *Annu. Rev. Biochem.* 2014. Vol. 83. P. 753–777.
 67. Dinger M.E. et al. Differentiating protein-coding and noncoding RNA: challenges and ambiguities // *PLoS Comput. Biol.* 2008. Vol. 4, № 11. P. e1000176.
 68. Couso J.-P., Patraquim P. Classification and function of small open reading frames // *Nat. Rev. Mol. Cell Biol.* 2017. Vol. 18, № 9. P. 575–589.
 69. Chen J. et al. Pervasive functional translation of noncanonical human open reading frames // *Science.* 2020. Vol. 367, № 6482. P. 1140–1146.
 70. Wright B.W. et al. The dark proteome: translation from noncanonical open reading frames // *Trends Cell Biol.* 2021.
 71. Huang J.-Z. et al. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth // *Mol. Cell.* 2017. Vol. 68, № 1. P. 171–184.e6.
 72. Johnstone T.G., Bazzini A.A., Giraldez A.J. Upstream ORFs are prevalent translational repressors in vertebrates // *EMBO J.* 2016. Vol. 35, № 7. P. 706–723.
 73. Churbanov A. et al. Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes // *Nucleic Acids Res. Oxford Academic,* 2005. Vol. 33, № 17. P. 5512–5520.
 74. Zhang H. et al. Determinants of genome-wide distribution and evolution of uORFs in eukaryotes // *Nat. Commun.* 2021. Vol. 12, № 1. P. 1076.
 75. Eisenberg E., Levanon E.Y. Human housekeeping genes, revisited // *Trends Genet.* 2013. Vol. 29, № 10. P. 569–574.
 76. Hayashi N. et al. Identification of *Arabidopsis thaliana* upstream open reading frames encoding peptide sequences that cause ribosomal arrest // *Nucleic Acids Res.* 2017. Vol. 45, № 15. P. 8844–8858.
 77. Lin M.F., Jungreis I., Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions // *Bioinformatics.* 2011. Vol. 27, № 13. P. i275–i282.
 78. Hartford C.C.R., Lal A. When Long Noncoding Becomes Protein Coding // *Mol. Cell. Biol.* 2020. Vol. 40, № 6.
 79. Kopp F., Mendell J.T. Functional Classification and Experimental Dissection of Long Noncoding RNAs // *Cell.* 2018. Vol. 172, № 3. P. 393–407.
 80. Ji Z. et al. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins // *Elife.* 2015. Vol. 4. P. e08890.

81. Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs // *Nat. Rev. Genet.* 2016. Vol. 17, № 10. P. 601–614.
82. Nelson B.R. et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle // *Science.* 2016. Vol. 351, № 6270. P. 271–275.
83. Hao Y. et al. SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci // *Brief. Bioinform.* 2018. Vol. 19, № 4. P. 636–643.
84. Lynch M., Marinov G.K. The bioenergetic costs of a gene // *Proc. Natl. Acad. Sci. U. S. A.* 2015. Vol. 112, № 51. P. 15690–15695.
85. Ruiz-Orera J. et al. Long non-coding RNAs as a source of new peptides // *Elife.* 2014. Vol. 3. P. e03523.
86. Aspden J.L. et al. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq // *Elife.* 2014. Vol. 3. P. e03528.
87. Kurihara Y. uORF Shuffling Fine-Tunes Gene Expression at a Deep Level of the Process // *Plants.* 2020. Vol. 9, № 5.
88. Pooggin M.M., Ryabova L.A. Ribosome Shunting, Polycistronic Translation, and Evasion of Antiviral Defenses in Plant Pararetroviruses and Beyond // *Front. Microbiol.* 2018. Vol. 9. P. 644.
89. van der Horst S. et al. Metabolite Control of Translation by Conserved Peptide uORFs: The Ribosome as a Metabolite Multisensor // *Plant Physiol.* 2020. Vol. 182, № 1. P. 110–122.
90. Arribere J.A., Gilbert W.V. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing // *Genome Res.* 2013. Vol. 23, № 6. P. 977–987.
91. Wu Q. et al. Translation of small downstream ORFs enhances translation of canonical main open reading frames // *EMBO J.* 2020. Vol. 39, № 17. P. e104763.
92. Brunet M.A. et al. OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes // *Nucleic Acids Res.* 2021. Vol. 49, № D1. P. D380–D388.
93. Mouilleron H., Delcourt V., Roucou X. Death of a dogma: eukaryotic mRNAs can code for more than one protein // *Nucleic Acids Res.* 2016. Vol. 44, № 1. P. 14–23.
94. Smith E. et al. Leaky ribosomal scanning in mammalian genomes: significance of histone H4 alternative translation in vivo // *Nucleic Acids Res.* 2005. Vol. 33, № 4. P. 1298–1308.
95. Bab I. et al. Biosynthesis of osteogenic growth peptide via alternative translational initiation at AUG85 of histone H4 mRNA // *J. Biol. Chem.* 1999. Vol. 274, № 20. P. 14474–14481.
96. Tavares L.S. et al. Antimicrobial activity of recombinant Pg-AMP1, a glycine-rich peptide from guava seeds // *Peptides.* 2012. Vol. 37, № 2. P. 294–300.
97. Tam J.P. et al. Antimicrobial Peptides from Plants // *Pharmaceuticals.* 2015. Vol. 8, № 4. P. 711–757.
98. Одинцова Т.И., Коростылева Т.В. Антимикробные пептиды пшеницы // журнал генетики и vavilov.elpub.ru, 2014.
99. Marmioli N., Maestri E. Plant peptides in defense and signaling // *Peptides.* 2014. Vol. 56. P. 30–44.
100. Stec B. Plant thionins--the structural perspective // *Cell. Mol. Life Sci.* 2006. Vol. 63, № 12. P. 1370–1385.
101. Finkina E.I. et al. Lipid Transfer Proteins As Components of the Plant Innate Immune System: Structure, Functions, and Applications // *Acta Naturae.* 2016. Vol. 8, № 2. P. 47–61.
102. Brand G.D. et al. Intragenic antimicrobial peptides (IAPs) from human proteins with potent antimicrobial and anti-inflammatory activity // *PLoS One.* 2019. Vol. 14, № 8. P. e0220656.
103. Ramada M.H.S. et al. Encrypted Antimicrobial Peptides from Plant Proteins // *Sci. Rep.* 2017. Vol. 7, № 1. P. 13263.
104. Tinoco A.B. et al. Leptins and leptin receptor expression in the goldfish (*Carassius auratus*). Regulation by food intake and fasting/overfeeding conditions // *Peptides.* 2012. Vol. 34, № 2. P. 329–335.
105. Vanhoye D. et al. Antimicrobial peptides from hylid and ranin frogs originated from a 150-million-year-old ancestral precursor with a conserved signal peptide but a hypermutable antimicrobial domain // *European Journal of Biochemistry.* 2003. Vol. 270, № 9. P. 2068–2081.
106. Brand G.D. et al. Probing protein sequences as sources for encrypted antimicrobial peptides // *PLoS One.* 2012. Vol. 7, № 9. P. e45848.
107. Ponti D. et al. An amphibian antimicrobial peptide variant expressed in *Nicotiana tabacum* confers resistance to phytopathogens // *Biochem. J.* 2003. Vol. 370, № Pt 1. P. 121–127.
108. Rivero M. et al. Stacking of antimicrobial genes in potato transgenic plants confers increased resistance to bacterial and fungal pathogens // *J. Biotechnol.* 2012. Vol. 157, № 2. P. 334–343.
