

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ НАУКИ ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ
СИБИРСКОГО ОТДЕЛЕНИЯ РАН**

На правах рукописи

СМЫШЛЯЕВ ГЕОРГИЙ АНДРЕЕВИЧ

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

03.02.07 – генетика

Диссертация на соискание ученой степени

кандидата биологических наук

Научный руководитель:

доктор биологических наук,

доцент Бугров Александр Геннадьевич

Новосибирск 2014

Оглавление:

Введение	5
Глава 1. Обзор литературы	10
1.1. Мобильные элементы	10
1.2. Non-LTR-ретротранспозоны	13
1.2.1. Структура non-LTR-ретротранспозонов	16
1.2.2. Механизм перемещения non-LTR-ретротранспозонов	19
1.2.3. Взаимодействие non-LTR-ретротранспозонов и организма хозяина	26
1.2.4. Регуляция с использованием малых РНК	28
1.2.5. Внутриклеточная защита с использованием белков семейства APOBEC ..	29
1.2.6. Особенности регуляции МЭ в геномах растений	30
1.2.7. Эволюция и классификация non-LTR-ретротранспозонов на основе домена обратной транскриптазы	31
1.2.8. Эволюция non-LTR-ретротранспозонов на основе рибонуклеазы Н	37
1.2.9. Модульная эволюция non-LTR-ретротранспозонов	39
1.2.10. Горизонтальный перенос non-LTR-ретротранспозонов	40
1.2.11. Эволюция и разнообразие non-LTR-ретротранспозонов растений	43
Глава 2. Материалы и методы	46
2.1. Биоинформатический анализ распространения и эволюции <i>L1</i> -non-LTR- ретротранспозонов в геномах растений	46
2.1.1. Поиск и филогенетический анализ <i>L1</i> -non-LTR-ретротранспозонов в геномах растений	46
2.1.2. Идентификация структурных характеристик <i>L1</i> -non-LTR- ретротранспозонов растений	49
2.1.3. Сравнительный анализ домена рибонуклеазы Н <i>L1</i> -non-LTR- ретротранспозонов	49
2.2. Экспериментальные исследования	50
2.2.1. Конструирование плазмид и очистка белка	50
2.2.2. Исследование активности рекомбинантной рибонуклеазы Н	51
Глава 3. Результаты	53

3.1. <i>NSLP</i> -элементы специфичны для несеменных растений.....	61
3.2. <i>BNR</i> -элементы специфичны для двудольных растений и содержат уникальный RRM-домен	61
3.3. <i>PUR</i> -элементы специфичны для двудольных растений и содержат PUR- домен	63
3.4. <i>Cin4</i> -элементы специфичны для геномов однодольных растений	66
3.5. <i>Tall</i> -элементы присутствуют в геномах как однодольных, так и двудольных растений и содержат домен рибонуклеазы H	68
3.6. Домен рибонуклеазы H <i>Tall</i> -ретротранспозонов растений	70
3.7. Активность домена рибонуклеазы H <i>Tall</i> -ретротранспозонов <i>in vitro</i>	76
Глава 4. Обсуждение. Модульная эволюция <i>L1</i> -non-LTR-ретротранспозонов растений.....	80
Заключение.....	83
Список используемой литературы.....	85

Список сокращений

МЭ – мобильные элементы

ДНК – дезоксирибонуклеиновая кислота

РНК – рибонуклеиновая кислота

кДНК – комплементарная ДНК

п.н. – пар нуклеотидов

LTR – длинный концевой повтор (Long Terminal Repeat)

RT – обратная транскриптаза (Reverse Transcriptase)

ORF – открытая рамка считывания (Open Reading Frame)

REL-endo – эндонуклеаза рестрикционного типа (Restriction Enzyme-Like endonuclease)

APE – апуриновая/апириmidiновая эндонуклеаза (APurinic/apyrimidinic Endonuclease)

ORFNp – белок кодируемый рамкой считывания N (Open Reading Frame N encoded protein)

RNH – рибонуклеаза H (RiboNuclease H)

RNP – рибонуклеопротеиновая частица (RiboNucleoprotein Particle)

RRM – домен узнавания РНК (RNA Recognition Motif)

UTR – нетранслируемый регион (UnTranslated Region)

мРНК – матричная РНК

polyA – полидезоксиденозиновая последовательность

TPRT – РНК-опосредованной интеграцией (Target Primed Reverse Transcription)

ГП – горизонтальный перенос

N-RRM – N-концевой RRM ретротранспозонов из группы BNR

PUR – пурин-богатый (PUrine Rich) домен

Введение

Актуальность

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 и 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 дубликации 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, ведь даже в свете наличия жестких механизмов регуляции пролиферации МЭ, многие геномы растений наполнены ретротранспозонами. Геном кукурузы, например, более чем на 60% состоит из ретротранспозонов. 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 эволюции и распространения ретротранспозонов в этих геномах.

Одними из наименее изученных в этом смысле 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 вносят новые данные в пользу представления об эволюции ретротранспозонов как о модульной эволюции.

3. Выявленное в данной работе распространение гомологов “архейной” RNH 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 RNH.

4. Разработан протокол синтеза и очистки домена RNH 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 RNH *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), «XLVI международная научная студенческая конференция» (Новосибирск, 2009), «XLVII международная научная студенческая конференция» (Новосибирск, 2010).

По теме диссертации опубликовано шесть работ.

1. **Smyshlyayev 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. U. S. A. - 2013. – V. 110. - No. 50. – P. 20140-20145.
2. Sormacheva I., **Smyshlyayev G.**, Mayorov V., Blinov A, Novikov A, Novikova O. Vertical evolution and horizontal transfer of *CR1* 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., **Smyshlyayev 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., **Smyshlyayev 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., **Smyshlyayev 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.

Структура и объем работы

Диссертация состоит из введения, 3-х глав, заключения и списка литературы. Работа изложена на 104 страницах, содержит 21 рисунок и 4 таблицы.

Глава 1. Обзор литературы

1.1. Мобильные элементы

Мобильные элементы (МЭ) были впервые описаны в работах Барбары Макклиток в начале 1950-х годов как генетические факторы, способные изменять свое положение на хромосоме номер 9 кукурузы и ассоциированные с вариативностью фенотипа её соматических клеток (McClintock, 1950). Позднее МЭ были найдены в большинстве из исследованных эукариотических геномов, и стало известно, что во многих видах МЭ занимают более половины состава этих геномов (Табл. 1).

Таблица 1. Размеры геномов некоторых организмов и содержание в них МЭ (согласно Viémont, Vieira, 2006). 1 пг. (пикограмм) = $978 \cdot 10^6$ п.н.

Вид	Размер генома, пг.	Содержание МЭ, %
<i>Rana esculenta</i>	5.6–8.0	77
<i>Zea mays</i>	5.0	60
<i>Homo sapiens</i>	3.5	45
<i>Mus musculus</i>	3.4	40
<i>Drosophila melanogaster</i>	0.18	15–22
<i>Caenorhabditis elegans</i>	0.1	12
<i>Saccharomyces cerevisiae</i>	0.012	3–5
<i>Escherichia coli</i>	0.0046	0.3

Тем не менее, в целом определение МЭ со времени открытия Макклиток практически не изменилось, и наиболее распространенным в литературе является классическое определение МЭ как последовательностей ДНК способных менять свою локализацию в геноме. Следует сразу отметить ряд ограничений данного определения. Во-первых, как это ни парадоксально, но не все те последовательности ДНК, которые часто называют МЭ, оказываются способными менять свою локализацию в геноме. Этот тезис основывается на том наблюдении, что, с одной стороны, во многих видах МЭ не перемещаются в геноме из поколения в поколение уже на протяжении миллионов лет. С другой стороны,

существует ряд исследований, в которых показана возможность реактивации перемещения таких МЭ под воздействием различных факторов (Caru *et al.*, 2000). Тем не менее, только в небольшом геноме *Arabidopsis thaliana* насчитывается несколько тысяч неподвижных МЭ (de la Chaux *et al.*, 2012), реактивировать экспериментально каждый из которых, а значит подтвердить их способность менять локализацию в геноме не представляется возможным. В связи с этим, во многих работах вместо термина МЭ используется термин “копия МЭ”. Данный термин позволяет определять любую последовательность ДНК генома, имеющую гомологию к МЭ с доказанной способностью перемещаться, как *копию* этого элемента. Во-вторых, многие МЭ не меняют свою локализацию в геноме, а создают новые копии, которые затем встраиваются в другое место в геноме. Благодаря этому свойству может происходить значительное увеличение количества копий МЭ в геноме. Процесс копирования и встройки МЭ, тем не менее, называют “перемещением” или “транспозицией” МЭ, хотя сам копируемый элемент при этом сохраняется в исходном сайте и свою локализацию в геноме не меняет. В-третьих, не все те последовательности генома, обладающие способностью менять локализацию в геноме являются МЭ. В последние годы развиваются представления о так называемом «мобиломе» - совокупности всех агентов, способных вызывать перемещения генетического материала как внутри одного, так и между различными геномами. Кроме МЭ к этим агентам относят плазмиды, вирусы и самосплайсирующиеся элементы (интроны группы II и др.) (Siefert, 2009). Кроме того, негомологичная рекомбинация может приводить к геномным перестройкам, то есть к изменению локализации определенных последовательностей ДНК в геноме. Учитывая все эти ограничения, классическое определение МЭ, тем не менее, позволяет подчеркнуть их главное свойство: подвижность в геноме.

Представление о МЭ подвергалось значительным изменениям со времени их открытия. Спустя несколько десятилетий после работ Барбары Макклинток, МЭ стали называть “паразитической” ДНК из-за их автономности, способности разрушать “полезные” гены генома, и в отсутствии какой-либо очевидной функции в геноме (Fedoroff, 2012). В нескольких статьях, опубликованных в 1980 г. в журнале *Nature*, развивалась идея о том, что большая часть ДНК эукариот является “мусорной” (Doolittle, Sapienza, 1980; Orgel *et al.*, 1980). Распространение

эгоистичной ДНК в геноме подобно распространению не слишком вредного паразита внутри хозяина (Orgel *et al.*, 1980). Первоначально концепция “мусорной” ДНК была предложена для объяснения так называемого С-парадокса, состоящего в отсутствии корреляции между размерами геномов и сложностью организмов (Gall, 1981). Согласно этому объяснению, большая часть генома наполнена “мусорной” ДНК, образованной вследствие перемещения МЭ и не несущей никакой функции. С-парадокс и концепция “мусорной” ДНК также связаны с решением проблемы так называемого “мутационного груза”. Если представить, что весь человеческий геном состоит из ДНК, имеющей функциональное значение и необходимой для организма, то, учитывая наблюдаемый уровень мутаций в ДНК генома, накопление вредных мутаций на поколение было бы слишком высоко и должно было бы привести к появлению все большего количества летальных мутаций (Eddy, 2012). Наличие значительной части генома, не несущей никаких функций, с одной стороны, решило бы эту проблему. С другой стороны, данная концепция не объясняла, каким образом геномы смогли накопить такое большое количество МЭ с учетом наличия эффективных методов устранения МЭ из генома, таких как гомологичная рекомбинация.

Дальнейшее развитие технологий секвенирования геномов позволило сформировать детальное понимание структуры и организации генома. К примеру, оказалось, что две трети генома человека и 85% генома кукурузы состоят из МЭ (Schnable *et al.*, 2009; de Koning *et al.*, 2011). Идеи о “мусорной” ДНК менялись: накапливалось все больше данных о регуляторной роли, выполняемой МЭ и их транскриптами (Zuckerkindl, Cavalli, 2007); появлялось все больше примеров “доместикации” МЭ и выполнения белками, кодируемыми в этих МЭ, функций “полезных” всему организму (Alzohairy *et al.*, 2013); развивались представления о роли МЭ как о важных факторах эволюционной пластичности геномов (Bennett, 2004). Тем не менее, известно, что в геноме человека МЭ не находятся под воздействием хоть сколько-нибудь заметного давления отбора (Eddy, 2012), что, по крайней мере, ставит под сомнение идею о потенциальной функциональной роли белков, кодируемых МЭ, в целом организме.

Споры о “мусорной” природе МЭ продолжают (Doolittle, 2013). Одной из господствующих теорий является представление о МЭ как о безвредном балласте,

который организм при случае мог бы использовать для собственных целей. А именно, в рамках этой теории предполагается, что часть эукариотического генома может состоять из ДНК, созданной в результате транспозиций МЭ. У этой части генома могут быть определенные функции, необходимые для организма, как, например, в случае с особыми МЭ дрозофилы, которые участвуют в образовании теломер. Главным же свойством этой ДНК является эволюционный потенциал: в результате эволюции часть её могла бы послужить нуждам организма (Eddy, 2012). В принципе, конкретный вид может избавиться и от всей этой ДНК, лишив себя строительного материала на случай необходимости поменять что-либо в своем функционировании. Недавно опубликованная последовательность компактного генома растения пузырчатки горбатой (*Utricularia gibba*) показывает, что почти полное отсутствие последовательностей МЭ не мешает нормальному развитию и размножению сложного эукариотического организма (Ibarra-Laclette *et al.*, 2013). Несмотря на это, большинство геномов, информация о последовательностях которых доступна на данный момент, все-таки содержат большое количество МЭ, и это говорит в пользу того, что наличие МЭ каким-то образом действительно влияет на организм.

В заключение этой темы стоит привести следующую цитату, которая является актуальной вне зависимости от того, нужны ли на самом деле МЭ организму или нет и выполняют ли в нем какие-либо функции: “если для конкретного участка ДНК возможно показать, что он развил стратегию (такую как транспозиция), обеспечивающую его выживание в геноме, то вообще говоря, отсутствует необходимость в каком-то дополнительном объяснении его существования” (Doolittle, Sapienza, 1980).

1.2. Non-LTR-ретротранспозоны

Механизмы перемещения в геноме различных типов МЭ совершенно различны, и появились они, скорее всего, независимо друг от друга. То есть, филогенетически МЭ не являются монофилетичной группой последовательностей ДНК. На основании филогенетических взаимоотношений и механизмов перемещения выделяют два основных класса МЭ: ДНК-транспозоны и ретротранспозоны. ДНК-транспозоны перемещаются с использованием механизма

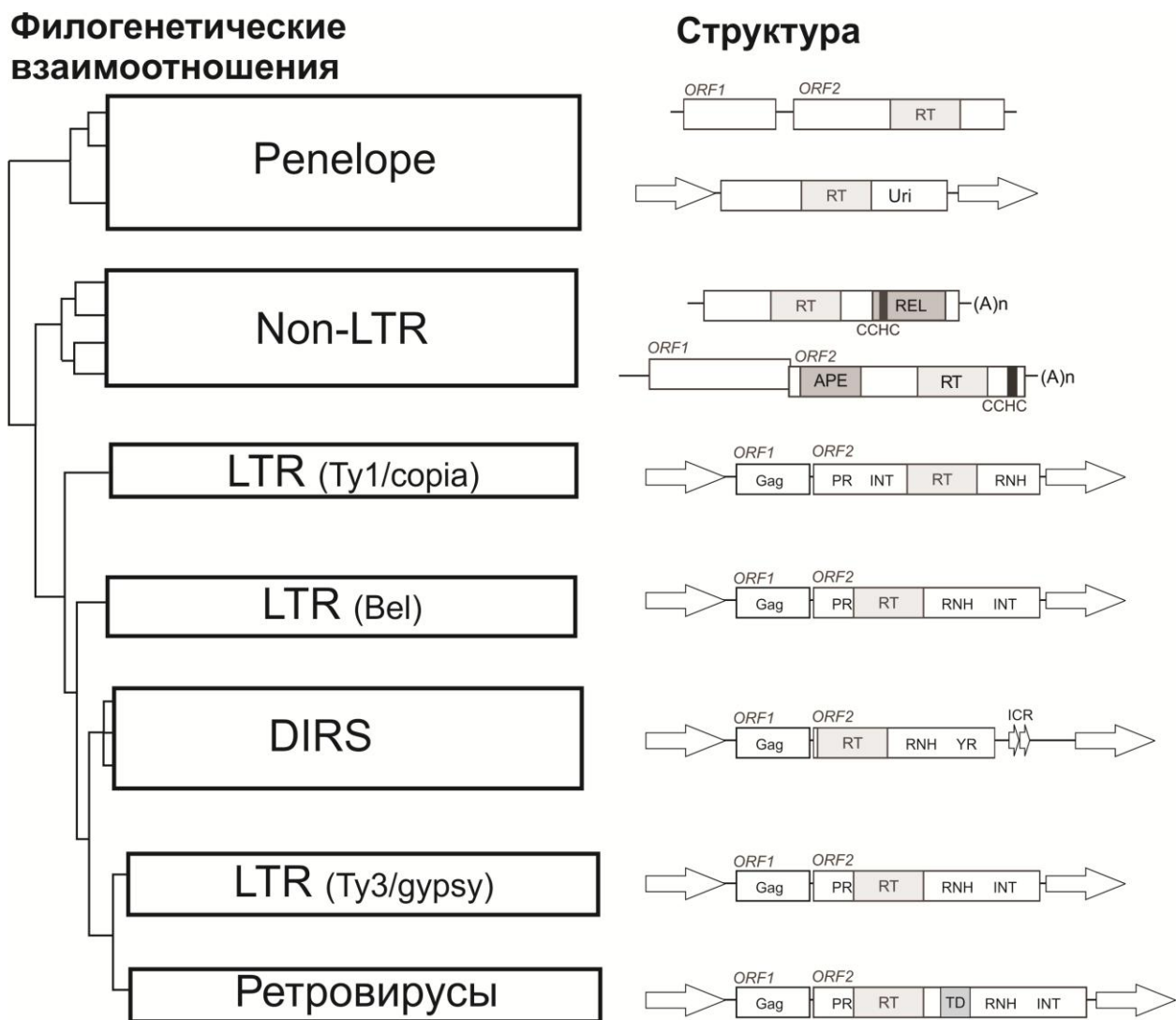


Рис. 1. Структура и филогенетические взаимоотношения различных групп ретротранспозонов. На левой части рисунка представлена схема филогенетического дерева ретротранспозонов, построенного на основе домена обратной транскриптазы (согласно Eickbush, Jamburuthugoda, 2008). Все элементы соответствующей группы обозначены прямоугольником, содержащим общепринятое название элемента. На правой части рисунка показаны схемы структур элементов из соответствующих групп. Открытые рамки считывания (ORF) изображены как горизонтальные прямоугольники. Длинные концевые повторы (LTRs) показаны в виде стрелок. Сокращения: TD - соединительный домен (tether domain); Uri - домен, гомологичный эндонуклеазе некоторых интронов группы I; YR - тирозиновая рекомбиназа; ICR - внутренние комплементарные повторы (internal complementary repeats); (A)_n - poly(A) хвост.

вырезания и встройки – “cut and paste” (Wicker *et al.*, 2007), в то время как ретротранспозоны для этого используют РНК-посредник и являются теми элементами, которые сохраняют свою исходную копию в геноме.

Для осуществления перемещения в ретротранспозоне обязательно должен быть закодирован фермент обратной транскриптазы (RT), который позволяет использовать РНК-посредник для синтеза кДНК и её встраивания в геном. RT — это РНК-зависимая ДНК-полимераза. Она является относительно древним и широко распространенным ферментом, о чем свидетельствует большое сходство её организации и аминокислотных последовательностей в геномах вирусов, прокариот и эукариот. Наличие же в этом ферменте доменов, гомологичных РНК-зависимой РНК-полимеразе вирусов, говорит о потенциальном общем происхождении этих двух полимераз (Xiong, Eickbush, 1990). Возможно, что время возникновения ретротранспозонов совпадает со временем перехода от РНК к ДНК, как основному носителю наследственной информации, то есть около 3.5 миллиардов лет назад. Более того, вполне вероятно, что на ранних стадиях развития для перехода от систем, основанных на РНК, к ДНК-системам, обратная транскриптаза играла ключевую роль (Heslop-Harrison, 2000).

На основе филогенетических взаимоотношений, деталей механизма перемещения и структурных характеристик, ретротранспозоны могут быть разделены на 4 больших группы (Рис. 1, Eickbush, Jamburuthugoda, 2008).

К первой группе относятся элементы несущие длинные концевые повторы (long terminal repeats — LTRs). LTR-ретротранспозоны не образуют монофилетичную группу, но имеют сходное строение (Рис. 1) и по своим структурным характеристикам и механизму перемещения сходны с ретровирусами (Wicker *et al.*, 2007).

Вторую группу образуют DIRS ретротранспозоны, которые не несут LTRs и для встраивания в геном используют тирозиновую рекомбиназу (Poulter, Goodwin, 2005).

В третью группу входят non-LTR-ретротранспозоны (или как их иначе называют LINE-элементы – long interspersed elements). Эти элементы не содержат LTRs и их механизм перемещения состоит в простом встраивании кДНК копии матричной РНК (мРНК) в геном с использованием кодируемой элементом

эндонуклеазы. Non-LTR-ретротранспозоны, судя по всему, являются наиболее древними ретротранспозонами и уже на протяжении сотен миллионов лет существуют в эукариотических геномах (Malik *et al.*, 1999).

Наконец, представители последней из описанных групп, Penelope, сходны по строению с non-LTR-ретротранспозонами, но занимают отдельное филогенетическое положение и характеризуются особым типом эндонуклеазы (Evgen'ev, Arkhipova, 2005).

