

На правах рукописи

**Смышляев Георгий Андреевич**

**СТРУКТУРНОЕ РАЗНООБРАЗИЕ И ЭВОЛЮЦИЯ NON-LTR-  
РЕТРОТРАНСПОЗОНОВ СУПЕРСЕМЕЙСТВА *L1* ИЗ ГЕНОМОВ  
РАСТЕНИЙ**

**03.02.07 – генетика**

**АВТОРЕФЕРАТ**  
**диссертации на соискание ученой степени**  
**кандидата биологических наук**

**Новосибирск 2014**

Работа выполнена в лаборатории молекулярно-генетических систем Федерального государственного бюджетного учреждения науки Института цитологии и генетики Сибирского отделения Российской академии наук, г. Новосибирск.

Научный руководитель: **Бугров Александр Геннадьевич**  
доктор биологических наук, доцент, профессор  
кафедры общей биологии и экологии НГУ

Официальные оппоненты: **Щербаков Дмитрий Юрьевич**  
доктор биологических наук, заведующий  
лабораторией геносистематики, Федеральное  
государственное бюджетное учреждение науки  
Лимнологический институт Сибирского отделения  
Российской академии наук, г. Иркутск

**Белякин Степан Николаевич**  
кандидат биологических наук, заведующий  
лабораторией геномики, Федеральное  
государственное бюджетное учреждение науки  
Институт молекулярной и клеточной биологии  
Сибирского отделения Российской академии наук, г  
Новосибирск

Ведущая организация: Федеральное государственное учреждение науки  
Государственный научный центр вирусологии и биотехнологии «Вектор» Федеральной  
службы по надзору в сфере защиты прав потребителей и благополучия человека.

Защита диссертации состоится «\_\_» \_\_\_\_\_ 201\_ г. на утреннем  
заседании диссертационного совета Д 003.011.01 по защите диссертаций на соискание  
ученой степени кандидата наук, на соискание ученой степени доктора наук в ИЦиГ  
СО РАН в конференц-зале Института по адресу:

пр. академика Лаврентьева 10, г. Новосибирск, 630090

тел/факс: (383) 363-49-06 (1321); e-mail: [dissov@bionet.nsc.ru](mailto:dissov@bionet.nsc.ru).

Факс: (383) 333-12-78

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте  
института [www.bionet.nsc.ru](http://www.bionet.nsc.ru).

Автореферат разослан «\_\_» \_\_\_\_\_ 201\_ г.

Ученый секретарь  
диссертационного совета,  
доктор биологических наук

Т.М. Хлебодарова

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность темы исследования

Ретротранспозоны – это мобильные элементы (МЭ), которые встраиваются в новое место в геноме с использованием РНК в качестве матрицы для синтеза новой копии ДНК. Ретротранспозоны присутствуют во многих геномах эукариот, и в том числе во всех без исключения исследованных на данный момент геномах растений (Ibarra-Laclette *et al.*, 2013). Значение ретротранспозонов для генома хозяина во многом остается недостаточно полно описанным. Известно, что, как и все МЭ, ретротранспозоны являются важнейшими факторами эволюции – они способны вызывать геномные перестройки и дупликации генов, а также своим встраиванием регулировать активность генов (Kumar, Bennetzen, 1999). В последнее десятилетие все большую актуальность приобретают вопросы, связанные с проблемами регуляции активности ретротранспозонов и других МЭ в геномах эукариотических организмов. Ярким примером в этой сфере является открытие механизмов инактивации МЭ у растений (Slotkin, 2010; Lisch, 2012). Возможно, благодаря этим активно действующим механизмам МЭ не получили широкого распространения у некоторых растений, в частности в геноме пузырчатки *Utricularia gibba*, геном которой всего на 3% состоит из МЭ (Ibarra-Laclette *et al.*, 2013).

Отсюда возникает вопрос, так ли нужны МЭ геному, если, как видно на примере пузырчатки, геном вполне может успешно существовать и без них? Сравнительный анализ растительных геномов показывает, что геном пузырчатки – скорее исключение – даже в свете наличия жестких механизмов регуляции пролиферации МЭ, многие геномы растений наполнены ретротранспозонами. Геном кукурузы, например, более чем на 60% состоит из ретроэлементов. В среднем, содержание МЭ в геномах растений составляет 10-60% (Kumar, Bennetzen, 1999). Для того чтобы понять какие принципы лежат в основе широкого распространения ретротранспозонов, на первом этапе необходимо оценить масштабы этого распространения и разнообразия форм ретротранспозонов. Сравнительный анализ распространения и разнообразия ретротранспозонов является одним из основных подходов в получении информации о важности ретротранспозонов для генома. Иными словами, в контексте получения информации о потенциальной роли и значении ретротранспозонов для геномов растений, ключевым аспектом является вопрос их эволюции и распространения в этих геномах.

Одной из наименее изученных групп ретротранспозонов растений являются non-LTR-ретротранспозоны суперсемейства *L1*. Элементы из этой группы широко представлены в геномах млекопитающих и, в том числе, человека

(Heitkam, Schmidt, 2009). *L1*-элементы человека к настоящему времени детально изучены, в то время как у растений описано всего несколько отдельных представителей *L1*-ретротранспозонов. Более того, до сих пор не было предпринято попыток комплексной идентификации основных филогенетических групп *L1*-ретротранспозонов растений, что значительно упростило бы работу с информацией относительно этих элементов и расширило наши представления об эволюции этой группы МЭ.

### **Цели и задачи исследования**

Цель: выявить разнообразие и реконструировать эволюцию *L1*-non-LTR-ретротранспозонов растений.

Задачи:

1. Поиск *L1*-non-LTR-ретротранспозонов в геномах растений и анализ их структур с помощью биоинформатических методов.
2. Реконструкция филогенетических взаимоотношений и определение основных филогенетических групп non-LTR-ретротранспозонов растений.
3. Сравнительный анализ структурных характеристик найденных элементов с помощью биоинформатических и экспериментальных методов исследований.

### **Научная новизна работы**

В работе впервые проведен комплексный биоинформатический анализ *L1*-ретротранспозонов в широком круге растительных геномов, в результате чего выявлено 5 филогенетических групп этих элементов, представители каждой из которых обладают специфичными для данной группы структурными характеристиками. На основании анализа этих характеристик сделан вывод о важной роли модульного строения исследуемых ретротранспозонов в эволюции этих элементов. Кроме того было показано, что рибонуклеаза H (RNH) этих non-LTR-ретротранспозонов филогенетически ближе к “архейной” RNH, чем к RNH других non-LTR-ретротранспозонов и, по-видимому, была приобретена *L1*-элементами независимо от остальных non-LTR-ретротранспозонов. В результате биохимического исследования домена RNH *L1b*-ретротранспозона, относящегося к *L1*-элементам, было продемонстрировано, что этот домен ферментативно активен *in vitro*. В целом, проведенный анализ домена RNH *L1* позволил сделать вывод о потенциальной роли ретротранспозонов в распространении RNH в геномах растений, бактерии и архей.

