

УДК 581.1

***Physcomitrium patens* – МОДЕЛЬ ДЛЯ ИЗУЧЕНИЯ ЭВОЛЮЦИИ БЕЛКОВ С ЛЕКТИНОВЫМИ ДОМЕНАМИ У РАСТЕНИЙ¹**© 2024 г. А. Р. Агълямова^{а,*}, А. Р. Хакимова^а, О. В. Горшков^а, Т. А. Горшкова^а^аКазанский институт биохимии и биофизики Федерального исследовательского центра

“Казанский научный центр Российской академии наук”, Казань, Россия

*e-mail: aliaglyamova@yandex.ru

Поступила в редакцию 30.08.2023 г.

После доработки 03.10.2023 г.

Принята к публикации 04.10.2023 г.

Мох *Physcomitrium* (ранее *Physcomitrella*) *patens* (Hedw.) Mitt. — бессемянное и бессосудистое растение с расшифрованным геномом, представитель наиболее древних из ныне живущих таксонов наземных растений — удобная модель для изучения эволюционного развития растений. С целью изучения формирования набора и функций углеводов-связывающих белков у растений в ходе эволюции проведен полногеномный скрининг генов, кодирующих белки с лектиновыми доменами, в геноме *P. patens*, и проанализирована их экспрессия в различных клетках и тканях мха. Выявлен 141 ген, кодирующий белки из 15 семейств, набор и число представителей которых существенно отличались от проанализированных ранее покрытосеменных растений. У *P. patens* некоторые из белков с лектиновыми доменами обладали специфичной доменной архитектурой, не представленной у высших семенных растений. Кластеризация генов по уровню их экспрессии в различных тканях мха выявила три паттерна экспрессии генов белков с лектиновыми доменами, из которых третий кластер, представленный в клетках с концевым типом роста (в каулонеме, хлоронеме и ризоидах мха), характеризовался наибольшим количеством активно экспрессирующихся генов. Полученные результаты подтверждают идею о раннем появлении у растений генов, кодирующих лектины, и дальнейшем расширении семейств белков с лектиновыми доменами с усложнением организации растений.

Ключевые слова: *Physcomitrium patens*, лектины, транскриптом, филогения**DOI:** 10.31857/S0015330324020061, **EDN:** OBJKNV**ВВЕДЕНИЕ**

Лектины — углеводов-связывающие белки, которые являются универсальными участниками метаболизма живых организмов. Благодаря наличию доменов, способных специфически взаимодействовать с полисахаридами разнообразной структуры (т.н. лектиновых доменов), эти белки принимают участие в самых разных процессах жизнедеятельности, в которых задействованы углеводы. Многие растительные лектины являются частью различных сигнальных путей, участвуя тем самым в регуляции роста и развития растения, синтезе и деградации белков и других полимеров, реализации иммунной реакции в ответ на вторжение патогена, что делает их важными участниками процессов развития и адаптации растений [1].

Популярной моделью для изучения эволюционного развития растений является мох *Physcomitrium* (ранее *Physcomitrella*) *patens* (Hedw.) Mitt., благодаря наличию расшифрованного генома и простоте получения трансгенных растений. Несмотря на простое строение (отсутствие сосудов, настоящих корней, стеблей, листьев, цветков и семян), этот вид мха обладает многими сигнальными путями, характерными для высших сосудистых растений [2]. Мхи эволюционно представляют собой обособленную сестринскую группу по отношению к высшим семенным растениям, и потому являются эффективной моделью для отслеживания изменений в процессе эволюции наземных растений [3].

Ранее некоторые лектины *P. patens* упоминались в исследованиях, связанных с анализом разнообразия доменной организации растительных лектинов, либо с анализом эволюционного развития того или иного семейства, однако не проводилось полноценного исследования, нацеленного на характеристику всего набора генов лектинов *P. patens* [4–8]. Анализ генов, кодиру-

¹ Дополнительные материалы размещены в электронном виде по DOI статьи: 10.31857/S0015330324020061

Сокращения: CrRLK1L — белки, подобные рецептор-подобной киназе 1 *Catharantus roseus* (от англ. *Catharantus roseus* Receptor-Like Kinase 1-like).

ющих белки с лектиновыми доменами, в геноме этого мха представляет интерес с точки зрения изучения эволюции этих белков в контексте эволюции растений.

С целью изучения различий в наборе белков с лектиновыми доменами у мха *P. patens* и высших семенных растений, мы идентифицировали в его геноме гены, кодирующие такие белки из 15 различных семейств, и провели анализ их экспрессии в различных тканях. Охарактеризована доменная организация и филогения белков с лектиновыми доменами, выявлены особенности их структурной организации. Полученные результаты обсуждаются с точки зрения эволюции лектинов с развитием и усложнением организации растений.

МАТЕРИАЛЫ И МЕТОДЫ

Полногеномный скрининг. Гены мха *Physcomitrium patens* (Hedw.) Mitt., кодирующие предполагаемые белки с лектиновыми доменами, были идентифицированы в базе данных растительных геномов Phytozome v13 (сборка генома *Physcomitrium patens* v3.3) по номеру характерных pfam-доменов соответствующих белков, взятых из базы данных Pfam 34.0 [9, 10]. Для семейств CRA и EUL вместо номера pfam были использованы идентификационный номер базы данных консервативных доменов NCBI или идентификационный номер базы данных PANTHER соответственно [11, 12].

Анализ доменной организации и предсказание внутриклеточной локализации. Анализ доменной организации предсказанных белков проведен на основе аминокислотных последовательностей, взятых из базы данных Phytozome v13, с использованием веб-инструмента InterProScan [13]. Белковые последовательности некоторых генов были повторно предсказаны с использованием веб-сервиса Augustus [14]. Идентифицированные белковые последовательности проверялись на наличие сигнального пептида (веб-сервис SignalP-5.0) и трансмембранного домена (веб-сервис TMHMM v2.0) [15, 16]. Предполагаемая субклеточная локализация белков с лектиновыми доменами была предсказана с использованием веб-сервисов LocTree3 и DeepLoc-1.0 [17, 18]. Полные результаты анализа представлены в табл. 1. Дополнительные материалы.

Филогенетический анализ. Множественное выравнивание аминокислотных последовательностей было выполнено в веб-сервисах Clustal Omega и MUSCLE [19, 20]. Белковые последовательности резуховидки Таля (*Arabidopsis thaliana*) и кукурузы (*Zea mays*), использованные в анализе, были взяты из базы данных Phytozome v13 и ранее полученных данных [10, 21]. Выровненные последовательности использовались

для построения филогенетических деревьев в программе IQ-tree 1.6.12 с использованием метода максимального правдоподобия [22]. С помощью функции ModelFinder на основе рассчитываемого Байесовского критерия подобрана наиболее подходящая модель эволюции последовательностей для построения деревьев [23]. Для определения уровня поддержки ветвей произведено 10000 итераций “сверхбыстрого бутстреп-анализа” (ultrafast bootstrap) [24]. Только деревья со значением взвешенной дистанции Робинсона-Фулдса, равной не более 4, были признаны достоверными. Визуализация и оформление полученных филогенетических деревьев осуществлялись в веб-сервисе iTOL 6.8 и программе Adobe Illustrator CC [25]. Из построенных деревьев были удалены ветки с уровнем поддержки ниже 90 с целью избежать неточности в суждениях о времени разделения и гомологичности отдельных генов. Названия генов *A. thaliana* для оформления деревьев были взяты из информационного ресурса TAIR [26].