1.2.1. Структура non-LTR-ретротранспозонов

Non-LTR-ретротранспозоны являются древнейшей группой ретроэлементов. RT, кодируемая non-LTR-ретротранспозонами, имеет высокое сходство с RT, обнаруженной у бактериальных интронов группы II и теломеразой (Xiong, Eickbush, 1990). Кроме того, non-LTR-ретротранспозоны были найдены в геномах древнейших эукариотических организмов (Bhattacharya *et al.*, 2002), что также свидетельствует в пользу их древнего происхождения. Считается, что возникновение non-LTR-ретротранспозонов по времени совпадает с возникновением эукариот (Malik *et al.*, 1999).

Полноразмерный non-LTR-ретротранспозон обычно имеет размер 3-8 тыс. пар нуклеотидов (п.н.) и содержит две открытые рамки считывания (ORF – Open Reading Frame) – ORF1 и ORF2, которые иногда могут быть объединены в одну ORF. Схемы структур представителей некоторых суперсемейств non-LTR-ретротранспозонов представлены на Рис. 2.

ORF2 кодирует белок ORF2p, играющий ключевую роль в процессе перемещения non-LTR-ретротранспозонов. В этом белке присутствуют два основных функциональных домена, необходимых для транспозиции non-LTR-ретротранспозонов. Центральным доменом, без которого автономное перемещение элемента невозможно, является домен RT. RT осуществляет синтез кДНК по матрице РНК при встраивании элемента в геном. Второй домен, присутствующий фактически у всех non-LTR-ретротранспозонов – это домен эндонуклеазы (Feng *et al.*, 1996; Malik *et al.*, 1999; Yang *et al.*, 1999). Несмотря на широкое распространение, согласно современным представлениям о механизме перемещения non-LTR-ретротранспозонов, в целом эндонуклеазная активность не

R2 (*Bombyx mori*)



L1 (*Homo sapiens*)



I (*Drosophila melanogaster*)

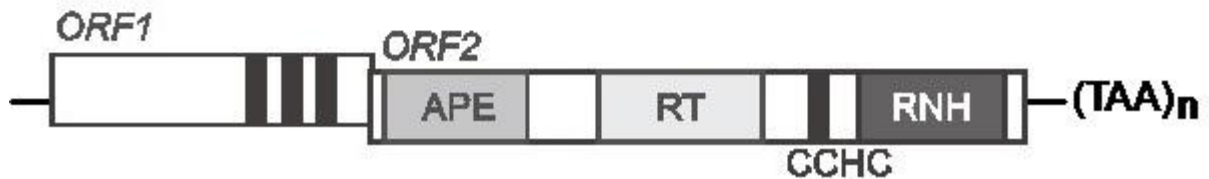


Рис. 2. Схема структурной организации некоторых non-LTR-ретротранспозонов (согласно Han, 2010, с модификациями). RT – обратная транскриптаза; REL – эндонуклеаза рестрикционного типа; ORF – открытая рамка считывания; CC – N-концевая суперспираль; RRM – домен узнавания РНК; CTD - C-концевой домен; APE – апуриновая/апиридиновая эндонуклеаза, RNH – рибонуклеаза H.

является абсолютно необходимой для ретротранспозиции non-LTR-элемента (Morrish *et al.*, 2002). Существует два типа эндонуклеаз non-LTR-ретротроэлементов: (1) эндонуклеаза рестрикционного типа (REL-endo - restriction enzyme-like endonuclease) и (2) апуриновая/апиридиновая эндонуклеаза (APE -apurinic/apurimidinic endonuclease). Сайт-специфичная REL-endo характерна для эволюционно более древних non-LTR-ретротранспозонов, в то время как более молодые, то есть позднее дивергировавшие элементы, кодируют APE, не имеющую выраженной сайт-специфичности (Malik *et al.*, 1999). Обычно non-LTR-ретротранспозон содержит только один домен эндонуклеазы. Исключением является элемент *Dualen* из водоросли *Chlamydomonas reinhardtii*, который содержит как APE, так и REL-endo и, по-видимому, является переходным звеном между “древними” и “молодыми” элементами (Kojima, Fujiwara, 2005). Кроме того, у некоторых семейств non-LTR-ретротранспозонов был найден домен рибонуклеазы H (RNH), расположенный в C-конце ORF2p (Malik, 2005), и

цистеиновые мотивы, расположенные как в N-, так и в С-концах кодируемых белков (Burke *et al.*, 1999).

Другая открытая рамка считывания (ORF1) кодирует белок ORF1p, имеющий РНК-связывающую (Hohjoh, Singer, 1997) и шаперонную (Martin, 2010) активности. ORF1p non-LTR-ретротранспозонов выполняет роль, подобную роли белков, кодируемых геном *gag* ретровирусов и LTR-ретротранспозонов, формируя вместе с ORF2p рибонуклеопротеиновую частицу (RNP - RiboNucleoprotein Particle) (Doucet *et al.*, 2010). Отметим, что в состав RNP вместо РНК non-LTR-ретротранспозона может включаться клеточная мРНК, что может приводить к появлению в геноме псевдогенов (Mandal *et al.*, 2013). Наиболее детально изучена структура ORF1p у *L1*-non-LTR-ретротранспозона человека. Этот белок состоит из трех доменов: N-концевой суперспирали (CC – coiled coil, Рис. 2), необходимой для функциональной тримеризации белка (Martin *et al.*, 2003; Khazina, Weichenrieder, 2009); центрального домена узнавания РНК (RRM – RNA recognition motif, Рис. 2) (Khazina, Weichenrieder, 2009); и С-концевого домена (CTD – C-Terminal Domain, Рис. 2), который участвует в связывании с РНК (Khazina *et al.*, 2011). Биоинформатические исследования показали, что многие ORF1p других non-LTR-ретротранспозонов также содержат один или несколько RRM и цистеиновых мотивов (Heitkam, Schmidt, 2009; Khazina, Weichenrieder, 2009; Капелинская и др., 2011).

5'- и 3'-нетранслируемые районы (UTRs) non-LTR-ретротранспозонов довольно вариабельны. Большинство non-LTR-ретротранспозонов в 5'-UTR имеют внутренний промотор для РНК-полимеразы II (Mizrokhi *et al.*, 1988; Schumann *et al.*, 1994; Ostertag, Kazazian, 2001). Наличие внутреннего промотора означает, что транскрибируемая мРНК сохранит последовательность промотора, которая в процессе обратной транскрипции будет включена в последовательность новой копии элемента (Takahashi, Fujiwara, 1999). Несмотря на то, что многие из охарактеризованных 5'-UTR non-LTR-ретротранспозонов содержат активный промотор, зачастую происходит замена одного промотора другим (Haas *et al.*, 2001), и поэтому их последовательности у различных элементов не всегда являются консервативными. Кроме того описан ряд элементов, у которых не было обнаружено внутренних промоторов. Например, *R2*-элемент насекомых,

встраивающийся рядом с рибосомальными генами, не содержит собственного промотора. Предполагается, что в данном случае происходит транскрипция РНК *R2-non-LTR*-ретротранспозона вместе с РНК рибосомального гена (George, Eickbush, 1999) с использованием промотора этого гена. 3'-UTR non-LTR-ретроэлементов обычно содержит специфичную последовательность, узнаваемую белком ORF2p (Kajikawa, Okada, 2002; Anzai *et al.*, 2005). Исключением являются *L1-non-LTR*-элементы млекопитающих (Moran *et al.*, 1996) и некоторых растений (Ohshima, 2012), у которых 3'-UTR не влияет на транспозицию. 3'-конец non-LTR-ретротранспозонов может содержать полидезоксиденозиновую последовательность (polyA) или тандемные повторы (Han, 2010). Кроме основных структурных доменов, некоторые non-LTR-ретротранспозоны имеют дополнительные домены. Например, ORF1p *CR1*-элемента содержит домен эстеразы (Kapitonov, Jurka, 2003), связывающийся с фосфолипидами и липосомами (Schneider *et al.*, 2013).

Многие геномные копии non-LTR-ретротранспозонов зачастую обрезаны с 5'-конца из-за неполного прохождения обратной транскрипции (Szak *et al.*, 2002). Только некоторые из копий non-LTR-элементов в геноме являются полноразмерными и активными. Например, из 500 тыс. копий *L1-non-LTR*-элементов в геноме человека всего около 7 тыс. целые, и из них только от 80 до 100 копий способны перемещаться (Brouha *et al.*, 2003).

1.2.2. Механизм перемещения non-LTR-ретротранспозонов

Общая схема перемещения non-LTR-ретротранспозона изображена на Рис. 3. После транскрипции и выхода мРНК элемента из ядра в цитоплазму происходит трансляция ORFs и сбор RNPs (Hohjoh, Singer, 1996; Seleme *et al.*, 2005). Для non-LTR-ретротранспозонов, транскрипция которых происходит с использованием промотора для РНК-полимеразы II, первая рамка, скорее всего, транслируется с использованием традиционной модели распознавания кэпа РНК. Этот тип трансляции (в противоположность трансляции с использованием IRES - внутреннего сайта связывания с рибосомой) подтверждается экспериментами на млекопитающих, в которых было продемонстрировано, что а) даже значительное

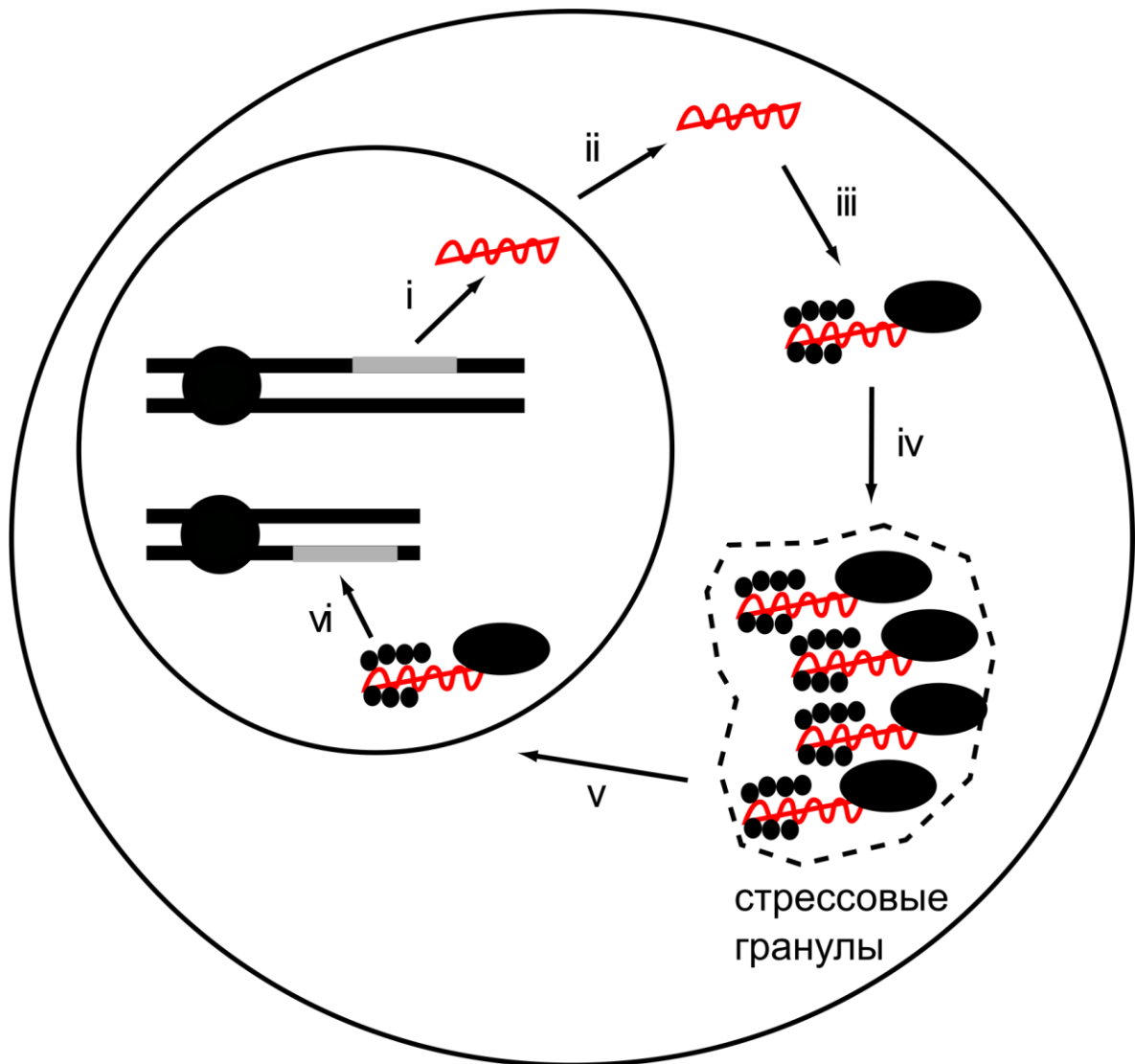


Рис. 3. Общая схема перемещения non-LTR-ретротранспозона (согласно Han, 2010). Цикл перемещения non-LTR-ретротранспозона в клетке: (i) Транскрипция полноразмерного активного элемента. (ii) экспорт мРНК из ядра. (iii) Трансляция белков ретротранспозона (iv) прохождение через цитоплазматическую гранулу. (v) Импорт рибонуклеиновой частицы (RNP) в ядро. (vi) интеграция с помощью механизма РНК-опосредованной интеграции (TPRT).

количество мутаций в 5'-UTR, 3'-UTR и последовательности между ORF1 и ORF2 не влияет на ретротранспозицию элементов (An *et al.*, 2011; Moran *et al.*, 1996); б) кодирующая последовательность может быть значительно изменена без влияния на ретротранспозицию (Han, Воеке, 2004); в) инициация трансляции с 5'UTR ORF1p *LI*-элемента человека зависит от наличия кэпа (Dmitriev *et al.*, 2007). Элементы, транскрибируемые РНК-полимеразой I вместе с рибосомальными генами с их

промоторов, такие как *R1/R2*-элементы, могут транслироваться с использованием IRES (George, Eickbush, 1999) – особых структур РНК, связывающихся с рибосомами и характерных для вирусов и некоторых клеточных мРНК (Pfungsten, Kieft, 2008). Для процессинга своей РНК, а именно её отделения от мРНК рибосомальных генов, эти non-LTR-ретротранспозоны кодируют специальный рибозим, катализирующий расщепление молекулы, в которой он находится (Eickbush, Eickbush, 2010). В транскриптах, содержащих отдельные открытые рамки считывания, трансляция ORF2 основывается на особом способе инициации. Исследования non-LTR-ретротранспозона *START1* шелкопряда и *L1*-элемента человека показали, что существует общий механизм инициации трансляции, при котором трансляция ORF2 сопряжена с трансляцией ORF1 (Alisch *et al.*, 2006; Kojima *et al.*, 2005). При этом происходит реинициация трансляции ORF2 либо той же рибосомой, что осуществляла трансляцию ORF1, либо с привлечением другой рибосомы. Например, в ORF2p некоторых non-LTR-ретротранспозонов присутствуют 2A-домен, который осуществляет пропуск синтеза рибосомой определенных аминокислотных последовательностей, что приводит к терминации синтеза полипептида и реинициации трансляции (Odon *et al.*, 2013). RNP non-LTR-ретротранспозонов в большинстве случаев локализуется с использованием ORF1p в качестве маркера в цитоплазме. Интересное исключение было обнаружено в клеточной культуре крыс линии хронического миелоидного лейкоза, где сигналы ORF1p *L1*-non-LTR-ретротранспозонов локализовались в ядре (Kirilyuk *et al.*, 2008). Причины такого аномального распространения белка в клетке не ясны. Существует мнение, что в клетках данной линии происходит синтез мутантной версии ORF1p, которая не задерживается в цитоплазме, а сразу поступает в ядро (Goodier *et al.*, 2007; Kirilyuk *et al.*, 2008). Стоит также отметить, что ORF1p non-LTR-ретротранспозонов *Drosophila melanogaster*, участвующих в достраивании теломер, также локализуется в ядре.

Детальное исследование RNP non-LTR-ретротранспозонов млекопитающих выявило их локализацию в так называемых стрессовых гранулах (Рис. 3) — цитоплазматических тел, ассоциированных с Р-тельцами (англ. P-body) (Goodier *et al.*, 2007). Р-тельца являются важной частью системы РНК-интерференции (Jakymiw *et al.*, 2005). Было показано, что стрессовые гранулы являются

резервуарами для нетранслируемой мРНК (Balagopal, Parker, 2009). Остается неясным, направляются ли RNP non-LTR-ретротранспозонов в стрессовые гранулы для последующего разрушения или прохождение через стрессовые гранулы является важным этапом в созревании RNP ретротранспозона. Тем не менее, известно, что потеря структурных компонент Р-телец в дрожжах приводит к уменьшению активности *Ty1*- и *Ty3*-LTR-ретротранспозонов, что говорит в пользу гипотезы активного участия стрессовых гранул в жизненном цикле ретротранспозонов (Aye *et al.*, 2004; Checkley *et al.*, 2010). Динамика RNP non-LTR-ретротранспозонов была наиболее тщательно изучена на примере *I*-элемента дрозофилы (Seleme *et al.*, 1999, 2005). RNPs *I*-элемента появляются в питающих клетках ооцита и затем, вместе с другими компонентами питающих клеток, перемещаются в цитоплазму ооцита. В ооците RNPs мигрируют из задней части ооцита в переднюю (Seleme *et al.*, 1999, 2005). Миграция RNP *I*-элемента напоминает транспорт RNP гена *bicoid* и происходит в направлении центра организации микротрубочек (Ferrandon *et al.*, 1994; Seleme *et al.*, 1999). Зависимость движения RNP от наличия микротрубочек, а также данные, подтверждающие взаимодействие ORF1p с промежуточными филаментами (Kirilyuk *et al.*, 2008), говорят в пользу гипотезы движения RNPs non-LTR-ретротранспозонов вдоль цитоскелета. В конечном счете, RNP non-LTR-ретротранспозона транспортируется в ядро, где и происходит процесс обратной транскрипции.

Оказавшись у подходящего сайта на хромосоме, non-LTR-ретротранспозон может начать копирование своей генетической информации в новый локус. Механизм этого процесса был назван “РНК-опосредованной интеграцией” (TPRT - target primed reverse transcription, схема механизма представлена на Рис. 4) и его начальные стадии наиболее убедительно были продемонстрированы *in vitro* в исследованиях *R2*-элемента из генома *Bombyx mori*. Очищенный белок, кодируемый единственной рамкой *R2*-элемента, смешанный с РНК *R2*-элемента и подходящей донорной молекулой ДНК, способен разрезать одну из цепей ДНК (Рис. 4, *ii*). Освобождающийся 3'-конец затем служит в качестве праймера обратной транскрипции РНК *R2*-элемента (синтез минус-цепи, Рис. 4, *iii*). Похожие данные были получены и для системы с *L1*-элемента человека, для которой, однако,

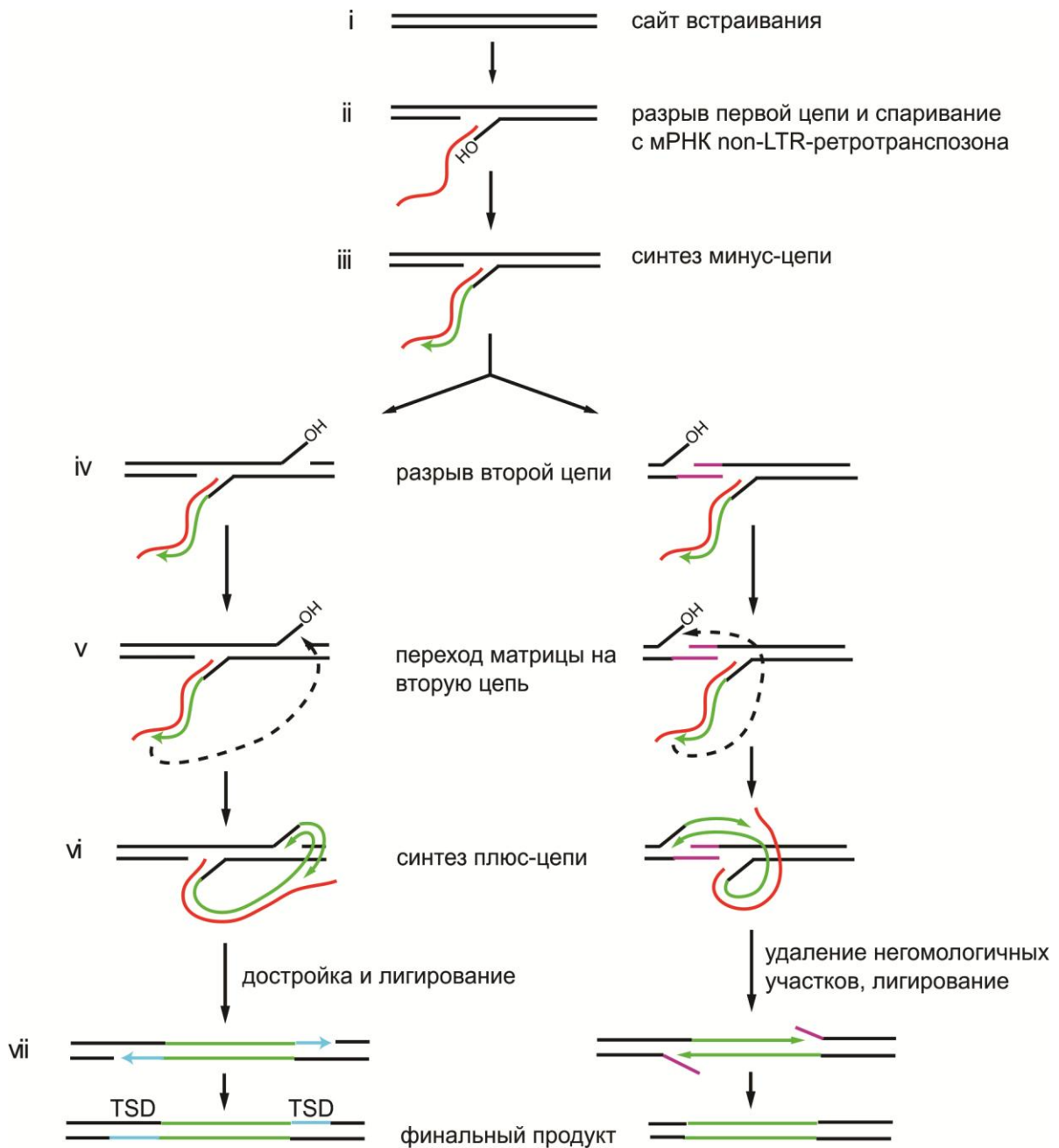


Рис. 4. Механизм РНК-опосредованной интеграции (по Нап, 2010, с модификациями). Основные этапы обозначены римскими цифрами. Объяснения даны в тексте. TSD - дупликация сайта встраивания.