### **Положение, выносимое на защиту**

Разнообразие структурных форм *L1*-non-LTR-ретротранспозонов растений возникло в результате приобретения данными элементами новых функциональных доменов.

## **Теоретическая и практическая значимость исследования**

1. Идентифицированные и классифицированные в данной работе элементы могут послужить в качестве референсной системы для поиска и анализа ретротранспозонов в *de novo* секвенированных геномах растений, что может существенно облегчить работу с этими геномами.

2. Разнообразие структурных вариантов *L1*-non-LTR-ретротранспозонов и связанные с этим разнообразием события в эволюции этих элементов представляют новые данные в пользу представления об эволюции ретротранспозонов как о модульной эволюции.

3. Выявленное в данной работе распространение гомологов “архейной” RNН как в составе *L1*-non-LTR-ретротранспозонов, так и в геномах растений, бактерий и архей, говорит о потенциальной роли *L1*-элементов в широком распространении этой RNН.

4. Разработан протокол синтеза и очистки домена RNН в *Escherichia coli*, который может быть использован для проведения последующих исследований этого домена, а также помочь в оптимизации условий для подобных исследований.

## **Вклад автора**

Основные результаты получены автором самостоятельно. Биоинформатический анализ проводился в Лаборатории молекулярно-генетических систем Института цитологии и генетики СО РАН. Экспериментальная часть работы по изучению биохимических характеристик домена RNН L1b-ретротранспозона проводилась в лаборатории Орсоли Барабас отдела Структурной и Вычислительной биологии EMBL (Barabas Lab, Structural and Computational Biology Unit, EMBL, Heidelberg).

## **Апробация работы**

Результаты работы были представлены на научных конференциях «The Sixth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2008)» (Новосибирск, 2008), «XLVII международная научная студенческая конференция» (Новосибирск, 2009), «XLVIII международная научная студенческая конференция» (Новосибирск, 2010).

По теме диссертации опубликовано шесть работ. Из них пять – в рецензируемых зарубежных журналах и одна – в рецензируемом отечественном журнале.

## **Структура и объем работы**

Диссертация состоит из введения, трех глав, заключения и списка литературы. Работа изложена на 104 страницах, содержит 21 рисунок и четыре таблицы.

## МАТЕРИАЛЫ И МЕТОДЫ

Для биоинформатического поиска и анализа *L1*-элементов в прочитанных последовательностях полных геномов растений и в базе данных RepBase (Jurka *et al.*, 2005) нами был разработан алгоритм поиска целых копий non-LTR-ретротранспозонов с использованием скрытых марковских моделей (НММ-профилей), представляющих собой статистическую модель аминокислотной последовательности региона обратной транскриптазы (RT) *L1*-элементов.

Для филогенетического анализа использовались методы объединения ближайших соседей из пакета программ MEGA5 (Tamura *et al.*, 2011) и метод максимального правдоподобия из пакета программ PhyML (Guindon *et al.*, 2010). Для поиска структурных доменов использовалась база данных консервативных доменов NCBI (Marchler-Bauer *et al.*, 2011), а также онлайн-ресурс детекции гомологии и предсказания структуры белков HHpred (Söding *et al.*, 2005).

Сравнительный анализ домена рибонуклеазы H (RNH) *L1*-non-LTR-ретротранспозонов проводился с использованием веб-сервиса PROMALS3D (<http://prodata.swmed.edu/promals3d/>), который осуществляет выравнивание с учетом предсказанных вторичных структур анализируемых полипептидов.

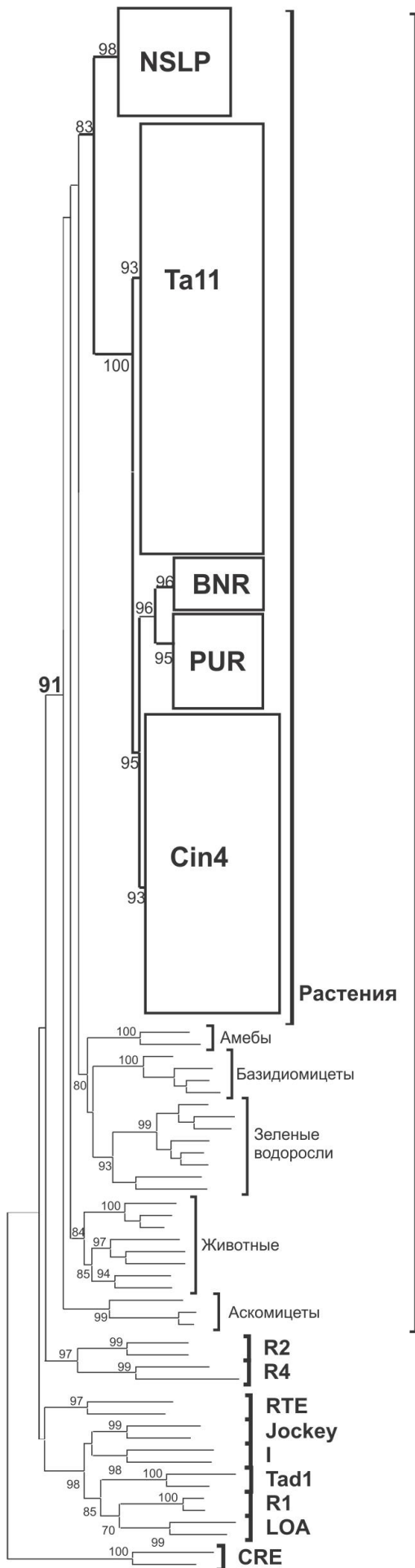
Экспериментальные исследования домена RNH non-LTR-ретротранспозона *L1b* из генома *Ipomoea batatas* проводились с использованием рекомбинантного белка, соответствующего этому домену, экспрессированного в *E. coli*. Для исследования активности этой рекомбинантной RNH оценивалась её способность разрезать РНК/ДНК-гибрид *in vitro*.