Транскриптомный анализ. Сравнение уровня экспрессии целевых генов проводили на основе данных ДНК-микрочипов анализа различных типов тканей *P. patens*, представленных как онлайн атлас экспрессии Physcomitrella eFP Browser, где значения экспрессии, полученные на основе RMA нормализации (Robust Multi-array Average), представлены в виде MEV (Microarray Expression Values) [27]. Типы исследуемых тканей *P. patens* включают каулоному, хлороному, ризоиды, гаметофоры, архегонии, спорофиты стадий развития S1, S2, S3, SM (зрелый спорофит) и споры. Среднее значение экспрессии было рассчитано для клеток с концевым типом роста, куда вошли каулонома, хлоронома и ризоиды, а также для спорофита, куда вошли данные экспрессии всех стадий развития спорофита. Конверсию аннотации генов, кодирующих белки с лектиновыми доменами, из версии 1.6 в версию 3.3 проводили в соответствии с файлом Ppatens_318_v3.3.synonym.txt (Phytozome v13, сборка генома *P. patens* v3.3). Расчет кластеров и их визуализацию осуществляли в среде R [28].

РЕЗУЛЬТАТЫ

Полногеномный скрининг генов, кодирующих белки с лектиновыми доменами, в геноме мха *P. patens* выявил 141 ген, кодирующий, предположительно, белки с лектиновым доменом, принадлежащие 15 разным семействам (табл. 1). Наиболее представленными в геноме мха семействами белков с лектиновым доменом являются семейства малектин-подобных белков и LysM (табл. 1). Наименьшее количество генов обнаруживается в семействах ризин В-подобных белков и циановиринов. Из встре-

чающихся у растений семейств лектинов не было найдено генов, кодирующих амарантины, агглютинины *Agaricus bisporus* (ABA) и хитиназа-связанные агглютинины (chitinase-related agglutinins, CRA). Гены амарантинов, хоть и распространены в растительном царстве, обнаруживаются в геноме не всех растений (отсутствуют в геноме риса и *A. thaliana*) (табл. 1). Гены лектинов семейства ABA ранее были обнаружены лишь в геноме некоторых мхов, в частности, у *M. polymorpha* [6]. Гены, кодирующие CRA, трудно идентифицируются в связи с их большой схожестью с каталитически активными хитиназами и практически не выявляются при полногеномном скрининге.

В геноме *P. patens* обнаружено всего 4 гена, кодирующих белки семейства Nictaba, в то время как у высших сосудистых растений это семейство представлено несколькими десятками генов (табл. 1). У мха также найдено 6 генов, кодирующих белки с лектиновым доменом С-типа. Лектиновый домен этого семейства наиболее распространен среди животных и в геноме сосудистых растений обнаруживается в единичных количествах. При выравнивании аминокислотных последовательностей лектиновых доменов нескольких генов *P. patens*, принадлежащих этому семейству, с доменами некоторых животных белков (номера белков из базы данных Uniprot Q9UJ71, O45824, Q9Z2H6, A5PMY6)

Таблица 1. Представленность лектинов по количеству генов в разных семействах в геноме мха *Physcomitrium patens* и четырех дополнительных видов: двудольных растений – *Arabidopsis thaliana* (резуховидка Таля) и *Linum usitatissimum* (лен); однодольных растений – *Oryza sativa* (рис) и *Zea mays* (кукуруза)

Семейство	ID (Pfam/ cd/ PTHR)	<i>P. patens</i> (0.43%)	<i>A. thaliana</i> (1.2%)	<i>L. usitatissimum</i> (0.93%)	<i>O. sativa</i> (1.04%)	<i>Z. mays</i> (0.77%)
Малектин-подобные	PF12819	30	75 ^a	47	61 ^b	32
LysM	PF01476	30	12	36	20	29
GNA	PF01453	16	49	112	134	81
Геветины	PF00187	12	10	17	10	10
Legume	PF00139	10	54	72	104	51
Жакалины	PF01419	9	50	4	30 ^b	20
Кальретикулины	PF00262	7	5 ^b	8	6 ^b	11
С-тип	PF00059	6	1 ^b	2	1	1
Nictaba	PF14299	4	30	37	20	18
Галактозосвязывающие	PF02140	4	13 ^b	23	12 ^b	13
Малектины	PF11721	4	15 ^a	14	23 ^b	9
EUL	PF14200, PTHR31257	4	1	1	5	8
Галектин-подобные	PF00337	3	6 ^b	11	10 ^b	17
Рицин В	PF00652	1	2	0	4	3
Циановирины	PF08881	1	0	0	0 ^г	0
Амарантины	PF07468	0	0	19	0	4
ABA	PF07367	0	0	0	0 ^г	0
CRA	cd02879	0	9	4	2	0
Общее количество генов:		141	332	407	442	307
Основная публикация:		Данная работа	[29]	[30]	[5]	[21]

Примечание. Под видовым названием указана доля генов лектинов среди всех белок-кодирующих генов в геноме данного растения. В отдельном столбце приведены идентификационные номера доменов лектинов, взятые из баз данных PFAM, PANTHER, NCBI CDD. Основные публикации, данные которой были использованы при создании таблицы, приведены в нижней строке, данные, взятые из дополнительной литературы, обозначены следующим образом: ^a[31]; ^b[32]; ^г[33]; ^гбыстрый скрининг через pfam в Phytozome v13.

сходство последовательностей составляло около 20–30%. В отличие от семенных растений, в геноме *P. patens* был идентифицирован ген *Pp3c15_13210*, кодирующий предполагаемый белок семейства циановиринов (табл. 1).

Доменная организация белков с лектиновыми доменами у мха *P. patens*. Растительные лектины часто представляют собой химерные белки, где лектиновый домен соединен с иным белковым доменом. Доменная организация белков с лектиновыми доменами мха, в целом, напоминает белки высших растений. Доменная архитектура типа лектиновой рецептор-подобной киназы, которая является характерной особенностью лектинов растений, уже обнаруживалась у *P. patens* [34]. Также встречаются характерные комбинации лектиновых доменов с F-box-доменом (семейства Nictaba и LysM), с доменом моторного белка кинезина (семейство малектинов), с каталитическими доменами галактозил-гидролаз и галактозил-трансфераз (семейства галактоза-связывающих белков и галектин-подобных белков, соответственно). Однако некоторые комбинации лектинового домена с другими доменами, обнаруженные при анализе, на сегодняшний день не описаны у семенных растений [4, 5]. Подобные комбинации были обнаружены у белков семейств гевеинов и LysM (рис. 1). Примечательно, что представители обоих семейств относятся к хитин-связывающим белкам, потенциально вовлеченным в защиту растений.

Хитин-связывающий домен белков семейства гевеинов у высших растений, как правило, связан с доменом хитиназы классов I, II и IV (GH19, PF00182). В то же время, в геноме *P. patens* был найден ген *Pp3c6_90*, у которого была обна-

ружена комбинация домена гевеина с хитиназой классов III и V (GH18, PF00704) (рис. 1). У некоторых белков семейства гевеинов был также обнаружен другой тип доменной организации, не охарактеризованный у высших семенных растений: у трех белков хитин-связывающий лектиновый домен связан с доменом двойного пси бета-бочонка (DPBB_1, PF03330) (рис. 1). Этот домен часто встречается в составе белков пыльцевых аллергенов (в частности экспансинов), при этом сходство последовательностей экспансинов *A. thaliana* (номера белков из базы данных Uniprot Q9C554 и Q9SKU2) с гевеинами *P. patens* составляло всего около 30%. Функциональное значение комбинации хитин-связывающего домена с DPBB_1 у мха остается неясным.