прямое отслеживание результатов проводимой реакции осложняется низким выходом её продукта и отсутствием строго определенного сайта разрыва ДНК (в отличие от *R2*-non-LTR-ретротранспозона). Сайт встройки non-LTR-ретротранспозона определяется типом домена эндонуклеазы, кодируемого конкретным элементом. Как уже говорилось, древние группы non-LTR-ретротранспозонов, такие как *CRE*, *R2* и *R4*, кодируют эндонуклеазу типа REL-endo

(Malik *et al.*, 1999), которая узнает специфичные последовательности. В более молодых группах этот домен был заменен на эндонуклеазу типа APE (Malik *et al.*, 1999), что, в большинстве случаев, привело к ослаблению или релаксации сайт-специфичности. Значение домена эндонуклеазы заключается в создании праймера для обратной транскрипции при разрезании донорной цепи ДНК. Если существуют дополнительные механизмы создания праймера, то транспозиция non-LTR-элемента может происходить и в отсутствие эндонуклеазной активности белка, кодируемого элементом (Morrish *et al.*, 2002). Например, перемещение *TART*-, *Het-A*- и *TARE*-non-LTR-ретротранспозонов дрозофилы, формирующих теломеры на концы хромосом, не требует эндонуклеазной активности (Levis *et al.*, 1993; Sheen, Levis, 1994; Pardue *et al.*, 2005). Похожий механизм эндонуклеазо-независимого перемещения был искусственно воспроизведен с использованием синтетического мутантного *L1*-non-LTR-элемента человека (Morrish *et al.*, 2007). Возможно, что древние non-LTR-ретротранспозоны, не содержали домен эндонуклеазы, а использовали по тем или иным причинам образовавшиеся свободные концы ДНК (Han, 2010). Кроме того, было показано влияние и других доменов белков, кодируемых non-LTR-ретротранспозонами, на процесс связывания RNP элемента с сайтом встраивания. В частности, полипротеин *R2Bm*-ретротранспозона на N-конце несет два цистеиновых мотива состава CCHC, а также Mub-связывающие мотивы (Burke *et al.*, 1999), отсутствие которых существенно влияет на способность полипротеина связывать ДНК (Christensen *et al.*, 2005).

Последующие этапы интеграции non-LTR-ретротранспозона остаются не совсем ясными, но, судя по всему, далее ретроинтеграция происходит согласно механизму, приведенному на Рис. 4. После или во время синтеза минус-цепи в сайте встраивания происходит разрыв второй цепи ДНК (Рис. 4, *iv*). В зависимости от конкретного non-LTR-ретротранспозона данный разрыв может происходить или в том же месте, ниже или выше места в комплементарной цепи, где произошел первый разрыв. Если второй разрыв произошел ниже первого разрыва в комплементарной цепи, то образуется дупликация сайта встраивания (левая часть Рис. 4, TSD – target site duplication); если выше, то делеция сайта встраивания; при разрыве того же сайта в комплементарной цепи происходит встраивание без делеций и дупликаций. Несмотря на отсутствие прямого экспериментального

доказательства, считается, что домен эндонуклеазы non-LTR-ретротранспозонов осуществляет разрыв и второй цепи ДНК. Было показано, что ORF2p R2-non-LTR-ретротранспозона функционирует как димер, каждая из субъединиц которого может катализировать один из разрывов (Yang, Eickbush, 1998; Christensen *et al.*, 2006). Остается непонятным, как происходит выбор места разрыва второй цепи, но, похоже, что этот процесс происходит по-разному у различных элементов. В некоторых случаях разрыв второй цепи может определяться доменом эндонуклеазы non-LTR-ретротранспозона. Для элементов млекопитающих существуют данные о том, что место разрыва второй цепи определяется уровнем его идентичности синтезируемой минус-цепи элемента (Symer *et al.*, 2002; Zingler *et al.*, 2005). Тем не менее, для большинства случаев, механизм того, как именно происходит разрыв второй цепи, остается неизвестным.

Разрыв второй цепи создает праймер для синтеза плюс-цепи ДНК встраиваемого non-LTR-ретротранспозона (Рис. 4, v-vi). То есть, синтезированная минус-цепь ДНК non-LTR-ретротранспозона вместо РНК ретротранспозона теперь связывается со второй цепью ДНК сайта встраивания. Этот переход минус-цепи ДНК non-LTR-ретротранспозона, по всей видимости, происходит с использованием шаперонной активности ORF1p (Martin, 2010). Исследования *in vitro* показали, что С-конец белка L1Tc, содержащий два цистеиновых мотива, выполняет роль молекулярного шаперона, способствуя отжигу комплементарных участков ДНК и препятствуя плавлению дуплексов ДНК (Heras *et al.*, 2005). Соответственно, синтез плюс-цепи non-LTR-ретротранспозона начинается в месте разрыва второй цепи сайта встраивания. Для того чтобы удалить РНК, связанную с минус цепью ДНК элемента, некоторые non-LTR-ретротранспозоны приобрели домен RNH (Blesa, Martínez-Sebastián, 1997; Malik *et al.*, 1999). Но многие другие ретротранспозоны либо совсем не содержат домен RNH, либо потеряли его в результате эволюции (Novikova *et al.*, 2009). Возможно, что для удаления РНК эти ретротранспозоны используют белок RNH, кодируемый хозяйским организмом и функционирующим в ядре в процессе репликации (Malik *et al.*, 1999; Cerritelli, Crouch, 2009), или же РНК просто смещается в результате синтеза плюс-цепи элемента (Kurzynska-Kokorniak *et al.*, 2007). После того, как синтез ДНК

заканчивается, происходит лигирование оставшихся разрывов и завершение транспозиции (Рис. 4, *vii*).

1.2.3. Взаимодействие non-LTR-ретротранспозонов и организма хозяина

Считается, что в процессе перемещения non-LTR-элемент использует ряд клеточных факторов, необходимых для его правильного прохождения (Beauregard *et al.*, 2008). Такая точка зрения возникла по многим причинам. Во-первых, не было найдено ни одного non-LTR-ретротранспозона, который бы кодировал белки со всеми необходимыми для ретротранспозиции активностями. Например, разрывы, создаваемые в результате TPRT, должны быть защищены после синтеза ДНК. При этом до сих пор не был найден non-LTR-элемент, который бы кодировал белок с лигазной активностью (Han, 2010). В случае удаления сайта встраивания неясно, какие нуклеазы осуществляют удаление неспаренной ДНК (правая часть Рис. 4, *vii*). Кроме того, полный процесс ретротранспозиции пока не был воспроизведен *in vitro* с использованием очищенных субстратов и белков, кодируемых non-LTR-ретротранспозоном. Вполне возможно, что ретротранспозицию *in vitro* не удастся воспроизвести по причине сложной многошаговой природы перемещения non-LTR-ретротранспозона и невозможности подобрать подходящие условия. Тем не менее, проведенный анализ последовательностей сайтов встраивания non-LTR-элементов в геном предполагает вовлечение клеточных механизмов репарации в этот процесс (Zingler *et al.*, 2005; Ichiyanagi *et al.*, 2007). И, наконец, экспериментальные данные говорят в пользу зависимости транспозиции non-LTR-ретротранспозонов от клеточных факторов репарации ДНК. Трансфекция плазмиды, содержащей *L1*-non-LTR-ретротранспозон человека, в клетки млекопитающих приводит к эндонуклеазо-зависимым разрывам в двойной цепи ДНК, и успешное прохождение транспозиции в этих клеточных культурах зависит от АТМ-киназы, вовлеченной в процессы гомологичной рекомбинации и негомологичного соединения концов при репарации (Gasior *et al.*, 2006). Кроме того, в клетках, лишенных таких факторов как Ku70, Artemis и LigIV, участвующих в негомологичном соединении концов, транспозиция *L1*-non-LTR-ретротранспозонов не происходит (Suzuki *et al.*, 2009). Аналогичное исследование бактериальных интронов группы II выявило участие в их функционировании многочисленных

факторов ДНК-репарации (Smith *et al.*, 2005). Бактериальные интроны имеют общего с non-LTR-ретротранспозонами предка (Malik *et al.*, 1999) и поэтому изучение этих генетических элементов играет большую роль в понимании механизмов перемещения non-LTR-ретротранспозонов. В целом, все эти данные говорят о важной роли клеточных факторов в перемещении non-LTR-ретротранспозонов. В этом отношении non-LTR-ретротранспозоны не являются уникальной группой МЭ, поскольку многим МЭ других классов на многих этапах их перемещения также необходимы клеточные факторы.

Для того чтобы копии МЭ передавались из поколения в поколение, имеет смысл предположить, что МЭ должны активно перемещаться в клетках зародышевого пути, ведь именно из этих клеток будут произведены последующие поколения. Напротив, перемещения в соматических клетках будут эволюционно непродуктивными, так как новые копии ретротранспозонов будут утеряны при гибели организма. В соответствии с такой точкой зрения оказалось, что non-LTR-ретротранспозоны действительно экспрессируются и перемещаются в клетках зародышевой линии (Trelogan, Martin, 1995; Seleme *et al.*, 1999). В то время как “успех” транспозона зависит от его перемещения в клетках зародышевой линии, большое число встраиваний новых элементов в геном может оказать негативный эффект на хозяйский организм. Поэтому клетки обычно стараются подавлять активность МЭ, и именно в клетках зародышевого пути происходит борьба между МЭ и организмом хозяина. Потеря контроля над активностью МЭ может привести к серьезным последствиям для организма. Известным примером является синдром гибридного дисгенеза, наблюдаемый при скрещивании самок *Drosophila melanogaster*, в геноме которых не содержится активных МЭ, с самцами, имеющими активно перемещающиеся МЭ. Потомство от такого скрещивания часто стерильно, и их геном характеризуется разнообразными мутациями и хромосомными перестройками (Kidwell *et al.*, 1977). Этот синдром вызван неспособностью организма контролировать перемещение МЭ. Кроме того, недавно был показан разрушительный эффект неконтролируемого увеличения числа транспозиций в клетках зародышевого пути млекопитающих, в которых, по-видимому, именно повышенная экспрессия *LI*- и *IAP*-элементов приводит к стерильности (Aravin *et al.*, 2007; Kuramochi-Miyagawa *et al.*, 2008).

1.2.4. Регуляция с использованием малых РНК

Открытие РНК-интерференции (RNAi) (Fire *et al.*, 1998) привело к возникновению гипотезы о том, что малые РНК вовлечены в регуляцию транспозонов посредством метилирования ДНК и деградации РНК МЭ. Генетические исследования выявили, что мутации RNAi на самом деле приводят к реактивации МЭ (Ketting *et al.*, 1999; Tabara *et al.*, 1999). Сейчас известно, что в клетках экспрессируются различные варианты малых РНК, самые изученные из которых это: микро РНК (miRNA), эндогенные малые интерферирующие РНК (endo-siRNA) и РНК, взаимодействующие с белками Piwi (piRNA). Значительная часть endo-siRNA и piRNA соответствуют последовательностям транспозонов, что позволило сделать предположение о том, что механизмы репрессии МЭ основаны на узнавании их последовательностей (Aravin *et al.*, 2003; Houwing *et al.*, 2007; Ghildiyal *et al.*, 2008). Для осуществления своих регуляторных функций эти РНК взаимодействуют с белками семейства Argonaut (Hutvagner, Simard, 2008). piRNA специфичны для клеток зародышевого пути, в то время как endo-siRNA присутствуют во всех клетках организма (Han, 2010).

Модель регуляции МЭ посредством piRNA в клетках зародышевого пути была разработана в экспериментах на дрозофиле и мышах. В этой модели два класса piRNA взаимодействуют с различными белками семейства Argonaut. Класс piRNA, соответствующий прямой ориентации транскрипта МЭ, взаимодействует с белками Piwi и Ago3 (в мышах – MILI) (Aravin *et al.*, 2007; Brennecke *et al.*, 2007), а класс piRNA, комплементарный последовательности МЭ, взаимодействует с белком Aubergine (MILI2 у мышей). Эти классы piRNA нарабатываются с использованием механизма “пинг-понг”, в рамках которого прямая цепь piRNA направляет разрезание и синтез обратной цепи и наоборот. Обратная цепь piRNA происходит от длинной молекулы-предшественника, транскрибируемой с небольшого количества piRNA-кластеров в геноме (Brennecke *et al.*, 2007; Aravin *et al.*, 2008), которая затем разрезается нуклеазой Zucchini (Voigt *et al.*, 2012).

Мутации, приводящие к нарушению функционирования белков семейства Argonaute, приводят к повышению активности транспозиции всех типов МЭ (Meister, 2013). Повышенная экспрессия МЭ, в свою очередь, связана с дефектами в развитии клеток зародышевого пути, что подчеркивает важность регуляции

перемещения транспозонов (Carmell *et al.*, 2007). Считается, что в клетках зародышевого пути Piwi-регуляция работает не только на уровне трансляции, осуществляя разрезание мРНК транспозона, но и на уровне транскрипции, так как осуществляет метилирование ДНК (Di Giacomo *et al.*, 2013). В этом заключается эволюционный смысл регуляции с помощью Piwi: нарабатываются транспозон-специфичные piRNA, которые осуществляют метилирование ДНК МЭ в клетках зародышевого пути. Это метелирование затем передается всем последующим клеткам, образующимся в результате митоза (Law, Jacobsen, 2010).

Давно известно, что геномные копии МЭ часто метилированы и это метилирование играет важную роль в долгосрочной регуляции активности МЭ (Yoder *et al.*, 1997). Было обнаружено, что нарушения в функционировании Piwi-белков приводят, в том числе, и к нарушению в метилировании промоторных районов *LINE-1*-элементов у мыши (De Fazio *et al.*, 2011). Каким образом это происходит, остается не до конца ясным (Han, 2010), но считается, что этот механизм может быть подобен механизму образования гетерохроматина в различных организмах (Grewal, Elgin, 2007).

1.2.5. Внутриклеточная защита с использованием белков семейства АРОВЕС

Впервые антивирусная активность цитидин-диаминазы АРОВЕС3G была показана при обнаружении её способности ингибировать ретровирус ВИЧ (Mangeat *et al.*, 2003). Дезаминирование цитозина в синтезированной цепи кДНК ВИЧ при помощи АРОВЕС3G приводит к гипермутациям гуанинов в аденины в кодирующей последовательности вируса. Поскольку ретроэлементы похожим образом синтезируют цепь кДНК, появилась гипотеза о том, что белки семейства АРОВЕС также регулируют перемещение ретротранспозонов. Действительно, было установлено, что белки семейства АРОВЕС3 ингибируют активность ретротранспозонов (Dutko *et al.*, 2005; Muckenfuss *et al.*, 2006). Однако стоит отметить, что в литературе присутствуют некоторые неясности относительно роли АРОВЕС3G в ингибировании *L1*-элементов млекопитающих (Kinomoto *et al.*, 2007). Было выявлено, что при коэкспрессии белков АРОВЕС3 с non-LTR-ретротранспозонами не происходит гипермутаций во встраиваемом элементе, и активность цитидин-диаминазы в большинстве, но не во всех случаях, оказалась несущественной для антиретротранспозонного эффекта этих белков (Stenglein,

Harris, 2006). Скорее всего, существует по крайней мере два различных механизма ингибирования перемещения non-LTR-ретротранспозонов при помощи белков семейства APOBEC: деаминазо-зависимый и деаминазо-независимый. Механизм деаминазо-независимой регуляции non-LTR-ретротранспозонов пока остается неясным, тем не менее, локализация белков APOBEC3 в клетке дает о нем некоторое представление.

Ингибиторы APOBEC3A, B, C и H *L1*-non-LTR-ретротранспозонов человека локализуются, по крайней мере, частично в ядре (Bogerd *et al.*, 2006; Kinomoto *et al.*, 2007), что говорит в пользу того, что эти факторы, скорее всего, участвуют в блокировании интеграции *L1*-элементов. Действительно, оказалось, что APOBEC3C, взаимодействуя с ORF1p, снижает эффективность процесса обратной транскрипции *L1*-элементов (Horn *et al.*, 2014). APOBEC3D, E, F и G чаще всего оказываются в цитоплазме. Они могут играть роль в цитоплазматическом изолировании RNP. Было показано, что APOBEC3G локализуется в Р-тельцах и стрессовых гранулах (Gallois-Montbrun *et al.*, 2007). Учитывая присутствие *L1*-RNPs в этих структурах, можно предположить, что туда их направляет для изолирования APOBEC3G.

1.2.6. Особенности регуляции МЭ в геномах растений

Растения, так же как и животные, используют механизм регуляции активности МЭ, связанный с использованием малых РНК (Wassenegger *et al.*, 1994). В этом процессе, помимо компонент RNAi, которые работают у животных, растения используют дополнительные факторы (Law, Jacobsen, 2010), в том числе и специфичные для растений РНК-полимеразы IV и V (Matzke *et al.*, 2009). В целом считается, что растения развили более мощную, чем у животных, систему эпигенетической регуляции активности МЭ путем метилирования ДНК (Fedoroff, 2012). У млекопитающих метилирование цитозина происходит главным образом в CG-динуклеотидах, в то время как у растений метилирование происходит в любом контексте (Law, Jacobsen, 2010). Кроме того, в отличие от животных, клетки зародышевого пути у растений закладываются на более поздних этапах развития, что накладывает более строгие ограничения на сохранение генетической целостности, так как любые из развивающихся клеток растения могут оказаться перепрограммированы на развитие репродуктивных клеток. По всей видимости,

именно развитие эпигенетических механизмов сделало возможным такой вариант функционирования растительных организмов (Fedoroff, 2012).

1.2.7. Эволюция и классификация non-LTR-ретротранспозонов на основе домена обратной транскриптазы

Единственным доменом, присутствующим во всех без исключения автономных non-LTR-ретротранспозонах, является домен RT. Поэтому, а также из-за высокой консервативности домена RT по сравнению с другими белковыми доменами, кодируемыми non-LTR-ретротранспозонами, этот домен является наиболее удобным для использования в реконструкции филогенетических взаимоотношений (Xiong, Eickbush, 1990).

Наиболее детальной работой по анализу эволюции non-LTR-ретротранспозонов, проведенной на данный момент и широко цитируемой в научной литературе, является работа Малика и др. (Malik *et al.*, 1999). Несмотря на то, что с момента публикации данного исследования прошло достаточно много времени, результаты, приведенные в ней, во многом остаются единственными, посвященными анализу эволюции всей группы non-LTR-ретротранспозонов. Проведенный тогда анализ позволил выделить среди non-LTR-ретротранспозонов одиннадцать филогенетических групп или *суперсемейств* non-LTR-ретротранспозонов. В последнее десятилетие были выявлены дополнительные суперсемейства этих МЭ, и их число достигло двадцати трех (Рис. 5) (Kojima, Fujiwara, 2005; Новикова, Блинов, 2009; Novikova *et al.*, 2009). В рамках проведенного филогенетического анализа, помимо классификации, были выявлены ключевые моменты в эволюции non-LTR-ретротранспозонов в виде приобретения дополнительных или замены «старых» функциональных доменов (подобные процессы легли в основу представления об эволюции non-LTR-ретротранспозонов как о *модульной* эволюции, подробнее см. раздел 1.2.6.). Эти события, по-видимому, давали преимущества при ретротранспозиции МЭ, такие как, например, релаксация сайт-специфичности (Malik *et al.*, 1999).

