ГЕНЕТИКА, 2020, том 56, № 5, с. 561–570

ГЕНЕТИКА РАСТЕНИЙ

УДК 575.22

ИДЕНТИФИКАЦИЯ *in silico* СЕМЕЙСТВ ГЕНОВ ЦИСТЕИН-БОГАТЫХ ПЕПТИДОВ *Solanum lycopersicum* L.[1]

© 2020 г. Е. А. Истомина[1], М. П. Слезина[1], А. С. Ковтун[2], Т. И. Одинцова[1, *]

[1]*Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия*

[2]*Московский физико-технический институт, Московская область, Долгопрудный, 141701 Россия*

**e-mail: odintsova2005@rambler.ru*

Поступила в редакцию 08.05.2019 г.
После доработки 05.07.2019 г.
Принята к публикации 23.07.2019 г.

Цистеин-богатым пептидам (ЦБП) растений принадлежит ключевая роль в защите растений от патогенов, абиотического стресса, симбиозе с бактериями, а также в развитии и размножении растений. Состав и роль ЦБП томата *Solanum lycopersicum* – важнейшей овощной культуры – практически не изучены. В настоящей работе с использованием двух подходов – скрытых марковских моделей и регулярных выражений – проведена идентификация *in silico* семейств ЦБП томата. В результате биоинформатического анализа баз данных NCBI был выявлен 191 предшественник ЦБП. Обнаруженные пептиды относятся к семействам антимикробных и регуляторных пептидов. Из антимикробных пептидов обнаружены дефензины и дефензиноподобные пептиды, неспецифические липид-переносящие белки, тионины, снакины, гевеино- и ноттиноподобные пептиды. Из регуляторных пептидов выявлены пептиды семейств RALF, Ole e 1, Ole e 6 и MEG. Кроме того, найдены пептиды с новыми цистеиновыми мотивами. Все выявленные пептиды синтезируются в виде пре- или препробелков. Предполагаемые зрелые пептиды охарактеризованы по цистеиновому мотиву, антимикробной активности и доменной структуре. Таким образом, впервые с использованием биоинформатических подходов получены системные данные об арсенале ЦБП в геноме томата, что создает базу для дальнейших функциональных исследований этих пептидов и последующего использования в сельском хозяйстве для разработки новых стратегий повышения устойчивости томата к патогенам, а также в медицине для создания лекарственных препаратов нового поколения.

Ключевые слова: иммунитет растений, томат *Solanum lycopersicum*, цистеин-богатые пептиды, цистеиновые мотивы, антимикробные пептиды, регуляторные пептиды, регулярные выражения, скрытые марковские модели.

DOI: 10.31857/S0016675820050069

Пептиды – молекулы, длина которых не превышает 100 аминокислотных остатков, обнаружены у огромного числа самых разнообразных живых существ, в том числе у растений. Совокупность пептидов целого организма составляет его пептидом. Пептиды различаются по структуре, функциям и путям биосинтеза. Часть пептидов растений образуется при метаболизме белков (структурных, транспортных белков, ферментов и пр.) и не играет функциональной роли; другие, получившие название биологически активных пептидов, участвуют в процессах роста, развития и размножения растений, а также в симбиотических взаимоотношениях и ответных реакциях на стресс [1].

Среди биологически активных пептидов особая роль принадлежит так называемым цистеин-богатым пептидам (**ЦБП**), которые содержат от 4 до 16 остатков цистеина. ЦБП подразделяют на семейства, которые различаются по числу и расположению остатков цистеина – цистеиновому мотиву, длине полипептидной цепи, первичной и пространственной структуре [2]. ЦБП синтезируются в виде предшественников (пре- или препробелков). Большинство ЦБП относится к антимикробным пептидам (**АМП**), однако среди них есть и регуляторные пептиды, участвующие в процессах развития и размножения растений. Кроме того, многочисленными исследованиями установлено, что среди биологически активных пептидов есть те, которые обладают ценными фармакологическими свойствами: они подавляют рост патогенных микроорганизмов человека, устойчивых к современным антибиотикам, ингибируют проли-

[1] Дополнительная информация для этой статьи доступна по doi 10.31857/S0016675820050069 для авторизованных пользователей.

ферацию раковых клеток человека и ангиогенез, подавляют размножение вируса иммунодефицита человека, разрушающего иммунную систему. Тем не менее исследования пептидома растений в настоящее время находятся лишь на начальной стадии развития, так как процессы выделения пептидов и их структурно-функциональной характеристики являются трудоемкими и малоэффективными. Принципиально новые возможности для анализа пептидома растений открылись с развитием методов полногеномного секвенирования (NGS) и совершенствованием методов биоинформатики [3].

Томат *Solanum lycopersicum* L. (или *Lycopersicon esculentum*) – одна из важнейших и наиболее ценных овощных культур, обладающая высокими вкусовыми качествами и богатая питательными веществами. Помимо пищевой ценности томат обладает ценными фармакологическими свойствами, проявляя антиоксидантные, противовоспалительные, иммуностимулирующие, диуретические, желчегонные и солнцезащитные свойства. До недавнего времени считалось, что основным биологически активным соединением, определяющим полезные для человека свойства томата, является каротиноид ликопин, обладающий антиканцерогенными и антиоксидантными свойствами [4, 5]. Однако накапливаются данные о том, что не только ликопин, но и другие соединения, в частности белковой природы, присутствующие в плодах, обусловливают ценные свойства томатов [6].

Цель настоящей работы состояла в характеристике семейств генов ЦБП томата *S. lycopersicum* с использованием биоинформатических подходов. Подобное исследование важно для формирования системного представления о пептидоме этой важнейшей сельскохозяйственной культуры. Кроме того, оно создаст основу для практического использования этих знаний для повышения продуктивности, товарных свойств томата и его устойчивости к болезням.