109. Ryan C.A., Pearce G. Systemin: a polypeptide signal for plant defensive genes // *Annu. Rev. Cell Dev. Biol.* 1998. Vol. 14. P. 1–17.
110. Beloshistov R.E. et al. Phytaspase-mediated precursor processing and maturation of the wound hormone systemin // *New Phytol.* 2018. Vol. 218, № 3. P. 1167–1178.

111. Wang L. et al. The systemin receptor SYR1 enhances resistance of tomato against herbivorous insects // *Nat Plants*. 2018. Vol. 4, № 3. P. 152–156.
112. Huffaker A., Pearce G., Ryan C.A. An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response // *Proc. Natl. Acad. Sci. U. S. A.* 2006. Vol. 103, № 26. P. 10098–10103.
113. Yamaguchi Y. et al. PEPR2 is a second receptor for the Pep1 and Pep2 peptides and contributes to defense responses in *Arabidopsis* // *Plant Cell*. 2010. Vol. 22, № 2. P. 508–522.
114. Ziemann S. et al. An apoplastic peptide activates salicylic acid signalling in maize // *Nat Plants*. 2018. Vol. 4, № 3. P. 172–180.
115. Schmelz E.A. et al. Fragments of ATP synthase mediate plant perception of insect attack // *Proc. Natl. Acad. Sci. U. S. A.* 2006. Vol. 103, № 23. P. 8894–8899.
116. Schmelz E.A. et al. An amino acid substitution inhibits specialist herbivore production of an antagonist effector and recovers insect-induced plant defenses // *Plant Physiol*. 2012. Vol. 160, № 3. P. 1468–1478.
117. Steinbrenner A.D. et al. A receptor for herbivore-associated molecular patterns mediates plant immunity // *bioRxiv*. 2019. P. 679803.
118. Pearce G. et al. A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes // *Proc. Natl. Acad. Sci. U. S. A.* 2010. Vol. 107, № 33. P. 14921–14925.
119. Chen Y.-L. et al. Quantitative peptidomics study reveals that a wound-induced peptide from PR-1 regulates immune signaling in tomato // *Plant Cell*. 2014. Vol. 26, № 10. P. 4135–4148.
120. Chien P.-S., Nam H.G., Chen Y.-R. A salt-regulated peptide derived from the CAP superfamily protein negatively regulates salt-stress tolerance in *Arabidopsis* // *J. Exp. Bot.* 2015. Vol. 66, № 17. P. 5301–5313.
121. Choi H.W., Klessig D.F. DAMPs, MAMPs, and NAMPs in plant innate immunity // *BMC Plant Biol*. 2016. Vol. 16, № 1. P. 232.
122. Pearce G. et al. A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins // *Science*. 1991. Vol. 253, № 5022. P. 895–897.
123. Murphy E., Smith S., De Smet I. Small signaling peptides in *Arabidopsis* development: how cells communicate over a short distance // *Plant Cell*. 2012. Vol. 24, № 8. P. 3198–3217.
124. Takahashi F. et al. A small peptide modulates stomatal control via abscisic acid in long-distance signalling // *Nature*. 2018. Vol. 556, № 7700. P. 235–238.
125. Chater C.C. et al. Origin and function of stomata in the moss *Physcomitrella patens* // *Nat Plants*. 2016. Vol. 2. P. 16179.
126. Hara K. et al. The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule // *Genes Dev*. 2007. Vol. 21, № 14. P. 1720–1725.
127. Hara K. et al. Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in *Arabidopsis* leaves // *Plant Cell Physiol*. 2009. Vol. 50, № 6. P. 1019–1031.
128. Sugano S.S. et al. Stomagen positively regulates stomatal density in *Arabidopsis* // *Nature*. 2010. Vol. 463, № 7278. P. 241–244.
129. Takata N. et al. Evolutionary relationship and structural characterization of the EPF/EPFL gene family // *PLoS One*. 2013. Vol. 8, № 6. P. e65183.
130. Whitford R. et al. Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells // *Proc. Natl. Acad. Sci. U. S. A.* 2008. Vol. 105, № 47. P. 18625–18630.
131. Ito Y. et al. Dodeca-CLE peptides as suppressors of plant stem cell differentiation // *Science*. 2006. Vol. 313, № 5788. P. 842–845.
132. Goad D.M., Zhu C., Kellogg E.A. Comprehensive identification and clustering of CLV3/ESR-related (CLE) genes in plants finds groups with potentially shared function // *New Phytol*. 2017. Vol. 216, № 2. P. 605–616.
133. Whitewoods C.D. et al. CLAVATA Was a Genetic Novelty for the Morphological Innovation of 3D Growth in Land Plants // *Curr. Biol*. 2018. Vol. 28, № 15. P. 2365–2376.e5.
134. Oelkers K. et al. Bioinformatic analysis of the CLE signaling peptide family // *BMC Plant Biol*. 2008. Vol. 8. P. 1.
135. Banks J.A. et al. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants // *Science*. 2011. Vol. 332, № 6032. P. 960–963.
136. Bowman J.L. et al. Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome // *Cell*. 2017. Vol. 171, № 2. P. 287–304.e15.
137. Murphy E., De Smet I. Understanding the RALF family: a tale of many species // *Trends Plant Sci*. 2014. Vol. 19, № 10. P. 664–671.
138. Morato do Canto A. et al. Biological activity of nine recombinant AtRALF peptides: Implications for their perception and function in *Arabidopsis* // *Plant Physiol. Biochem*. 2014. Vol. 75. P. 45–54.

139. Galindo-Trigo S., Gray J.E., Smith L.M. Conserved Roles of CrRLK1L Receptor-Like Kinases in Cell Expansion and Reproduction from Algae to Angiosperms // *Front. Plant Sci.* 2016. Vol. 7. P. 1269.
140. Campbell L., Turner S.R. A Comprehensive Analysis of RALF Proteins in Green Plants Suggests There Are Two Distinct Functional Groups // *Front. Plant Sci.* 2017. Vol. 8. P. 37.
141. Honkanen S. et al. The Mechanism Forming the Cell Surface of Tip-Growing Rooting Cells Is Conserved among Land Plants // *Curr. Biol.* 2017. Vol. 27, № 20. P. 3224.
142. Neme R. et al. Random sequences are an abundant source of bioactive RNAs or peptides // *Nat Ecol Evol.* 2017. Vol. 1, № 6. P. 0217.
143. Bao Z. et al. Identification of Novel Growth Regulators in Plant Populations Expressing Random Peptides // *Plant Physiol.* 2017. Vol. 175, № 2. P. 619–627.
144. Ren Y. et al. Coding of Non-coding RNA: Insights Into the Regulatory Functions of Pri-MicroRNA-Encoded Peptides in Plants // *Front. Plant Sci.* 2021. Vol. 12. P. 641351.
145. van Heesch S. et al. The Translational Landscape of the Human Heart // *Cell.* 2019. Vol. 178, № 1. P. 242–260.e29.
146. Wu P. et al. Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA // *Mol. Cancer.* 2020. Vol. 19, № 1. P. 22.
147. Rubtsova M. et al. Protein encoded in human telomerase RNA is involved in cell protective pathways // *Nucleic Acids Res.* 2018. Vol. 46, № 17. P. 8966–8977.
148. Guo B. et al. Humanin peptide suppresses apoptosis by interfering with Bax activation // *Nature.* 2003. Vol. 423, № 6938. P. 456–461.
149. Pandita D. Plant MIRnome: miRNA Biogenesis and Abiotic Stress Response // *Plant Abiotic Stress Tolerance.* 2019. P. 449–474.