## РЕЗУЛЬТАТЫ

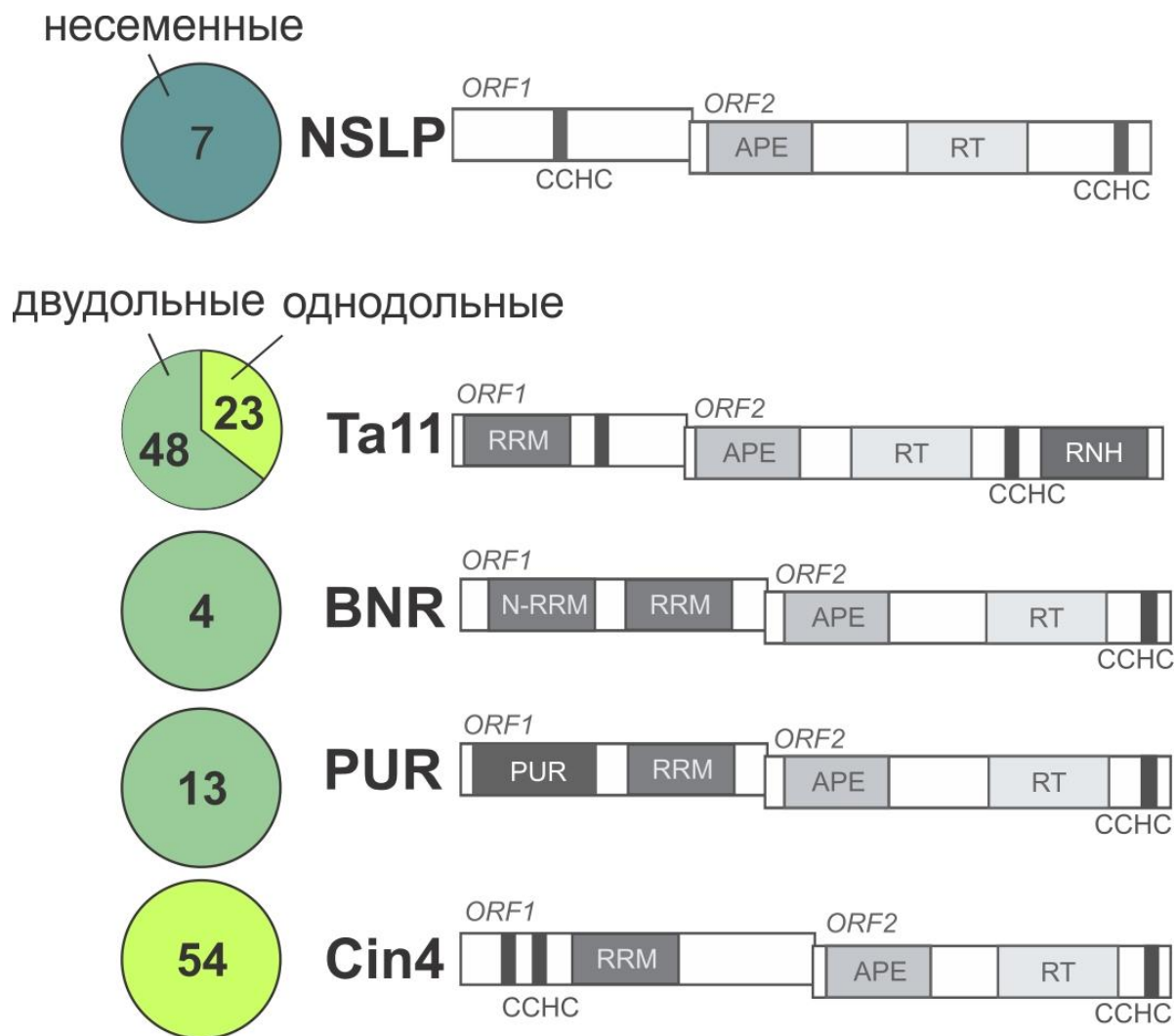
Для выявления разнообразия *L1*-ретротранспозонов растений был проведён биоинформатический анализ 29 геномов наземных растений.

В рамках этого анализа был проведен поиск регионов обратной транскриптазы (RT) *L1*-элементов либо в целых геномных последовательностях, либо в базах данных RepBase и GenBank с помощью НММ-профиля. Все найденные в целых геномных последовательностях RT-регионы *L1*-элементов были использованы для филогенетического анализа и, на основе результатов такого анализа, были отнесены к копиям конкретного элемента. Всего было идентифицировано 149 различных элементов.

Для выявления основных эволюционных групп *L1*-ретротранспозонов растений на основе найденных RT-последовательностей проводился филогенетический анализ с использованием всех найденных элементов (Рис. 1).



**Рис. 1.** Схема филогенетического дерева non-LTR-ретротранспозонов, реконструированного с помощью метода максимального правдоподобия на основе аминокислотных последовательностей доменов обратной транскриптазы. Суперсемейства non-LTR-ретротранспозонов (*CRE*, *R2*, *R4*, *L1*, *RTE*, *Jockey*, *I*, *Tad1*, *R1*, *LOA*) обозначены справа от дерева. Внутри суперсемейства *L1* выделены филогенетические группы, относящиеся к определенным таксонам (растения, животные и т.д.). Кроме того, для компактности, прямоугольниками обозначены семейства *L1*-non-LTR-ретротранспозонов растений (*NSLP*, *Ta11*, *BNR*, *PUR* и *Cin4*). В качестве корня использовались последовательности *CRE*-элементов. Слева от некоторых узлов дерева указана их статистическая достоверность согласно aLRT (для узла соответствующему суперсемейству *L1* значение достоверности выделено жирным шрифтом).



**Рис. 2.** Структура и распространение элементов из пяти семейств *L1*-non-LTR-ретротранспозонов растений (Smyshlyaev *et al.*, 2013, с модификациями). Количество и распространение элементов идентифицированных в нашем исследовании показано в круговых графиках. Цифры отражают количество элементов, относящихся к семейству в выделенной группе растений (темно-зеленый – несеменные, зеленый – двудольные, светло-зеленый – однодольные растения). ORF – открытая рамка считывания, RRM – домен узнавания РНК, N-RRM – N-концевой RRM ретротранспозонов из группы *BNR*, APE – апуриновая/апиримидиновая эндонуклеаза, RT – обратная транскриптаза, RNH – рибонуклеаза H, PUR – пуриин-богатый домен.

В результате данного исследования было выявлено пять основных филогенетических групп *L1*-элементов растений: *NSLP*, *Ta11*, *BNR*, *PUR* и *Cin4* (Рис. 1). Для каждой группы была установлена общая структура входящих в её состав элементов (Рис. 2).

Как и большинства других *L1*-элементов, оказалось, что растительные *L1*-non-LTR-ретротранспозоны всех пяти групп содержат APE-, RT- и CCHC-домены в своих ORF2 (Рис. 2). Относительно RNH-домена было установлено,



что многие элементы группы *Tall* содержат этот домен, в то время как у других групп растительных *L1*-элементов такого домена найдено не было.

Далее было показано, что элементы каждой из пяти идентифицированных групп обладают уникальной структурой ORF1 и содержат в своем составе различные РНК-связывающие домены, такие как N-RRM, RRM, ССНС и, обнаруженный впервые в ретротранспозонах, PUR (Рис. 2).