Домен LysM обнаружен в комбинации с каталитическим доменом пептидазы M23 (PF01551), характерной для бактерий и обладающей гидролитической активностью по отношению к пептидогликану (рис. 1). Вероятнее всего, лектиновый домен в составе этих пептидов служит для ориентации и нацеливания фермента на пептидогликан клеточной стенки бактерий. Ген *Pp3s155_60*, помимо двух доменов LysM на С-конце, имеет также домен хитиназы класса III и V (GH18) на N-конце (рис. 1). Белки с похожей доменной организацией ранее были описаны у других споровых растений, в частности у хвоща *Equisetum arvense* [35]. Другие специфические комбинации в семействе LysM были обнаружены в белковой последовательности гена *Pp3c13_19830*, имеющей дополнительный липид-связывающий домен с N-конца (NT-C2, PF10358), у гена *Pp3c5_17160* с дополнительным доменом глутаминазы (Destabilase, PF05497) и у двух генов со

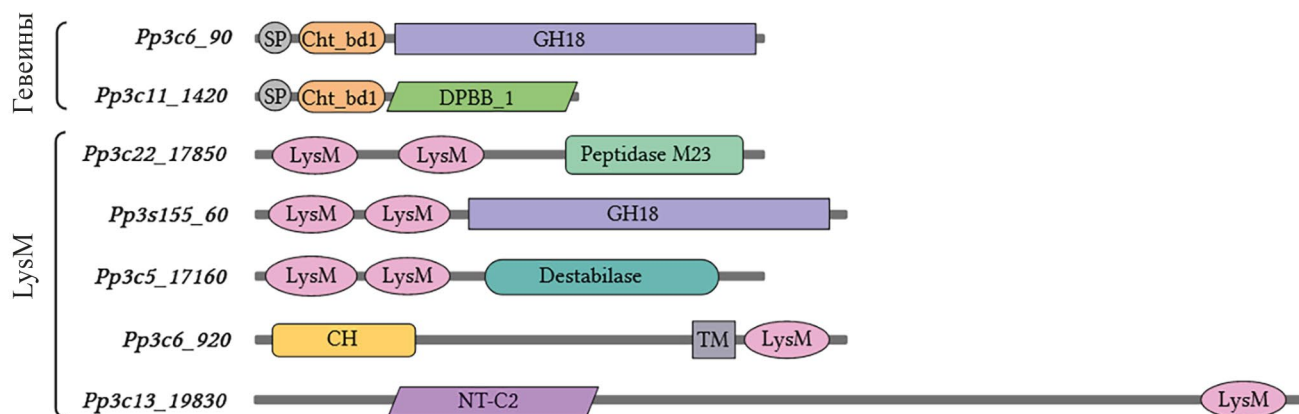


Рис. 1. Схематичное изображение специфичной доменной организации некоторых белков мха *P. patens* семейств гевеинов и LysM (номера соответствующих генов приведены слева). Во избежание путаницы названия всех доменов взяты из базы данных InterPro и поэтому приведены на английском. Обозначения: SP – сигнальный пептид; Cht_bd1 (сокращенно от chitin_bind_1) – хитин-связывающий домен гевеинов (PF00187); GH18 – домен хитиназы классов III и V (PF00704); DPBB_1 – домен двойного пси бета-бочонка (PF03330); LysM – лектиновый домен семейства LysM (PF01476); Peptidase M23 – домен пептидазы M23 (PF01551); Destabilase – домен глутаминазы (PF05497); CH – гомологичный кальпониу домен (PF00307); TM – трансмембранный домен; NT-C2 – N-концевой C2 домен (PF10358). Изображение было создано с помощью BioRender.com.

сходной аминокислотной последовательностью в Phytozome v13 (*Pp3c6_920*, *Pp3c6_900*), имеющими N-концевой гомологичный кальпониновый домен (CH, PF00307) (рис. 1). Для верификации результатов анализа доменной организации были проведены повторные предсказания аминокислотной последовательности этих генов, полученные результаты при этом были аналогичны представленным в базе данных.

Филогенетический анализ белков с лектиновыми доменами. Филогенетические деревья были построены для 9 семейств белков с лектиновым

доменом (рис. 2; рис. 1–7, Дополнительные материалы). Белки мха *P. patens* на полученных деревьях зачастую представлены отдельной кладой, более или менее обособленной от генов высших растений (*Z. mays* и *A. thaliana*), что хорошо проиллюстрировано на примере семейства жакалинов (рис. 2).

Для генов малектинов и малектин-подобных белков, ввиду схожести их последовательностей, было сконструировано общее филогенетическое дерево (рис. 1, Дополнительные материалы). Обособленной кладой здесь выделяется