МАТЕРИАЛЫ И МЕТОДЫ

Поиск предшественников ЦБП проводили с использованием сборки генома *Solanum lycopersicum* (RefSeq ID GCF_000188115.4). Аминокислотные последовательности аннотированных белков были загружены из базы данных NCBI (https://www.ncbi.nlm.nih.gov/). Для определения ЦБП в аминокислотных последовательностях были использованы цистеиновые мотивы по классификации Сильверштейна [2] и разработанный нами алгоритм, основанный на двух подходах: скрытых марковских моделях и регулярных выражениях [7].

Для всех найденных последовательностей АМП были определены сигнальный пептид и значение изоэлектрической точки (**PI**) с использованием программ SignalP v4.1 [8] и Isoelectric Point Calculator [9]. Антимикробные свойства пептидов предсказывали при помощи программ AMPPred [10] и CAMPR3 [11]. Свойства пептидов, важные для проявления антимикробной активности, анализировали с помощью программ: ProtParam [12] – заряд молекулы, молекулярная масса, алифатический индекс, индекс нестабильности, SOSUI [13] – гидрофобность, APD3 [14] – индекс Боумана. Выявление доменов в структуре пептида осуществляли при помощи программы InterProScan [15]. Для аннотации пептидов использовали программу BLAST и базы данных GenBank. Процедуру выравнивания последовательностей и построение филогенетических деревьев проводили с помощью программы Vector NTI Advance 9.

РЕЗУЛЬТАТЫ

С использованием разработанного ранее алгоритма [7] в геноме томата нами обнаружен 191 ген, кодирующий предшественники антимикробных и регуляторных ЦБП (см. Приложение). Кроме того, обнаружены ЦБП с новыми, ранее не известными мотивами. Выравнивания аминокислотных последовательностей предшественников ЦБП томата и филогенетические деревья, построенные на их основе, представлены в Приложении.

АМП-подобные пептиды

Дефензины – древнее и наиболее изученное семейство АМП [16]. Представители этого семейства обнаружены практически у всех многоклеточных организмов. Основной структурный элемент этих АМП – цистеин-стабилизированный αβ-мотив (CSαβ-мотив) с восьмью или десятью остатками цистеина, образующими дисульфидные связи, который выявлен также в токсинах беспозвоночных животных. По первичной структуре растительные дефензины существенно различаются между собой, что и определяет наличие у них различных биологических активностей – антифунгальной, антибактериальной и инсектицидной [17]. Кроме того, ряд дефензинов ингибируют ферменты – протеиназы и амилазы, а некоторые являются блокаторами ионных каналов. Все перечисленные выше активности связаны с защитной ролью дефензинов от патогенов. Однако накапливаются данные о том, что дефензины также связаны с устойчивостью к абиотическому стрессу (солям тяжелых металлов и холодовому стрессу) [17]. Более того, некоторые из них участвуют в процессах развития растений [18]. Все дефензины растений синтезируются в виде предшественни-

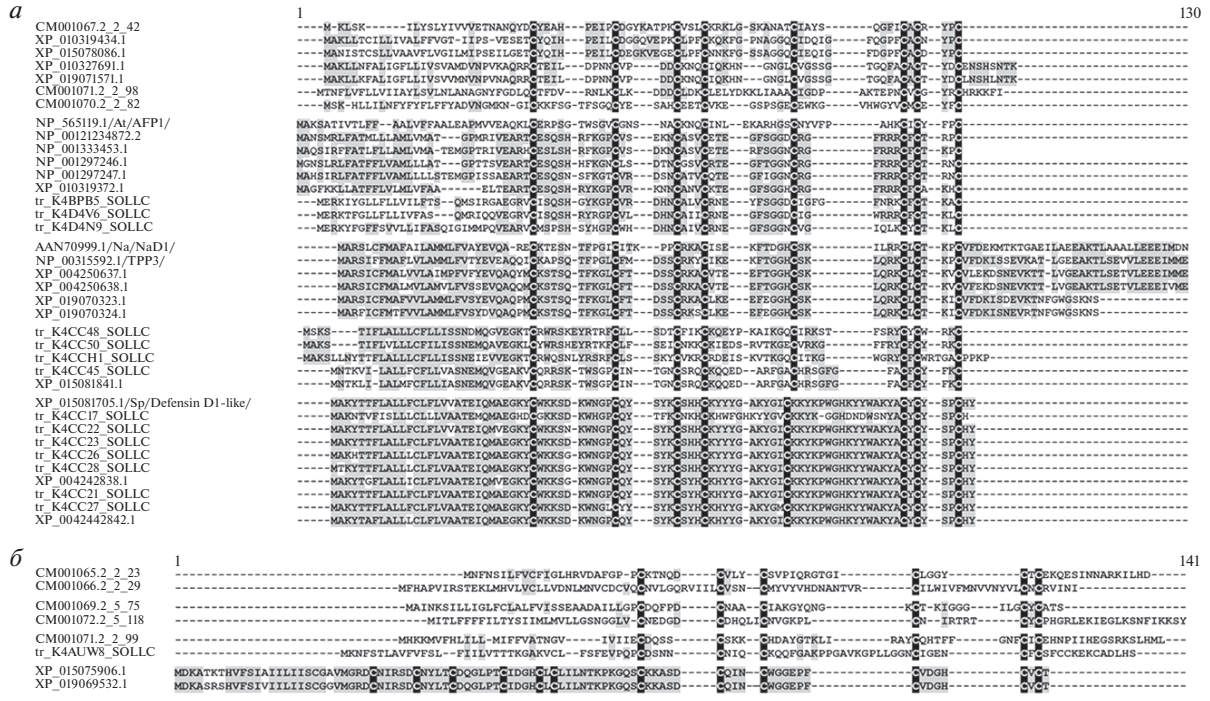

Рис. 1. Выравнивание аминокислотных последовательностей предшественников дефензинов и дефензиноподобных пептидов *S. lycopersicum* и некоторых дефензинов растений из базы данных NCBI: *а* – с 8-цистеиновым мотивом; *б* – с 6-цистеиновым мотивом. Консервативные аминокислотные остатки выделены серым цветом; остатки цистеина в зрелом пептиде выделены белым цветом на сером фоне; группы сходных пептидов разделены пустой строкой. At – *Arabidopsis thaliana*, Na – *Nicotiana alata*, Sp – *Solanum pennellii*.