### ***NSLP*-элементы специфичны для несеменных растений**

Эти элементы, как следует из названия (*NSLP* – NonSeed Plant-Specific), были обнаружены в геномах несеменных растений: мха *Physomitrella patens* и плауна *Selaginella moellendorffii*. Всего в этих геномах было обнаружено семь элементов, которые на основе филогенетического анализа легли в эту монофилетическую группу (Рис. 1).

### ***BNR*-элементы специфичны для двудольных растений и содержат уникальный RRM-домен**

Эти элементы не были найдены ни в одном из исследованных геномов злаков. Поэтому можно говорить о том, что данная группа специфична для двудольных растений. Однако ввиду отсутствия информации относительно геномов незлаковых однодольных в свободном доступе, очевидно, что для проверки этой гипотезы требуется дополнительная информация.

Ранее была проведена работа, в которой были описаны основные характеристики представителей *BNR*-элементов (Heitkam *et al.*, 2009). Особенности элементов этой группы является присутствие высоко консервативного N-концевого RRM, центрального домена и отсутствие цистеинового мотива ССНС в ORF1. RRM является одним из самых распространенных РНК-связывающих мотивов в non-LTR-ретротранспозонах и характеризуется консервативной вторичной структурой. Сравнительный анализ выявил, что центральный домен структурно сходен с N-RRM и другими описанными RRM. В пользу этого предположения говорит наличие многочисленных RRM в ORF1 некоторых non-LTR-ретротранспозонов (Kapelinskaia *et al.*, 2011). ORF2 у *BNR*-элементов стандартная для *L1* и содержит домены APE и RT, а также цистеиновый мотив ССНС в С-конце (Рис. 2).

### ***PUR*-элементы специфичны для двудольных растений и содержат PUR-домен**

В рамках проведенного исследования нам удалось выявить новую группу non-LTR-ретротранспозонов растений, специфичную для геномов двудольных растений. Эта группа получила название *PUR* (Рис. 1, 2). Наличие *PUR*-элементов в одном из самых примитивных двудольных *Aquilegia coerulea* из порядка Лютикоцветных говорит в пользу происхождения этой группы до

появления так называемых основных эудикотов (англ. core eudicots), то есть более 110 млн лет назад (Chaw *et al.*, 2004). Тринадцать элементов, объединяющихся в монофилетическую группу *PUR*-элементов на основе эволюции RT-домена, филогенетически близки к группе *BNR*-элементов (Рис. 1), но отличаются от них структурно (Рис. 2). Как и *BNR*-элементы, они содержат два домена, которые предположительно функционально отвечают за связывание с РНК, но вместо N-RRM-домена, как у *BNR*-элементов, в ORF1 *PUR*-элементов был обнаружен консервативный домен, структурно похожий на пурин-богатый (*PUR* - purine-rich) домен.

*PUR*-домены являются высоко консервативными ДНК- и РНК-связывающими доменами и входят в состав многих регуляторных клеточных белков, которые играют значительную роль в клетке, в частности обеспечивают функционирование механизмов репарации (White *et al.*, 2009). Вторым РНК-связывающим доменом является RRM, имеющий сходство на уровне первичной аминокислотной последовательности с центральным RRM группы *BNR*-элементов, что говорит в пользу общего происхождения этих RRM в *L1*-элементах растений.

#### ***Cin4*-элементы специфичны для геномов однодольных растений**

Элементы из этой группы были обнаружены в геномах всех злаковых растений, исследованных в данной работе. При анализе структуры ORF1 элементов из группы *Cin4* был выявлен новый тип ORF1 *L1*-элементов растений (Рис. 2). ORF1 элементов *Cin4* имеет RRM сходный с RRM *BNR*- и *PUR*-элементов, но в отличие от них элементы *Cin4* в добавок к этому RRM несут два цистеиновых ССНС мотива в первой рамке, которые располагаются в N-конце относительно RRM.

#### ***Tall*-элементы присутствуют в геномах как однодольных, так и двудольных растений и содержат домен RNH**

Элементы группы *Tall* оказались наиболее распространенными у растений по сравнению с другими найденными группами. Они были обнаружены в большинстве из исследованных геномов. Таким образом, *Tall*-элементы оказались единственной группой, представленной в геномах как однодольных, так и двудольных растений. Такое широкое распространение этой группы говорит о её древнем происхождении и возникновении ещё до разделения однодольных и двудольных растений, то есть более 150 млн. лет назад (Chaw *et al.*, 2004).

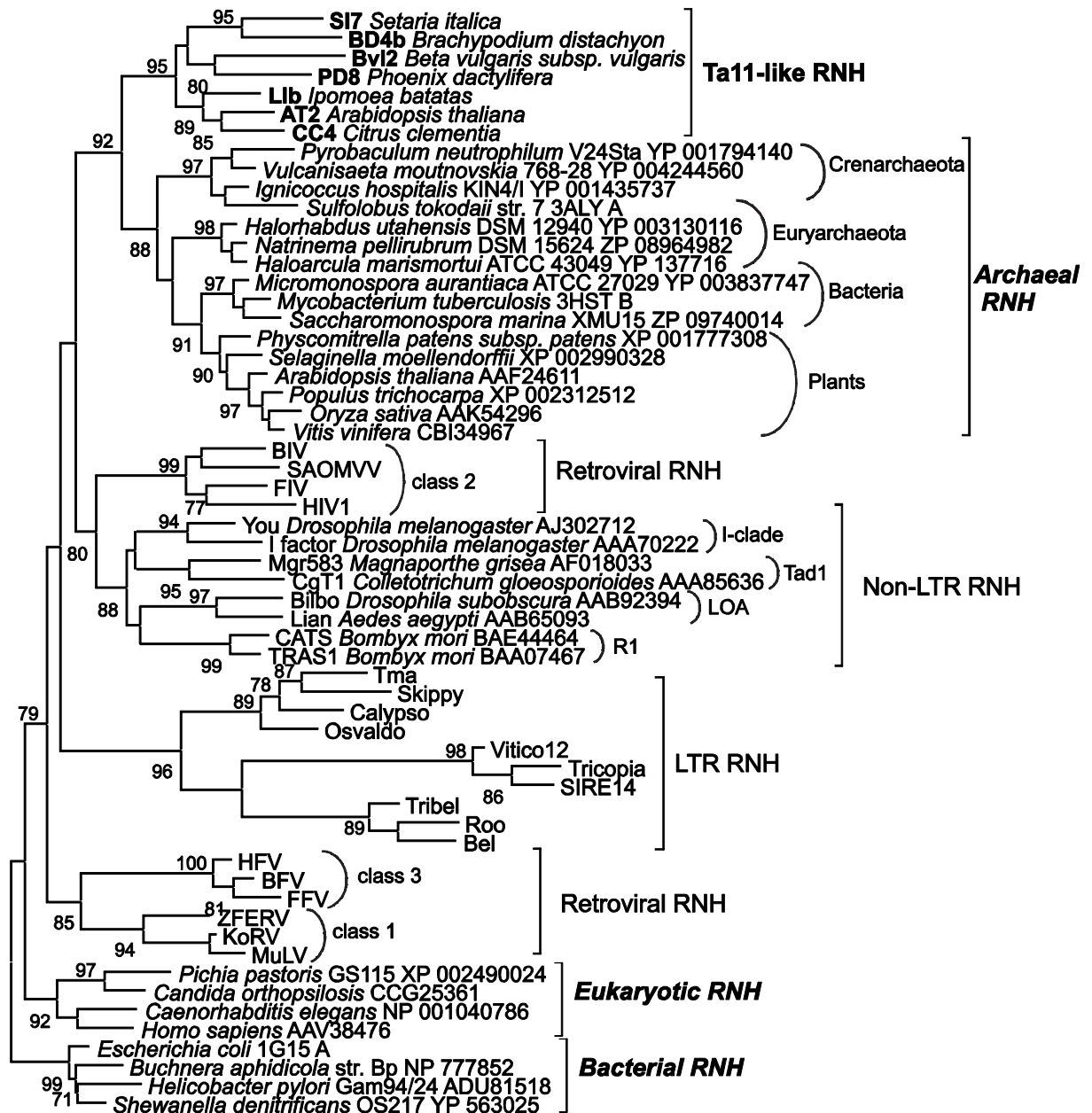
ORF1 *Tall*-элементов содержит RRM, за которым следует цистеиновый мотив ССНС (Рис. 2). Сравнительный анализ аминокислотных последовательностей RRM всех обнаруженных *L1*-ретротранспозонов растений выявил высокий уровень консервативности RRM у *BNR*-, *PUR*- и *Tall*-