150. Laressergues D. et al. Primary transcripts of microRNAs encode regulatory peptides // *Nature.* 2015. Vol. 520, № 7545. P. 90–93.
151. Sharma A. et al. Primary transcript of miR858 encodes regulatory peptide and controls flavonoid biosynthesis and development in Arabidopsis // *Nat Plants.* 2020. Vol. 6, № 10. P. 1262–1274.
152. Couzigou J.-M. et al. Use of microRNA-encoded peptide miPEP172c to stimulate nodulation in soybean // *New Phytol.* 2016. Vol. 211, № 2. P. 379–381.
153. Chen Q.-J. et al. A miRNA-Encoded Small Peptide, vvi-miPEP171d1, Regulates Adventitious Root Formation // *Plant Physiol.* 2020. Vol. 183, № 2. P. 656–670.
154. Couzigou J.-M. et al. miRNA-encoded peptides (miPEPs): A new tool to analyze the roles of miRNAs in plant biology // *RNA Biol.* 2015. Vol. 12, № 11. P. 1178–1180.
155. Anderson D.M. et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance // *Cell.* 2015. Vol. 160, № 4. P. 595–606.
156. Matsumoto A. et al. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide // *Nature.* 2017. Vol. 541, № 7636. P. 228–232.
157. Polycarpou-Schwarz M. et al. The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation // *Oncogene.* 2018. Vol. 37, № 34. P. 4750–4768.
158. Guo B. et al. Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression // *EMBO J.* 2020. Vol. 39, № 1. P. e102190.
159. Chugunova A. et al. LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism // *Proc. Natl. Acad. Sci. U. S. A.* 2019. Vol. 116, № 11. P. 4940–4945.
160. Niu L. et al. A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation // *Sci Adv.* 2020. Vol. 6, № 21. P. eaaz2059.
161. Immarigeon C. et al. Identification of a micropeptide and multiple secondary cell genes that modulate *Drosophila* male reproductive success // *Proc. Natl. Acad. Sci. U. S. A.* 2021. Vol. 118, № 15.
162. Casson S.A. et al. The POLARIS Gene of Arabidopsis Encodes a Predicted Peptide Required for Correct Root Growth and Leaf Vascular Patterning // *Plant Cell. Oxford Academic.* 2002. Vol. 14, № 8. P. 1705–1721.
163. Narita N.N. et al. Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in *Arabidopsis thaliana* // *Plant J.* 2004. Vol. 38, № 4. P. 699–713.
164. Guo P. et al. Comparative analysis of the RTFL peptide family on the control of plant organogenesis // *J. Plant Res.* 2015. Vol. 128, № 3. P. 497–510.
165. Röhrig H. et al. Soybean ENOD40 encodes two peptides that bind to sucrose synthase // *Proc. Natl. Acad. Sci. U. S. A. National Academy of Sciences.* 2002. Vol. 99, № 4. P. 1915–1920.
166. Blanvillain R. et al. The Arabidopsis peptide kiss of death is an inducer of programmed cell death // *EMBO*

- J. 2011. Vol. 30, № 6. P. 1173–1183.
167. Frank M.J., Cartwright H.N., Smith L.G. Three Brick genes have distinct functions in a common pathway promoting polarized cell division and cell morphogenesis in the maize leaf epidermis // *Development*. 2003. Vol. 130, № 4. P. 753–762.
168. Dong X. et al. Zm908p11, encoded by a short open reading frame (sORF) gene, functions in pollen tube growth as a profilin ligand in maize // *J. Exp. Bot.* 2013. Vol. 64, № 8. P. 2359–2372.
169. Wang D. et al. Zm401p10, encoded by an anther-specific gene with short open reading frames, is essential for tapetum degeneration and anther development in maize // *Funct. Plant Biol.* 2009. Vol. 36, № 1. P. 73–85.
170. De Coninck B. et al. Mining the genome of *Arabidopsis thaliana* as a basis for the identification of novel bioactive peptides involved in oxidative stress tolerance // *J. Exp. Bot.* 2013. Vol. 64, № 17. P. 5297–5307.
171. Chilley P.M. et al. The POLARIS peptide of *Arabidopsis* regulates auxin transport and root growth via effects on ethylene signaling // *Plant Cell*. 2006. Vol. 18, № 11. P. 3058–3072.
172. Frank M.J., Smith L.G. A small, novel protein highly conserved in plants and animals promotes the polarized growth and division of maize leaf epidermal cells // *Curr. Biol.* 2002. Vol. 12, № 10. P. 849–853.
173. Calvo S.E., Pagliarini D.J., Mootha V.K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans // *Proc. Natl. Acad. Sci. U. S. A.* 2009. Vol. 106, № 18. P. 7507–7512.
174. Srivastava A.K. et al. UTR-Dependent Control of Gene Expression in Plants // *Trends Plant Sci.* 2018. Vol. 23, № 3. P. 248–259.
175. Spealman P. et al. Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data // *Genome Res.* 2018. Vol. 28, № 2. P. 214–222.
176. Gao X. et al. Quantitative profiling of initiating ribosomes in vivo // *Nat. Methods*. 2015. Vol. 12, № 2. P. 147–153.
177. Duncan C.D.S., Mata J. The translational landscape of fission-yeast meiosis and sporulation // *Nat. Struct. Mol. Biol.* 2014. Vol. 21, № 7. P. 641–647.
178. Fritsch C. et al. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting // *Genome Research*. 2012. Vol. 22, № 11. P. 2208–2218.
179. Hsu P.Y. et al. Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis* // *Proceedings of the National Academy of Sciences*. 2016. Vol. 113, № 45. P. E7126–E7135.
180. Zhang H. et al. Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during *Drosophila* development // *PLoS Biol.* 2018. Vol. 16, № 7. P. e2003903.
181. Kurihara Y. et al. Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis* // *Proc. Natl. Acad. Sci. U. S. A.* 2018. Vol. 115, № 30. P. 7831–7836.
182. Gerashchenko M.V., Lobanov A.V., Gladyshev V.N. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress // *Proc. Natl. Acad. Sci. U. S. A.* 2012. Vol. 109, № 43. P. 17394–17399.
183. Andreev D.E. et al. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression // *Elife*. 2015. Vol. 4. P. e03971.
184. Hanfrey C. et al. A dual upstream open reading frame-based autoregulatory circuit controlling polyamine-responsive translation // *J. Biol. Chem.* 2005. Vol. 280, № 47. P. 39229–39237.
185. Wiese A. et al. A conserved upstream open reading frame mediates sucrose-induced repression of translation // *Plant Cell*. 2004. Vol. 16, № 7. P. 1717–1729.
186. Laing W.A. et al. An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in *Arabidopsis* // *Plant Cell*. 2015. Vol. 27, № 3. P. 772–786.
187. Orr M.W. et al. Alternative ORFs and small ORFs: shedding light on the dark proteome // *Nucleic Acids Res.* 2020. Vol. 48, № 3. P. 1029–1042.
188. Lorenzo-Orts L. et al. Concerted expression of a cell cycle regulator and a metabolic enzyme from a bicistronic transcript in plants // *Nat Plants*. 2019. Vol. 5, № 2. P. 184–193.
189. Bazzini A.A. et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation // *EMBO J. Wiley*, 2014. Vol. 33, № 9. P. 981–993.
190. Mackowiak S.D. et al. Extensive identification and analysis of conserved small ORFs in animals // *Genome Biol.* 2015. Vol. 16. P. 179.
191. Schrader M., Schulz-Knappe P., Fricker L.D. Historical perspective of peptidomics // *EuPA Open Proteomics*. 2014. Vol. 3. P. 171–182.