ков, в зависимости от структуры которых они подразделяются на два класса: предшественники дефензинов класса I состоят из сигнального пептида и зрелого пептида, а класса II содержат еще и С-концевой продомен [16]. Этот домен необходим для транспорта предшественника в вакуоли и для защиты клеток самого растения от токсического действия пептида. По классификации Сильверштейна для дефензинов и дефензиноподобных пептидов характерны четыре типа цистеиновых мотивов – с восьмью, шестью и четырьмя остатками цистеина [2].

В геноме томата нами обнаружено 46 последовательностей с характерными цистеиновыми мотивами (Приложение). 34 последовательности имеют 8-цистеиновый мотив классических дефензинов, среди них пять последовательностей имеют С-концевой продомен, характерный для дефензинов класса II, встречающихся в цветках растений семейства пасленовых (рис. 1). Одна последовательность имеет 8-цистеиновый мотив с тремя расположенными подряд остатками цистеина (CCC). У шести последовательностей выявлен 6-цистеиновый αβ-мотив; в последовательностях еще двух предшественников (XP_019069532.1 и XP_015075906.1) этот мотив дуплицирован (рис. 1). В последовательности одного предшественника – 4-цистеиновый мотив. Кроме того, обнаружены две последовательности, содержащие помимо 4-цистеинового мотива дополнительные остатки цистеина (6–7), одна из которых (tr_K4BXU0_SOLLC) с 10 остатками цистеина сходна по мотиву с 10-цистеиновым дефензином PhD2 *Petunia hybrida*. Из всех обнаруженных последовательностей 33 аннотируются как дефензиноподобные белки (степень идентичности аминокислотных последовательностей 40–100%), четыре последовательности как “hypothetical” или “uncharacterized” белки *Solanum pennellii* и *Capsicum chinense* с высокой степенью сходства – 67–100%. Девять последовательностей не имеют гомологов в базах данных NCBI, т.е. являются новыми белками *S. lycopersicum*.

PI дефензинов и дефензиноподобных пептидов томата варьируют от 3.6 до 10.0. Как известно, большинство АМП являются катионными пептидами, однако описаны и анионные АМП, обладающие антимикробной активностью [19]. 32 пептида, по всей видимости, обладают антимикробными свойствами, о чем свидетельствуют результаты анализа пептидов с помощью программы AMPPred [10].

Неспецифические липид-переносящие белки. Липид-переносящие белки (**ЛПБ**) образуют большое семейство, представители которого обнару-

жены у всех наземных растений [20]. Это небольшие (менее 10 кДа) полипептиды, содержащие восемь консервативно расположенных остатков цистеина. Молекулы ЛПБ состоят из четырех–пяти параллельно расположенных α-спиральных участков, стабилизированных дисульфидными связями и образующих туннелеподобную гидрофобную полость для связывания и последующего транспорта липидов. ЛПБ синтезируются в виде предшественников, содержащих сигнальный пептид, обеспечивающий секрецию в апопласт, и зрелый пептид. Роль ЛПБ в растениях разнообразна: они участвуют в синтезе защитных полимеров, таких как суберин и воск, в защите от патогенов и абиотического стресса. Кроме того, ЛПБ принимают участие в развитии пыльцы и семян, растяжении клеток, образовании клубеньков и др.

Биоинформатический анализ позволил выявить 81 ген предшественников ЛПБ-подобных пептидов у *S. lycopersicum*, среди которых можно выделить 11 групп родственных пептидов (Приложение). 62 пептида обладали сходством (57–100%) с неспецифическими липид-переносящими белками *S. lycopersicon*, *S. pennellii, S. tuberosum*, *Nicotiana attenuata, N. sylvestris, N. tabacum, Capsicum annuum* и *C. baccatum*. 14 пептидов аннотировались как пролин-богатые пептиды, пять как “uncharacterized” или “hypothetical” белки. PI обнаруженных пептидов варьируют от 3.6 до 9.9. По данным программы AMPPred [10] 43 пептида обладают антимикробными свойствами. Другие, по всей видимости, выполняют сигнальные функции.

Гевеиноподобные пептиды получили свое название от гевеина – антимикробного пептида из латекса гевеи. Эти небольшие (примерно 40 аминокислотных остатков) ЦБП различаются по числу остатков цистеина в молекуле (6, 8 или 10) [21]. Пространственная структура гевеиноподобных пептидов включает α-спиральный участок и три или четыре антипараллельных β-тяжа. Все гевеиноподобные пептиды содержат хитин-связывающий сайт, состоящий из нескольких консервативных остатков гидрофобных аминокислот, которые связывают хитин и родственные полисахариды клеточных стенок патогенных грибов. Считается, что связывание этих пептидов с кончиками растущих гиф гриба препятствует элонгации гиф и тормозит рост гриба, что лежит в основе антифунгальной активности гевеиноподобных пептидов [22]. Однако выявлены и иные механизмы действия. Так, для 10-цистеиновых гевеиноподобных пептидов пшеницы показано, что они являются специфическими ингибиторами фунгализина – секретируемой металлопротеиназы фузариевых грибов, мишень действия которой – защитные хитиназы растения [23]. Известные гевеиноподобные АМП синтезируются в виде препробелков. Некоторые пептиды этого семейства образуются в результате протеолитического процессинга хитиназ класса 1 [22].

Анализ баз данных позволил выявить десять предшественников гевеиноподобных пептидов у томата с характерным 8-цистеиновым мотивом, среди которых можно выделить три группы родственных пептидов (Приложение). Пять имеют свойственную гевеиноподобным пептидам трехдоменную структуру: сигнальный пептид, зрелый пептид и С-концевой продомен. У двух предшественников С-концевой продомен отсутствует. Три предшественника, обладающих сходством с лектинами, имеют более одного цистеин-богатого домена. Все обнаруженные предшественники обладают высоким сходством (55–98%) с хитин-связывающими белками *S. lycopersicum*, *S. pennellii*, *S. tuberosum, C. baccatum* и *C. annuum*. PI зрелых пептидов варьируют в широких пределах: от 3.3 до 8.5. Семь пептидов предположительно обладают антимикробной активностью.