элементов, что говорит об общем происхождении RRM всех *L1*-non-LTR-ретротранспозонов из семенных растений, за исключением N-RRM *BNR*-элементов, которые, по-видимому, имеют иное филогенетическое происхождение.

ORF2, кроме стандартных для *L1*-элементов APE- и RT-доменов и ССНС-мотива, кодирует ещё и домен RNН (Рис. 2). Появление RNН в non-LTR-ретротранспозонах считается более поздним событием: впервые белок RNН у non-LTR-ретротранспозонов появляется в филогенетически более молодом семействе *Jockey* (Malik, 2005). Поэтому, было предположено, что RNН была приобретена в группе *Tall*-элементов растений независимо. В пользу этого предположения говорит и тот факт, что у остальных non-LTR-элементов, обладающих RNН, ССНС мотив располагается после RNН, в то время как у *Tall*-элементов в С-конце располагается именно RNН (Рис. 2).

### ***Домен рибонуклеазы Н Tall-ретротранспозонов растений***

Гомологи рибонуклеазы Н (RNН) обнаруживаются практически во всех организмах, начиная от вирусов и заканчивая высшими эукариотами (Malik *et al.*, 1999; Cerritelli, Crouch, 2009), и делятся на RT-ассоциированные и неассоциированные (клеточные) RNН. Фермент RNН осуществляет специфический гидролиз РНК-цепи в РНК/ДНК-гибриде, и роль клеточной RNН в клетке заключается в участии в репликации генома (Cerritelli, Crouch, 2009). На основании аминокислотных последовательностей и наличия определенных функциональных сайтов все клеточные RNН разделяют на бактериальные, эукариотические и архейные (Marchler-Bauer *et al.*, 2011). В зависимости от происхождения, RT-ассоциированные RNН подразделяются на RNН LTR-ретротранспозонов, non-LTR-ретротранспозонов и ретровирусов. Считается, что non-LTR-ретротранспозонам необходима ферментативная активность RNН в процессе перемещения, однако наличие собственного домена не является обязательным условием, поскольку процесс обратной транскрипции у non-LTR-ретротранспозонов происходит в ядре, где эти элементы могут использовать RNН, кодируемую хозяйской клеткой. Тем не менее, многие non-LTR-ретротранспозоны кодируют собственный домен RNН. По одной из гипотез считается, что домен RNН был приобретен один раз общим предком «молодых» групп APE-содержащих non-LTR-ретротранспозонов. Эти элементы появились значительно позже *L1*-элементов. В этом контексте обнаружение RNН в эволюционно древнем суперсемействе *L1*-элементов является неожиданной находкой.



**Рис. 3.** Филогенетическое дерево различных типов RNH, построенное методом максимального правдоподобия. На дереве присутствует информация о номерах последовательностей, использованных для анализа, в базе данных GenBank. Названия трех типов клеточных RNH (не ассоциированных с RT) выделены жирным курсивом. Согласно Smyshlyaev *et al.*, 2013.

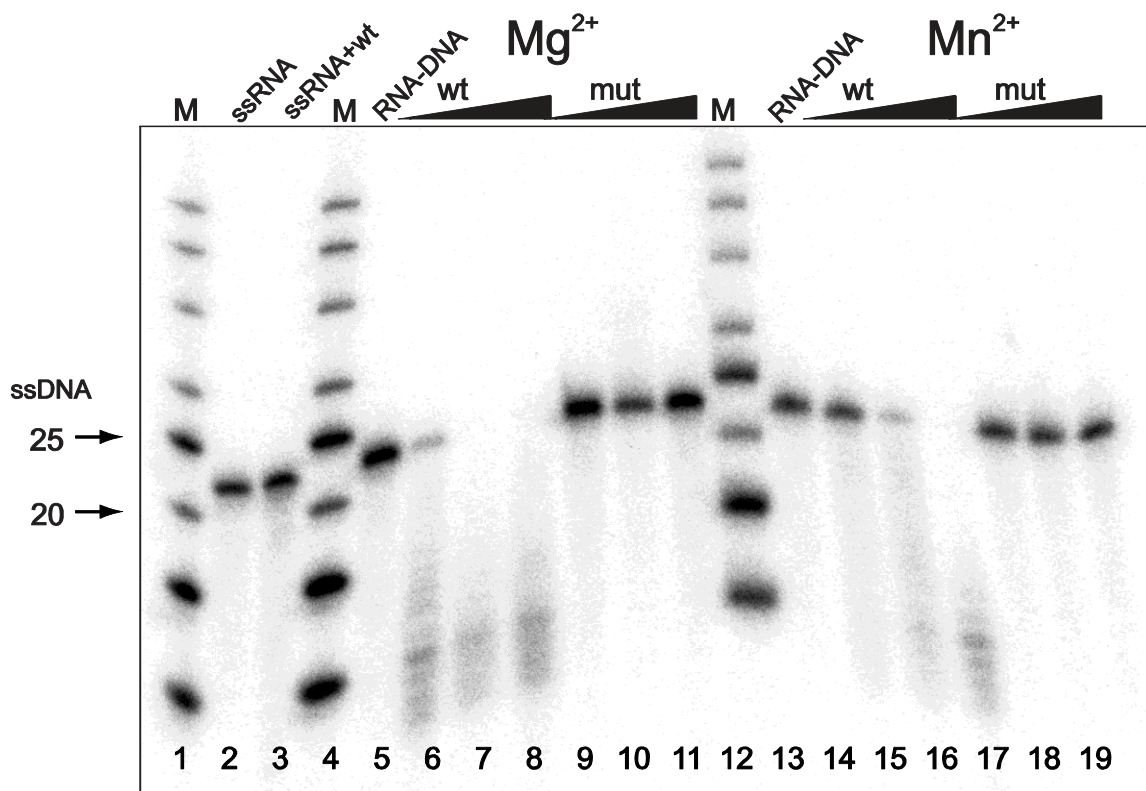
В связи с таким необычным положением домена RNH как в составе ORF2, так и в эволюционном контексте, для выяснения деталей его происхождения мы провели сравнительный филогенетический анализ домена RNH *Ta11*-ретротранспозонов. Для этого было произведено множественное выравнивание RNH из различных ретротранспозонов, ретровирусов, а также клеточных RNH, и реконструировано их филогенетическое дерево (Рис. 3).