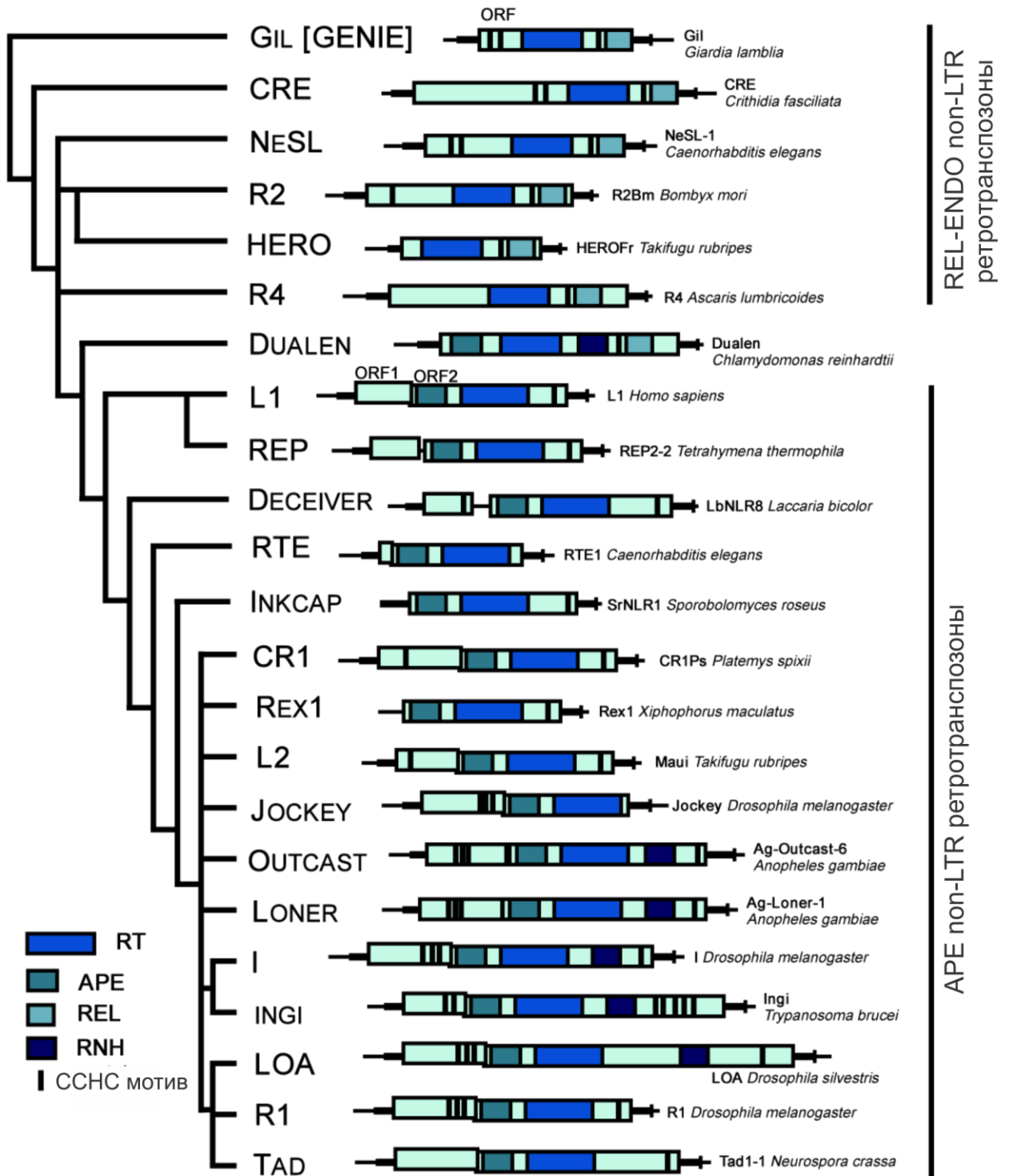


Рис. 5. Эволюция и структурное разнообразие non-LTR-ретротранспозонов. RT – обратная транскриптаза; APE – апуриновая/апириமிдиновая эндонуклеаза; REL – рестриктазо-подобная эндонуклеаза; RNH – рибонуклеаза H. Рисунок взят с сайта <http://retrotransposons.org/> с разрешения автора (Новикова О.С).

REL-endo non-LTR-ретротранспозоны

Согласно филогенетическому дереву на основе RT выделяют шесть наиболее древних филогенетических групп: *CRE*, *R2*, *R4*, *Gil* (или *GENIE*) и *NeSL* (Рис. 5) (Malik *et al.*, 1999; Burke *et al.*, 2002; Kojima, Fujiwara, 2005). Non-LTR-ретротранспозоны, относящиеся к этим группам, чаще всего имеют в своем составе единственную ORF. На С-конце ORFp этих элементов, после домена RT, всегда идентифицируется домен эндонуклеазы с рестриктазной активностью (REL-endo) (Malik *et al.*, 1999; Yang *et al.*, 1999). Как уже говорилось ранее (раздел 1.2.2), считается, что эндонуклеаза non-LTR-ретротранспозонов определяет выбор сайта встраивания элемента. Поэтому, учитывая, что REL-endo относится к эндонуклеазам рестрикционного типа, неудивительно, что древние non-LTR-ретротранспозоны обладают сайт-специфичностью встраивания, то есть внедрение новой копии элемента происходит по определенным сайтам в геноме. Основными мишенями для встраивания REL-endo-элементов являются ген 28S рибосомальной РНК (Burke *et al.*, 1995), микросателлиты (Kojima, Fujiwara, 2004), а также теломерные повторы (Okazaki *et al.*, 1995), то есть высокоповторенные участки генома.

Gil/ GENIE

Наиболее древними из известных на сегодняшний день non-LTR-ретротранспозонов считаются элементы филогенетической группы *Gil* (или *GENIE*). Эти non-LTR-ретротранспозоны описаны как теломер-специфичные ретротранспозоны простейшего *Giardia lamblia* (Arkhipova, Morrison, 2001). Других элементов, относящихся к этой группе, пока не найдено. Non-LTR-элементы этой группы несут одну ORF, кодирующую белок, содержащий домены RT и REL-endo, а также два цистеиновых мотива ССНН на N-конце, которые предположительно отвечают за связывание с РНК (Arkhipova, Morrison, 2001).

CRE

Элементы филогенетической группы *CRE* долгое время были известны только для трипаносом (Teng *et al.*, 1995). Однако, позднее *CRE*-элементы были обнаружены и в грибах (Goodwin, Poulter, 2001; Novikova *et al.*, 2009). Общая структура этих элементов идентична структуре *Gil/GENIE*-элементов.

NeSL-1

NeSL-1-non-LTR-ретротранспозоны были впервые обнаружены у нематоды *Caenorhabditis elegans* (Malik, Eickbush, 2000). В дальнейшем, элементы этой группы были обнаружены у костных рыб (Kojima, Fujiwara, 2004). Уникальной особенностью структуры этих элементов является наличие домена протеазы (Malik, Eickbush, 2000), характерного для жизненного цикла LTR-ретротранспозонов и отсутствующего у большинства других non-LTR-элементов. Значение этого домена для перемещения *NeSL-1*-элементов остается неясным.

R4

Элементы группы *R4* встраиваются специфично в гены рибосомальных РНК (Burke *et al.*, 1995) и микросателлитные повторы (Kojima, Fujiwara, 2004), и описаны у широкого круга животных, таких как насекомые (Kojima, Fujiwara, 2004), костные рыбы (Volff *et al.*, 2001), ящерицы (Novick *et al.*, 2009), нематоды *Ascaris lumbricoides* (Burke *et al.*, 1995), а также у амебы *Entamoeba histolytica* (Mandal *et al.*, 2004). ORF у этих элементов также единственная, и она содержит RT и REL-endo.

R2

Наиболее изученными среди древних non-LTR-ретротранспозонов, содержащих REL-endo являются элементы группы *R2*. Именно с использованием *R2*-элемента *Bombyx mori* была получена большая часть данных о механизме TPRT перемещения non-LTR-ретротранспозонов (Yang *et al.*, 1999; Christensen *et al.*, 2005, 2006; Eickbush, Eickbush, 2010). *R2*-элементы впервые были обнаружены у видов рода *Drosophila*, и они являются сайт-специфичными для генов рибосомальных РНК (Eickbush *et al.*, 1997). Впоследствии эти non-LTR-ретротранспозоны были описаны для многих животных (Luchetti, Mantovani, 2013). ORFp, кодируемый этими транспозонами, обычно содержит домены RT, REL-endo, а также N-концевой домен, который содержит два цистеиновых мотива ССНН, Муб-мотив и, предположительно, отвечает за связывания с ДНК-сайтом встраивания элемента (Christensen *et al.*, 2005).

HERO

Эти элементы описаны в геномах *Branchiostoma floridae* и *Danio rerio* (Kapitonov, Jurka, 2009). Их отличительной особенностью является отсутствие ССНС-мотивов в N-конце кодируемого протеина (Kojima, Fujiwara, 2004).

ARE non-LTR-ретротранспозоны

В дальнейшем, в процессе эволюции non-LTR-ретротранспозонов, REL-endo была замещена другой эндонуклеазой, которая, как считается, была приобретена из генома “хозяина” и родственна апурин/апириимидин ДНК-эндонуклеазе (ARE) (Malik *et al.*, 1999). За исключением RTE- и *Inkcap*-элементов, все другие филогенетические группы, обладающие доменом ARE, содержат ORF1 (Рис. 5). Подавляющее большинство ARE ретроэлементов не являются сайт-специфичными, а значит потенциально имеют больше возможностей для распространения в геноме хозяина. Тем не менее, исследования большого числа ARE-содержащих non-LTR-ретротранспозонов показало, что, даже не имея такой строгой сайт-специфичности, как элементы с REL-endo, эти элементы интегрируются в геном по определенным сайтам, которые, по-видимому, позволяют внедрившимся элементам избегать давления отбора (Zingler *et al.*, 2005).

Среди ARE-ретротранспозонов выделяют семнадцать филогенетических групп, одна из которых (*Dualen*) занимает промежуточное положение между REL-endo- и ARE-non-LTR-ретротранспозонами (Рис. 5).

Dualen

Одним из самых необычных суперсемейств non-LTR-ретротранспозонов являются элементы группы *Dualen*, обнаруженные в геноме зеленой водоросли *Chlamydomonas reinhardtii*. Оказалось, что кроме домена RT, эти элементы содержат как ARE так и REL-endo, а также домен протеазы (Kojima, Fujiwara, 2005). Значение присутствия сразу двух различных доменов эндонуклеаз, так же как и значение домена протеазы, остается неясным. Кроме того, элементы *Dualen* являются первыми элементами, которые приобрели домен RNH.

L1

Филогенетически суперсемейство *L1* является наиболее древним из ARE-ретротранспозонов. Некоторые элементы этого кластера строго сайт-специфичны (*Tx1* из генома лягушки *Xenopus laevis*, *Zepp* из генома зеленой водоросли *Chlorella vulgaris*) (Higashiyama *et al.*, 1997; Kojima, Fujiwara, 2004), но у большинства элементов этого суперсемейства нет ограничений на место встраивания в геном.

L1-элементы получили очень широкое распространение в геномах эукариот. К ним относятся практически все элементы человека и других млекопитающих.

Кроме того, *L1*-кластер включает в себя элементы из других групп эукариот – грибов аскомицетов (*Zorro*-элементы из генома *Candida albicans*; *Ylli*-элемент из генома *Yarrowia lipolytica*), насекомых (множество элементов из *D. melanogaster*) и растений (Wright *et al.*, 1996; Higashiyama *et al.*, 1997; Malik *et al.*, 1999; Goodwin *et al.*, 2001; Casaregola *et al.*, 2002).

Структура этих элементов хорошо изучена на примере *LINE-1*-элемента человека. Эти элементы содержат две рамки считывания. Структурные характеристики белка ORF1p *L1*-элемента человека детально описаны в разделе 1.2.1. данной работы. ORF2p содержит домены APE, RT (Рис. 5) и в некоторых случаях (*L1*-элементы растений) домен RNH.

RTE

Элементы этого суперсемейства примечательны тем, что среди них до сих пор не описаны элементы с двумя рамками считывания. Как и *L1*-элементы, *RTE*-non-LTR-ретротранспозоны получили широкое распространение и присутствуют в геномах различных растений и животных. Считается, что все известные *RTE*-элементы имеют обширные делеции с 5'-конца и именно из-за этого у них до сих пор не удалось обнаружить ORF1. Однако известно, что ORF1 элементов суперсемейства *L1* не имеет ничего общего с ORF1 элементов из других суперсемейств и, скорее всего, была приобретена ими независимо. Таким образом, возможно, что *RTE*-элементы в действительности содержат только одну рамку считывания (Zupunski *et al.*, 2001).

Еще одной отличительной особенностью этой группы является то, что один из немногих описанных случаев горизонтального переноса non-LTR-ретротранспозонов был зафиксирован именно в этой группе (Zupunski *et al.*, 2001).

Структура ORF2p *RTE*-элементов стандартная для APE-non-LTR-ретротранспозонов и содержит домены RT и APE. RNH-домена в этих non-LTR-ретротранспозонах обнаружено не было, что, скорее всего, говорит в пользу того, что он был потерян у ближайшего предка этих элементов (Novikova *et al.*, 2009).

Inkcap и *Deciever*

Эти две группы non-LTR-ретротранспозонов обнаружены в грибах. Структурно они схожи с молодыми группами non-LTR-ретротранспозонов и

отличаются количеством цистеиновых мотивов в своих ORF1p (Novikova *et al.*, 2009).

Молодые суперсемейства non-LTR-ретротранспозонов

К этим ретротранспозонам относят элементы суперсемейств *Jockey*, *Rex1*, *CR1*, *L2*, *Outcast*, *Loner*, *Tad*, *LOA*, *R1*, *I*, *ingi*, образующих одну монофилетическую группу и имеющих сходное строение (Рис. 5). Обычно эти элементы специфичны для определенной таксономической группы организмов. В частности, элементы *Jockey* были обнаружены только в геномах членистоногих (Zingler *et al.*, 2005), а *Rex1*-элементы представлены в геномах рыб (Volf *et al.*, 2000). Филогенетические группы *CR1* и *L2* получили более широкое распространение среди эукариот и описаны у многих животных (Zingler *et al.*, 2005; Novikova *et al.*, 2009). *Outcast*- и *Loner*-non-LTR-ретротранспозоны обнаружены исключительно в геноме комара *Anopheles gambiae* (Biedler, Tu, 2003). *R1*- и *LOA*-элементы присутствуют только у насекомых, а группа *Tad1* известна только для грибов (Malik *et al.*, 1999; Novikova *et al.*, 2009). *R1*-элементы являются сайт-специфичными и могут выполнять функцию формирования теломер у некоторых насекомых (Fujiwara *et al.*, 2005; Zingler *et al.*, 2005). *I*-элементы характерны для насекомых, и у дрозофилы эти элементы ответственны за явление гибридного дисгенеза (Bucheton *et al.*, 1992). И, наконец, *ingi*-элементы обнаружены только у трипаносом (Bringaud *et al.*, 2004).

1.2.8. Эволюция non-LTR-ретротранспозонов на основе рибонуклеазы H

Кроме RT, филогенетический анализ отдельных групп non-LTR-ретротранспозонов также можно проводить на основе других доменов, свойственных данным группам. Прежде всего, к таким доменам относят те, что обладают достаточно высоким уровнем консервативности (эндонуклеаза, RNH), который позволил бы оценить эволюционные расстояния между доменами даже из отдаленных групп non-LTR-ретротранспозонов. Построение филогенетических деревьев на основе этих доменов дает дополнительную информацию об эволюции, происхождении и формировании различных групп non-LTR-ретротранспозонов. Более того, в некоторых случаях удастся получить более достоверное разрешение филогенетических взаимоотношений и, к примеру, выявить события «обмена» отдельными доменами между неродственными ретротранспозонами, что было бы

невозможно на основе только лишь домена RT. Комплексный подход заключается в комбинировании доменов, на основе последовательностей которых возможно построение более достоверных филогенетических деревьев, а также использование различных подходов и методов реконструирования филогенетических взаимоотношений. Тем не менее, за исключением некоторых случаев, в целом топология филогенетических деревьев, построенных на основе других доменов, отличных от RT, оказывается сходной, что является дополнительным подтверждением единой схемы эволюции non-LTR-ретротранспозонов. К исключениям относится, например, ретроэлемент *DRE* (Schumann *et al.*, 1994), который принадлежит к *L1*-филогенетической группе на основе RT филогении. Однако, на дереве, построенном на основе APE-домена, этот элемент кластеризуется вместе с *Jockey*-элементами.

Как отмечалось выше, некоторые группы non-LTR-ретротранспозонов в процессе эволюции приобрели домен RNH. RNH обладает эндонуклеазной активностью и разрезает цепь РНК в РНК/ДНК-гибриде. Кроме как в ретротранспозонах RNH присутствует в геноме как отдельный ген, продукт которого осуществляет удаление фрагментов Оказаки в процессе репликации и участвует в репарации ДНК (Cerritelli, Crouch, 2009). Такая клеточная RNH присутствует в геномах большинства организмов, начиная от бактерий и заканчивая человеком. Среди клеточных RNH наиболее полно исследованной является RNH *Escherichia coli*, как с точки зрения функций, так и с точки зрения структурных особенностей, определяющих её ферментативную активность. Было установлено, что ключевые аминокислоты, вовлеченные в каталитический механизм RNH, являются чрезвычайно консервативными и сохраняются у RNH даже из различных царств живых организмов (Davies *et al.*, 1991). В связи с этим домен RNH часто используется наряду с доменом RT для реконструкции филогенетических взаимоотношений ретротранспозонов (Malik, 2005).

Кроме non-LTR-ретротранспозонов, домен RNH есть и у LTR-ретротранспозонов и ретровирусов. Считается, что домены RNH LTR- и non-LTR-ретротранспозонов имеют общее происхождение, в отличие от RNH ретровирусов, которая филогенетически ближе к эукариотической клеточной RNH (Malik, 2005). Предполагается, что домен RNH в процессе эволюции non-LTR-ретротранспозонов

был приобретен всего один раз. Это подтверждается филогенетическими исследованиями, указывающими на монофилетичность RNH non-LTR-ретротранспозонов (Malik, 2005).

Долгое время считалось, что домен RNH был приобретен общим предком «молодых» групп APE-содержащих non-LTR-ретротранспозонов (Malik, 2005). Этот домен обнаружен у некоторых элементов из филогенетических групп *Tad1*, *R1*, *LOA*, *I* и *ingi*. Отсутствие этого домена у других элементов из тех же самых групп, а также из близкородственных групп, таких как *CR1* и *Jockey*, приводит к выводу о его потере в этих группах (Malik *et al.*, 1999; Zingler *et al.*, 2005; Novikova *et al.*, 2009). Однако, присутствие этого домена у элементов *Dualen*, которые являются филогенетически гораздо более древними по сравнению с элементами из вышеперечисленных суперсемейств, находится в противоречии с этой теорией. Филогенетический анализ показал, что домен RNH элементов *Dualen* родственен домену RNH других non-LTR-ретроэлементов, и, соответственно, этот результат отодвигает приобретение домена RNH non-LTR-ретротранспозонами на гораздо более ранний срок. Вероятно, домен RNH был приобретен гораздо раньше, чем сформировался кластер молодых APE-содержащих non-LTR-ретротранспозонов и был потерян другими элементами (Kojima, Fujiwara, 2005). Альтернативной гипотезой является рекомбинация элементов *Dualen* с non-LTR-ретротранспозонами из уже сформировавшейся группы молодых non-LTR-ретротранспозонов, к этому времени получивших RNH, и перенос этого домена в состав элемента *Dualen*. Подобные события не являются редкостью для МЭ и легли в основу представления эволюции МЭ как *модульной* эволюции (Roberts, Mullany, 2009).

1.2.9. Модульная эволюция non-LTR-ретротранспозонов

Впервые теория модульной эволюции была сформулирована для вирусов и в её основе лежит идея о функционировании вируса как группы скоординированных независимых функциональных модулей. Главным движущим фактором эволюции таких систем являются события обмена/потери/приобретения модулей (Botstein, 1980). Позднее данная теория была распространена и на МЭ (Lerat *et al.*, 1999). Считается, что ввиду отсутствия давления отбора на МЭ (Eddy, 2012) и

подверженности мутагенным регуляторным механизмам хозяйского организма, МЭ очень быстро мутируют и теряют способность к перемещению. Обмен функциональными модулями, в свою очередь, позволяет МЭ продуцировать копии, способные на перемещение, а также создавать новые структурные варианты.

Недавние исследования показали, что бактериальные МЭ могут приобретать отдельные функциональные модули (Roberts, Mullany, 2009) и, кроме того, были выявлены многочисленные случаи приобретения отдельных функциональных единиц ретротранспозонами (Malik, Eickbush, 1999; Khazina, Weichenrieder, 2009). Таким образом, идея модульной эволюции по отношению к МЭ заключается в том, что приобретение дополнительной функциональной единицы в составе МЭ (такой, как RNH или протеаза) или обмен одного модуля на другой (замена REL-endo на ARE) может привести к повышению жизнеспособности элемента, к повышению шансов на выживание в условиях жесткой регуляции со стороны хозяйского организма и, в конечном счете, к появлению новой группы МЭ.

Исследования модульной эволюции non-LTR-ретротранспозонов выявили широкое структурное разнообразие этих элементов. В частности, изучение ORF1p non-LTR-элементов привело к обнаружению многочисленных событий приобретения новых функциональных доменов (Khazina, Weichenrieder, 2009). Приобретение дополнительного RRM в состав ORF1p привело к появлению новой группы non-LTR-ретротранспозонов растений (Heitkam, Schmidt, 2009). Таким образом, модульная эволюция является одним из основных источников структурного разнообразия non-LTR-ретротранспозонов.