Снакины (Snakin/Gibberellic acid stimulated-like (GASA)) – цистеин-богатые пептиды длиной около 65 аминокислотных остатков. В молекулах этих пептидов предположительно два α-спиральных участка и 12 остатков цистеина. Установлено, что дисульфидные связи, образованные остатками цистеина, важны для поддержания пространственной структуры пептидов и их биологической активности [24]. Эти пептиды были впервые обнаружены у картофеля (St-SN1 и St-SN2), а затем и у других видов растений. Снакины синтезируются в виде пре- или препробелков. Экспрессия ряда генов снакинов усиливается в ответ на стрессовые воздействия, а также под действием гиббереллина и подавляется абсцизовой кислотой [25]. Снакины подавляют рост широкого круга грибов и бактерий в системе *in vitro* [26]. Гиперэкспрессия генов снакинов в трансгенных растениях ряда видов приводит к повышению устойчивости к грибным и бактериальным патогенам [27]. Предполагают также, что снакины участвуют в процессах развития, поскольку сайленсинг гена St-SN1 картофеля приводит к уменьшению размеров листьев, а также влияет на состав их клеточных стенок [28].

У томата нами обнаружено 17 генов предшественников снакинов (Приложение). Один из них идентичен по последовательности снакину-1 *S. tuberosum*, другие совпадают с снакинами *S. lycopersicum*, в том числе с наиболее изученным снакином томата SN-2. Все, кроме одного, снакины положительно заряжены: их PI находятся в диапазоне 6.9–9.8, и все предположительно обладают антимикробной активностью.

Ноттиноподобные пептиды (Knottin-type peptides). В молекулах этих пептидов длиной 30–50 аминокислотных остатков шесть остатков цистеина, образующих три дисульфидные связи, при этом одна из них "пронизывает" макрокольцо, образованное двумя другими дисульфидными связями [29]. Такая структура получила название цистинового узла. Она придает молекулам ноттинов необычайную стабильность к действию протеиназ и экстремальных температур. Пространственная структура ноттинов включает три антипараллельных β-тяжа и соединяющие их петли различной длины и аминокислотного состава [29]. Ноттиноподобные пептиды синтезируются в виде препробелков. Ноттины обладают различными биологическими активностями – антифунгальной, инсектицидной, антиканцерогенной [29]. Ноттины растений семейства Solanaceae являются ингибиторами карбоксипептидаз [30].

Биоинформатический поиск позволил выявить два ингибитора металлокарбоксипептидаз *S. lycopersicum,* которые сходны по последовательности с ингибиторами карбоксипептидаз другого представителя семейства пасленовых – картофеля (Приложение). По данным программы AMPPred [10] оба пептида томата не обладают антимикробными свойствами.

Тионины представляют собой цистеин-богатые пептиды, в молекулах которых шесть или восемь остатков цистеина. По пространственной структуре тионины напоминают букву Г, в которой длинное плечо образовано двумя антипараллельными α-спиралями, а короткое – двумя параллельными тяжами β-структуры [31]. Тионины токсичны для различного рода клеток – грибов, бактерий и насекомых. Кроме того, они подавляют рост опухолевых клеток человека. Цитотоксичность тионинов связана с образованием пор в мембранах, что приводит к разрушению клеток. Тионины синтезируются в виде предшественников, содержащих сигнальный пептид, зрелый пептид и С-концевой продомен с шестью консервативно расположенными остатками цистеина, который, как полагают, нейтрализует литическую активность тионинов при транспорте предшественника в вакуоли.

У предшественников тионинов Сильверштеин выделяет несколько типов цистеиновых мотивов [2]. Биоинформатический поиск последовательностей с такими мотивами позволил выявить пять последовательностей у томата, причем лишь одна из них (CM001066.2_6_6) аннотируется как тиониноподобный пептид *S. lycopersicum*, а четыре другие – как "uncharacterized proteins" *S. lycopersicum* и *S. pennellii* (Приложение). В то же время анализ этих последовательностей с помощью программы InterProScan [15] показал, что все они содержат тиониноподобный домен (Приложение). Сходство последовательностей предшественников обнаруженных тиониноподобных пептидов томата с аннотированными белками из баз данных NCBI высокое и составляет 96–100%. Расчетные PI зрелых тиониноподобных пептидов составляют 7.1–8.1, а антимикробными свойствами предположительно обладают три пептида.

Регуляторные цистеин-богатые пептиды

RALFs. Пептиды RALF, вызывающие быстрое защелачивание среды в суспензионной культуре клеток (rapid alkalinization of the medium) и молекулярная масса которых составляет примерно 5 кДа (примерно 50 аминокислотных остатков), были впервые обнаружены у табака [32]. Родственные пептиды были позднее обнаружены и у других видов растений. Пептиды RALF образуются из С-концевой области препробелка. В области пропоследовательности пептиды RALF содержат сайт с двумя остатками основных аминокислот (dibasic site) RRXL, который узнается субтилизиноподобной протеазой, осуществляющей посттрансляционный процессинг предшественника. Зрелый пептид имеет обычно четыре остатка цистеина, которые образуют две дисульфидные связи, важные для активности. Кроме того, пептиды RALF содержат консервативный мотив YISY вблизи N-конца, участвующий в связывании рецептора, мотив GASYY (между двумя первыми цистеинами) и мотив PYXRGCS [33]. Пептиды RALF относятся к сигнальным пептидам, которые регулируют процессы роста и развития растений [34]. Они подавляют рост корневых волосков и элонгацию клеток гипокотиля и корня, ингибируют образование клубеньков и регулируют рост пыльцевых трубок.

В базах данных нами найдено восемь последовательностей предшественников RALF у томата, среди которых выделяются три группы родственных пептидов (Приложение). Все, кроме двух последовательностей, содержат сайт RRXL, который узнается протеазой. У четырех предшественников в мотиве YISY есть замены. Все обнаруженные предшественники аннотируются в базах данных NCBI как RALF или RALF-подобные белки *S. lycopersicum* и *C. baccatum* с 66–100% сходства. Предполагаемые зрелые пептиды RALF являются высокоосновными. Более того, согласно программе AMPPred [10] их молекулы обладают свойствами АМП.