Проведенный анализ продемонстрировал, что RNH *Tall*-ретротранспозонов и клеточная архейная RNH объединяются в монофилетическую группу. Кроме того, из проведенного анализа следует еще один вывод о том, что RNH *Tall*-ретротранспозонов не родственна RNH других non-LTR-ретротранспозонов из всех суперсемейств, а гораздо ближе к архее-подобной клеточной RNH из различных царств живых организмов: растений, бактерий и архей. Остается неясным только источник такой гомологии. Близкое филогенетическое положение клеточной RNH этих организмов и RNH *Tall*-ретротранспозонов содержит потенциальное объяснение такому распространению, а именно горизонтальный перенос (ГП) RNH между этими группами организмов посредством *Tall*-ретротранспозонов.

Ранее было высказано мнение о том, что гомологи архейной RNH в бактериях и растениях могли появиться в этих организмах путем ГП из архей посредством ретровирусов (Ohtani *et al.*, 2004). Non-LTR-ретротранспозоны, хоть и в меньшей степени, чем LTR-ретротранспозоны и ретровирусы, также способны перемещаться горизонтально и, поэтому, могли являться средством перемещения гена RNH между геномами архей, растений и бактерий. Для того чтобы косвенно проверить эту гипотезу мы решили протестировать биохимическую активность домена RNH *Tall*-ретротранспозонов *in vitro*. В археях, бактериях и растениях клеточная архее-подобная RNH не ассоциирована с RT, поэтому, только в случае сохранения RNH-активности этого домена как отдельного полипротеина вне ORF2, можно было бы говорить о возможности выполнения этим доменом функций характерных для клеточной RNH. Кроме того, проверка активности RNH *Tall*-ретротранспозонов могла бы подтвердить представление об эволюции растительных *L1*-non-LTR-ретротранспозонов как о *модульной* эволюции, поскольку в ней, под модулем понимается домен функционально независимый от других доменов и способный свободно выполнять свою функцию.

#### **Активность домена RNH *Tall*-ретротранспозонов *in vitro***

Для проверки биохимической активности изолированного RNH-домена *Tall*-ретротранспозонов мы провели *in vitro* анализ активности рекомбинантной RNH *Llb*-ретротранспозона (*Llb*RNH) из генома *Ipomoea batatas*. Этот ретротранспозон был выбран, поскольку для него была показана активность перемещения в геноме (Yamashita, Tahara, 2006), что скорее всего означает, что все структурные домены белков этого ретротранспозона функционально активны. Для проведения этого анализа рекомбинантный полипротеин *Llb*RNH и его мутантная версия *Llb*RNHD1326N были экспрессированы в *E. coli*.