192. Peng J. et al. Peptidomic analyses: The progress in enrichment and identification of endogenous peptides //

- Trends Analyt. Chem. 2020. Vol. 125. P. 115835.
193. Peng J., Wu R. 'an. Metal-organic frameworks in proteomics/peptidomics-A review // *Anal. Chim. Acta*. 2018. Vol. 1027. P. 9–21.
 194. Zheng X., Baker H., Hancock W.S. Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer // *J. Chromatogr. A*. 2006. Vol. 1120, № 1-2. P. 173–184.
 195. Goodwin R.J.A. Sample preparation for mass spectrometry imaging: small mistakes can lead to big consequences // *J. Proteomics*. 2012. Vol. 75, № 16. P. 4893–4911.
 196. Albalat A., Mischak H., Mullen W. Clinical application of urinary proteomics/peptidomics // *Expert Rev. Proteomics*. 2011. Vol. 8, № 5. P. 615–629.
 197. Dufresne J. et al. A method for the extraction of the endogenous tryptic peptides (peptidome) from human EDTA plasma // *Anal. Biochem*. 2018. Vol. 549. P. 188–196.
 198. Kawashima Y. et al. High-yield peptide-extraction method for the discovery of subnanomolar biomarkers from small serum samples // *J. Proteome Res*. 2010. Vol. 9, № 4. P. 1694–1705.
 199. Secher A. et al. Analytic framework for peptidomics applied to large-scale neuropeptide identification // *Nat. Commun*. 2016. Vol. 7. P. 11436.
 200. Núñez C. et al. Novel functionalized nanomaterials for the effective enrichment of proteins and peptides with post-translational modifications // *J. Proteomics*. 2018. Vol. 181. P. 170–189.
 201. Piovesana S. et al. Graphitized Carbon Black Enrichment and UHPLC-MS/MS Allow to Meet the Challenge of Small Chain Peptidomics in Urine // *Anal. Chem*. 2019. Vol. 91, № 17. P. 11474–11481.
 202. Kyte J., Doolittle R.F. A simple method for displaying the hydropathic character of a protein // *J. Mol. Biol*. 1982. Vol. 157, № 1. P. 105–132.
 203. Kanlaya R., Thongboonkerd V. Quantitative peptidomics of endogenous peptides involved in TGF- β 1-induced epithelial mesenchymal transition of renal epithelial cells // *Cell Death Discov*. 2018. Vol. 4. P. 9.
 204. Azkargorta M. et al. Human Basal Tear Peptidome Characterization by CID, HCD, and ETD Followed by in Silico and in Vitro Analyses for Antimicrobial Peptide Identification // *J. Proteome Res*. 2015. Vol. 14, № 6. P. 2649–2658.
 205. Sasaki K., Osaki T., Minamino N. Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry // *Mol. Cell. Proteomics*. 2013. Vol. 12, № 3. P. 700–709.
 206. Zeng Q. et al. Gastrin, Cholecystokinin, Signaling, and Biological Activities in Cellular Processes // *Front. Endocrinol.* . 2020. Vol. 11. P. 112.
 207. Czyzewicz N. et al. Message in a bottle: small signalling peptide outputs during growth and development // *J. Exp. Bot*. 2013. Vol. 64, № 17. P. 5281–5296.
 208. Parrish J.R., Gulyas K.D., Finley R.L. Jr. Yeast two-hybrid contributions to interactome mapping // *Curr. Opin. Biotechnol*. 2006. Vol. 17, № 4. P. 387–393.
 209. Paiano A. et al. Yeast two-hybrid assay to identify interacting proteins // *Curr. Protoc. Protein Sci*. Wiley, 2019. Vol. 95, № 1. P. e70.
 210. Chang I.-F. Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes // *Proteomics*. 2006. Vol. 6, № 23. P. 6158–6166.
 211. Richards A.L., Eckhardt M., Krogan N.J. Mass spectrometry-based protein-protein interaction networks for the study of human diseases // *Mol. Syst. Biol. EMBO*, 2021. Vol. 17, № 1. P. e8792.
 212. D’Lima N.G. et al. A human microprotein that interacts with the mRNA decapping complex // *Nat. Chem. Biol*. 2017. Vol. 13, № 2. P. 174–180.
 213. Arnoult N. et al. Regulation of DNA Repair pathway choice in S/G2 by the NHEJ inhibitor CYREN Europe PMC Funders Group // *Nature*. 2017.
 214. Agou F., Ye F., Véron M. In vivo protein cross-linking // *Methods Mol. Biol*. 2004. Vol. 261. P. 427–442.
 215. Mellacheruvu D. et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data // *Nat. Methods*. 2013. Vol. 10, № 8. P. 730–736.
 216. Collonnier C. et al. CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss *Physcomitrella patens* // *Plant Biotechnol. J. Wiley*, 2017. Vol. 15, № 1. P. 122–131.
 217. Aslanidis C., de Jong P.J. Ligation-independent cloning of PCR products (LIC-PCR) // *Nucleic Acids Res*. 1990. Vol. 18, № 20. P. 6069–6074.
 218. Kubo M. et al. System for stable β -estradiol-inducible gene expression in the moss *Physcomitrella patens* // *PLoS One*. 2013. Vol. 8, № 9. P. e77356.
 219. Schaefer D.G., Bisztray G., Zrýd J.P. Genetic Transformation of the Moss *Physcomitrella patens* // *Plant Protoplasts and Genetic Engineering V / ed. Bajaj Y.P.S. Berlin, Heidelberg: Springer Berlin Heidelberg*,

1994. P. 349–364.
220. Faurobert M., Pelpoir E., Chaïb J. Phenol extraction of proteins for proteomic studies of recalcitrant plant tissues // *Methods Mol. Biol.* 2007. Vol. 355. P. 9–14.
221. Laemmli U.K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4 // *Nature*. Nature Publishing Group, 1970. Vol. 227, № 5259. P. 680–685.
222. Schägger H. Tricine–SDS–PAGE // *Nat. Protoc.* Nature Publishing Group, 2006. Vol. 1, № 1. P. 16–22.
223. Ye J. et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction // *BMC Bioinformatics*. 2012. Vol. 13. P. 134.
224. Lanfear R. et al. MinIONQC: fast and simple quality control for MinION sequencing data // *Bioinformatics*. 2019. Vol. 35, № 3. P. 523–525.
225. Lang D. et al. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution // *Plant J. Wiley Online Library*, 2018. Vol. 93, № 3. P. 515–533.
226. Li H. Minimap2: pairwise alignment for nucleotide sequences // *Bioinformatics*. 2018. Vol. 34, № 18. P. 3094–3100.
227. Li H. et al. The Sequence Alignment/Map format and SAMtools // *Bioinformatics*. 2009. Vol. 25, № 16. P. 2078–2079.
228. Quinlan A.R. BEDTools: the Swiss-army tool for genome feature analysis // *Curr. Protoc. Bioinformatics*. Wiley Online Library, 2014. Vol. 47, № 1. P. 11–12.
229. Kovaka S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2 // *Genome Biol.* 2019. Vol. 20, № 1. P. 278.
230. Perteza G., Perteza M. GFF Utilities: GffRead and GffCompare // *F1000Res*. F1000 Research Limited, 2020. Vol. 9, № 304. P. 304.
231. Tang A.D. et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns // *Nat. Commun.* 2020. Vol. 11, № 1. P. 1438.
232. Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features // *Bioinformatics*. 2014. Vol. 30, № 7. P. 923–930.
233. Tyanova S., Temu T., Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics // *Nat. Protoc.* 2016. Vol. 11, № 12. P. 2301–2319.