Ole e 1 и Ole e 6 являются аллергенами пыльцы оливы европейской *Olea europaea*, откуда и происходит название белков этого семейства [35].

Ole e 1 представляет собой белок с молекулярной массой 20 кДа различной степени гликозили-

рования. Ole e 1 состоит из единственной полипептидной цепи, состоящей из 145 аминокислотных остатков и обладающей активностью ингибитора трипсина [36]. Пространственная структура полипептида не установлена, однако показано, что в нем 22% α-спирали, 38% β-структуры и 40% неупорядоченной структуры (random coil). Обнаружено сходство аминокислотной последовательности Ole e 1 с белками пыльцы других видов растений: кукурузы, томата, риса, березы и арабидопсиса.

У томата нами обнаружено пять последовательностей полипептидов семейства Ole e 1, три из которых относятся к LAT52-подобным белкам (Приложение). Анализ InterProScan [15] показал, что все они содержат домен, характерный для Ole e 1. PI зрелых пептидов составляют 4.3–7.6. Антимикробная активность предсказана для трех зрелых пептидов.

Ole e 6 – кислый (PI = 4.2), 50-членный полипептид с высоким содержанием заряженных аминокислотных остатков и шестью остатками цистеина. В отличие от Ole e 1, в Ole e 6 не обнаружены сайты N-гликозилирования. В основе пространственной структуры Ole e 6 – две антипараллельные α-спирали, соединенные короткой петлей, и длинный неструктурированный С-конец [37]. Биологическая роль Ole e 6 и Ole e 1 неизвестна.

В геноме томата обнаружено 11 последовательностей с характерным для Ole e 6 мотивом. Однако лишь шесть из них аннотируются как основной аллерген пыльцы (major pollen allergen) *S. lycopersicum* и *C. baccatum*, и лишь четыре из них являются отрицательно заряженными полипептидами. Один не имеет гомологов в базах данных NCBI. Другие аннотируются как “uncharacterized” или “hypothetical” белки *S. lycopersicum, S. pennellii* и *C. chinense*. Три из обнаруженных предшественников предположительно являются АМП.

MEG1 – это ген, специфически экспрессирующийся в базальном слое транспортных клеток эндосперма кукурузы [38]. Родственные гены обнаружены у других видов злаков. У кукурузы предшественник MEG1 состоит из сигнального и зрелого пептида, в котором восемь остатков цистеина. Полагают, что он может выполнять роль структурного белка в транспортной базальной области эндосперма, либо выступать в качестве защитной молекулы, либо играть регуляторную роль в процессах транспорта питательных веществ из материнских клеток к развивающемуся зародышу. Для белков семейства MEG описаны два цистеиновых мотива [2].

У томата обнаружены две последовательности MEG (Приложение). Они обладают 100%-ным сходством с “uncharacterized” белками *S. lycopersicum*. Зрелые пептиды предположительно обладают антимикробной активностью, что свидетельствует в пользу их защитной роли.

Цистеин-богатые пептиды с новыми мотивами

В ходе поиска были обнаружены четыре пептида с новыми, ранее неизвестными мотивами (Приложение). Все они имеют сходство (31–94%) с “uncharacterized” белками растений семейства Solanaceae. Расчетные молекулярные массы этих пептидов варьируют от 7275 до 10029 Да. Нами были проанализированы свойства этих пептидов, важные для проявления антимикробной активности (заряд молекулы, PI, алифатический индекс, индекс нестабильности, гидрофобность, индекс Боумана, наличие известных доменов) (Приложение). Хотя согласно программе AMPPred [10] лишь один из этих пептидов обладает антимикробными свойствами, определенные нами характеристики для этих пептидов, а также данные программы предсказания антимикробных свойств CAMPr3 [11] свидетельствуют в пользу того, что все они обладают антимикробной активностью.

ОБСУЖДЕНИЕ

Исследования последних лет свидетельствуют о том, что пептиды выполняют в растениях чрезвычайно важные функции. Однако у томата изучены свойства лишь отдельных цистеин-богатых пептидов. Так, из семейства дефензинов был изучен дефензин TPP3 (tomato pistil predominant 3), относящийся к классу II [39]. Исследование механизмов действия дефензина TPP3 показало, что он образует димеры и специфически взаимодействует с фосфатидилинозитол-4,5-дифосфатом (PIP2) мембран клеток грибов, что индуцирует увеличение проницаемости мембран и последующий лизис клеток [40]. В другом исследовании была показана антимикробная активность 17-членного пептида, который соответствовал последовательности γ-кора (универсального для всех АМП структурного элемента) одного из дефензинов томата, а также, что этот пептид является эффективным ингибитором роста ряда патогенных бактерий человека [41]. О функциях ЛПБ томата также известно немного. Так, было показано, что один из ЛПБ томата участвует в процессе размягчении плодов, влияя на активность основного фермента деградации пектина – полигалактуроназы [42]. ЛПБ семейства DIR, обнаруженные в соке флоэмы растений томата, предположительно выступают в роли сигнальных молекул при развитии системной приобретенной устойчивости – феномена повышенной устойчивости растения к широкому кругу