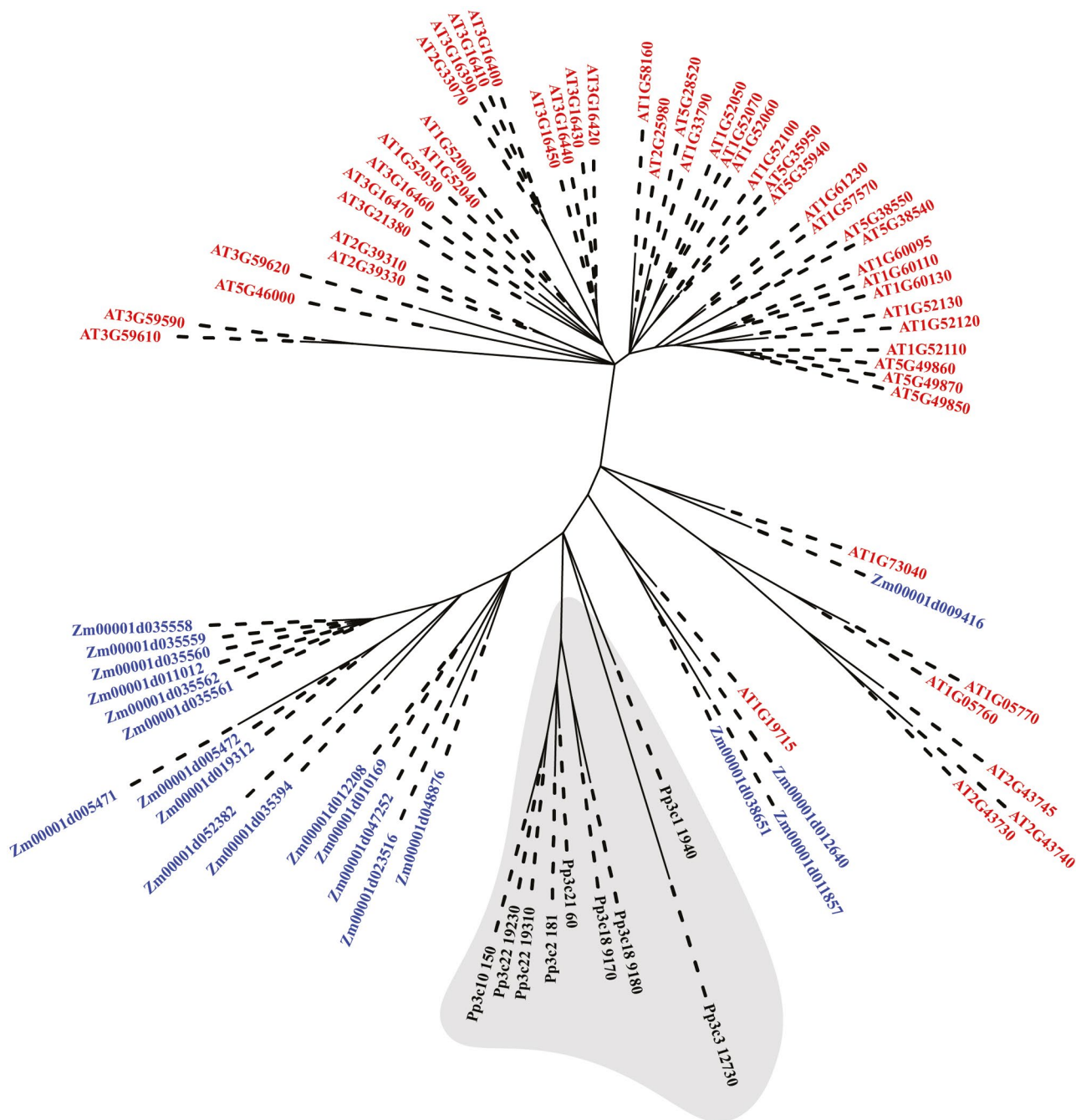


Рис. 2. Филогенетическое дерево представителей семейства жакалинов мха *P. patens*, кукурузы (*Z. mays*) и *A. thaliana*. Дистанции Робинсона-Фулдса для данного дерева была равна 0. Из дерева были удалены точки ветвления с уровнем поддержки ниже 90. Серым цветом выделена обособленная клада жакалинов мха.

был разделен на три кластера (рис. 5). Гены первого кластера экспрессировались на относительно одинаковом уровне во всех исследуемых образцах с небольшим увеличением значений экспрессии в спорах. Наиболее представлены в этом кластере гены семейств гевеинов, LysM и малектин-подобных белков. Уровень экспрессии генов второго кластера с преобладающей долей пред-

ставителей семейства малектин-подобных белков в значительной степени повышен в гаметофорах, архегониях и на ранних стадиях развития спорофита. Самый многочисленный третий кластер (47 генов) в основном включает в себя гены семейств Legume, GNA, LysM и малектин-подобных белков, экспрессия которых повышена в клетках, удлиняющихся путем концевой роста.

Ген	Доменная организация	Субклеточная локализация (по LocTree3)	Спорофит	Споры	Каулонема	Хлоронема	Гаметофоры	Ризоиды	Архегония
Семейство малектин-подобных белков									
<i>Pr3c9_10760</i>	MLD, PK	Плазмалемма	9589	4119	11364	7697	11359	10965	11398
<i>Pr3c9_10730</i>	MLD, PK	Плазмалемма	5074	2465	9847	6334	8937	14998	6677
<i>Pr3c9_10640</i>	MLD, PK	Плазмалемма	5085	2622	8060	5487	8176	11271	6527
<i>Pr3c20_16970</i>	MLD, PK	Плазмалемма	346	623	1132	518	414	1752	364
<i>Pr3c20_17290</i>	MLD, PK	Плазмалемма	3273	2648	5374	3143	4139	2035	3090
<i>Pr3c4_2350</i>	MLD, LRR, PK	Плазмалемма	5422	707	283	677	16664	4115	7723
<i>Pr3c4_18780</i>	MLD, LRR, PK	Плазмалемма	424	1353	6567	7000	4420	6241	4112
<i>Pr3c3_28930</i>	MLD, LRR, PK	Плазмалемма	4461	15	243	287	3343	14	8518
<i>Pr3c16_5720</i>	MLD, LRR	Цитоплазма	2697	858	5548	1932	2866	1728	3428
<i>Pr3c2_37290</i>	MLD, LRR	Мембрана ЭПР	1138	10096	7484	1998	917	1730	646
<i>Pr3c9_5240</i>	MLD, LRR	Внеклеточная	2923	2733	3808	1352	1169	1600	1221
<i>Pr3c9_7920</i>	MLD, LRR	Внеклеточная	3293	18	16	303	3318	245	5125
<i>Pr3c4_18360</i>	MLD, LRR, PK	Плазмалемма	905	21	3745	3656	783	3259	1037
<i>Pr3c7_20780</i>	MLD, LRR	Ядро	984	885	5647	1455	419	2540	1404
<i>Pr3c3_10220</i>	MLD, LRR	Плазмалемма	1377	18	5070	3857	233	610	59
<i>Pr3c1_9180</i>	MLD, LRR, PK	Плазмалемма	335	1558	1224	2001	573	1261	317
<i>Pr3c15_1540</i>	MLD, LRR, PK	Плазмалемма	140	262	1659	1462	930	1732	410
<i>Pr3c27_5920</i>	MLD, LRR	Плазмалемма	890	14	549	104	1263	32	1437
<i>Pr3c3_25390</i>	MLD, LRR, PK	Плазмалемма	1202	40	84	443	512	369	160
<i>Pr3c12_16180</i>	MLD, LRR, PK	Плазмалемма	853	9	32	317	613	389	235
<i>Pr3c1_23040</i>	MLD, LRR, PK	Плазмалемма	285	2353	41	76	30	1153	95
<i>Pr3c15_3950</i>	MLD, LRR	Плазмалемма	647	171	328	318	416	254	465
<i>Pr3c7_25750</i>	MLD, LRR, PK	Плазмалемма	255	1115	358	957	175	132	121
<i>Pr3c1_23020</i>	MLD, LRR, PK	Плазмалемма	235	1877	45	102	54	287	112
<i>Pr3c13_11270</i>	MLD, LRR	Плазмалемма	23	13	2354	118	52	18	10
<i>Pr3c2_18520</i>	MLD, PK	Плазмалемма	52	135	698	282	354	306	247
<i>Pr3c26_12890</i>	MLD, LRR	Плазмалемма	91	44	604	303	38	553	33
<i>Pr3c10_2160</i>	MLD	Внеклеточная	11	21	12	13	10	22	12
Семейство Nictaba									
<i>Pr3c19_22630</i>	F-box, PP2	Ядро	9305	7455	5881	6580	7165	4547	7837
<i>Pr3c24_14280</i>	F-box, PP2	Цитоплазма	1955	1507	3939	699	2236	4290	615
<i>Pr3c16_11450</i>	F-box, PP2	Хлоропласт	1505	7482	7394	4797	2638	1869	1358
<i>Pr3c26_1510</i>	PP2	Хлоропласт	277	29	3738	158	143	40	29
Гевеины со специфической доменной организацией									
<i>Pr3c6_90</i>	Chitin_bind_1, GH18	Вакуоль	9750	35	1045	525	2711	19	4285
<i>Pr3c11_1420</i>	Chitin_bind_1, DPBB_1	Вакуоль	879	182	42	70	49	115	123
<i>Pr3c7_25120</i>	Chitin_bind_1, DPBB_1	Вакуоль	101	2038	43	33	35	63	75
Белки семейства LysM со специфической доменной организацией									
<i>Pr3c13_19830</i>	NT-C2, LysM	Ядро	912	42	369	955	1413	611	705
<i>Pr3c6_920</i>	CH, LysM	Цитоплазма	1041	789	403	1435	2039	1992	452
<i>Pr3c1_33030</i>	LysM, LysM, Peptidase_M23	Цитоплазма	4493	3577	10181	7853	2476	6482	4581
<i>Pr3c1_21480</i>	LysM, LysM, Peptidase_M23	Цитоплазма	792	22	3804	477	668	799	110
<i>Pr3c21_17920</i>	LysM, LysM, LysM, Peptidase_M23	Хлоропласт	2114	3643	4030	6612	5082	3266	3168
<i>Pr3c5_17160</i>	LysM, LysM, Destabilase	Хлоропласт	1851	1089	4162	5421	3647	4630	956
<i>Pr3c2_19990</i>	LysM, LysM, Peptidase_M23	Цитоплазма	214	22	1729	505	394	683	224