234. Nesvizhskii A.I. Proteogenomics: concepts, applications and computational strategies // *Nat. Methods*. 2014. Vol. 11, № 11. P. 1114–1125.
235. Szcześniak M.W. et al. CANTATAdb 2.0: Expanding the Collection of Plant Long Noncoding RNAs // *Methods Mol. Biol.* 2019. Vol. 1933. P. 415–429.
236. Paytuví-Gallart A., Sanseverino W., Aiese Cigliano R. A Walkthrough to the Use of GreenNC: The Plant lncRNA Database // *Methods Mol. Biol.* 2019. Vol. 1933. P. 397–414.
237. Simopoulos C.M.A., Weretilnyk E.A., Golding G.B. Molecular Traits of Long Non-protein Coding RNAs from Diverse Plant Species Show Little Evidence of Phylogenetic Relationships // *G3*. 2019. Vol. 9, № 8. P. 2511–2520.
238. Hanada K. et al. sORF finder: a program package to identify small open reading frames with high coding potential // *Bioinformatics*. 2010. Vol. 26, № 3. P. 399–400.
239. Tarailo-Graovac M., Chen N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences // *Curr. Protoc. Bioinformatics*. 2009. Vol. 25, № 1.
240. Fesenko I. et al. Alternative splicing shapes transcriptome but not proteome diversity in *Physcomitrella patens* // *Sci. Rep.* 2017. Vol. 7, № 1. P. 2698.
241. Calviello L. et al. Detecting actively translated open reading frames in ribosome profiling data // *Nat. Methods*. 2016. Vol. 13, № 2. P. 165–170.
242. Samandi S. et al. Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins // *Elife*. 2017. Vol. 6.
243. Zhu M., Gribskov M. MiPepid: MicroPeptide identification tool using machine learning // *BMC Bioinformatics*. 2019. Vol. 20, № 1. P. 559.
244. Jenuth J.P. The NCBI // *Bioinformatics Methods and Protocols* / ed. Misener S., Krawetz S.A. Totowa, NJ: Humana Press, 1999. P. 301–312.
245. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants // *Nature*. 2019. Vol. 574, № 7780. P. 679–685.
246. Haas B.J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis // *Nat. Protoc.* 2013. Vol. 8, № 8. P. 1494–1512.
247. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // *Nucleic Acids Res.* 2004. Vol. 32, № 5. P. 1792–1797.

248. Finn R.D., Clements J., Eddy S.R. HMMER web server: interactive sequence similarity searching // *Nucleic Acids Res.* 2011. Vol. 39, № Web Server issue. P. W29–W37.
249. Suyama M., Torrents D., Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments // *Nucleic Acids Res.* 2006. Vol. 34, № Web Server issue. P. W609–W612.
250. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences // *J. Mol. Evol.* 1980. Vol. 16, № 2. P. 111–120.
251. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood // *Mol. Biol. Evol.* 2007. Vol. 24, № 8. P. 1586–1591.
252. Murrell B. et al. Gene-wide identification of episodic selection // *Mol. Biol. Evol.* 2015. Vol. 32, № 5. P. 1365–1371.
253. Nguyen L.-T. et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies // *Mol. Biol. Evol.* 2015. Vol. 32, № 1. P. 268–274.
254. Thissen D., Steinberg L., Kuang D. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons // *J. Educ. Behav. Stat. American Educational Research Association*, 2002. Vol. 27, № 1. P. 77–83.
255. Eaton D.A.R. Toytree: A minimalist tree visualization and manipulation library for Python // *Methods Ecol. Evol.* Wiley, 2020. Vol. 11, № 1. P. 187–191.
256. Van Rossum G., Drake F.L. Jr. Python tutorial: Centrum voor Wiskunde en Informatica Amsterdam // *The Netherlands*. 1995.
257. Virtanen P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python // *Nat. Methods*. 2020. Vol. 17, № 3. P. 261–272.
258. Waskom M. seaborn: statistical data visualization // *J. Open Source Softw. The Open Journal*, 2021. Vol. 6, № 60. P. 3021.
259. McKinney W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. “O’Reilly Media, Inc.,” 2012. 466 p.
260. Lex A. et al. UpSet: Visualization of Intersecting Sets // *IEEE Trans. Vis. Comput. Graph.* 2014. Vol. 20, № 12. P. 1983–1992.
261. Quevillon E. et al. InterProScan: protein domains identifier // *Nucleic Acids Res.* 2005. Vol. 33, № Web Server issue. P. W116–W120.
262. Armenteros J.J.A. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks // *Nat. Biotechnol.* Nature Publishing Group, 2019. Vol. 37, № 4. P. 420–423.
263. Krogh A. et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes // *J. Mol. Biol.* 2001. Vol. 305, № 3. P. 567–580.
264. Jumper J. et al. Highly accurate protein structure prediction with AlphaFold // *Nature*. 2021. Vol. 596, № 7873. P. 583–589.
265. Moretti S. et al. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods // *Nucleic Acids Res.* Oxford University Press, 2007. Vol. 35, № suppl_2. P. W645–W648.
266. Hong-Bo S., Zong-Suo L., Ming-An S. LEA proteins in higher plants: structure, function, gene expression and regulation // *Colloids Surf. B Biointerfaces*. 2005. Vol. 45, № 3-4. P. 131–135.
267. Nelson C.J. et al. Proteins with high turnover rate in barley leaves estimated by proteome analysis combined with in planta isotope labeling // *Plant Physiol.* 2014. Vol. 166, № 1. P. 91–108.
268. Khanna-Chopra R. Leaf senescence and abiotic stresses share reactive oxygen species-mediated chloroplast degradation // *Protoplasma*. 2012. Vol. 249, № 3. P. 469–481.
269. Dure L., Galau G.A. Developmental Biochemistry of Cottonseed Embryogenesis and Germination : XIII. REGULATION OF BIOSYNTHESIS OF PRINCIPAL STORAGE PROTEINS // *Plant Physiol.* 1981. Vol. 68, № 1. P. 187–194.
270. Battaglia M. et al. The enigmatic LEA proteins and other hydrophilins // *Plant Physiol.* 2008. Vol. 148, № 1. P. 6–24.
271. Cui S. et al. Proteome analysis of *Physcomitrella patens* exposed to progressive dehydration and rehydration // *J. Exp. Bot.* 2012. Vol. 63, № 2. P. 711–726.
272. Furuki T. et al. Effects of Group 3 LEA protein model peptides on desiccation-induced protein aggregation // *Biochim. Biophys. Acta*. 2012. Vol. 1824, № 7. P. 891–897.
273. Lehtonen M.T. et al. Protein secretome of moss plants (*Physcomitrella patens*) with emphasis on changes induced by a fungal elicitor // *J. Proteome Res.* 2014. Vol. 13, № 2. P. 447–459.
274. Hoernstein S.N.W. et al. Host Cell Proteome of *Physcomitrella patens* Harbors Proteases and Protease

- Inhibitors under Bioproduction Conditions // *J. Proteome Res.* 2018. Vol. 17, № 11. P. 3749–3760.
275. Filippova A. et al. Salicylic acid influences the protease activity and posttranslational modifications of the secreted peptides in the moss *Physcomitrella patens* // *J. Pept. Sci.* 2019. Vol. 25, № 2. P. e3138.
276. Magalhães B. et al. Reviewing Mechanistic Peptidomics in Body Fluids Focusing on Proteases // *Proteomics.* 2018. Vol. 18, № 18. P. e1800187.