патогенов после первичного заражения [43]. Следует также отметить, что ЛПБ томата, как и ЛПБ других видов растений, являются пищевыми аллергенами [44]. Из семейства снакинов у томата изучен снакин SN2. Показано, что в системе *in vitro* он подавляет рост различных бактерий и грибов за счет образования пор в мембранах [45]. Ген этого пептида экспрессируется конститутивно, причем при поранении, заражении патогенами и обработке метилжасмонатом его экспрессия усиливается во всех органах растения. Абиотический стресс (экстремальные температуры и дегидратация) не влияет на экспрессию гена SN2 [46], в то время как суперэкспрессия этого гена увеличивала устойчивость трансгенных растений томата к патогенной бактерии *Clavibacter michiganensis* [45]. Из семейства ноттинов один пептид, обладающий свойствами ингибитора металлокарбоксипептидаз, был выделен из плодов томата. Было показано, что он накапливается в листьях при поранении, т.е. выполняет в растениях томата защитную роль [47]. Недавно было показано, что ингибиторы металлокарбоксипептидаз томата играют также регуляторную роль при развитии плодов [48]. Ноттины плодов томата обладают ценными для человека свойствами, в частности антиангиогенной активностью, что делает эти молекулы перспективными для фармакологии при разработке противоопухолевых препаратов [6]. Из сигнальных пептидов у томата исследована биологическая роль одного из пептидов RALF. Добавление синтетического пептида к прорастающим семенам томата и арабидопсиса приводило к прекращению элонгации первичных корней и роста корневых волосков [32]. Из семейства Ole e 1 для пептида LAT52 томата показано, что он играет существенную роль в гидратации пыльцы и/или прорастании [49]. Таким образом, лишь для единичных представителей семейств ЦБП томата имеются сведения об их функциях.

Цель настоящего исследования заключалась в поиске ЦБП различных семейств у томата биоинформатическим методом. Следует отметить, что *in silico* поиск предшественников ЦБП у растений является нетривиальной задачей, поскольку сходство их последовательностей, за исключением относительно консервативного цистеинового мотива, может быть весьма низким, а между пептидами разных семейств и вообще отсутствовать [50], в то же время в пределах одного мотива число аминокислотных остатков между соседними остатками цистеина может варьировать. В связи с этим в настоящей работе для идентификации семейств ЦБП были использованы два подхода – скрытые марковские модели и регулярные выражения, эффективность которых была подтверждена нами ранее [7].

В результате проведенного анализа у томата *S. lycopersicum* был выявлен 191 предшественник ЦБП. Они относятся к важнейшим семействам АМП растений: дефензинам, ЛПБ, тионинам, снакинам, гевеино- и ноттиноподобным пептидам, а также к семействам пептидов, не связанных с защитой: RALF, Ole e 1 и Ole e 6, а также MEG. Кроме того, обнаружен ряд пептидов с новыми, ранее не описанными цистеиновыми мотивами. Все выявленные ЦБП синтезируются в виде предшественников – пре- или препробелков. Каждое семейство состоит из групп родственных пептидов, что указывает на то, что дупликации генов играли существенную роль в эволюции ЦБП растений. Наиболее многочисленными у томата являются семейства ЛПБ и дефензинов, что, вероятно, связано с разнообразием выполняемых этими пептидами функций.

Таким образом, нами впервые с использованием биоинформатических подходов получены системные данные об арсенале генов ЦБП у томата, что создает базу для дальнейших функциональных исследований этих пептидов и вносит вклад в представление об иммунной системе этого вида растений. Обнаруженные в ходе функциональных исследований биологически активные пептиды смогут найти практическое применение в медицине для создания новых лекарственных препаратов и в сельском хозяйстве для повышения устойчивости растений к патогенам.

Работа выполнена при финансовой поддержке гранта РНФ № 16-16-00032.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Farrokhi N., Whitelegge J.P., Brusslan J.A.* Plant peptides and peptidomics // Plant Biotechnol. J. 2008. V. 6. № 2. P. 105–134. https://doi.org/10.1111/j.1467-7652.2007.00315.x
2. *Silverstein K.A., Moskal W.A., Jr., Wu H.C. et al.* Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants // Plant J. 2007. V. 51. № 2. P. 262–280. https://doi.org/10.1111/j.1365-313X.2007.03136.x
3. *Koehbach J., Jackson K.A.V.* Unravelling peptidomes by *in silico* mining // Peptidomics. 2015. V. 2. № 1. P. 17–25. https://doi.org/10.1515/ped-2015-0002

4. *Choksi P.M., Joshi V.Y.* A review on lycopene – extraction, purification, stability and applications // Int. J. Food Properties. 2007. V. 10. P. 289–298. https://doi.org/10.1080/10942910601052699

5. *Rowles J.L. 3rd, Ranard K.M., Smith J.W. et al.* Increased dietary and circulating lycopene are associated with reduced prostate cancer risk: a systematic review and meta-analysis // Prostate Cancer Prostatic Dis. 2017. V. 20. № 4. P. 361–377. https://doi.org/10.1038/pcan.2017.25

6. *Cavallini C., Trettene M., Degan M. et al.* Anti-angiogenic effects of two cystine-knot miniproteins from tomato fruit // Br. J. Pharmacol. 2011. V. 162. № 6. P. 1261–1273. https://doi.org/10.1111/j.1476-5381.2010.01154.x

7. *Odintsova T.I., Slezina M.P., Istomina E.A. et al.* Defensin-like peptides in wheat analyzed by whole-transcriptome sequencing: a focus on structural diversity and role in induced resistance // PeerJ. 2019. V. 7. e6125. https://doi.org/10.7717/peerj.6125

8. *Bendtsen J.D., Nielsen H., von Heijne G., Brunak S.* Improved prediction of signal peptides: SignalP 3.0 // J. Mol. Biol. 2004. V. 340. № 4. P. 783–795. https://doi.org/10.1016/j.jmb.2004.05.028

9. *Kozlowski L.P.* IPC – Isoelectric Point Calculator // Biology Direct. 2016. V. 11:55. https://doi.org/10.1186/s13062-016-0159-9

10. *Porto W.F., Pires Á.S., Franco O.L.* CS-AMPPred: an updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides // PLoS One. 2012. V. 7. № 12. e51444. https://doi.org/10.1371/journal.pone.0051444

11. *Waghu F.H., Barai R.S., Gurung P., Idicula-Thomas S.* CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides // Nucl. Acids Res. 2016. V. 44. № D1. P. D1094–D1097. https://doi.org/10.1093/nar/gkv1051

12. *Gasteiger E., Hoogland C., Gattiker A. et al.* Protein identification and analysis tools on the ExPASy server // The Proteomics Protocols Handbook / Ed. Walker John M. Humana Press, 2005. P. 571–607.