1.2.10. Горизонтальный перенос non-LTR-ретротранспозонов

Кроме модульной эволюции, одним из важных факторов, определяющих разнообразие МЭ в живых организмах, является горизонтальный перенос (ГП). Термином ГП обозначают события, которые не могут быть объяснены вертикальным наследованием генетического материала и включают в себя его перемещение из одного организма в другой посредством различных векторов. ГП широко распространен у прокариот (Koonin *et al.*, 2001), в то время как у эукариот этот феномен является гораздо более редким событием, и главным объектом ГП у эукариот являются как раз МЭ (Schaack *et al.*, 2010). В случае МЭ, ГП достаточно

часто использовался для объяснения либо присутствия схожих элементов у филогенетически далеких организмов, либо несоответствия между эволюцией организмов-хозяев и МЭ (Burke *et al.*, 1993). ГП приводит к увеличению темпов пролиферации МЭ в новом геноме (Silva *et al.*, 2004), и некоторые исследователи считают его необходимым этапом жизненного цикла МЭ (Janicki *et al.*, 2011). Впервые ГП МЭ в эукариотическом организме был продемонстрирован на примере *P*-элемента дрозофилы, который оказался способен колонизировать новые геномы посредством ГП (Daniels *et al.*, 1990). Предполагается, что ГП является одним из механизмов, который позволяет МЭ избежать элиминации под действием защитных механизмов хозяйского организма путем перемещения в более «толерантную» к конкретному МЭ среду (Schaack *et al.*, 2010). Таким образом, в дополнение к модульной эволюции, ГП является вторым фундаментальным способом противодействия МЭ регуляторным механизмам хозяина: в то время как модульная эволюция позволяет быстро изменить сам МЭ и подстроить его под окружающую среду, ГП просто меняет эту среду.

Различные МЭ, тем не менее, с различной частотой используют ГП в своем жизненном цикле (Табл. 2). Наибольшее число событий ГП (103 случая) описано для ДНК-транспозонов, чуть меньше - для LTR-ретротранспозонов (97 случаев), и совсем немного для non-LTR-ретротранспозонов (14 случаев, Табл. 2) (Schaack *et al.*, 2010). Такая разница в частоте ГП объясняется отсутствием ДНК-копии non-LTR-элемента в цитоплазме в процессе его перемещения. Как было описано ранее, в ходе перемещения non-LTR-ретротранспозона в цитоплазме клетки присутствует только РНК элемента, связанная с белками в RNP. В процессе перемещения как ДНК-транспозонов, так и LTR-ретротранспозонов в цитоплазме находится гораздо более устойчивая двуцепочечная ДНК этих элементов (Schaack *et al.*, 2010). Кроме того, известно, что ДНК-транспозоны не нуждаются в дополнительных клеточных факторах для осуществления своего перемещения, в то время как LTR-ретротранспозоны, и особенно non-LTR-ретротранспозоны, во многом полагаются на кодируемые хозяйским организмом белки (см. раздел 1.2.3.) (Schaack *et al.*, 2010).

Таблица 2. Количество описанных случаев горизонтального переноса (ГП) для различных типов МЭ (согласно Schaack *et al.*, 2010).

Класс	Суперсемейство	Количество описанных событий ГП
Non-LTR-ретротранспозоны	Jockey	3
	RTE	6
	CR1	1
	Rex1	1
	Tad	2
LTR-ретротранспозоны	Ty3/gypsy	70
	Ty1/copia	16
Penelope-ретротранспозоны	Penelope	11
ДНК-транспозоны	P	28
	Tc1/mariner	34
	PIF	3
	Mutator	1
	hAT	38
	IS5	1
	PiggyBac	2
Всего:		218

Тем не менее, ряд случаев ГП non-LTR-ретротранспозонов был все же показан. Отметим, что в большинстве описанных случаев ГП non-LTR-ретротранспозонов удалось обнаружить ассоциированный с ним вектор переноса. Это также говорит в пользу того, что, в отличие от ДНК-транспозонов и LTR-ретротранспозонов, non-LTR-ретротранспозоны, ввиду озвученных выше причин, не способны на самостоятельное перемещение из одного организма в другой. Наиболее детально охарактеризован случай ГП non-LTR-ретротранспозона *Bov-B* из суперсемейства *RTE* (Zupunski *et al.*, 2001). Чрезвычайно высокий уровень

гомологии между двумя элементами *Bov-B* из геномов чешуйчатых рептилий и жвачных животных, который не вписывался в представление о строгой вертикальной эволюции этих non-LTR-ретротранспозонов, привел к выводу о его ГП (Kordis, Gubensek, 1999). Данная гипотеза впоследствии получила подтверждение, так как удалось найти потенциальный вектор ГП этих элементов между различными видами, содержащий следы встройки элемента — им оказался поксвирус, заражающий млекопитающих (Piskurek, Okada, 2007). Другой случай, в котором также удалось обнаружить потенциальный вектор переноса - это перемещение *CR1*-элемента между различными родами бабочек. В этом случае ГП *CR1*-элемента совпал с ГП ДНК-транспозона *Tcl/mariner*. Также удалось обнаружить фрагменты *CR1*, встроенные в *Tcl/mariner*, что привело к гипотезе о ГП *CR1* посредством ГП ДНК-транспозона *Tcl/mariner* (Sormacheva *et al.*, 2012). Кроме того, ГП non-LTR-ретротранспозонов также был описан между различными видами дрозофил (Sánchez-Gracia *et al.*, 2005) и грибов (Novikova *et al.*, 2009).

1.2.11. Эволюция и разнообразие non-LTR-ретротранспозонов растений

Ретротранспозоны широко распространены и многочисленны в геномах растений. К примеру, более 60% генома кукурузы состоит из ретротранспозонов различных групп. В отличие от млекопитающих, у которых non-LTR-ретротранспозоны являются основным типом МЭ в геноме, в растениях гораздо большую представленность получили LTR-ретротранспозоны (Kumar, Bennetzen, 1999) Тем не менее, и non-LTR-ретротранспозоны получили распространение, и на данный момент в геномах растений обнаружены non-LTR-ретротранспозоны относящиеся к трем суперсемействам: *Dualen*, *RTE* и *L1*.

Dualen-элементы впервые были описаны в геноме *Chlamydomonas reinhardtii*, а затем обнаружены в других зеленых водорослях (Kapitonov *et al.*, 2009) и, по-видимому, специфичны только для них. Эти non-LTR-ретротранспозоны занимают промежуточное положение между REL-endo и APE-содержащими non-LTR-ретротранспозонами и содержат оба типа эндонуклеаз. Сравнительный анализ этих эндонуклеаз показал, что ключевые аминокислоты, выполняющие каталитические функции этих доменов сохранены и это означает, что обе эти эндонуклеазы сохраняют свои функциональные свойства (Kojima,

Fujiwara, 2005). Кроме этих двух доменов элементы *Dualen* содержат в своем составе RT, RNH и протеазу (Рис. 5).

RTE-элементы были найдены в растениях при поиске гомологов элемента из генома японской медаки *Oryzias latipes*. Структура этих элементов такая же, как у представителей *RTE*-суперсемейства из других видов (см. раздел 1.2.7). В связи с чрезвычайно высоким уровнем гомологии, наблюдаемым между этими элементами из растений и рыб, было сделано предположение, что данная ситуация представляет собой еще один случай ГП non-LTR-ретротранспозонов (Zupunski *et al.*, 2001).

Из всех non-LTR-ретротранспозонов представители суперсемейства *L1* получили наиболее широкое распространение в геномах растений. Первым найденным в растениях non-LTR-элементом был *Cin4*-элемент генома кукурузы, отнесенный к *L1*-суперсемейству (Schwarz-Sommer *et al.*, 1987). Затем *L1*-элементы были обнаружены в широком круге растительных геномов: от геномов водорослей (Higashiyama *et al.*, 1997) и до геномов покрытосеменных растений (Noma *et al.*, 1999). Кроме того, было показано, что *L1*-non-LTR-ретротранспозоны активно перемещаются в геномах *Oryza sativa* и *Ipomoea batatas* (Komatsu *et al.*, 2003; Yamashita, Tahara, 2006).

Хотя четкая классификация *L1*-элементов растений отсутствует, однако, описано несколько структурных вариантов этих non-LTR-ретротранспозонов, покрывающих все разнообразие исследованных ORF1p и ORF2p (Heitkam, Schmidt, 2009; Wenke *et al.*, 2009). Во-первых, *BvL1*-элемент из генома *Beta vulgaris* филогенетически близкий к *L1*-non-LTR-ретротранспозонам из других двудольных растений содержит ORF1p, несущий цистеиновый мотив CCHC, и ORF2p с доменами RT, APE и RNH (Heitkam, Schmidt, 2009; Wenke *et al.*, 2009). Во-вторых, у гомологов *Cin4*-элемента кукурузы, по сравнению с *BvL1*-элементом, в ORF1p присутствует два CCHC-мотива, а в ORF2p отсутствует домен RNH (Schwarz-Sommer *et al.*, 1987). В-третьих, *BNR*-элемент из *B. vulgaris* также относится к *L1*-non-LTR-ретротранспозонам, но вместо цистеинового мотива CCHC в ORF1p имеет RRM домен, и в его ORF2p, также как у *Cin4*-элемента, отсутствует домен RNH (Heitkam, Schmidt, 2009). И, наконец, теломер-специфичный *Zepp*-non-LTR-ретротранспозон из водоросли *Chlorella vulgaris* имеет структуру, напоминающую

структуру *BvLI*-ретротранспозона, но не несет RNН (Higashiyama *et al.*, 1997). Интересно, что *LI*-элементы животных не содержат ни ССНС, ни RNН (Khazina, Weichenrieder, 2009) и наличие этих структур у некоторых *LI*-элементов растений, по всей видимости, является специфичным для представителей этой группы ретротранспозонов в растительных геномах.

Несмотря на то, что в целом считается, что растения развили более мощную, чем у животных систему эпигенетической регуляции активности МЭ путем метилирования ДНК, ретротранспозоны получили широкое распространение в геномах растений и в среднем их содержание достигает 10-60% генома (Kumar, Bennetzen, 1999). Для того чтобы понять какие принципы лежат в основе такого разнообразия в распространении ретротранспозонов, на первом этапе необходимо оценить масштабы этого разнообразия. Посредством сравнительного анализа распространения и разнообразия ретротранспозонов становится возможным получение 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 эволюции и распространения ретротранспозонов в этих геномах и, в том числе, роли в этих процессах таких механизмов эволюции как модульная эволюция и ГП.

Одними из наименее изученных среди ретротранспозонов растений являются *LI*-non-LTR-ретротранспозоны. Элементы из этой группы широко представлены в геномах млекопитающих и, в том числе, человека. *LI*-элементы человека к настоящему времени детально изучены, в то время как у растений описано всего несколько отдельных представителей *LI*-ретротранспозонов. Кроме получения информации об отдельных семействах, до сих пор не было предпринято попыток комплексной идентификации основных филогенетических групп *LI*-ретротранспозонов растений, что значительно упростило бы работу с информацией относительно этих элементов и расширило бы наши представления об эволюции этой группы МЭ.

Глава 2. Материалы и методы

2.1. Биоинформатический анализ распространения и эволюции *L1*-non-LTR-ретротранспозонов в геномах растений

2.1.1. Поиск и филогенетический анализ *L1*-non-LTR-ретротранспозонов в геномах растений

Для биоинформатического поиска *L1*-элементов в прочитанных последовательностях полных геномов растений и в базах данных RepBase (<http://www.girinst.org/repbase/>, Jurka *et al.*, 2005) и Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) (Табл. 3 и 4) нами был разработан алгоритм поиска целых копий non-LTR-ретротранспозонов (Рис. 6).

Таблица 3. Список исследуемых видов. *Таксономия согласно NCBI Taxonomy (<http://www.ncbi.nlm.nih.gov/taxonomy>); **Для этих видов были исследованы ретротранспозоны только из Repbase; ***Для этих видов последовательности ретротранспозонов были взяты из базы данных Genbank.

	Порядок	Семейство	Вид*	Размер генома, 10 ⁶ п.н.	Количество найденных копий доменов RT	Количество различных <i>L1</i> -ретротранспозонов
Bryopsida	Funariales	Funariaceae	<i>Physcomitrella patens</i>	480	63	2
Isoetopsida	Selaginellales	Selaginellaceae	<i>Selaginella moellendorffii</i>	110	138	5
Lilopsida	Alismatales	Hydrocharitaceae	<i>Hydrilla verticillata</i> ***	-	-	1
	Liliales	Liliaceae	<i>Lilium speciosum</i> ***	-	-	1
	Poales	Poaceae	<i>Brachypodium distachyon</i>	272	1421	11
			<i>Hordeum vulgare</i> **	-	-	2
			<i>Oryza sativa</i>	372	750	12
			<i>Setaria italica</i>	515	1423	8
			<i>Sorghum bicolor</i> **	-	-	14
			<i>Triticum aestivum</i> **	-	-	2
<i>Triticum</i>	-	-	1			

			<i>monococcum</i> **			
			<i>Zea mays</i> **	-	-	10
	Arecales	Areaceae	<i>Phoenix dactylifera</i>	650	767	15
Eudicotyledons	Lamiales	Phrymaceae	<i>Mimulus guttatus</i>	430	649	7
	Caryophyllales	Amaranthaceae	<i>Beta vulgaris</i> **	-	-	2
	Solanales	Solanaceae	<i>Solanum lycopersicum</i>	950	944	3
		Convolvulaceae	<i>Ipomoea batatas</i> ***	-	-	1
	Brassicales	Brassicaceae	<i>Arabidopsis thaliana</i>	135	277	3
		Caricaceae	<i>Carica papaya</i>	135	549	2
	Sapindales	Rutaceae	<i>Citrus clementia</i>	296	701	4
	Myrtales	Myrtaceae	<i>Eucalyptus grandis</i>	691	2813	2
	Fabales	Fabaceae	<i>Medicago truncatula</i> **	-	-	8
	Cucurbitales	Cucurbitaceae	<i>Cucumis sativus</i>	203	108	5
	Rosales	Rosaceae	<i>Malus domestica</i> **	-	-	5
		Cannabaceae	<i>Cannabis sativa</i> ***	-	-	1
	Malpighiales	Euphorbiaceae	<i>Manihot esculenta</i>	533	431	3
		Salicaceae	<i>Populus trichocarpa</i> **	-	-	2
	Vitales	Vitaceae	<i>Vitis vinifera</i>	487	2878	8
	Ranunculales	Ranunculaceae	<i>Aquilegia coerulea</i>	302	1133	9
						Всего: 149

Для этого на первом этапе для поиска регионов RT в целых геномах растений и в базах данных использовался инструмент HMMERsearch (HMMER2) из программного обеспечения UGENE (<http://ugene.unipro.ru>). Этот инструмент позволяет проводить анализ с помощью скрытых марковских моделей (HMM профилей), представляющих собой статистическую модель разыскиваемой аминокислотной последовательности (Eddy, 1998). Мы использовали уже готовый

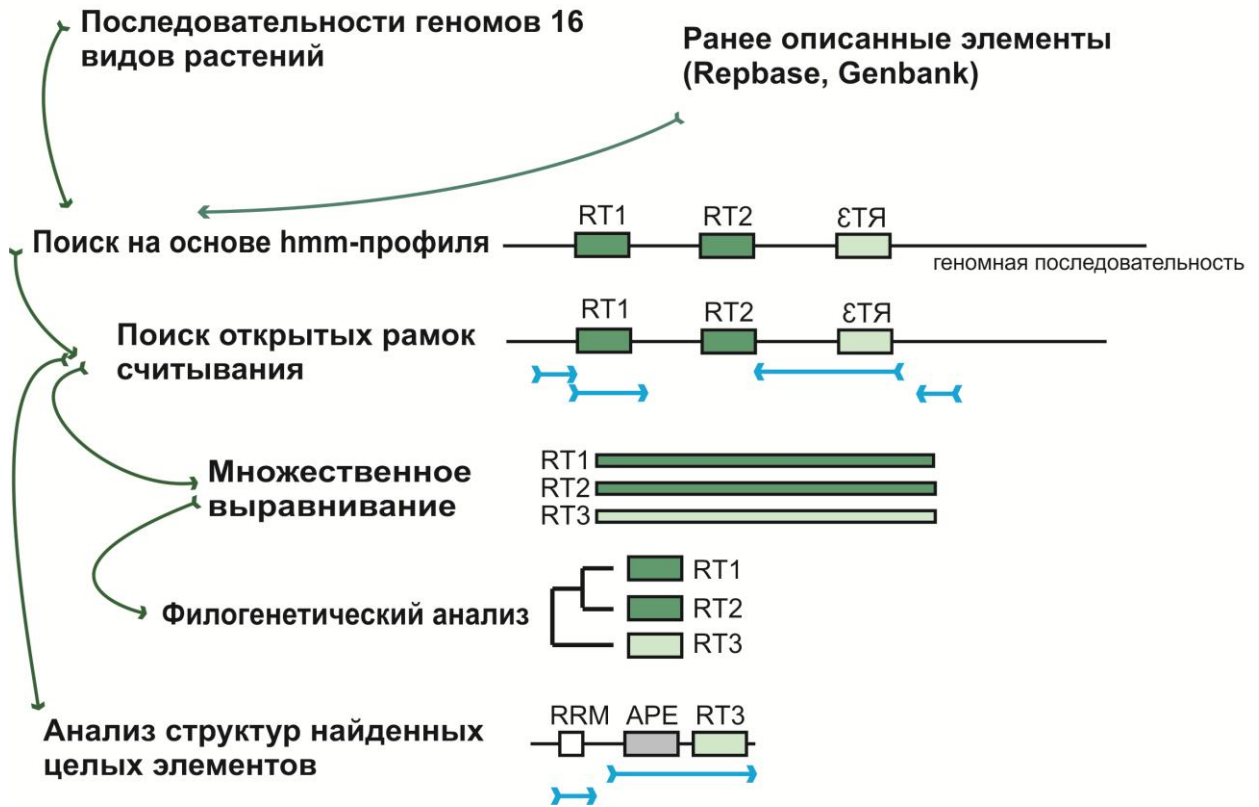


Рис. 6. Алгоритм идентификации и анализа non-LTR-ретротранспозонов. RT – обратная транскриптаза, RRM – домен узнавания РНК, APE – апуриновая/апиримидиновая эндонуклеаза.

HMM профиль для региона RT *L1*-элементов, доступный вместе с пакетом программ MGEScan-non-LTR (Rho, Tang, 2009) и отбирали результаты с показателем “score” > 100. Далее, после обнаружения RT регионов *L1*-элементов в конкретном геноме, проводилось выравнивание и филогенетический анализ найденных последовательностей с целью выявления копий одного и того же элемента. Выравнивание проводилось с использованием алгоритма MUSCLE (Edgar, 2004). Для филогенетического анализа использовался метод объединения ближайших соседей из пакета программ MEGA5 (Tamura *et al.*, 2011). Тестирование топологии дерева проводилось с помощью бутстреп метода (100-500 репликаций в зависимости о количества копий в геноме). Для каждого из элементов, выбиралась одна «мастер-копия» из всех представленных в геноме копий, которая использовалась для последующего филогенетического анализа. Всего для этого анализа было использовано 149 различных элементов: 99 из них были идентифицированы в геномах, и 50 взяты из баз данных Rebase и Genbank

(Табл. 3). После идентификации всех элементов проводился ещё один филогенетический анализ с использованием аминокислотных последовательностей RT всех найденных «мастер-копий» элементов, а также аминокислотных последовательностей RT элементов из базы данных RepBase. В рамках этого анализа построение филогенетического дерева проводилось с помощью программы PhyML, использующей для реконструкции филогенетических отношений метод максимального правдоподобия (Guindon *et al.*, 2010). Анализ проводился с использованием LG-модели относительных скоростей замен аминокислот. Для статистического подтверждения топологии дерева использовался тест aLRT (approximate Likelihood Ratio Test - тест аппроксимированного правдоподобия, Anisimova, Gascuel, 2006).

2.1.2. Идентификация структурных характеристик *L1-non-LTR*-ретротранспозонов растений

Проводилось исследование структурных характеристик найденных элементов. Для этого найденные в геноме районы RT расширялись на 5 тыс. п.н. в обе стороны. Далее, в этих последовательностях и в элементах из базы данных RepBase разыскивались открытые рамки считывания и структурные домены, характерные для non-LTR-ретротранспозонов. Для поиска структурных доменов использовалась база данных консервативных доменов в NCBI (Marchler-Bauer *et al.*, 2011), а также онлайн-ресурс детекции гомологии и предсказания структуры белков HHpred (Söding *et al.*, 2005).

2.1.3. Сравнительный анализ домена рибонуклеазы H *L1-non-LTR*-ретротранспозонов

На первом этапе проводился поиск гомологов RNH, обнаруженных в ретротранспозонах из группы *Ta11* с использованием программы blastp (<http://www.ncbi.nlm.nih.gov/>). Эти последовательности далее анализировались вместе с RNH, описанными ранее (Malik, 2005). Аминокислотные последовательности обнаруженных доменов выравнивались с использованием веб-сервера PROMALS3D (<http://prodata.swmed.edu/promals3d/>), который осуществляет выравнивание с учетом предсказанных вторичных структур анализируемых полипептидов. Филогенетический анализ RNH проводился так же как при анализе

RT с помощью программы PhyML с LG-моделью относительных скоростей замен аминокислот.