277. Klein J. et al. Proteasix: a tool for automated and large-scale prediction of proteases involved in naturally occurring peptide generation // *Proteomics.* Wiley, 2013. Vol. 13, № 7. P. 1077–1082.
278. Song J. et al. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites // *PLoS One.* 2012. Vol. 7, № 11. P. e50300.
279. Thomas E.L., van der Hoorn R.A.L. Ten Prominent Host Proteases in Plant-Pathogen Interactions // *Int. J. Mol. Sci.* 2018. Vol. 19, № 2.
280. Guo R. et al. Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape // *BMC Genomics.* 2013. Vol. 14. P. 554.
281. Timotijević G.S. et al. Ubiquitous aspartic proteinase as an actor in the stress response in buckwheat // *J. Plant Physiol.* 2010. Vol. 167, № 1. P. 61–68.
282. Ramírez V. et al. An extracellular subtilase switch for immune priming in *Arabidopsis* // *PLoS Pathog.* 2013. Vol. 9, № 6. P. e1003445.
283. Stacey G. et al. Peptide transport in plants // *Trends Plant Sci.* 2002. Vol. 7, № 6. P. 257–263.
284. Abele R., Tampé R. Moving the Cellular Peptidome by Transporters // *Front Cell Dev Biol.* 2018. Vol. 6. P. 43.
285. Mamaeva A. et al. The role of chloroplast protein remodeling in stress responses and shaping of the plant peptidome // *New Phytol.* 2020. Vol. 227, № 5. P. 1326–1334.
286. Üstün S. et al. The Proteasome Acts as a Hub for Plant Immunity and Is Targeted by *Pseudomonas* Type III Effectors // *Plant Physiol.* 2016. Vol. 172, № 3. P. 1941–1958.
287. Engelsdorf T. et al. The plant cell wall integrity maintenance and immune signaling systems cooperate to control stress responses in *Arabidopsis thaliana* // *Sci. Signal.* 2018. Vol. 11, № 536.
288. Xiao L. et al. Transcriptome of protoplasts reprogrammed into stem cells in *Physcomitrella patens* // *PLoS One.* 2012. Vol. 7, № 4. P. e35961.
289. Prins A. et al. Cysteine proteinases regulate chloroplast protein content and composition in tobacco leaves: a model for dynamic interactions with ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) vesicular bodies // *J. Exp. Bot.* 2008. Vol. 59, № 7. P. 1935–1950.
290. Attaran E. et al. Temporal Dynamics of Growth and Photosynthesis Suppression in Response to Jasmonate Signaling // *Plant Physiol.* 2014. Vol. 165, № 3. P. 1302–1314.
291. Phukan U.J., Jeena G.S., Shukla R.K. WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants // *Front. Plant Sci.* 2016. Vol. 7. P. 760.
292. Feussner I., Wasternack C. The lipoxygenase pathway // *Annu. Rev. Plant Biol.* 2002. Vol. 53. P. 275–297.
293. Ponce De León I. et al. *Physcomitrella patens* activates reinforcement of the cell wall, programmed cell death and accumulation of evolutionary conserved defence signals, such as salicylic acid and 12-oxo-phytodienoic acid, but not jasmonic acid, upon *Botrytis cinerea* infection // *Mol. Plant Pathol.* Wiley, 2012. Vol. 13, № 8. P. 960–974.
294. Schnell Ramos M. et al. Characterization of a transport activity for long-chain peptides in barley mesophyll vacuoles // *J. Exp. Bot. Oxford Academic,* 2011. Vol. 62, № 7. P. 2403–2410.
295. Kurepa J. et al. Proteasome regulation, plant growth and stress tolerance // *Plant Signal. Behav.* 2009. Vol. 4, № 10. P. 924–927.
296. Üstün S. et al. Bacteria Exploit Autophagy for Proteasome Degradation and Enhanced Virulence in Plants // *Plant Cell.* 2018. Vol. 30, № 3. P. 668–685.
297. Marcotullio M.C. Qualitative and Quantitative Analysis of Bioactive Natural Products 2018. MDPI, 2019. 374 p.
298. Chichkova N.V. et al. Phytaspase, a relocatable cell death promoting plant protease with caspase specificity // *EMBO J.* 2010. Vol. 29, № 6. P. 1149–1161.
299. Vartapetian A.B. et al. A plant alternative to animal caspases: subtilisin-like proteases // *Cell Death Differ.* 2011. Vol. 18, № 8. P. 1289–1297.
300. Krause C. et al. Plant secretome - from cellular process to biological activity // *Biochim. Biophys. Acta.* Elsevier BV, 2013. Vol. 1834, № 11. P. 2429–2441.
301. Farrokhi N., Whitelegge J.P., Brusslan J.A. Plant peptides and peptidomics // *Plant Biotechnol. J.* 2008. Vol. 6, № 2. P. 105–134.
302. Porto W.F., Pires A.S., Franco O.L. Computational tools for exploring sequence databases as a resource for

- antimicrobial peptides // *Biotechnol. Adv.* 2017. Vol. 35, № 3. P. 337–349.
303. Oh E., Seo P.J., Kim J. Signaling Peptides and Receptors Coordinating Plant Root Development // *Trends Plant Sci.* 2018. Vol. 23, № 4. P. 337–351.
304. Patel N. et al. Diverse Peptide Hormones Affecting Root Growth Identified in the *Medicago truncatula* Secreted Peptidome // *Mol. Cell. Proteomics.* 2018. Vol. 17, № 1. P. 160–174.
305. Parsons J. et al. A gene responsible for prolyl-hydroxylation of moss-produced recombinant human erythropoietin // *Sci. Rep.* 2013. Vol. 3. P. 3019.
306. Yamaguchi Y., Huffaker A. Endogenous peptide elicitors in higher plants // *Curr. Opin. Plant Biol.* 2011. Vol. 14, № 4. P. 351–357.
307. Guillot A. et al. Mass Spectrometry Analysis of the Extracellular Peptidome of *Lactococcus lactis*: Lines of Evidence for the Coexistence of Extracellular Protein Hydrolysis and Intracellular Peptide Excretion // *J. Proteome Res.* 2016. Vol. 15, № 9. P. 3214–3224.
308. Yuan N. et al. Peptidomics-based study reveals that GAPEP1, a novel small peptide derived from pathogenesis-related (PR) protein of cotton, enhances fungal disease resistance // *Molecular Breeding.* 2019. Vol. 39, № 10–11.
309. Harno E. et al. POMC: The Physiological Power of Hormone Processing // *Physiol. Rev.* 2018. Vol. 98, № 4. P. 2381–2430.
310. Cock P.J.A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics // *Bioinformatics.* 2009. Vol. 25, № 11. P. 1422–1423.
311. Wang S. et al. Large-Scale Discovery of Non-conventional Peptides in Maize and Arabidopsis through an Integrated Peptidogenomic Pipeline // *Mol. Plant.* 2020. Vol. 13, № 7. P. 1078–1093.
312. Fesenko I.A. et al. Specific pools of endogenous peptides are present in gametophore, protonema, and protoplast cells of the moss *Physcomitrella patens* // *BMC Plant Biol.* 2015. Vol. 15. P. 87.
313. Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells // *J. Mol. Biol.* 1987. Vol. 196, № 4. P. 947–950.
314. Carvunis A.-R. et al. Proto-genes and de novo gene birth // *Nature.* 2012. Vol. 487, № 7407. P. 370–374.
315. Van Oss S.B., Carvunis A.-R. De novo gene birth // *PLoS Genet.* 2019. Vol. 15, № 5. P. e1008160.
316. Vakirlis N. et al. De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences // *Nat. Commun.* 2020. Vol. 11, № 1. P. 781.