13. *Hirokawa T., Boon-Chieng S., Mitaku S.* SOSUI: classification and secondary structure prediction system for membrane proteins // Bioinformatics. 1998. V. 14. № 4. P. 378–379.

14. *Wang G., Li X., Wang Z.* APD3: the antimicrobial peptide database as a tool for research and education // Nucl. Acids Res. 2016. V. 44. № D1. P. D1087–D1093. https://doi.org/10.1093/nar/gkv1278

15. *Quevillon E., Silventoinen V., Pillai S. et al.* InterProScan: protein domains identifier // Nucl. Acids Res. 2005. V. 33. P. W116–W120. https://doi.org/10.1093/nar/gki442

16. *Lay F.T., Anderson M.A.* Defensins – components of the innate immune system in plants // Curr. Protein Pept. Sci. 2005. V. 6. № 1. P. 85–101. https://doi.org/10.2174/1389203053027575

17. *Parisi K., Shafee T.M.A., Quimbar P. et al.* The evolution, function and mechanisms of action for plant defensins // Semin. Cell Dev. Biol. 2019. V. 88. P. 107–118. https://doi.org/10.1016/j.semcdb.2018.02.004

18. *Stotz H.U., Spence B., Wang Y.* A defensin from tomato with dual function in defense and development // Plant Mol. Biol. 2009. V. 71. № 1–2. P. 131–143. https://doi.org/10.1007/s11103-009-9512-z

19. *Harris F., Dennison S.R., Phoenix D.A.* Anionic antimicrobial peptides from eukaryotic organisms // Curr. Protein Pept. Sci. 2009. V. 10. № 6. P. 585–606. https://doi.org/10.2174/138920309789630589

20. *Kader J.C.* Lipid-transfer proteins in plants // Annu. Rev. Plant Physiol. Plant Mol. Biol. 1996. V. 47. P. 627–654. https://doi.org/10.1146/annurev.arplant.47.1.627

21. *Slavokhotova A.A., Shelenkov A.A., Andreev Y.A., Odintsova T.I.* Hevein-like antimicrobial peptides of plants // Biochemistry (Mosc). 2017. V. 82. № 13. P. 1659–1674. https://doi.org/10.1134/S0006297917130065

22. *Van den Bergh K.P., Rougé P., Proost P. et al.* Synergistic antifungal activity of two chitin-binding proteins from spindle tree (*Euonymus europaeus* L.) // Planta. 2004. V. 219. № 2. P. 221–232. https://doi.org/10.1007/s00425-004-1238-1

23. *Slavokhotova A.A., Naumann T.A., Price N.P. et al.* Novel mode of action of plant defense peptides – hevein-like antimicrobial peptides from wheat inhibit fungal metalloproteases // FEBS J. 2014. V. 281. № 20. P. 4754–4764. https://doi.org/10.1111/febs.13015

24. *Harris P.W., Yang S.H., Molina A., et al.* Plant antimicrobial peptides snakin-1 and snakin-2: chemical synthesis and insights into the disulfide connectivity // Chemistry. 2014. V. 20. № 17. P. 5102–5210. https://doi.org/10.1002/chem.201303207

25. *Berrocal-Lobo M., Segura A., Moreno M. et al.* Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection // Plant Physiol. 2002. V. 128. № 3. P. 951–961. https://doi.org/10.1104/pp.010685

26. *Segura A., Moreno M., Madueño F. et al.* Snakin-1, a peptide from potato that is active against plant pathogens // Mol. Plant. Microbe Interact. 1999. V. 12. № 1. P. 16–23. https://doi.org/10.1094/MPMI.1999.12.1.16

27. *Almasia N.I., Bazzini A.A., Hopp H.E., Vazquez-Rovere C.* Overexpression of snakin-1 gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants // Mol. Plant Pathol. 2008. V. 9. № 3. P. 329–338.

28. *Nahirñak V., Almasia N.I., Fernandez P.V. et al.* Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition // Plant Physiol. 2012. V. 158. № 1. P. 252–263. https://doi.org/10.1104/pp.111.186544

29. *Tam J.P., Wang S., Wong K.H., Tan W.L.* Antimicrobial peptides from plants // Pharmaceuticals (Basel). 2015. V. 8. № 4. P. 711–757. https://doi.org/10.3390/ph8040711

30. *Cotabarren J., Tellechea M.E., Tanco S.M. et al.* Biochemical and MALDI-TOF mass spectrometric characterization of a novel native and recombinant cystine knot miniprotein from *Solanum tuberosum* subsp. *andigenum* cv. Churqueña // Int. J. Mol. Sci. 2018. V. 19. № 3. pii: E678. https://doi.org/10.3390/ijms19030678

31. *Stec B.* Plant thionins – the structural perspective // Cell. Mol. Life Sci. 2006. V. 63. № 12. P. 1370–1385. https://doi.org/10.1007/s00018-005-5574-5

32. *Pearce G., Moura D.S., Stratmann J., Ryan C.A., Jr.* RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development // Proc. Natl Acad. Sci. USA. 2001. V. 98. № 22. P. 12843–12847. https://doi.org/10.1073/pnas.201416998

33. *Bedinger P.A., Pearce G., Covey P.A.* RALFs: peptide regulators of plant growth // Plant Signal. Behav. 2010. V. 5. № 11. P. 1342–1346. https://doi.org/10.4161/psb.5.11.12954

34. *Chen L.Y.* Small peptides, big roles – RALFs regulate pollen tube growth and burst in plant reproduction // J. Genet. Genomics. 2018. V. 45. № 3. P. 121–123. https://doi.org/10.1016/j.jgg.2018.02.006

35. *Quiralte J., Palacios L., Rodríguez R. et al.* Modelling diseases: the allergens of *Olea europaea* pollen // J. Investig. Allergol. Clin. Immunol. 2007. V. 17. Suppl. 1. P. 76–82.

36. *Villalba M., Batanero E., Monsalve R.I. et al.* Cloning and expression of Ole e I, the major allergen from olive tree pollen. Polymorphism analysis and tissue specificity // J. Biol. Chem. 1994. V. 269. № 21. P. 15217–15222.