2.2. Экспериментальные исследования

2.2.1. Конструирование плазмид и очистка белка

Для конструирования рекомбинантной плазмиды, содержащей домен RNH non-LTR-ретротранспозона *Llb* из генома *Ipomoea batatas*, была использована кодон-оптимизированная для экспрессии в *Escherichia coli* версия нуклеотидной последовательности этого домена. Эта последовательность, соответствующая аминокислотам 1,205–1,366 ORF2p *Llb*-элемента (номер Genbank: BAE79382), была синтезирована с использованием программы GeneArt (www.lifetechnologies.com) и клонирована в pETM-22-вектор для экспрессии в *E. coli*, содержащий ген устойчивости к канамицину.

Полученная рекомбинантная плазида p22LlbRNHwt содержала ген, кодирующий RNH-домен *Llb*-ретротранспозона дикого типа в слиянии с thioredoxin-6×His-тегом. Этот тег содержал ген тиреодоксина, увеличивающего растворимость синтезируемого белка, и последовательность, кодирующую 6 остатков гистидина для последующей очистки и расщепления слитного белка. Другой сконструированный вектор p22LlbRNHmut содержал вариант гена RNH *Llb*-ретротранспозона, содержащий в положении 1,326 (D1326N) глицин вместо консервативного остатка аспарагиновой аминокислоты, что должно было приводить к инактивации ферментативной активности синтезируемого белка (Fisher, Pei, 1997).

Далее, клетки *E. coli* штамма BL21 трансформировали созданными рекомбинантными векторами экспрессии методом электропорации. Трансформированные клетки выращивали в среде ТВ (Terrific Broth) в присутствии 50 мкг/мл канамицина при температуре 37°C до показателей оптической плотности OD~0,6-0,7 ($\lambda = 600$ нм), после чего температуру понижали до 14°C, проводили индукцию экспрессии рекомбинантного белка с использованием 40μМ IPTG (изопропил-β-D-тиогалактопиранозид) и продолжали культивирование в течение 3 ч. Далее все процедуры проводили при температуре 4°C. Клетки собирали центрифугированием при 5000 × g, полученный осадок клеток ресуспендировали в

буфере, содержащем 100mM Tris (pH 7.5), 1M LiCl, 500 mM NaCl, 0.1mM трис(2-карбоксииэтил)фосфин (ТСЕР), 5mM имидазол и 5% глицерол, а затем лизировали с использованием ультразвука. Далее, лизированные клетки центрифугировали на $20000 \times g$ и полученный супернатант использовали для хроматографии на заряженной Ni^{2+} сефарозной колонке(GE Healthcare), используя в качестве элюента буфер, содержащий 100 mM Tris (pH 7.5), 500 mM NaCl, 0.1mM ТСЕР, 5% глицерол и 100 mM имидазол.

Поскольку рекомбинантный белок содержал 6 остатков гистидина, то он задерживался на Ni^{2+} сефарозной колонке при более высоких концентрациях имидазола, что позволило очистить его от основной массы белков *E. coli*. Затем thioredoxin-6 \times His tag удаляли с использованием протеазы Precission (предоставлена PepCore; EMBL, Heidelberg) и белок далее очищался с помощью гель-фильтрации на колонке S200 (16/60). Чистота рекомбинантного белка дополнительно оценивалась с помощью электрофореза в полиакриламидном геле в присутствии додецилсульфата натрия (SDS/PAGE).

2.2.2. Исследование активности рекомбинантной рибонуклеазы Н

Для исследования активности домена RNH non-LTR-ретротранспозона *Llb* из генома *I. batatas*, оценивалась способность экспрессированного в *E. coli* рекомбинантного полипротеина, соответствующего этому домену, разрезать РНК/ДНК гибрид *in vitro*. Для этого были синтезированы РНК-олигонуклеотид, состоящий из 20 аденинов, и комплементарный ему ДНК-олигонуклеотид (www.idtdna.com). Цепь РНК была мечена на 5'-конце фосфатом, содержащим радиоактивный фосфор P32.

Далее, с помощью гибридизации меченой цепи РНК на двукратный избыток ДНК-олигомеров приготавливался РНК/ДНК-гибрид (5'-P32-poly(rA)/poly(dT)). Для анализа расщепления цепи РНК в РНК/ДНК-гибриде *in vitro* исследуемый белок инкубировали с РНК/ДНК-гибридом в течение 1 часа при комнатной температуре в буфере, содержащем 50mM Tris·HCl (pH 7.0), 200mM NaCl, 0.1mM ТСЕР, 5% глицерол, 20 μ g/mL BSA, и 5 mM MgCl₂ или MnCl₂. Исследовался градиент концентраций белка в реакции (от 1 до 10 μ M) при постоянной концентрации субстрата (100nM).

Кроме того, для выявления специфичности расщепления доменом RNH *Llb*-ретротранспозона РНК/ДНК-гибрида, реакцию повторяли с использованием одноцепочечной меченой РНК вместо РНК/ДНК-гибрида. Все реакции останавливали с помощью добавления протеиназы К и кратковременной инкубации, и, далее, продукты реакции очищали осаждением в этаноле. Результаты анализировали на 12% ТВЕ (Tris-Borate-EDTA-urea) гель-электрофорезе и визуализировали с помощью фосфоримаджера Fuji FLA 7000.

Глава 3. Результаты

Для выявления разнообразия non-LTR-ретротранспозонов растений из суперсемейства *L1* был проведён биоинформатический анализ 29 геномов наземных растений, из которых 2 относятся к несемнным растениям, 11 к однодольным и 16 к двудольным (Табл. 3).

В рамках этого анализа был проведен поиск регионов обратной транскриптазы *L1*-элементов либо в целых геномных последовательностях, либо в базах данных RepBase и Genbank с помощью HMM-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 (Табл. 4). Далее, для представителей каждого элемента в районе найденных RT-регионов производился поиск открытых рамок считывания, и проверялось наличие консервативных структурных доменов. Кроме того, для выявления основных эволюционных групп *L1*-ретротранспозонов растений на основе найденных RT-последовательностей проводился филогенетический анализ с использованием всех найденных элементов из всех исследованных геномов, а также ранее описанных элементов. В результате данного исследования было выявлено пять основных филогенетических групп, или семейств, *L1*-элементов наземных растений: *NSLP*, *Ta11*, *BNR*, *PUR* и *Cin4* (Рис. 7). Для каждой из групп была установлена общая структура входящих в её состав элементов (Рис. 8). Идентифицированные и классифицированные в данной работе элементы могут послужить в качестве базы данных для поиска и анализа ретротранспозонов в заново отсеквенированных геномах растений, что может существенно облегчить работу с этими геномами.

Как и большинство других *L1*-элементов, оказалось, что растительные *L1*-non-LTR-ретротранспозоны всех пяти групп содержат APE-, RT- и ССНС-домены в своих ORF2p (Рис. 8). Кроме того, было установлено, что многие элементы *Ta11* группы содержат RNН-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

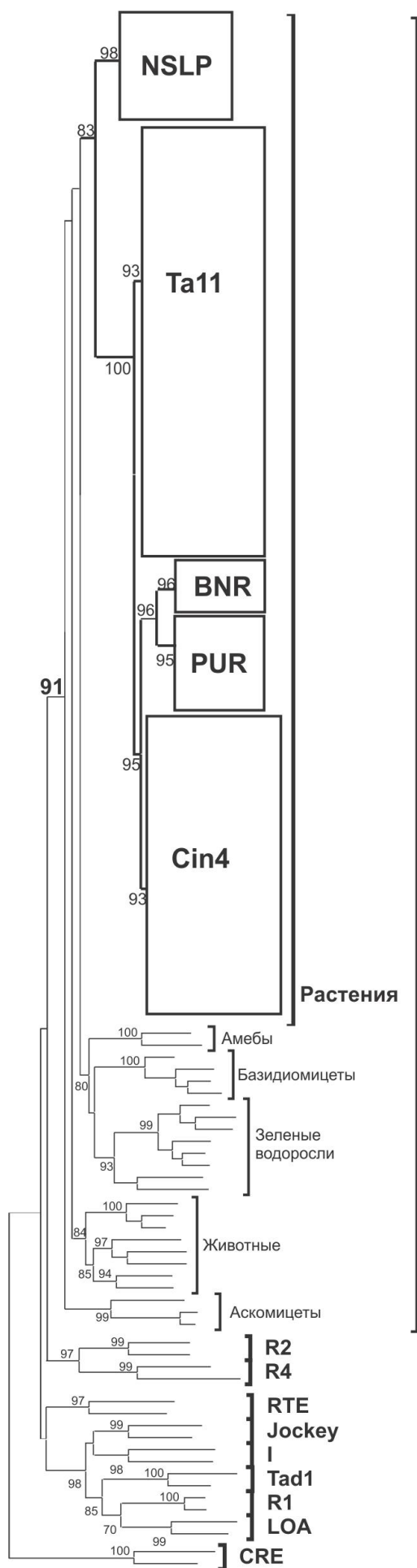


Рис. 7. Схема филогенетического дерева non-LTR-ретротранспозонов, **L1** реконструированного с помощью метода максимального правдоподобия на основе аминокислотных последовательностей обратных транскриптаз. Суперсемейства non-LTR-ретротранспозонов (*CRE*, *R2*, *R4*, *L1*, *RTE*, *Jockey*, *I*, *Tad1*, *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). Более детально дерево представлено в Приложении 1.

Таблица 4. Список ретротранспозонов, исследованных в данной работе. Источники геномов: 1 - JGI (<http://www.jgi.doe.gov/>), 2 - NCBI (<http://www.ncbi.nlm.nih.gov/>), 3 - Weill Cornell Medical College (<http://qatar-weill.cornell.edu/research/datepalmGenome/>), 4 - The International Tomato Genome Sequencing Consortium (http://solgenomics.net/organism/Solanum_lycopersicum/genome), 5 - TAIR (<http://www.arabidopsis.org/>), 6 - ASGPB (<http://asgpb.mhpc.hawaii.edu/papaya/>), 7 - International Citrus Genome Consortium (<http://www.citrusgenome.ucr.edu/>), 8 - Genoscope (<http://www.genoscope.cns.fr/spip/Vitis-vinifera-e.html>), 9 - Repbase (<http://www.girinst.org/repbase/index.html>). Элемент считался целым, если содержал узнаваемые ORF1 и ORF2.

Вид	Элемент	Источник генома	Позиция целого элемента в геноме или название элемента согласно Repbase	Длина, п.н.	Последовательность RT, использованная для реконструкции филогенетических взаимоотношений	Семейство
<i>Aquilegia coerulea</i>	AC1a	1	sc2_comp(8737255..8744064)	6810	sc2_comp(8740021..8741313)	Ta11
	AC1b	1	sc39_(1673911..1680280)	6370	sc39_(1677263..1678552)	Ta11
	AC2a	1	sc3_comp(2224800..2231967)	7168	sc3_comp(2226378..2227676)	Ta11
	AC2b	1	no intact	no	sc112_(3985..5283)	Ta11
	AC2c	1	no intact	no	sc57_comp(377826..379124)	Ta11
	AC3a	1	no intact	no	sc59_(352149..353405)	Ta11
	AC3b	1	no intact	no	sc83_(110032..111318)	Ta11
	AC3c	1	sc60_(1134739..1140639)	5901	sc60_(1137734..1139020)	Ta11
	AC4	1	sc10_compl(2207662..2214457)	6796	sc10_compl(2208917..2210200)	PUR
<i>Arabidopsis thaliana</i>	AT1	9	ATLINE1-10A	5670	ATLINE1-10A_(2805..4103)	Ta11
	AT2	9	ATLINE1-2	5814	ATLINE1-2_(2918..4213)	Ta11
	AT3	5	chr3_comp(1555417..1560693)	5277	chr3_compl(1556566..1557852)	Ta11
<i>Beta vulgaris</i>	BvL1-2	9	no intact	no	BvL1-2_(3967..5268)	Ta11
	BNR1	9	no intact	no	BNR1_(3610..4896)	BNR
<i>Brachypodium distachyon</i>	BD1a	1	Bd1_comp(42630246..42636851)	6606	Bd1_comp(42631311..42632597)	Cin4
	BD1b	1	no intact	no	Bd1_comp(45071375..45072661)	Cin4
	BD1c	1	no intact	no	Bd2_(53166536..53167822)	Cin4
	BD1d	1	no intact	no	Bd4_(1519534..1520823)	Cin4
	BD2a	1	no intact	no	Bd4_comp(2205174..2206460)	Cin4

	BD2b	1	no intact	no	Bd3_comp(41372254..41373513)	Cin4
	BD3a	1	no intact	no	Bd5_comp(16762125..16763411)	Cin4
	BD3b	1	no intact	no	Bd4_(19695386..19696672)	Cin4
	BD4a	1	no intact	no	Bd1_(21154322..21155629)	Ta11
	BD4b	1	no intact	no	Bd2_(52626443..52627840)	Ta11
	BD4c	1	no intact	no	Bd1_(40572839..40574134)	Ta11
<i>Cannabis sativa</i>	LINE-CS	2	no intact	no	AB013908_(2186..3472)	Ta11
<i>Carica papaya</i>	CP1	6	no intact	no	sc32_(901730..903010)	PUR
	CP2	6	no intact	no	sc263_compl(40027..41283)	Ta11
<i>Citrus clementia</i>	CC1	7	sc21_(3420704..3426728)	6025	sc21_(3423904..3425199)	Ta11
	CC2	7	sc19_(10..6808)	6770	sc19_(3835..5124)	Ta11
	CC3	7	sc147_(179891..185312)	5422	sc147_(182509..183810)	Ta11
	CC4	7	sc3_(5566745..5572694)	5950	sc3_(5569837..5571132)	Ta11
<i>Cucumis sativus</i>	CS1	1	no intact	no	sc.927_compl(86375..87658)	PUR
	CS2	1	no intact	no	sc.2242_(26974..28254)	Ta11
	CS3	1	no intact	no	sc.1106_compl(30921..32120)	PUR
	CS4	1	no intact	no	sc.1305_compl(301511..302686)	PUR
	CS5	1	no intact	no	sc.1221_(556365..557621)	PUR
<i>Eucalyptus grandis</i>	EG1	1	sc2_comp(20649023..20654942)	5920	sc2_compl(20650233..20651519)	Ta11
	EG2	1	sc2_(6081689..6087998)	6310	sc2_(6085125..6086423)	Ta11
<i>Hordeum vulgare</i>	TREP52	9	no intact	no	TREP52_(4237..5523)	Cin4
	TREP20	9	no intact	no	TREP20_(1335..2606)	Ta11
<i>Hydrilla verticillata</i>	no name	2	no intact	no	DV216735_compl(1..663)	Cin4
<i>Ipomoea batatas</i>	Llb	2	AB231841_(155..6486)	6332	AB231841_(3640..4935)	Ta11
<i>Lilium speciosum</i>	Del2	2	no intact	no	Z17425_(1696...2508)	Cin4
<i>Malus domestica</i>	L1-15 Mad	9	no intact	no	L1-15 Mad_(779..2077)	Ta11
	L1-13 Mad	9	no intact	no	L1-13 Mad_(2611.3900)	Ta11
	L1-11 Mad	9	L1-11 Mad	5843	L1-11 Mad_(2981..4276)	Ta11
	L1-12 Mad	9	no intact	no	L1-12 Mad_(2914..4203)	Ta11
	L1-1 Mad	9	L1-1 Mad	6675	L1-1 Mad_(4160..5446)	PUR
<i>Manihot esculenta</i>	ME1	1	no intact	no	sc.4782_(177194..178414)	Ta11
	ME2	1	no intact	no	sc.5497_(138350..139645)	Ta11
	ME3	1	no intact	no	sc.2306_(132371..133636)	Ta11
<i>Medicago truncatula</i>	LINE1H	9	no intact	no	LINE1H MT_(1054..2355)	Ta11
	MT					
	SHALINE1	9	SHALINE16 MT	5758	SHALINE16	Ta11
	6 MT				MT_(2892..4193)	
	SHALINE1	9	no intact	no	SHALINE10	Ta11
	0 MT				MT_(2580..3878)	

	SHALINE5 MT	9	no intact	no	SHALINE5 MT_(4473..5768)	Ta11
	SHALINE1 2 MT	9	SHALINE12 MT	5914	SHALINE12 MT_(3056..4348)	Ta11
	SHALINE7 MT	9	no intact	no	SHALINE7 MT _(2693..3982)	Ta11
	SHALINE1 1 MT	9	no intact	no	SHALINE11 MT_(1037..2329)	BNR
	LINE1A MT	9	no intact	no	LINE1A MT_(749..2038)	BNR
<i>Mimulus guttatus</i>	MG1	1	no intact	no	sc248_(5280..6575)	Ta11
	MG2	1	no intact	no	sc369_(53062..54348)	Ta11
	MG3	1	no intact	no	sc220_(84969..86264)	Ta11
	MG4	1	no intact	no	sc17_compl(294331..29562 6)	Ta11
	MG5	1	no intact	no	sc74_(334556..335827)	Ta11
	MG6	1	no intact	no	sc78_(515490..516776)	Ta11
	MG7	1	sc34_compl(303570..311308)	7739	sc34_compl(305662..30695 1)	Ta11
<i>Oryza sativa</i>	OS1a	9	LINE-05_OS	6739	LINE-05_OS_(4361..5647)	Cin4
	OS1b	2	no intact	no	chr12_compl(18879445..18 880737)	Cin4
	OS1c	2	no intact	no	chr7_(25180149..25181435)	Cin4
	OS1d	2	no intact	no	chr12_compl(23882786..23 884072)	Cin4
	OS1e	2	no intact	no	chr11_(23387969..2338925 5)	Cin4
	OS2	9	LINE-7_OS	7090	LINE-7_OS_(4736..6010)	Cin4
	OS3	9	no intact	no	RILN8_OS_(2737..4023)	Cin4
	OS4	2	no intact	no	chr5_compl(23612656..236 13939)	Cin4
	OS5	9	OSLINE1-3	6101	OSLINE1-3_(3479..4762)	Cin4
	OS6	9	OSLINE1-2	7128	OSLINE1-2_(4655..5941)	Cin4
	OS7	2	chr10_(10925263..10931688)	6426	chr10_(10928761..1093006 8)	Ta11
OS8	9	OSLINE1_2	6317	OSLINE1_2_(3422..4720)	Ta11	
<i>Phoenix dactylifera</i>	PD1	3	no intact	no	PDK_30s1148721_(2650..38 97)	Ta11
	PD2	3	no intact	no	PDK_30s902441_(8908..101 10)	Ta11
	PD3	3	no intact	no	PDK_30s1101531_comp(86 65..9993)	Ta11
	PD4	3	no intact	no	PDK_30s6550992_comp(27 968..29041)	Ta11
	PD5	3	no intact	no	PDK_30s1010751_(30113..3 1285)	Ta11
	PD6a	3	no intact	no	PDK_30s973551_(31346..32 509)	Cin4
	PD6b	3	no intact	no	PDK_30s9031_41compl(563 1..6797)	Cin4

	PD6c	3	no intact	no	PDK_30s658771_compl(134 17..14667)	Cin4
	PD7	3	no intact	no	PDK_30s804151_compl(171 02..18391)	Ta11
	PD8	3	PDK_30s1060841_(8851..15231)	6381	PDK_30s1060841_(12375..1 3670)	Ta11
	PD9	3	no intact	no	PDK_30s791021_(2..1054)	Ta11
	PD10	3	no intact	no	PDK_30s1126011_(11028..1 2278)	Ta11
	PD11	3	no intact	no	PDK_30s1195861_compl(3.. 842)	Ta11
	PD12	3	no intact	no	PDK_30s1020141_compl(36 835..38025)	Ta11
	PD13	3	no intact	no	PDK_30s963311_(2767..372 9)	Ta11
<i>Physcomitrel la patens</i>	PP1	1	no intact	no	sc175_(27625..28836)	NSLP
	PP2	1	sc127_compl(136674..144400)	7727	sc127_compl(139155..1404 35)	NSLP
<i>Populus trichocarpa</i>	L1-1 PTr	9	L1-1 PTr	6095	L1-1 PTr_(3585..4871)	Ta11
	L1-2 PTr	9	L1-2 PTr	6167	L1-2 PTr_(3662..4945)	BNR
<i>Selaginella moellendorfi i</i>	SM1a	1	sc78_compl(240908..249177)	8270	sc78_compl(242854..24416 4)	NSLP
	SM1b	1	sc76_compl(618206..625856)	7651	sc76_compl(619807..62111 7)	NSLP
	SM1c	1	sc2_compl(4452410..4460142)	7733	sc2_compl(4454378..44556 88)	NSLP
	SM2a	1	sc153_compl(64802..72341)	7540	sc153_compl(66672..67988)	NSLP
	SM2b	1	sc103_(70199..78068)	7870	sc103_(74870..76183)	NSLP
<i>Setaria italica</i>	SI1	1	PROV_1_(33354788..33362351)	7564	PROV_1(33359999..333612 85)	Cin4
	SI2	1	no intact	no	PROV_3_comp(44504320.4 4505606)	Cin4
	SI3	1	no intact	no	PROV_1_(9692309..969360 1)	Cin4
	SI4	1	PROV_2_comp(40195086..40201787)	6702	PROV_2_comp(40196227..4 0197513)	Cin4
	SI5	1	no intact	no	PROV_2_comp(4776645..47 77919)	Cin4
	SI6	1	PROV_5_comp(39116340..39124037)	7903	PROV_5_comp(39117456..3 9118742)	Cin4
	SI7	1	PROV_4_comp(21919686..21925467)	5782	PROV_4_comp(21921220.2 1922515)	Ta11
	SI8	1	PROV_8_(20517477..20525038)	7562	PROV_8_(20522468..20523 754)	Cin4
<i>Solanum lycopersicum</i>	SL1	4	ch01_comp(90218152..90225449)	7298	ch01_comp(90220063..902 21355)	Ta11
	SL2	4	no intact	no	ch00_compl(10162107..101 63399)	Ta11
	SL3	4	no intact	no	ch00_(12020151..12021437	Ta11