317. Lange A. et al. Structural and functional characterization of a putative de novo gene in *Drosophila* // *Nat. Commun.* 2021. Vol. 12, № 1. P. 1667.
318. Hezroni H. et al. A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes // *Genome Biol.* 2017. Vol. 18, № 1. P. 162.
319. Ingolia N.T. et al. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments // *Nat. Protoc.* 2012. Vol. 7, № 8. P. 1534–1550.
320. Guttman M. et al. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins // *Cell.* 2013. Vol. 154, № 1. P. 240–251.
321. Prabakaran S. et al. Quantitative profiling of peptides from RNAs classified as noncoding // *Nat. Commun.* 2014. Vol. 5. P. 5429.
322. Verheggen K. et al. Noncoding after All: Biases in Proteomics Data Do Not Explain Observed Absence of lncRNA Translation Products // *J. Proteome Res.* 2017. Vol. 16, № 7. P. 2508–2515.
323. Gallaher S.D. et al. Widespread polycistronic gene expression in green algae // *Proc. Natl. Acad. Sci. U. S. A.* 2021. Vol. 118, № 7.
324. Delcourt V. et al. The Protein Coded by a Short Open Reading Frame, Not by the Annotated Coding Sequence, Is the Main Gene Product of the Dual-Coding Gene MIEF1 // *Mol. Cell. Proteomics.* 2018. Vol. 17, № 12. P. 2402–2411.
325. Brunet M.A. et al. The FUS gene is dual-coding with both proteins contributing to FUS-mediated toxicity // *EMBO Rep.* 2021. Vol. 22, № 1. P. e50640.
326. Loughran G. et al. Unusually efficient CUG initiation of an overlapping reading frame in POLG mRNA yields novel protein POLGARF // *Proc. Natl. Acad. Sci. U. S. A.* 2020. Vol. 117, № 40. P. 24936–24946.
327. Chang C.-Y., Lin W.-D., Tu S.-L. Genome-Wide Analysis of Heat-Sensitive Alternative Splicing in *Physcomitrella patens* // *Plant Physiol.* 2014. Vol. 165, № 2. P. 826–840.
328. Lloyd J.P.B. The evolution and diversity of the nonsense-mediated mRNA decay pathway // *F1000Res.* 2018. Vol. 7. P. 1299.
329. Karousis E.D., Nasif S., Mühlemann O. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact // *Wiley Interdiscip. Rev. RNA.* Wiley, 2016. Vol. 7, № 5. P. 661–682.
330. Ge Y., Porse B.T. The functional consequences of intron retention: alternative splicing coupled to NMD as

- a regulator of gene expression // *Bioessays*. 2014. Vol. 36, № 3. P. 236–243.
331. Lloyd J.P.B. et al. The loss of SMG1 causes defects in quality control pathways in *Physcomitrella patens* // *Nucleic Acids Res.* 2018. Vol. 46, № 11. P. 5822–5836.
332. Seo P.J. et al. Competitive inhibition of transcription factors by small interfering peptides // *Trends Plant Sci.* 2011. Vol. 16, № 10. P. 541–549.
333. Staudt A.-C., Wenkel S. Regulation of protein function by “microProteins” // *EMBO Rep.* 2011. Vol. 12, № 1. P. 35–42.
334. Eguen T. et al. MicroProteins: small size – big impact // *Trends in Plant Science.* 2015. Vol. 20, № 8. P. 477–482.
335. Basile W. et al. High GC content causes orphan proteins to be intrinsically disordered // *PLoS Comput. Biol.* 2017. Vol. 13, № 3. P. e1005375.
336. Heames B., Schmitz J., Bornberg-Bauer E. A Continuum of Evolving De Novo Genes Drives Protein-Coding Novelty in *Drosophila* // *J. Mol. Evol.* 2020. Vol. 88, № 4. P. 382–398.
337. Wilson B.A. et al. Young genes are highly disordered as predicted by the preadaptation hypothesis of de novo gene birth // *Nature Ecology & Evolution.* 2017. Vol. 1, № 6. P. 0146.
338. Li Y.-R., Liu M.-J. Prevalence of alternative AUG and non-AUG translation initiators and their regulatory effects across plants // *Genome Res.* 2020. Vol. 30, № 10. P. 1418–1433.
339. Hazarika R.R. et al. ARA-PEPs: a repository of putative sORF-encoded peptides in *Arabidopsis thaliana* // *BMC Bioinformatics.* 2017. Vol. 18, № 1. P. 37.
340. Mergner J. et al. Mass-spectrometry-based draft of the *Arabidopsis* proteome // *Nature.* 2020. Vol. 579, № 7799. P. 409–414.
341. Hezroni H. et al. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species // *Cell Rep.* 2015. Vol. 11, № 7. P. 1110–1122.
342. Ruiz-Orera J., Albà M.M. Conserved regions in long non-coding RNAs contain abundant translation and protein–RNA interaction signatures // *NAR Genom Bioinform.* Oxford Academic, 2019. Vol. 1, № 1. P. e2–e2.
343. Wissler L., Godmann L. Evolutionary dynamics of simple sequence repeats across long evolutionary time scale in genus *Drosophila* // *Trends Evol. Biol.* pagepress.org, 2012.
344. Keeling D.M. et al. Philosophy of Biology: The meanings of 'function' in biology and the problematic case of de novo gene emergence // *Elife.* eLife Sciences Publications Limited, 2019. Vol. 8. P. e47014.
345. Chaudhary R. et al. Prosurvival long noncoding RNA PINCR regulates a subset of p53 targets in human colorectal cancer cells by binding to Matrin 3 // *Elife.* 2017. Vol. 6.
346. Managadze D. et al. The vast, conserved mammalian lincRNome // *PLoS Comput. Biol.* 2013. Vol. 9, № 2. P. e1002917.
347. Mortz M. et al. Comparative genomic analysis identifies small open reading frames (sORFs) with peptide-encoding features in avian 16S rDNA // *Genomics.* 2020. Vol. 112, № 2. P. 1120–1127.
348. Shabalina S.A. et al. Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes // *Mol. Biol. Evol.* 2010. Vol. 27, № 8. P. 1745–1749.
349. Shabalina S.A. et al. Evolution at protein ends: major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals // *Nucleic Acids Res.* 2014. Vol. 42, № 11. P. 7132–7144.
350. Li X.L. et al. A small protein encoded by a putative lincRNA regulates apoptosis and tumorigenicity in human colorectal cancer cells // *Elife.* 2020. Vol. 9.
351. Bazin J. et al. Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation // *Proc. Natl. Acad. Sci. U. S. A.* 2017. Vol. 114, № 46. P. E10018–E10027.
352. Zeng C., Fukunaga T., Hamada M. Identification and analysis of ribosome-associated lincRNAs using ribosome profiling data // *BMC Genomics.* 2018. Vol. 19, № 1. P. 414.
353. Toll-Riera M. et al. Role of low-complexity sequences in the formation of novel protein coding sequences // *Mol. Biol. Evol.* 2012. Vol. 29, № 3. P. 883–886.
354. Radó-Trilla N., Albà M. Dissecting the role of low-complexity regions in the evolution of vertebrate proteins // *BMC Evol. Biol.* 2012. Vol. 12. P. 155.
355. Jarnot P. et al. LCR-BLAST—A New Modification of BLAST to Search for Similar Low Complexity Regions in Protein Sequences // *Man-Machine Interactions 6.* Springer International Publishing, 2020. P. 169–180.