37. *Treviño M.A., García-Mayoral M.F., Barral P. et al.* NMR solution structure of Ole e 6, a major allergen from olive tree pollen // J. Biol. Chem. 2004. V. 279. № 37. P. 39035–39041. https://doi.org/10.1074/jbc.M406045200

38. *Gutiérrez-Marcos J.F., Costa L.M., Biderre-Petit C. et al. maternally expressed gene1* Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression // Plant Cell. 2004. V. 16. № 5. P. 1288–1301. https://doi.org/10.1105/tpc.019778

39. *Milligan S.B., Gasser C.S.* Nature and regulation of pistil-expressed genes in tomato // Plant Mol. Biol. 1995. V. 28. № 4. P. 691–711.

40. *Baxter A.A., Richter V., Lay F.T. et al.* The tomato defensin TPP3 binds phosphatidylinositol (4,5)-bisphosphate via a conserved dimeric cationic grip conformation to mediate cell lysis // Mol. Cell. Biol. 2015. V. 35. № 11. P. 1964–1978. https://doi.org/10.1128/MCB.00282-15

41. *Rigano M.M., Romanelli A., Fulgione A. et al.* A novel synthetic peptide from a tomato defensin exhibits antibacterial activities against *Helicobacter pylori* // J. Pept. Sci. 2012. V. 18. № 12. P. 755–762. https://doi.org/10.1002/psc.2462

42. *Tomassen M.M., Barrett D.M., van der Valk H.C., Woltering E.J.* Isolation and characterization of a tomato non-specific lipid transfer protein involved in polygalacturonase-mediated pectin degradation // J. Exp. Bot. 2007. V. 58. № 5. P. 1151–1160. https://doi.org/10.1093/jxb/erl288

43. *Mitton F.M., Pinedo M.L., de la Canal L.* Phloem sap of tomato plants contains a DIR1 putative ortholog // J. Plant Physiol. 2009. V. 166. № 5. P. 543–547.

44. *Volpicella M., Leoni C., Fanizza I. et al.* Expression and characterization of a new isoform of the 9 kDa allergenic lipid transfer protein from tomato (variety San Marzano) // Plant Physiol. Biochem. 2015. V. 96. P. 64–71. https://doi.org/10.1016/j.plaphy.2015.07.019

45. *Balaji V., Smart C.D.* Over-expression of snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to *Clavibacter michiganensis* subsp. *michiganensis* in transgenic tomato (*Solanum lycopersicum*) // Transgenic Res. 2012. V. 21. № 1. P. 23–37. https://doi.org/10.1007/s11248-011-9506-x

46. *Herbel V., Sieber-Frank J., Wink M.* The antimicrobial peptide snakin-2 is upregulated in the defense response of tomatoes (*Solanum lycopersicum*) as part of the jasmonate-dependent signaling pathway // J. Plant Physiol. 2017. V. 208. P. 1–6. https://doi.org/10.1016/j.jplph.2016.10.006

47. *Díez-Díaz M., Conejero V., Rodrigo I. et al.* Isolation and characterization of wound-inducible carboxypeptidase inhibitor from tomato leaves // Phytochemistry. 2004. V. 65. № 13. P. 1919–1924. https://doi.org/10.1016/j.phytochem.2004.06.007

48. *Molesini B., Rotino G.L., Dusi V. et al.* Two metallocarboxypeptidase inhibitors are implicated in tomato fruit development and regulated by the Inner No Outer transcription factor // Plant Sci. 2018. V. 266. P. 19–26. https://doi.org/10.1016/j.plantsci.2017.10.011

49. *Muschietti J., Dircks L., Vancanneyt G., McCormick S.* LAT52 protein is essential for tomato pollen development: pollen expressing antisense *LAT52* RNA hydrates and germinates abnormally and cannot achieve fertilization // Plant J. 1994. V. 6. № 3. P. 321–338.

50. *Shelenkov A.A., Slavokhotova A.A., Odintsova T.I.* Cysmotif searcher pipeline for antimicrobial peptide identification in plant transcriptomes // Biochemistry (Mosc). 2018. V. 83. № 11. P. 1424–1432. https://doi.org/10.1134/S0006297918110135

In silico Identification of Gene Families Encoding Cysteine-Rich Peptides in *Solanum lycopersicum* L.

E. A. Istomina[a], M. P. Slezina[a], A. S. Kovtun[b], and T. I. Odintsova[a, *]

[a] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia*

[b] *Moscow Institute of Physics and Technology, Moscow oblast, Dolgoprudny, 141701 Russia*

**e-mail: odintsova2005@rambler.ru*

Cysteine-rich plant peptides (CRP) play a key role in protection of plants from pathogens, abiotic stress, symbiosis with bacteria, and in plant development and reproduction. The composition and role of CRP in tomato *Solanum lycopersicum*, the most important vegetable, has not been studied so far. In this paper, using two approaches – hidden Markov models and regular expressions, in silico identification of tomato CRP families was performed. A bioinformatic research in NCBI databases revealed 191 CRP precursors. The peptides found belong to the families of antimicrobial and signaling peptides. Among the antimicrobial peptides,defensins and defensin-like peptides, nonspecific lipid-transfer proteins, thionins, snakins, hevein- and knottin-like peptides were found. Among signaling peptides, the peptides belonging to the RALF, Ole e 1, Ole e 6 and MEG families were discovered. In addition, peptides with new cysteine motifs were identified. All discovered CRPs are synthesized as pre- or preproproteins. The predicted mature peptides were characterized by cysteine motifs, antimicrobial activity and domain structure. Thus, for the first time using bioinformatic approaches, systemic data on the arsenal of CRP in the tomato genome were obtained, which creates a basis for further functional studies of these peptides and their subsequent use in agriculture to develop new strategies of increasing pathogen resistance in tomato, as well as in medicine to create next-generation drugs.

Keywords: plant immunity, tomato *Solanum lycopersicum*, cysteine-rich peptides, cysteine motifs, antimicrobial peptides, signaling peptides, regular expressions, hidden Markov models.