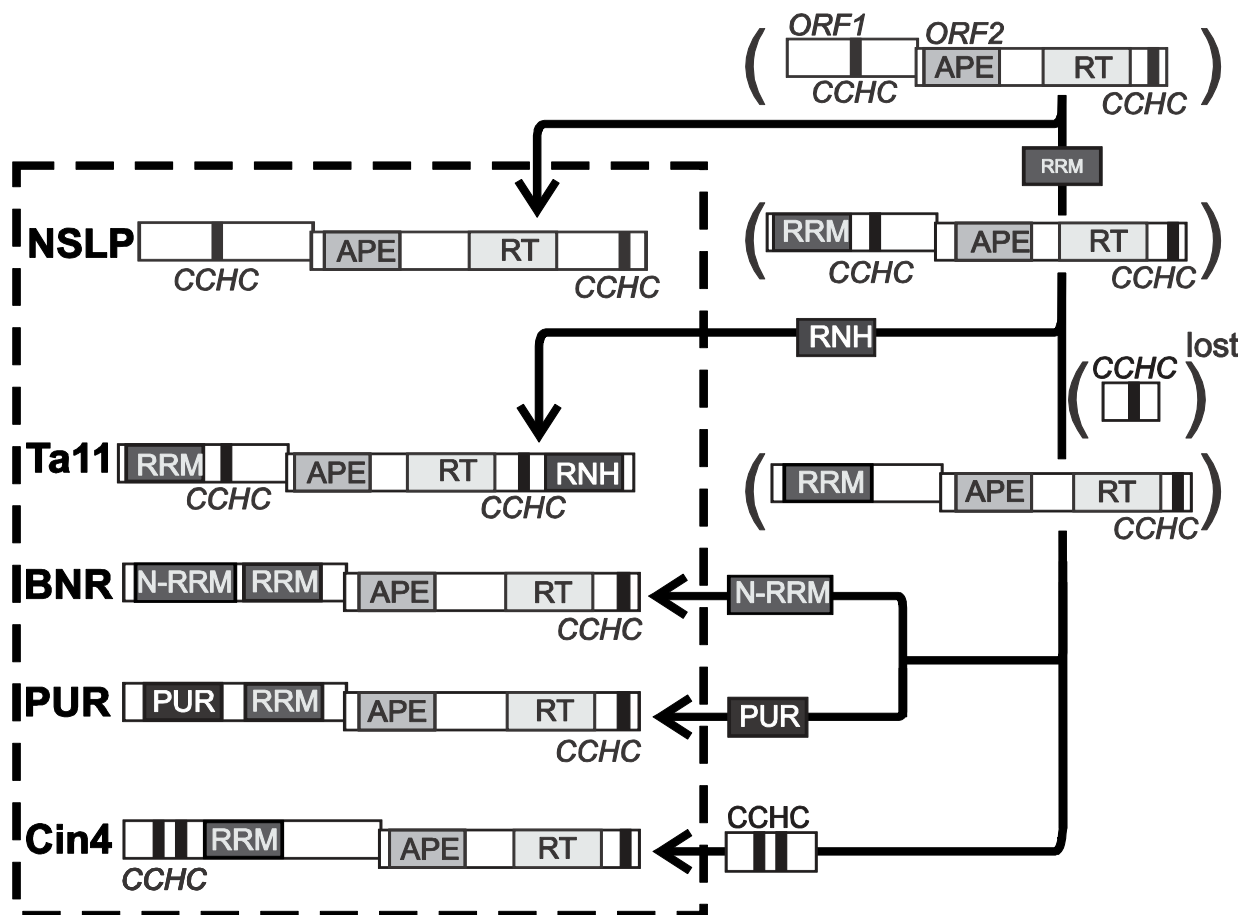


**Рис. 4.** Ферментативная активность домена RNH *Llb*-ретротранспозона; ‘wt’ обозначает LlbRNH дикого типа; ‘mut’ обозначает мутантный LlbRNHD1326N белок, содержащий замену в консервативной позиции RNH. Образцы, содержащие одноцепочечную РНК (ssRNA) и ssRNA вместе с LlbRNH в присутствии Mg<sup>2+</sup> показаны в рядах 2 и 3 соответственно. При добавлении LlbRNH к ssRNA субстрату расщепления не наблюдается. Ряды 5 и 13 представляют РНК/ДНК-гибрид без добавления фермента. Расщепление РНК/ДНК-гибрида при добавлении белка LlbRNH представлено в рядах 6-8 и 14-16 (в присутствии Mg<sup>2+</sup> или Mn<sup>2+</sup> соответственно). Отсутствие расщепления наблюдается при добавлении мутантного белка LlbRNHD1326N (ряды 9-11 и 17-19 в присутствии Mg<sup>2+</sup> или Mn<sup>2+</sup> соответственно). Согласно Smyshlyaev *et al.*, 2013.

Далее проверяли способность обоих белков разрезать РНК-цепь в гибриде 5’-P32-poly(rA)/poly(dT). При инкубации LlbRNH с избытком субстрата (100nM), РНК-цепь в РНК/ДНК-гибриде разрезалась в присутствии Mg<sup>2+</sup> или Mn<sup>2+</sup>, причем каталитический мутант LlbRNHD1326N не проявлял подобной активности (Рис. 4). Более того, обнаруженная активность оказалась специфической для РНК/ДНК-гибрида, инкубация LlbRNH с одноцепочечной РНК не приводила к гидролизу субстрата.

## Модульная эволюция *L1*-non-LTR ретротранспозонов растений

Теория модульной эволюции вирусов, впервые сформулированная Ботштейном, гласит, что результатом эволюции вирусов является не отдельный вирус, но семейство обмениваемых генетических элементов, каждый из которых обладает определенной биологической функцией (Botstein, 1980). Поэтому модульность, или “кластеризация эпистатических взаимодействий” является определяющим свойством вирусоподобных систем.



**Рис. 5.** Модульная эволюция *L1*-non-LTR-ретротранспозонов растений. Схемы структур современных растительных *L1*-элементов показаны в пунктирной рамке. События элиминации или приобретения функциональных доменов показаны снаружи рамки. Предполагаемые схемы структур элементов-предшественников заключены в скобки. ORF – открытая рамка считывания, RRM – домен узнавания РНК, N-RRM – N-концевой RRM ретротранспозонов из группы *BNR*, APE – апуриновая/апиримидиновая эндонуклеаза, RT – обратная транскриптаза, RNH – рибонуклеаза H, PUR – пурин-богатый домен. Согласно Smyshlyaev *et al.*, 2013.

Считается, что подобным свойством обладают и МЭ (Lerat *et al.*, 1999), а значит и их эволюция может характеризоваться как модульная эволюция. Недавние работы показали, что бактериальные МЭ могут захватывать отдельные функциональные модули (Roberts, Mullany, 2009), и кроме того известны многочисленные примеры приобретения новых модулей ретротранспозонами (Malik, Eickbush, 1999; Heitkam, Schmidt, 2009; Kramerov, Vassetzky, 2011). Обмен модулями между МЭ и между МЭ и геномом способен производить новые структурные варианты МЭ с потенциальными эволюционными преимуществами и впоследствии, при амплификации такого МЭ, может привести к появлению нового семейства МЭ (Lerat *et al.*, 1999; Roberts, Mullany, 2009).

В результате проведенной работы, было описано пять семейств *L1*-ретротранспозонов, каждое из которых характеризуется особой структурной организацией входящих в неё элементов. В рамках биоинформатического исследования структурного разнообразия было выявлено семь структурных доменов, составляющих многообразие форм *L1*-non-LTR-ретротранспозонов растений: два типа RRM, PUR-домен, ССНС-мотив, и домены APE, RT, и RNH. Причем для домена RNH удалось подтвердить его способность функционировать как отдельная единица, а значит как функциональный модуль в терминологии Ботштейна. Такое структурное разнообразие позволяет предположить, что, возможно, и эволюцию растительных *L1*-non-LTR-ретротранспозонов можно рассматривать как модульную, в связи с чем далее была предпринята попытка выявить основные события в модульной эволюции этих элементов (Рис. 5).

За основу эволюции *L1*-non-LTR-ретротранспозонов растений была взята реконструированная с помощью домена RT филогения этих элементов (Рис. 1). В соответствии с этой филогенией, наиболее древней из выявленных групп *L1*-элементов растений, является группа *NSLP*, специфичная для мхов и плаунов. Как было сказано ранее, в ORF1 этих элементов не было идентифицировано никаких РНК-связывающих доменов за исключением ССНС-мотива (Рис. 2). *L1*-non-LTR-ретротранспозоны семенных растений, напротив, содержат, по крайней мере, один RRM. По этой причине мы предположили, что *NSLP*-элементы могли сохранить структуру предшественника *L1*-ретротранспозонов семенных растений, которые затем, в процессе эволюции, приобрели RRM. Поиск гомологов с помощью HHPRED-анализа выявил, что RRM *L1*-ретротранспозонов растений, возможно, является структурным гомологом нуклеолина и фактора сплайсинга UNAF(65), которые присутствуют, в том числе, и в геномах растений (Domon *et al.*, 1998; Tajrishi *et al.*, 2011), что позволяет предположить, что именно эти растительные гены были источником RRM в *L1*-элементах растений. Приобретение этого модуля является первым



шагом в реконструированной нами схеме модульной эволюции *L1*-non-LTR-ретротранспозонов растений (Рис. 5).