356. Wootton J.C., Federhen S. Statistics of local complexity in amino acid sequences and sequence databases // *Computers & Chemistry.* 1993. Vol. 17, № 2. P. 149–163.
357. Singh S., Mittal A. Transmembrane Domain Lengths Serve as Signatures of Organismal Complexity and

- Viral Transport Mechanisms // *Sci. Rep.* 2016. Vol. 6. P. 22352.
358. Hershberg R., Petrov D.A. Evidence that mutation is universally biased towards AT in bacteria // *PLoS Genet.* 2010. Vol. 6, № 9. P. e1001115.
359. Boschiero C. et al. Identification and functional investigation of genome-encoded, small, secreted peptides in plants // *Curr. Protoc. Plant Biol.* Wiley, 2019. Vol. 4, № 3. P. e20098.
360. Colling J. et al. Overexpression of the Arabidopsis thaliana signalling peptide TAXIMIN1 affects lateral organ development // *J. Exp. Bot.* 2015. Vol. 66, № 17. P. 5337–5349.
361. Edqvist J. et al. Plant lipid transfer proteins: are we finally closing in on the roles of these enigmatic proteins? // *J. Lipid Res.* 2018. Vol. 59, № 8. P. 1374–1382.
362. Parker M.T. et al. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification // *Elife.* 2020. Vol. 9.
363. Zhang S. et al. New insights into Arabidopsis transcriptome complexity revealed by direct sequencing of native RNAs // *Nucleic Acids Res.* 2020. Vol. 48, № 14. P. 7700–7711.
364. Kirov I. et al. Nanopore RNA Sequencing Revealed Long Non-Coding and LTR Retrotransposon-Related RNAs Expressed at Early Stages of Triticale SEED Development // *Plants.* 2020. Vol. 9, № 12.
365. Michel A.M., Andreev D.E., Baranov P.V. Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning // *BMC Bioinformatics.* 2014. Vol. 15. P. 380.
366. Wu H., Yang L., Chen L.-L. The Diversity of Long Noncoding RNAs and Their Generation // *Trends Genet.* 2017. Vol. 33, № 8. P. 540–552.
367. Palazzo A.F., Koonin E.V. Functional Long Non-coding RNAs Evolve from Junk Transcripts // *Cell.* 2020.
368. Zhang J., Yang J.-R. Determinants of the rate of protein sequence evolution // *Nat. Rev. Genet.* 2015. Vol. 16, № 7. P. 409–420.
369. Ruiz Cuevas M.V. et al. Most non-canonical proteins uniquely populate the proteome or immunopeptidome // *Cell Rep.* 2021. Vol. 34, № 10. P. 108815.
370. Miravet-Verde S. et al. Unraveling the hidden universe of small proteins in bacterial genomes // *Mol. Syst. Biol.* 2019. Vol. 15, № 2. P. e8290.
371. Torres M.D.T. et al. Mining for encrypted peptide antibiotics in the human proteome // *Nat Biomed Eng.* 2022. Vol. 6, № 1. P. 67–75.
372. Torrent M. et al. AMPA: an automated web server for prediction of protein antimicrobial regions // *Bioinformatics.* 2012. Vol. 28, № 1. P. 130–131.
373. Blank L. et al. Variables Associated with Severity of Bacterial Canker and Wilt Caused by *Clavibacter michiganensis* subsp. *michiganensis* in Tomato Greenhouses // *Phytopathology.* 2016. Vol. 106, № 3. P. 254–261.
374. Bergsson G. et al. Isolation and identification of antimicrobial components from the epidermal mucus of Atlantic cod (*Gadus morhua*) // *FEBS J.* 2005. Vol. 272, № 19. P. 4960–4969.
375. Fernandes J.M.O., Smith V.J. A novel antimicrobial function for a ribosomal peptide from rainbow trout skin // *Biochem. Biophys. Res. Commun.* 2002. Vol. 296, № 1. P. 167–171.
376. Vidali L., Bezanilla M. *Physcomitrella patens*: a model for tip cell growth and differentiation // *Curr. Opin. Plant Biol.* 2012. Vol. 15, № 6. P. 625–631.
377. Jang G., Dolan L. Auxin promotes the transition from chloronema to caulonema in moss protonema by positively regulating PpRSL1 and PpRSL2 in *Physcomitrella patens* // *New Phytol.* 2011. Vol. 192, № 2. P. 319–327.
378. Kavi Kishor P.B. et al. Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny // *Front. Plant Sci.* 2015. Vol. 6. P. 544.
379. Pearce G. Systemin, hydroxyproline-rich systemin and the induction of protease inhibitors // *Curr. Protein Pept. Sci.* 2011. Vol. 12, № 5. P. 399–408.
380. Singh A., Vivek A.T., Kumar S. AlnC: An extensive database of long non-coding RNAs in angiosperms // *PLoS One.* 2021. Vol. 16, № 4. P. e0247215.
381. Tran L.T., Taylor J.S., Constabel C.P. The polyphenol oxidase gene family in land plants: Lineage-specific duplication and expansion // *BMC Genomics.* 2012. Vol. 13. P. 395.
382. Fujimoto M., Tsutsumi N. Dynamin-related proteins in plant post-Golgi traffic // *Front. Plant Sci.* 2014. Vol. 5. P. 408.
383. Jilly R. et al. Dynamin-Like Proteins Are Potentially Involved in Membrane Dynamics within Chloroplasts and Cyanobacteria // *Front. Plant Sci.* 2018. Vol. 9. P. 206.
384. Kwon C., Lee J.-H., Yun H.S. SNAREs in Plant Biotic and Abiotic Stress Responses // *Mol. Cells.* 2020. Vol. 43, № 6. P. 501–508.

385. Lycett G. The role of Rab GTPases in cell wall metabolism // *J. Exp. Bot.* 2008. Vol. 59, № 15. P. 4061–4074.
386. Ma Q.-H. Small GTP-binding Proteins and their Functions in Plants // *J. Plant Growth Regul.* 2007. Vol. 26, № 4. P. 369–388.
387. Bhui T., Roy J.K. Rab proteins: the key regulators of intracellular vesicle transport // *Exp. Cell Res.* 2014. Vol. 328, № 1. P. 1–19.
388. Orr R.G. et al. Orchestrating cell morphology from the inside out--using polarized cell expansion in plants as a model // *Curr. Opin. Cell Biol.* Elsevier, 2020. Vol. 62. P. 46–53.
389. Rensing S.A. et al. The Moss *Physcomitrium* (*Physcomitrella*) patens: A Model Organism for Non-Seed Plants // *Plant Cell.* 2020. Vol. 32, № 5. P. 1361–1376.
390. Minamino N., Ueda T. RAB GTPases and their effectors in plant endosomal transport // *Curr. Opin. Plant Biol.* 2019. Vol. 52. P. 61–68.
391. Jelínková A. et al. Probing plant membranes with FM dyes: tracking, dragging or blocking? // *Plant J.* 2010. Vol. 61, № 5. P. 883–892.
392. Šamaj J. et al. The endocytic network in plants // *Trends Cell Biol.* 2005. Vol. 15, № 8. P. 425–433.
393. Pleskot R. et al. Turnover of Phosphatidic Acid through Distinct Signaling Pathways Affects Multiple Aspects of Pollen Tube Growth in Tobacco // *Front. Plant Sci.* 2012. Vol. 3. P. 54.
394. Bove J. et al. Magnitude and direction of vesicle dynamics in growing pollen tubes using spatiotemporal image correlation spectroscopy and fluorescence recovery after photobleaching // *Plant Physiol.* 2008. Vol. 147, № 4. P. 1646–1658.