Рис. 4. Экспрессия генов мха *P. patens*, кодирующих белки семейств малектин-подобных белков и Nictaba, а также отдельных представителей семейств гевеинов и LysM со специфической доменной организацией. Представлена их экспрессия в различных типах тканей *P. patens* (по данным [27]). Для спорофита указано среднее значение экспрессии для всех стадий развития спорофита. Приведена доменная организация и предсказанная *in silico* субклеточная локализация белков.

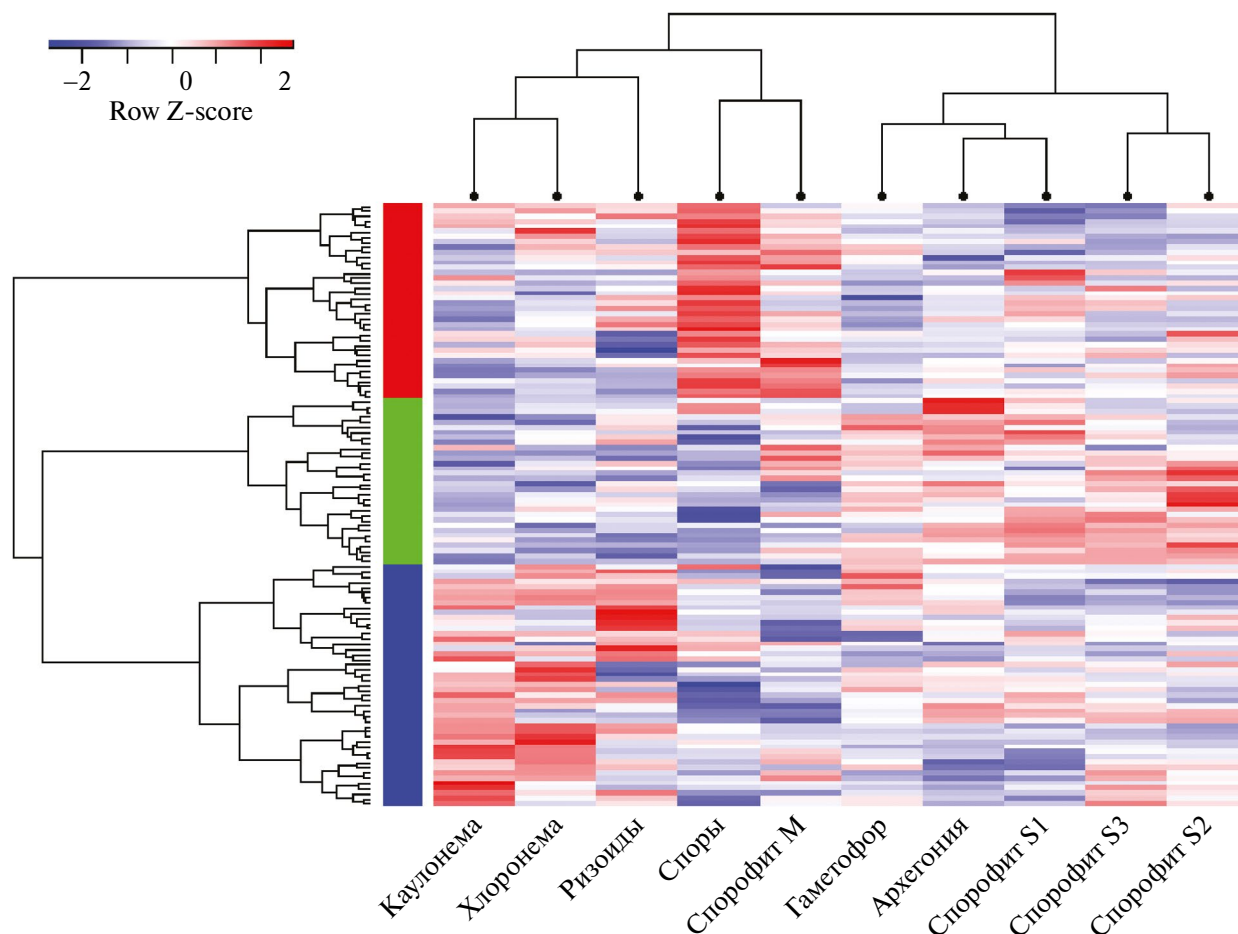


Рис. 5. Результаты иерархического кластерного анализа и тепловая карта уровней экспрессии генов, кодирующих белки с лектиновыми доменами в различных типах тканей мха *P. patens*. Тепловая карта отражает относительный уровень экспрессии каждого гена (строки) в каждом образце (ряды). Уровень экспрессии в кластерах представлен в виде трансформированного в Z-score логарифмированного значения экспрессии (цветовая шкала в верхнем левом углу рисунка). Три выявленных кластера отображены в левой части в виде вертикальных прямоугольников (красный, зеленый, синий).

ОБСУЖДЕНИЕ

Нами был проведен полногеномный скрининг генов, кодирующих белки с углевод-связывающими лектиновыми доменами в геноме модельного спорового растения — *P. patens*. При сравнении представленности генов лектинов в геноме мха с геномами некоторых сосудистых растений было выявлено, что количество генов, кодирующих белки с лектиновыми доменами, у мха в 2 раза меньше, чем у сосудистых. Доля генов лектинов среди всех белок-кодирующих генов в геноме сосудистых растений составляет примерно 1%, тогда как у *P. patens* — всего 0.43%. Хорошо известно, что многие белки с лектиновыми доменами так или иначе вовлечены в защитные реакции организма растения против абиотических (засуха, засоление и др.) и биотических стрессовых факторов (поражение патогенными бактериями, грибами, насекомыми) [1]. Вероятно, в процессе эволюции у растений в ответ на изменяющиеся условия

окружающей среды и усложнения организации сигнальных систем количество генов, кодирующих лектины, увеличилось в несколько раз.

У *P. patens* наблюдались отличия в представленности различных семейств белков с лектиновыми доменами. Наряду с малектин-подобными белками, наиболее представленные семейства лектинов у сосудистых растений — семейства GNA и Legume, в то время как у мха *P. patens* наиболее обширны семейства малектин-подобных белков и LysM (табл. 1). Белки семейств Legume и GNA по большей части представляют собой рецептор-подобные киназы и рецептор-подобные белки, локализованные в плазмалемме. Как правило, это гены, экспрессия которых в обычных условиях остается низкой и повышается в условиях стресса. Они вовлечены в процессы взаимодействия растений с различными микроорганизмами, включая симбиоз и защитные реакции в ответ на атаку патогена [34]. В целом, доля генов, кодирующих рецептор-подоб-

ные киназы с лектиновыми доменами, от генов всех белков с лектиновыми доменами у *P. patens* ниже (около 1/3), чем у высших растений (1/2) (табл. 1, Дополнительные материалы).