Следующим эволюционным шагом к появлению современных форм *L1*-ретротранспозонов, возможно, стала потеря у некоторых из них С-концевого ССНС-мотива в ORF1 и его замена на другой тип РНК-связывающего домена. Те же элементы, которые сохранили исходный ССНС-мотив, присутствующий еще у *NSLP*-элементов, дали начало *Tall*-элементам. Тот факт, что только *Tall*-элементы представлены в геномах как двудольных, так и однодольных растений, находится в соответствии с такой реконструкцией, которая предполагает, что элементы с соответствующей структурой являются наиболее древними из четырех оставшихся форм (Рис. 5). Впоследствии эти элементы приобрели домен RNH. Следующим событием в модульной эволюции в рамках нашей модели стала замена потерянного ССНС-мотива новыми РНК-связывающими доменами. Приобретение RRM-, PUR-домена или двойного N-концевого ССНС-мотива произошло соответственно в наиболее ранних общих предках *BNR*-, *PUR*- и *Cin4*-элементов (Рис. 5). Наличие RRM домена во всех *L1*-non-LTR-ретротранспозонах семенных растений говорит о важности его функций в этих элементах, тем не менее, обязательное наличие дополнительных структур в ORF1, предположительно также отвечающих за связывание с РНК (таких как ССНС-мотивы, PUR-домен и N-RRM), говорит о том, что самого по себе RRM не достаточно для выполнения функций ORF1.

### **Заключение**

Несмотря на то, что за последние десятилетия в геномах растений было обнаружено достаточно большое количество *L1*-non-LTR-ретротранспозонов, к настоящему времени не было проведено детального анализа их распространения, разнообразия и эволюции для широкого круга растений. В данной работе мы провели идентификацию и эволюционный анализ *L1*-non-LTR-ретротранспозонов из 29 геномов растений из различных таксонов. Мы обнаружили многочисленные новые *L1*-элементы и выявили многообразие их строения. Нам удалось показать, что на основании филогенетических взаимоотношений и структурных особенностей, все найденные *L1*-ретротранспозоны могут быть объединены в пять основных групп. Кроме того была предпринята попытка проследить за тем, какие события в эволюции non-LTR-ретротранспозонов растений могли привести к появлению структурных вариантов современных элементов. В целом, в данной работе мы продемонстрировали, что эволюция растительных *L1*-ретротранспозонов может быть описана в виде серии событий приобретения и замены различных функциональных доменов и может быть охарактеризована как модульная эволюция. Приобретение и потеря РНК-связывающих доменов, а также приобретение функционально активной археоподобной RNH – все это привело к

появлению новых филогенетических групп *L1*-non-LTR-ретротранспозонов растений. Кроме того, мы показали, что приобретенный домен RNH автономно активен *in vitro* и предположили, что *L1*-non-LTR-ретротранспозоны могли сыграть роль в ГП и распространении RNH между растениями, археями и бактериями.

## ВЫВОДЫ

1. В результате проведенного биоинформатического анализа нуклеотидных последовательностей, относящихся к геномам 29 видов растений, было идентифицировано 149 различных *L1*-non-LTR-ретротранспозонов.

2. Сравнительный анализ идентифицированных элементов показал, что *L1*-non-LTR-ретротранспозоны растений образуют пять филогенетических групп: *NSLP*, *Ta11*, *Cin4*, *BNR* и *PUR*. Каждая из этих групп характеризуется уникальной структурой входящих в неё элементов, выраженной наличием специфичных функциональных доменов: N-RRM, RRM, PUR, CCHC и RNH. *Ta11*-элементы содержат RRM и CCHC-мотив, *Cin4*-элементы – RRM и два CCHC-мотива, *BNR*-элементы – N-RRM и RRM, *PUR*-элементы – PUR-домен и RRM и, наконец, *NSLP*-элементы не содержат ни один из перечисленных функциональных доменов.

3. Проведенный филогенетический анализ показал, что домен RNH, приобретенный в процессе эволюции *L1*-non-LTR-ретротранспозонами растений из семейства *Ta11*, филогенетически ближе к архейной клеточной RNH, чем к RNH других non-LTR-ретротранспозонов.

4. Было установлено, что домен RNH *Llb*-ретротранспозона функционально активен *in vitro*, и сделано предположение о том, что *L1*-non-LTR-ретротранспозоны могут являться векторами переноса RNH между растениями, археями и бактериями.

5. Сравнительный анализ структурной организации идентифицированных элементов подтвердил модульный характер эволюции *L1*-non-LTR-ретротранспозонов растений, которая сопровождалась приобретением и заменой отдельных функциональных доменов.

## СПИСОК ОСНОВНЫХ РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Smyshlyaeв G.**, Voigt F., Blinov A., Barabas O., Novikova O. Acquisition of an Archaea-like ribonuclease H domain by plant *L1* retrotransposons supports modular evolution // Proc. Natl Acad. Sci. USA. - 2013. – V. 110. - No. 50. – P. 20140-20145.
2. Sormacheva I., **Smyshlyaeв G.**, Mayorov V., Blinov A, Novikov A, Novikova O. Vertical evolution and horizontal transfer of *CRI* non-LTR retrotransposons and *Tc1/mariner* DNA transposons in Lepidoptera species // Molecular biology and evolution. - 2012. - V. 29. – No. 12. - P. 3685–3702.
3. Novikov A., **Smyshlyaeв G.**, Novikova O. Evolutionary history of LTR retrotransposon chromodomains in plants // Int. J. Plant Genomics. - 2012. – V. 874743.
4. **Смышляев Г. А.**, Блинов А. Г. Эволюция и разнообразие *L1*-ретротранспозонов в геномах покрытосеменных растений // Вавиловский журнал генетики и селекции. – 2011. – Т. 15. - № 3. – С. 72-78.
5. Novikova O., **Smyshlyaeв G.**, Blinov A. Evolutionary genomics revealed interkingdom distribution of *Tcn1*-like chromodomain-containing *Gypsy* LTR retrotransposons among fungi and plants // BMC Genomics. – 2010. – V. 11:231.
6. Novikova O., Mayorov V., **Smyshlyaeв G.**, Fursov M., Adkison L., Pisarenko O., Blinov A.. Novel clades of chromodomain-containing *Gypsy* LTR retrotransposons from mosses (Bryophyta) // Plant J. – 2008 – V. 56 – No. 4. – P. 562-574.