)	
<i>Sorghum bicolor</i>	LINE1-54 SBi	9	no intact	no	LINE1-54 SBi_(2..1294)	Ta11
	LINE1-60	9	LINE1-60 SBi	5576	LINE1-60 SBi_(2738..4027)	Ta11
	LINE1-62B	9	LINE1-62B SBi	7283	LINE1-62B SBi _(4317..5624)	Ta11
	LINE1-51	9	LINE1-51 SBi	7234	LINE1-51 SBi_(4756..6039)	Cin4
	LINE1-27	9	no intact	no	LINE1-27 SBi_(2910..4193)	Cin4
	LINE1-73	9	no intact	no	LINE1-73 SBi_(166..1452)	Cin4
	LINE1-8	9	LINE1-8 SBi	6649	LINE1-8 SBi_(4288..5574)	Cin4
	LINE1-53	9	LINE1-53 SBi	7629	LINE1-53 SBi_(5218..6510)	Cin4
	LINE1-40	9	no intact	no	LINE1-40 SBi_(4752..6041)	Cin4
	LINE1-16	9	LINE1-16 SBi	6892	LINE1-16 SBi_(4557..5825)	Cin4
	L1-1	9	L1-1 SBi	6698	L1-1 SBi_(4306..5592)	Cin4
	LINE1-38	9	no intact	no	LINE1-38 SBi_(4755..6041)	Cin4
	LINE1-58	9	LINE1-58 SBi	7672	LINE1-58 SBi_(5174..6460)	Cin4
	LINE1-36	9	LINE1-36 SBi	6762	LINE1-36 SBi_(4274..5560)	Cin4
<i>Triticum aestivum</i>	TAR2 TA	9	no intact	no	TAR2 TA_(580..1809)	Ta11
	KARIN TA	9	no intact	no	KARIN TA_(548..1834)	Cin4
<i>Triticum monococcum</i>	ISABELLE TM	9	no intact	no	ISABELLE TM_(5878..7167)	Cin4
<i>Vitis vinifera</i>	VV1	8	chr6_(6728195..6734544)	6350	chr6_(6732054..6733340)	PUR
	VV2	8	no intact	no	chr2_(14041764..14043062)	Ta11
	VV3	8	no intact	no	chr1_(45595..46770)	Ta11
	VV4	8	no intact	no	chr8_compl(11243781..112 45073)	PUR
	VV5	8	chr16_compl(4508413..4514510)	6098	chr16_compl(4509568..451 0860)	PUR
	VV6	8	chr1_compl(21309887..21316295)	6409	chr1_compl(21311223..213 12515)	PUR
	VV7	8	chr5_(20524966..20531062)	6097	chr5_(20528594..20529886)	PUR
	VV8	8	chr19_(19453058..19459151)	6094	chr19_(19456769..1945806 1)	PUR
<i>Zea mays</i>	LINE1-20	9	LINE1-20 ZM	7117	LINE1-20 ZM_(4718..6004)	Cin4
	LINE1-41	9	no intact	no	LINE1-41 ZM_(2..1177)	Cin4
	LINE1-30	9	no intact	no	_(909..2192)	Cin4
	LINE1-37	9	no intact	no	LINE1-37 ZM_(4541..5833)	Cin4
	Cin4	9	CIN4E ZM	6822	CIN4E ZM_(4484..5776)	Cin4
	LINE1-39	9	LINE1-39 ZM	6690	LINE1-39 ZM_(4287..5573)	Cin4
	LINE1-33	9	LINE1-33 ZM	5829	LINE1-33 ZM_(3473..4759)	Cin4
	LINE1-13	9	no intact	no	LINE1-13 ZM_(1282..2568)	Cin4
	LINE1-32	9	LINE1-32 ZM	6297	LINE1-32 ZM_(3635..4921)	Cin4
	LINE1-40	9	LINE1-40 ZM	7257	LINE1-40 ZM_(4572..5858)	Cin4

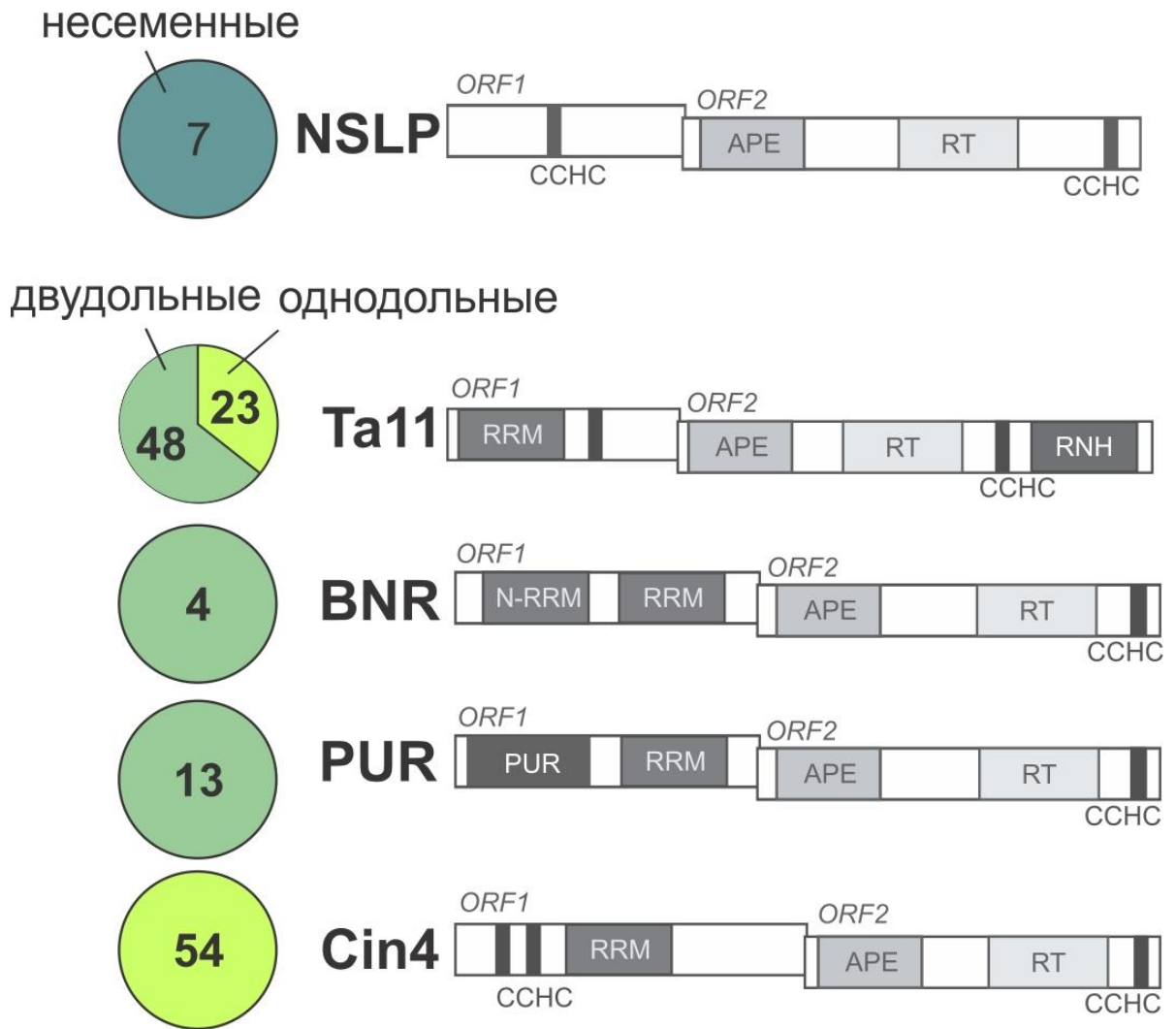


Рис. 8. Структура и распространение элементов из пяти семейств *L1* non-LTR-ретротранспозонов растений (Смышляев, Блинов, 2011; Smyshlyayev *et al.*, 2013, с модификациями). Количество и распространение элементов идентифицированных в нашем исследовании показано в круговых графиках. Цифры отражают количество элементов, относящихся к семейству в выделенной группе растений (темнозеленый – несеменные, зеленый – двудольные, светлозеленый – однодольные растения). ORF – открытая рамка считывания, RRM – домен узнавания РНК, CCHC – 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.

ORF1p и содержат в своем составе различные РНК-связывающие домены, такие как RRM, CCHC и обнаруженный впервые в ретротранспозонах PUR-домен (более подробно описан в разделе 3.3). Наличие в ORF1p разнообразных РНК-

связывающих доменов подтверждает предположение о том, что белок ORF1 необходим для формирования рибонуклеопротеиновой частицы элемента.

3.1. *NSLP*-элементы специфичны для несеменных растений

Эти элементы, как следует из названия (*NSLP* – NonSeed Plant-Specific), были обнаружены в геномах несеменных растений, мха *Physcomitrella patens* и плауна *Selaginella moellendorffii*. Всего в этих геномах было обнаружено 7 элементов, которые на основе филогенетического анализа легли в эту монофилетическую группу (Рис. 9). В ORF1p этих ретротранспозонов не удалось обнаружить никаких РНК-связывающих доменов, кроме ССНС-мотива (Рис. 8). ORF2p *NSLP*-элементов стандартный, содержит АРЕ-, RT- и ССНС-домены и структурно сходен с ORF2p остальных растительных *L1*-элементов, кроме ретротранспозонов из группы *Ta11*.

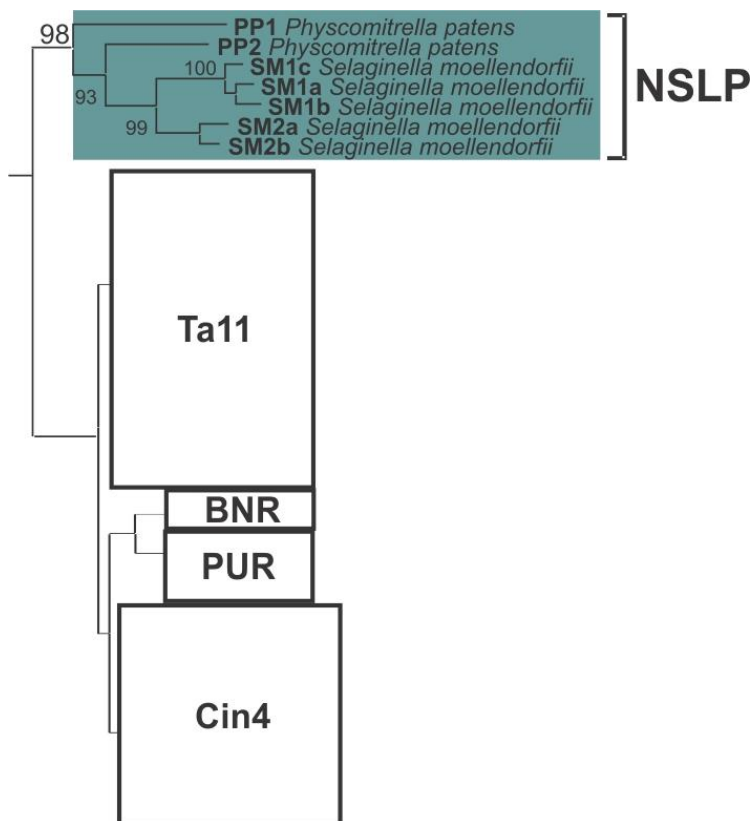


Рис. 9. Элементы *NSLP* на схеме филогенетического дерева *L1*-non-LTR-ретротранспозонов растений. Всего было обнаружено 7 элементов, относящихся к этому семейству. Все эти элементы найдены в геномах несеменных растений (выделены темнозеленым цветом).

3.2. *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 злаковых однодольных в свободном доступе, очевидно, что для проверки этой гипотезы требуются дополнительные данные.

Ранее была проведена работа, в которой были описаны основные характеристики представителей этой группы (Heitkam *et al.*, 2009). Особенности элементов этой группы является присутствие высоко консервативного N-концевого RRM (N-RRM), центрального домена и отсутствие цистеинового мотива в ORF1. RRM является одним из самых распространенных РНК-связывающих мотивов в non-LTR-ретротранспозонах и характеризуется консервативной вторичной структурой.

Наш сравнительный анализ и анализ структур с помощью HHPRED выявил, что несмотря на отсутствие заметной гомологии центрального домена и N-RRM на основании аминокислотной последовательности, центральный домен структурно сходен с N-RRM (Рис. 12) и другими описанными RRM. Таким образом, был сделано предположение о том, что центральный мотив – это второй RRM домен, имеющий отличное от N-RRM происхождение. В пользу этого предположения говорит наличие многочисленных RRM в ORF1p некоторых non-LTR-ретротранспозонов (Капелинская и др., 2011).

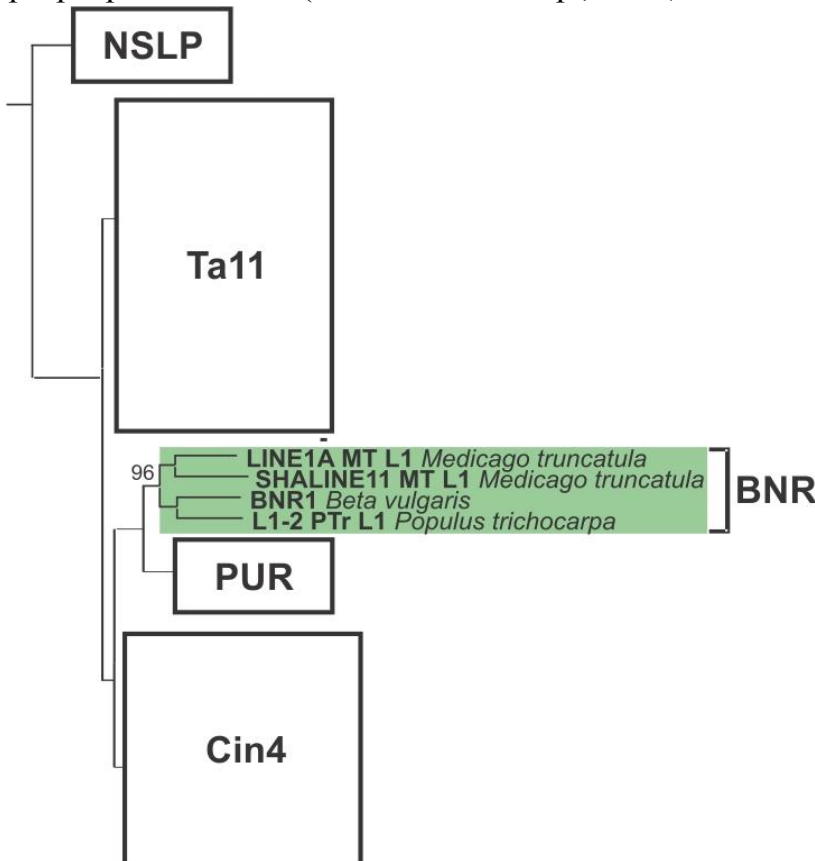


Рис. 10. Элементы *BNR* на схеме филогенетического дерева *L1*-non-LTR-ретротранспозонов растений. Всего было обнаружено 4 элемента, относящихся к этому семейству. Все эти элементы найдены в геномах двудольных растений (выделены зеленым цветом).

ORF2 у *BNR*-элементов стандартная для *LI*-элементов и содержит домены APE и RT, а также цистеиновый мотив CCHC в С-конце. В проведенном исследовании мы также обнаружили, что представленные в базе данных Repbase элементы *SHALINE11* и *LINE1A* из генома *Medicago truncatula* также относятся к этой группе (Рис. 10).

3.3. *PUR*-элементы специфичны для двудольных растений и содержат *PUR*-домен

В рамках проведенного исследования нам удалось выявить новую группу non-LTR-ретротранспозонов растений, также как и *BNR*-элементы специфичную для геномов двудольных растений (Рис. 11). Эта группа получила название *PUR*. 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).

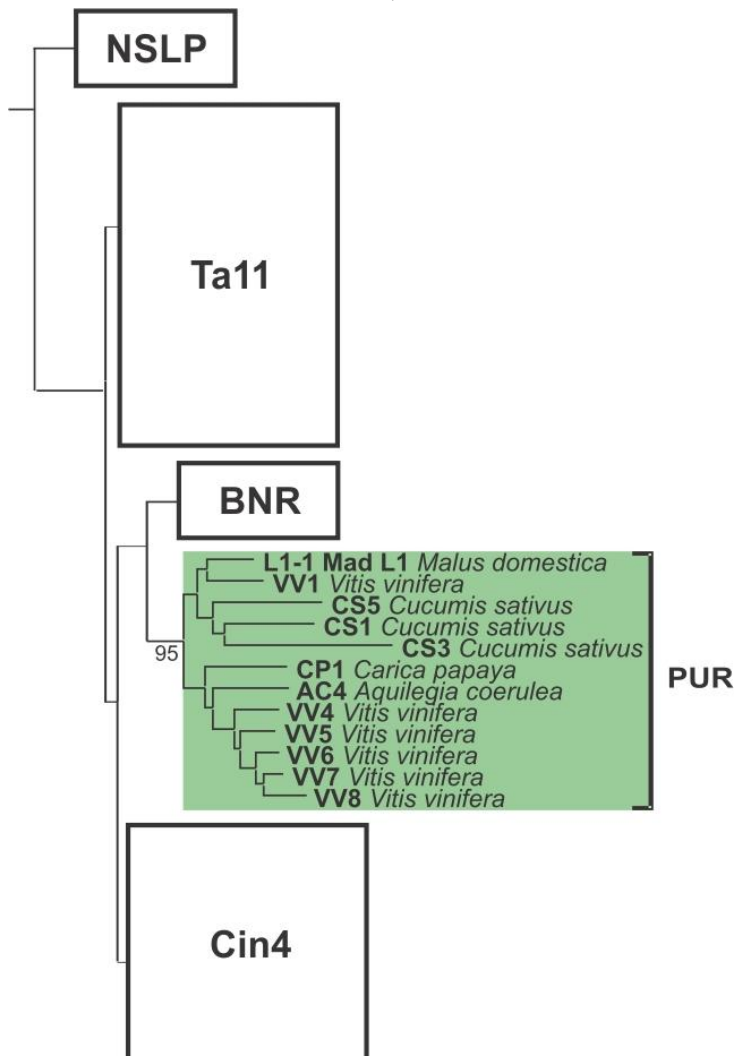


Рис. 11. Элементы *PUR* на схеме филогенетического дерева *LI*-non-LTR-ретротранспозонов растений. Всего было обнаружено 13 элементов, относящихся к этому семейству. Все эти элементы найдены в геномах двудольных растений (выделены зеленым цветом).