Рецептор-подобные киназы с лектиновыми доменами в последние десять лет заслужили особое внимание со стороны научного общества в связи с их участием в различных сигнальных путях и способностью к восприятию разнообразных внеклеточных сигналов, включая микроб-ассоциированные, патоген-ассоциированные и дистресс-ассоциированные молекулярные паттерны [34, 37]. Помимо этого, предполагается, что углевод-связывающие домены некоторых подобных белков способны непосредственно взаимодействовать с углеводами клеточной стенки. Другими словами, предполагается их роль в восприятии и мониторинге состояния клеточной стенки в процессе жизнедеятельности растений [38]. Доля белков *P. patens*, которые потенциально способны взаимодействовать с углеводами клеточной стенки (мембраносвязанные белки с внеклеточным лектиновым доменом и непосредственно белки клеточной стенки) составила около 65%, тогда как у высших растений это число достигает 80% и выше (табл. 1, Дополнительные материалы).

Состав клеточной стенки *P. patens* существенно отличается от клеточной стенки покрытосеменных растений: помимо целлюлозы, основной его структурный компонент – маннаны [39]. Потому набор углевод-связывающих белков (лектинов) у этого мха может также иметь свои особенности. Однако основные семейства маннозосвязывающих лектинов – GNA и жакалины либо количественно менее представлены, либо сравнимы по числу представителей с покрытосеменными растениями. Для белков семейств С-типа, циановиринов и EUL была также показана возможная специфичность по отношению к высокоманнозным N-гликанам, но предполагается, что эти лектины взаимодействуют с различными гликопротеинами [6].

Растительные лектины С-типа представляют собой рецептор-подобные киназы, локализованные на плазмалемме. У *P. patens* также обнаруживаются белки С-типа с лейциновыми повторами, предсказанные в веб-сервисе LocTree3 как секретируемые белки. Это семейство было хорошо охарактеризовано у животных, однако на сегодняшний день остается мало изученным у растений [40]. Циановирины, в основном, характерны для грибов, однако присутствуют у некоторых низших и высших споровых растений и полностью отсутствуют в геноме семенных растений. Сходный белок, выделенный из цианобактерий, показал специфичность по отношению к высокоманнозным N-гликанам на поверхности гликопротеина ви-

руса иммунодефицита человека [6]. Ген *P. patens* *Pp3c15_13210* кодирует белок, предсказанный веб-сервисами DeepLoc-1.0 и LocTree3 как белок клеточной стенки, его функция неясна, экспрессия гена практически отсутствует в проанализированных образцах (табл. 1, Дополнительные материалы).

Семейство Nictaba у *P. patens* представлено всего 4 генами. Некоторые белки этого семейства также известны как PP2 (Phloem Protein 2) и представляют собой структурные компоненты проводящей ткани высших сосудистых растений. Проводящие элементы мхов, выполняющие аналогичную функцию, структурно отличаются от флоэмы [41]. Вероятнее всего, у более низкоорганизованных растений белки Nictaba не обладают схожей функциональной активностью и, по всей видимости, это семейство белков с лектиновым доменом получило расширение вместе с развитием проводящих тканей у сосудистых растений. Три гена Nictaba, имеющие дополнительный домен F-box (считается, что он участвует в убиквитинировании), конститутивно экспрессируются во всех проанализированных частях растений, тогда как экспрессия гена с исключительно лектиновым доменом (*Pp3c26_1510*) наиболее выражена в каулонеме [42].

Филогенетический анализ белков, кодируемых искомыми генами, показал наличие существенных различий в белковых последовательностях проанализированных организмов (спорового растения мха *P. patens* и покрытосеменных растений *A. thaliana* и *Z. mays*) – в большинстве случаев белки мха локализовались в обособленной кладе. При этом, на примере семейства жакалинов можно заметить, что белки мха филогенетически ближе к белкам однодольных растений (кукурузы) (рис. 2). Семейство жакалинов состоит из разнообразных белков химерной структуры, и доменная организация белков однодольных и двудольных имеет существенные отличия. Действительно, доменная организация белков жакалинов мха ближе к жакалинам однодольных, среди которых также, к примеру, часто встречается комбинация домена жакалина с киназным доменом; у двудольных такая комбинация доменов не выявлена. Роль этих белков в растительной клетке до сих пор не выяснена, и также неясно, могут ли они выполнять схожие функции у таких разных групп растений, как мхи и однодольные.

Пять белков семейства малектин-подобных *P. patens* оказались гомологичны белкам CrRLK1L *A. thaliana* (рис. 3). Эта группа рецептор-подобных киназ относительно хорошо охарактеризована у *A. thaliana*. Известными лигандами белков CrRLK1L являются короткие

внеклеточные пептиды RALF (rapid alkalization factor — фактор быстрого ощелачивания). За счет взаимодействия со своим пептидным лигандом, эти рецепторы опосредуют передачу сигналов и регуляцию разнообразных процессов развития растения, морфогенеза, размножения, иммунного ответа и реакции на стресс [1]. Отдельные представители этого семейства у *A. thaliana* были также охарактеризованы как участники регуляции и поддержания целостности клеточной стенки [36]. Количество генов, кодирующих CrRLK1L у *P. patens* многократно меньше, чем у *A. thaliana* и кукурузы (рис. 3). Вероятно, механизм поддержания целостности клеточной стенки у исследуемого мха развит слабо, что может быть связано с отличиями в наборе полисахаридов клеточной стенки у мхов и сосудистых растений. Характерным полисахаридом клеточной стенки у эволюционно ранних таксонов растений является маннан, составляющий основу всех клеточных стенок во всем организме [39, 43]. С усложнением организации растений и увеличением разнообразия полисахаридов клеточной стенки — потенциальных углеводных лигандов белков с лектиновыми доменами, количество членов этого семейства у высших растений значительно увеличилось.

У *P. patens* гены белков с лектиновыми доменами активно экспрессируются во всех частях растения, что подчеркивает важность их функционирования для обеспечения жизнедеятельности клеток. Экспрессия индивидуальных генов часто дифференциальна; выявлено три основных паттерна экспрессии (рис. 4). По количеству экспрессирующихся генов выделяются семейства малектин-подобных белков, LysM, GNA и гевеины. Среди них малектин-подобные белки представляют собой по большей части мембранные белки, вовлеченные в регуляцию самых разных процессов жизнедеятельности растений. Семейства LysM и гевеины, как лектины со специфичностью к олигомерам хитина, участвуют в иммунных ответах. Семейство GNA слабо изучено, однако также часто рассматривается как часть системы защиты растения от биотического и абиотического стресса [37]. Наиболее активная экспрессия этих лектинов наблюдается в кластере, включающем каулоному, хлоронему и ризоиды (рис. 4; табл. 1, Дополнительные материалы). Эти филаментные образования растут концевым ростом, соответственно, для них характерны активные перестройки клеточной стенки и формирование дополнительной поверхности контакта с внешней средой [44]. Это служит аргументом в пользу того, что уже на ранних этапах эволюции наземных растений лектины могли участвовать как компоненты сигнальных систем, задействованных в мониторинге состояния клеточной стенки и в реакциях

на биотические и абиотические факторы внешней среды.

Лектины возникли рано в процессе эволюции живых организмов. С развитием растений и усложнением их организации семейства растительных лектинов расширялись и, вероятно, обретали новые функции. При этом у мхов отмечается характерная только для них доменная организация белков с лектиновыми доменами. Наблюдаются некоторые различия в доменной организации белков мха и высших семенных растений, в частности, в группе белков, потенциально вовлеченных в реализацию иммунного ответа растений (например, семейство гевеинов и LysM). Хотя набор семейств белков с лектиновыми доменами у мха практически не отличается от представленных в геномах высших семенных растений, относительная доля генов, кодирующих лектины, в геноме *P. patens* существенно меньше (табл. 1). Увеличение числа белков с лектиновыми доменами в ходе эволюции может свидетельствовать о выполнении ими более специализированных функций у высших сосудистых растений.

Несмотря на то, что углевод-связывающая активность (по крайней мере, в отношении гликозилированных белков) и функции некоторых растительных лектинов достаточно хорошо описаны, остается неясным, имеют ли похожие функции белки мха. Из-за методических сложностей пока не охарактеризованы полисахаридные лиганды многочисленных белков с лектиновыми доменами, локализованных в клеточной стенке, в которой у растений сконцентрировано огромное разнообразие углеводных структур. Дальнейшее изучение этих белков может стать важным шагом в изучении эволюции наземных растений.

Работа выполнена при поддержке Российского научного фонда (грант № 20-64-47036) (сопоставление с лектинами высших растений), а также Государственного задания Федерального исследовательского центра “Казанский научный центр Российской академии наук” (идентификация в геноме, филогения, анализ экспрессии).

Настоящая статья не содержит каких-либо исследований с участием людей и животных в качестве объектов. Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. De Coninck T., Van Damme E.J.M. The multiple roles of plant lectins // Plant Sci. 2021. V. 313. P. 111096. <https://doi.org/10.1016/j.plantsci.2021.111096>
2. Cove D.J., Perroud P.F., Charron A.J., McDaniel S.F., Khandelwal A., Quatrano R.S. The moss *Physcomitrella patens*: a novel model system for plant development

- and genomic studies // Cold Spring Harb. Protoc. 2009. V. 2009.pdb.emo115.
<https://doi.org/10.1101/pdb.emo115>
3. *Rensing S.A., Lang D., Zimmer A.D., Terry A., Salamov A., Shapiro H., Nishiyama T., Perroud P.F., Lindquist E.A., Kamisugi Y., Tanahashi T., Sakakibara K., Fujita T., Oishi K., Shin-I T. et al.* The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants // Science. 2008. V. 319. P. 64.
<https://doi.org/10.1126/science.1150646>
4. *Van Holle S., Van Damme E.J.M.* Messages from the past: new insights in plant lectin evolution // Front. Plant Sci. 2019. V. 10. P. 36.
<https://doi.org/10.3389/fpls.2019.00036>
5. *Van Holle S., De Schutter K., Eggermont L., Tsaneva M., Dang L., Van Damme E.J.M.* Comparative study of lectin domains in model species: new insights into evolutionary dynamics // Int. J. Mol. Sci. 2017. V. 18. P. 1136.
<https://doi.org/10.3390/ijms18061136>
6. *Van Damme E.J.M., Lannoo N., Peumans W.J.* Plant lectins // Adv. Bot. Res. 2008. V. 48. P. 107.
[https://doi.org/10.1016/S0065-2296\(08\)00403-5](https://doi.org/10.1016/S0065-2296(08)00403-5)
7. *Dinant S., Clark A.M., Zhu Y., Vilaine F., Palauqui J.C., Kusiak C., Thompson G.A.* Diversity of the superfamily of phloem lectins (phloem protein 2) in angiosperms // Plant Physiol. 2003. V. 131. P. 114.
<https://doi.org/10.1104/pp.013086>
8. *Fouquaert E., Peumans W.J., Vandekerckhove T., Onghenaert M., Van Damme E.J.* Proteins with an Eucalyptus lectin-like domain are ubiquitous in Embryophyta // BMC Plant Biol. 2009. V. 9. P. 136.
<https://doi.org/10.1186/1471-2229-9-136>
9. *Lang D., Ullrich K.K., Murat F., Fuchs J., Jenkins J., Haas F.B., Piednoel M., Gundlach H., Van Bel M., Meyberg R., Vives C., Morata J., Symeonidi A., Hiss M., Muchero W. et al.* The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution // Plant J. 2018. V. 93. P. 515.
<https://doi.org/10.1111/tpj.13801>
10. *Goodstein D.M., Shu S., Howson R., Neupane R., Hayes R.D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar D.S.* Phytozome: a comparative platform for green plant genomics // Nucleic Acids Res. 2012. V. 40. P. D1178.
<https://doi.org/10.1093/nar/gkr944>
11. *Wang J., Chitsaz F., Derbyshire M.K., Gonzales N.R., Gwartz M., Lu S., Marchler G.H., Song J.S., Tanaka N., Yamashita R.A., Yang M., Zhang D., Zheng C., Lanczycki C.J., Marchler-Bauer A.* The conserved domain database in 2023 // Nucleic Acids Res. 2022. V. 51. P. D384.
<https://doi.org/10.1093/nar/gkac1096>
12. *Thomas P.D., Ebert D., Muruganujan A., Mushayaha T., Albou L.P., Mi H.* PANTHER: making genome-scale phylogenetics accessible to all // Protein Sci. 2022. V. 31. P. 8.
<https://doi.org/10.1002/pro.4218>
13. *Blum M., Chang H.Y., Chuguransky S., Grego T., Kandasaamy S., Mitchell A., Nuka G., Paysan-Lafosse T., Qureshi M., Raj S., Richardson L., Salazar G.A., Williams L., Bork P., Bridge A.* The InterPro protein families and domains database: 20 years on // Nucleic Acids Res. 2021. V. 49. P. D344.
<https://doi.org/10.1093/nar/gkaa977>
14. *Keller O., Kollmar M., Stanke M., Waack S.* A novel hybrid gene prediction method employing protein multiple sequence alignments // Bioinformatics. 2011. V. 27. P. 757.
<https://doi.org/10.1093/bioinformatics/btr010>
15. *Almagro Armenteros J.J., Tsirigos K.D., Sønderby C.K., Petersen T.N., Winther O., Brunak S., von Heijne G., Nielsen H.* SignalP 5.0 improves signal peptide predictions using deep neural networks // Nat. Biotechnol. 2019. V. 37. P. 423.
<https://doi.org/10.1038/s41587-019-0036-z>
16. *Krogh A., Larsson B., Von Heijne G., Sonnhammer E.L.* Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes // J. Mol. Biol. 2001. V. 305. P. 567.
<https://doi.org/10.1006/jmbi.2000.4315>
17. *Goldberg T., Hecht M., Hamp T., Karl T., Yachdav G., Ahmed N., Altermann U., Angerer P., Ansorge S., Balasz K., Bernhofer M., Betz A., Cizmadija L., Do K.T., Gerke J.* LocTree3 prediction of localization // Nucleic Acids Res. 2014. V. 42. P. W350.
<https://doi.org/10.1093/nar/gku396>
18. *Almagro Armenteros J.J., Sønderby C.K., Sønderby S.K., Nielsen H., Winther O.* DeepLoc: prediction of protein subcellular localization using deep learning // Bioinformatics. 2017. V. 33. P. 3387.
<https://doi.org/10.1093/bioinformatics/btx431>
19. *Madeira F., Park Y.M., Lee J., Buso N., Gur T., Madhusoodanan N., Basutkar P., Tivey A.R.N., Potter S.C., Finn R.D., Lopez R.* The EMBL-EBI search and sequence analysis tools APIs in 2019 // Nucleic Acids Res. 2019. V. 47. P. W636.
<https://doi.org/10.1093/nar/gkz268>
20. *Edgar R.C.* MUSCLE: multiple sequence alignment with high accuracy and high throughput // Nucleic Acids Res. 2004. V. 32. P. 1792.
<https://doi.org/10.1093/nar/gkh340>
21. *Aglyamova A., Petrova N., Gorshkov O., Kozlova L., Gorshkova T.* Growing maize root: lectins involved in consecutive stages of cell development // Plants. 2022. V. 11. P. 1799.
<https://doi.org/10.3390/plants11141799>
22. *Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies // Mol. Biol. Evol. 2015. V. 32. P. 268.
<https://doi.org/10.1093/molbev/msu300>
23. *Kalyaanamoorthy S., Min B.Q., Wong T.K., Von Haeseler A., Jermiin L.S.* ModelFinder: fast model selection for accurate phylogenetic estimates // Nat. Methods. 2017. V. 14. P. 587.
<https://doi.org/10.1038/nmeth.4285>
24. *Minh B.Q., Nguyen M.A.T., Von Haeseler A.* Ultrafast approximation for phylogenetic bootstrap // Mol.

- Biol. Evol. 2013. V. 30. P. 1188.
<https://doi.org/10.1093/molbev/mst024>
25. Letunic I., Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation // Nucleic Acids Res. 2021. V. 49. P. W293.
<https://doi.org/10.1093/nar/gkab301>
 26. Garcia-Hernandez M., Berardini T., Chen G., Crist D., Doyle A., Huala E., Knee E., Lambrecht M., Miller N., Mueller L.A., Mundodi S., Reiser L., Rhee S.Y., Scholl R., Tacklind J. TAIR: a resource for integrated *Arabidopsis* data // Funct. Integr. Genomics. 2002. V. 2. P. 239.
<https://doi.org/10.1007/s10142-002-0077-z>
 27. Ortiz-Ramírez C., Hernandez-Coronado M., Thamm A., Catarino B., Wang M., Dolan L., Feijó J.A., Becker J.D. A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants // Mol. Plant. 2016. V. 9. P. 205.
<http://dx.doi.org/10.1016%2Fj.molp.2015.12.002>
 28. R Development Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.r-project.org/>
 29. Eggermont L., Verstraeten B., Van Damme E.J.M. Genome-wide screening for lectin motifs in *Arabidopsis thaliana* // The plant genome. 2017. V. 10. plantgenome2017.02.0010.
<https://doi.org/10.3835/plantgenome2017.02.0010>
 30. Petrova N., Nazipova A., Gorshkov O., Mokshina N., Patova O., Gorshkova T. Gene expression patterns for proteins with lectin domains in flax stem tissues are related to deposition of distinct cell wall types // Front. Plant Sci. 2021. V. 12. P. 634594.
<https://doi.org/10.3389/fpls.2021.634594>
 31. Yang H., Wang D., Guo L., Pan H., Yvon R., Garman S., Wu H., Cheung A.Y. Malectin/Malectin-like domain-containing proteins: a repertoire of cell surface molecules with broad functional potential // Cell Surf. 2021. V. 7. P. 100056.
<https://doi.org/10.1016/j.tcs.2021.100056>
 32. Jiang S.Y., Ma Z., Ramachandran S. Evolutionary history and stress regulation of the lectin superfamily in higher plants // BMC Evol. Biol. 2010. V. 10. P. 1.
<https://doi.org/10.1186/1471-2148-10-79>
 33. Jing X.Q., Shalmani A., Zhou M.R., Shi P.T., Muhammad I., Shi Y., Sharif R., Li W., Liu W., Chen K.M. Genome-wide identification of malectin/malectin-like domain containing protein family genes in rice and their expression regulation under various hormones, abiotic stresses, and heavy metal treatments // J. Plant Growth Regul. 2020. V. 39. P. 492.
<https://doi.org/10.1007/s00344-019-09997-8>
 34. Bellande K., Bono J.J., Savelli B., Jamet E., Canut H. Plant lectins and lectin receptor-like kinases: how do they sense the outside? // Int. J. Mol. Sci. 2017. V. 18. P. 1164.
<https://doi.org/10.3390/ijms18061164>
 35. Inamine S., Onaga S., Ohnuma T., Fukamizo T., Taira T. Purification, cDNA cloning, and characterization of LysM-containing plant chitinase from horsetail (*Equisetum arvense*) // Biosci. Biotechnol. Biochem. 2015. V. 79. P. 1296.
<https://doi.org/10.1080/09168451.2015.1025693>
 36. Franck C.M., Westermann J., Boisson-Dernier A. Plant malectin-like receptor kinases: from cell wall integrity to immunity and beyond // Annu. Rev. Plant Biol. 2018. V. 69. P. 301.
<https://doi.org/10.1146/annurev-arplant-042817-040557>
 37. De Coninck T., Van Damme E.J.M. Plant lectins: handymen at the cell surface // Cell Surf. 2022. V. 8. P. 100091.
<https://doi.org/10.1016/j.tcs.2022.100091>
 38. Ji D., Chen T., Zhang Z., Li B., Tian S. Versatile roles of the receptor-like kinase *feronia* in plant growth, development and host-pathogen interaction // Int. J. Mol. Sci. 2020. V. 21. P. 7881.
<https://doi.org/10.3390/ijms21217881>
 39. Popper Z.A., Fry S.C. Primary cell wall composition of bryophytes and charophytes // Ann. Bot. 2003. V. 91. P. 1.
<https://doi.org/10.1093/aob/mcg013>
 40. Mayer S., Raulf M.K., Lepenies B. C-type lectins: their network and roles in pathogen recognition and immunity // Histochem. Cell Biol. 2017. V. 147. P. 223.
<https://doi.org/10.1007/s00418-016-1523-7>
 41. Ligrone R., Duckett J.G., Renzaglia K.S. Conducting tissues and phyletic relationships of bryophytes // Philos. Trans. R. Soc. 2000. V. 355. P. 795.
<https://doi.org/10.1098/rstb.2000.0616>
 42. Stefanowicz K., Lannoo N., Van Damme E.J.M. Plant F-box proteins—judges between life and death // Crit. Rev. Plant Sci. 2015. V. 34. P. 523.
<https://doi.org/10.1080/07352689.2015.1024566>
 43. Ye Z.H., Zhong R. Cell wall biology of the moss *Physcomitrium patens* // J. Exp. Bot. 2022. V. 73. P. 4440.
<https://doi.org/10.1093/jxb/erac122>
 44. Vidali L., Bezanilla M. *Physcomitrella patens*: a model for tip cell growth and differentiation // Curr. Opin. Plant Biol. 2012. V. 15. P. 625.
<https://doi.org/10.1016/j.pbi.2012.09.008>