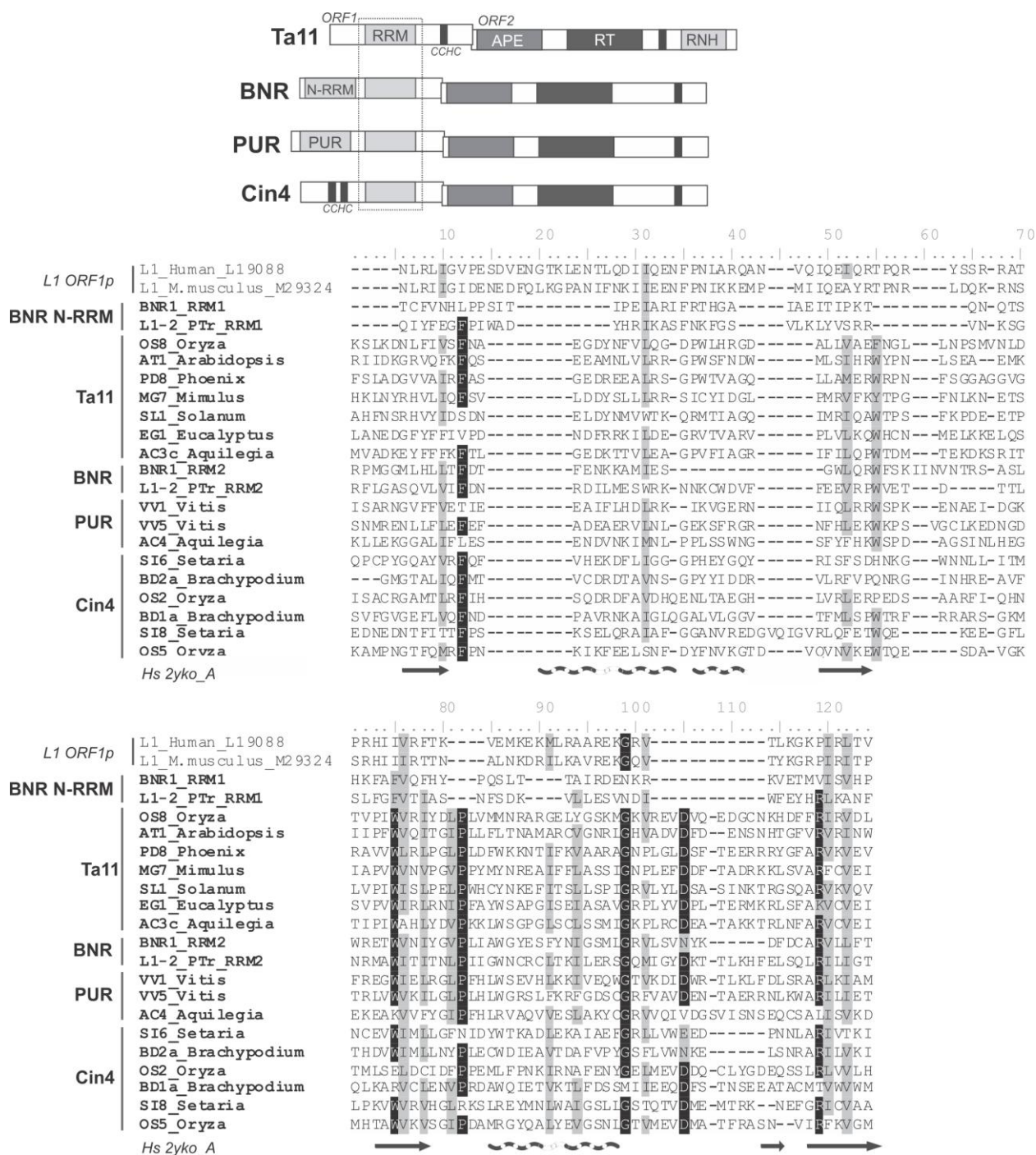


Рис. 12. Выравнивание аминокислотных последовательностей домена узнавания РНК (RRM – RNA recognition motif) *L1*-non-LTR-ретротранспозонов из различных групп вместе с RRM *L1*-элемента млекопитающих (Smyshlyayeva *et al.*, 2013, с модификациями). Положение RRM в исследуемых элементах показано сверху от выравнивания пунктирным прямоугольником. Выравнивание производилось с использованием веб-сервера PROMALS3D, который осуществляет выравнивание с использованием предсказанных вторичных структур белка. Консервативные позиции доменов выделены черным и серым. Вторичная структура RRM *L1*-

ретротранспозона человека [Protein Data Bank (PDB) ID код 2LDY_A] показана внизу выравнивания: альфа-спирали в виде волнистых линий и бета-цепи в виде стрелок. 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.

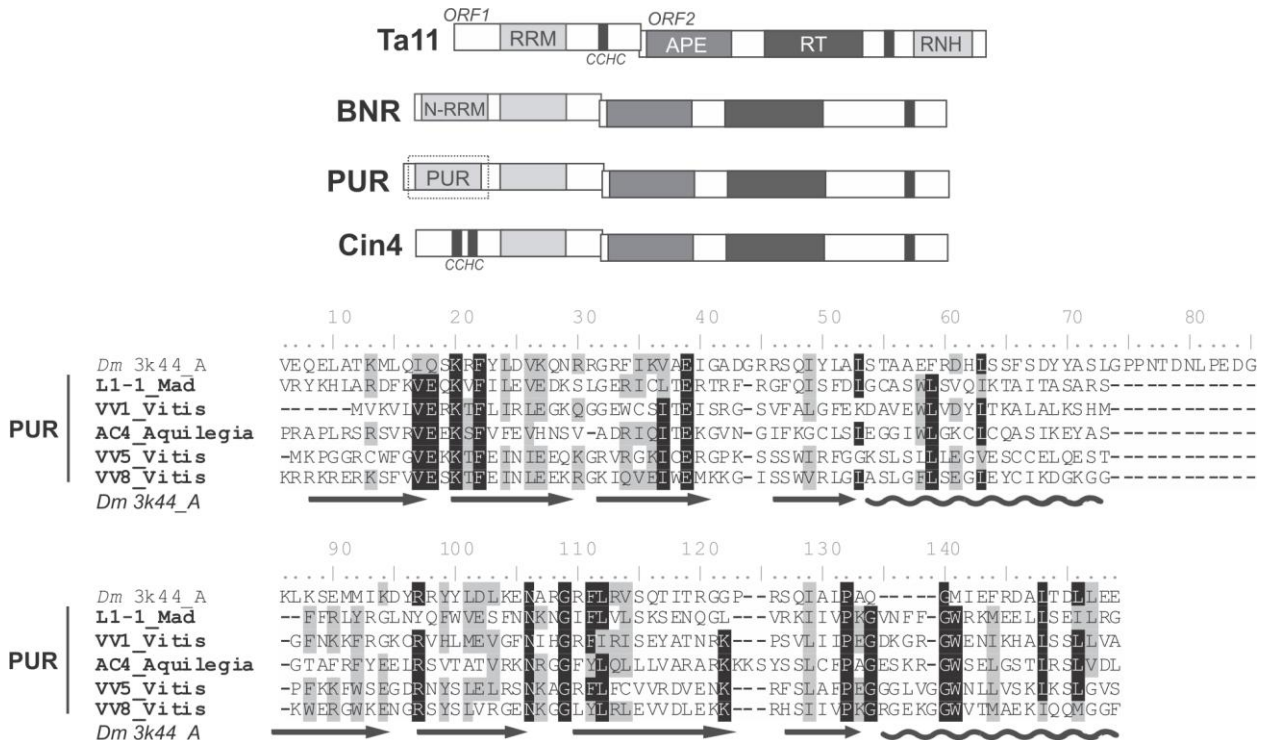


Рис. 13. Выравнивание аминокислотных последовательностей PUR-домена (пурин-богатого), вместе с доменом PUR-α дрозофилы (Smyshlyaev *et al.*, 2013, с модификациями). Консервативные позиции доменов выделены черным и серым. Вторичная структура PUR-α дрозофилы (PDB ID код 3k44_A) показана внизу выравнивания: альфа-спирали в виде волнистых линий и бета-цепи в виде стрелок. 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.

Тринадцать элементов, объединяющихся в монофилетичную группу *PUR*-элементов на основе эволюции RT домена (Рис. 11), филогенетически близки к группе *BNR*-элементов (Рис. 7), но структурно от них отличаются (Рис. 8). 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 ORF1p *PUR*-элементов был обнаружен консервативный домен, структурно похожий на так называемый пурин-богатый (*PUR* - purine-rich) домен (Рис. 13).

PUR-домены являются высококонсервативными 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, на уровне первичной аминокислотной последовательности имеющий сходство с центральным RRM группы *BNR*-элементов, что говорит в пользу общего происхождения этих RRM в *L1*-элементах растений. Особенно многочисленны *PUR*-элементы в геноме винограда (*Vitis vinifera*). Более 2 тысяч *L1*-non-LTR-ретротранспозонов было найдено в этом геноме (Табл. 3), большая часть которых относится именно к *PUR*-элементам (Табл. 4).

3.4. *Cin4*-элементы специфичны для геномов однодольных растений

Элементы из этой группы были обнаружены в геномах всех злаковых растений, исследованных в данной работе. Кроме того, к этой группе оказались отнесены элемент *del2* из *Lilium speciosum*, а также элементы из *Hydrilla verticillata* и *Phoenix dactylifera*, не относящихся к злакам, что говорит в пользу того, что *Cin4*-элементы произошли до появления злаковых растений.

В результате филогенетического анализа было выявлено несколько подгрупп элементов *Cin4*, элементы из которых, впрочем, имеют сходную структуру и, поэтому, были объединены в одну большую группу (Рис. 14). При анализе структуры ORF1 элементов из группы *Cin4* был выявлен новый тип ORF1 *L1*-элементов растений (Рис. 8). ORF1 элементов *Cin4* имеет RRM сходный с RRM *BNR*- и *PUR*-элементов (Рис. 12), но в отличие от них, элементы *Cin4*, вдобавок к этому RRM, несут два цистеиновых CCHC-мотива в первой рамке, которые располагаются в N-конце относительно RRM. Кроме того, ORF1 элементов *Cin4* продолжительна, её средняя длина составляет более 3 тыс. п.н. Для сравнения, ORF1 человека всего около 1 тыс. п.н. в длину (GenBank: U93570).

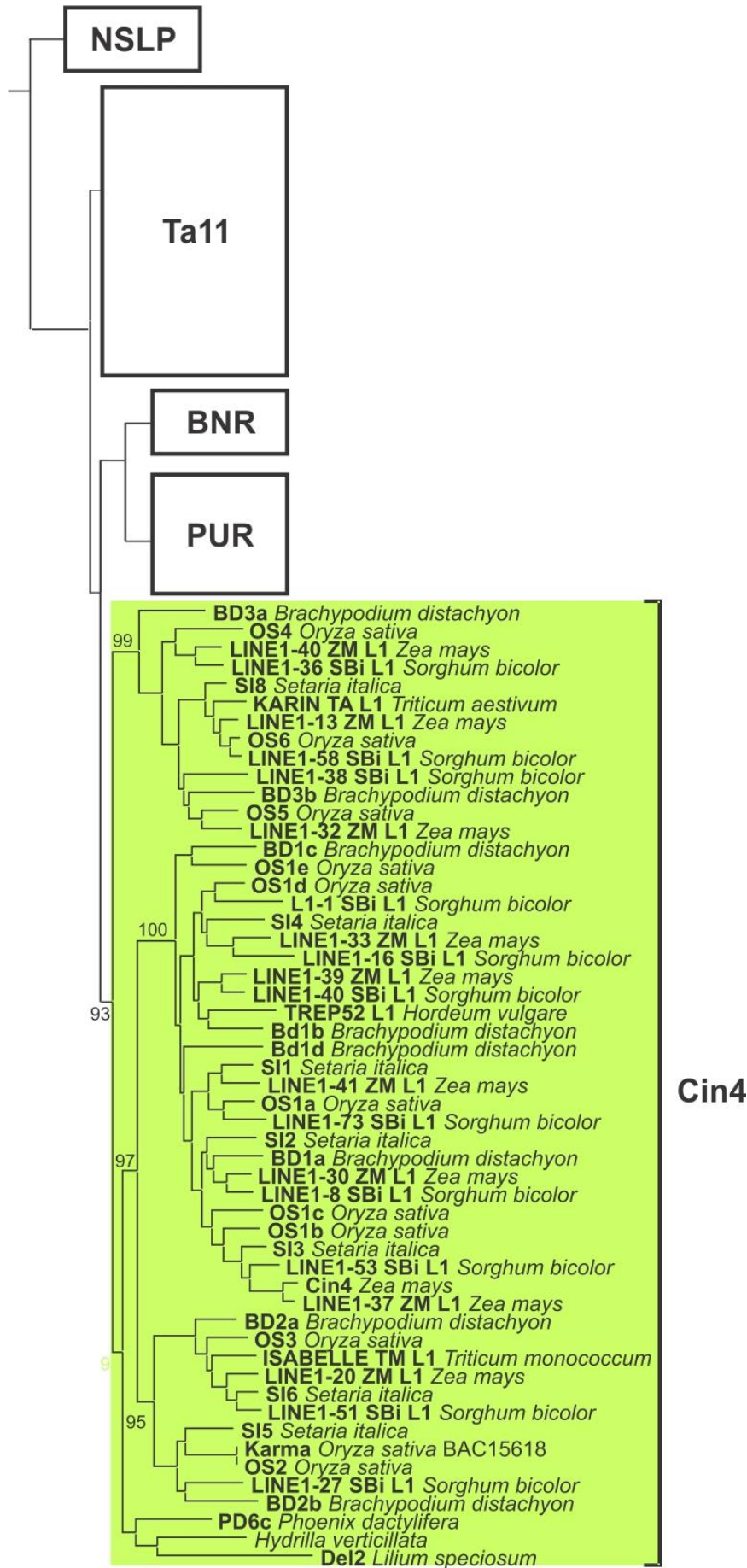


Рис.14. Элементы *Cin4* на схеме филогенетического дерева *L1*-non-LTR-ретротранспозонов растений. Всего было обнаружено 52 элемента, относящихся к этому семейству. Все эти элементы найдены в геномах однодольных растений (выделены светлозеленым цветом).

Большая часть ORF1, за исключением двух ССНС-мотивов и RRM, значительно варьирует по размеру между элементами группы, что позволяет предположить её функциональность в качестве стабилизатора пространственной структуры синтезируемого с ORF1 белка.

3.5. *Tall*-элементы присутствуют в геномах как однодольных, так и двудольных растений и содержат домен рибонуклеазы H

Элементы группы *Tall* оказались наиболее распространенными у растений, по сравнению с другими найденными группами. Они были обнаружены в большинстве из исследованных геномов. Из тех видов, для которых анализировалась полная геномная последовательность, только для *Zea mays* не удалось обнаружить копий *Tall*-элементов. Тем не менее, возможно, что геном *Z. mays* также содержит *Tall*-элементы, но в разрушенном виде и в меньшем количестве, из-за чего они пока не были обнаружены используемым нами алгоритмом. 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). ORF1p *Tall*-элементов содержит RRM, за которым следует цистеиновый мотив ССНС (Рис. 8). Сравнительный анализ аминокислотных последовательностей RRM всех обнаруженных RRM *L1*-ретротранспозонов растений выявил высокий уровень консервативности RRM *BNR*-, *PUR*- и *Tall*-элементов (Рис. 12), 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. RRM и цистеиновый мотив являются единственными консервативными регионами ORF1p *Tall*-элементов, остальная часть этого белка значительно варьирует между элементами, что, скорее всего, говорит об отсутствии ее прямого функционального значения.

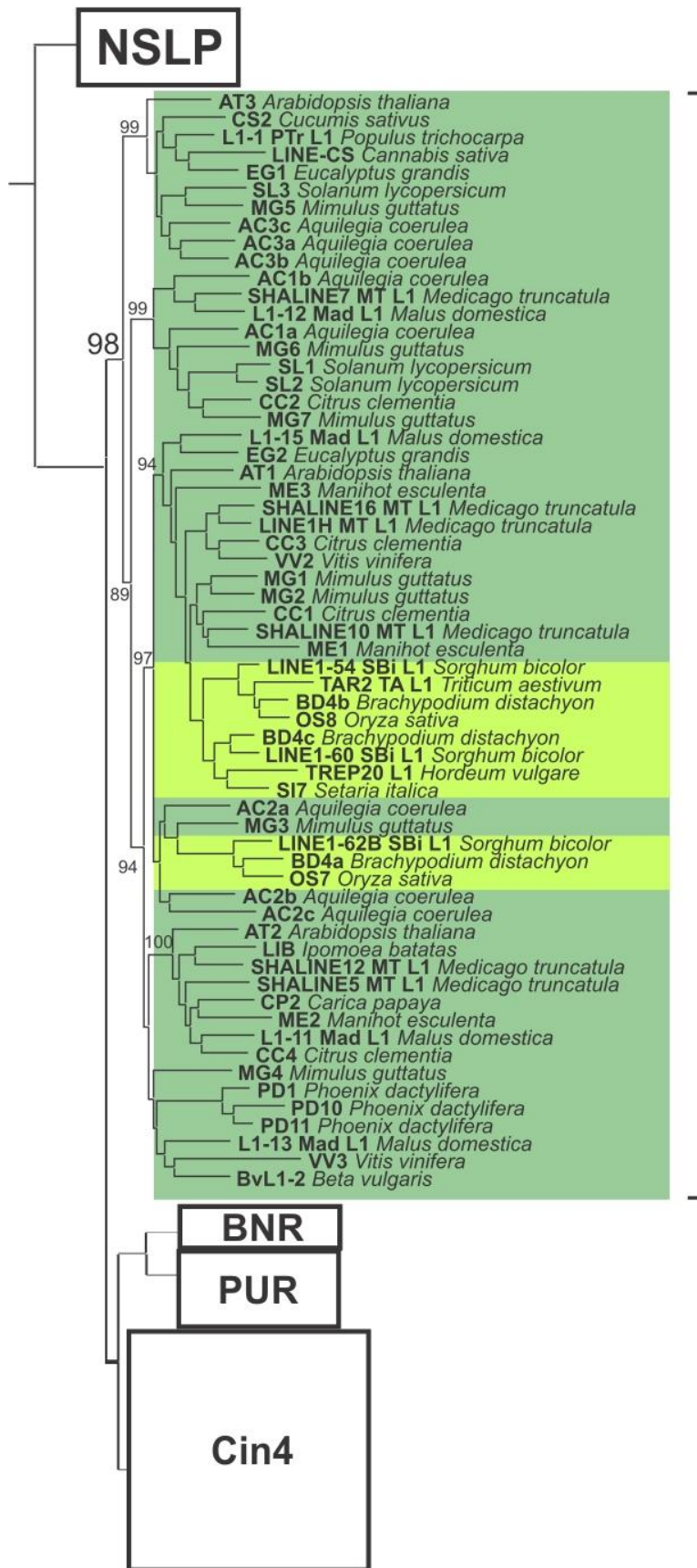


Рис. 15. Элементы *Ta11* на схеме филогенетического дерева *L1*-non-LTR-ретротранспозонов растений. Всего было обнаружено 80 элементов, относящихся к этому семейству. Эти элементы найдены в геномах как однодольных (выделены светлозеленым цветом) так и двудольных растений (выделены зеленым цветом).

ORF2p, кроме стандартных для *L1*-элементов APE- и RT- доменов и CCHC-мотива, кодирует ещё и белок RNH (Рис. 8). Ни одна из других ранее описанных групп *L1*-элементов эукариот не кодирует этот белок. Более того, появление RNH в

non-LTR-ретротранспозонах считается более поздним событием: впервые белок RNH у non-LTR-ретротранспозонов появляется в филогенетически более молодом семействе *Jockey* (Malik, 2005). Поэтому, мы предполагаем, что RNH была приобретена в группе *Ta11*-элементов растений независимо. В пользу этого предположения говорит и тот факт, что у остальных non-LTR-элементов, обладающих RNH, ССНС-мотив располагается после RNH (Рис. 5), в то время как у *Ta11*-элементов в С-конце располагается именно RNH (Рис. 8). Тем не менее, наличие RNH характерно не для всех элементов группы *Ta11*. Некоторые элементы, такие как *BD4a* из генома *Brachypodium distachyon*, *OS7* из генома *Oryza sativa* и *LINE1-62B* из генома *Sorghum bicolor*, не имеют RNH. Скорее всего, этот домен был потерян у этих элементов, как это происходило в эволюции других non-LTR-ретротранспозонов (Novikova *et al.*, 2009).

3.6. Домен рибонуклеазы Н *Ta11*-ретротранспозонов растений

Гомологи рибонуклеазы Н (RNH) 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) RNH. Фермент RNH осуществляет специфический гидролиз РНК-цепи в РНК/ДНК-гибриде, и роль клеточной RNH в клетке заключается в участии в репликации генома (Cerritelli, Crouch, 2009). На основании аминокислотных последовательностей и наличия определенных функциональных сайтов все клеточные RNH разделяют на бактериальные, эукариотические и архейные (Marchler-Bauer *et al.*, 2011). Тем не менее, названия этих RNH не всегда совпадают с реальным распространением RNH в живых организмах. Например, так называемая архейная RNH была обнаружена и в бактериальных, и в растительных геномах (Ohtani *et al.*, 2004).

Как уже отмечалось выше, в ретровирусах и ретротранспозонах RNH обычно кодируется в составе одного полипротеина вместе с доменом RT. В данном случае функция RNH заключается в удалении матрицы РНК вируса или ретроэлемента после процесса обратной транскрипции. В зависимости от происхождения, RT-ассоциированные RNH подразделяются на RNH 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. Этим non-LTR-ретротранспозоны отличаются от LTR-ретротранспозонов и ретровирусов, у которых обратная транскрипция происходит в цитоплазме, а значит им необходимо иметь собственную RNН.

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. Эти элементы появились значительно позже *L1*-элементов, поэтому появление RNН у *L1*-элементов должно иметь совсем другое происхождение. В этом контексте обнаружение RNН в эволюционно достаточно древнем суперсемействе *L1*-элементов является неожиданной находкой. Более того, как видно на Рис. 5, во всех элементах из «молодых» групп APE-содержащих non-LTR-ретротранспозонов домен RNН находится перед С-концевым ССНС-доменом, в то время как в *Tall*-non-LTR-ретротранспозонах из суперсемейства *L1*-RNН находится в самом конце ORF2p после ССНС (Рис. 8).

В связи с таким необычным положением домена RNН как в составе ORF2p так и в эволюционном контексте, для выяснения деталей его происхождения мы провели сравнительный филогенетический анализ домена RNН *Tall*-ретротранспозонов. Для этого было произведено множественное выравнивание RNН из различных ретротранспозонов, ретровирусов, а так же клеточных RNН (Рис. 16).

Проведенный сравнительный анализ выявил, что RNН *Tall*-ретротранспозонов не содержит консервативной аминокислоты His, характерной для многих RNН, включая клеточные прокариотическую и эукариотическую RNН, и, что самое удивительное, включая RNН других non-LTR-ретротранспозонов.

Вместо этого RNН *Tall*-ретротранспозонов содержит в этой позиции аминокислоту Arg, что сближает ее с клеточной RNН архей (Рис. 16). Консервативность этой позиции RNН объясняется тем, что аминокислотные остатки Arg или His находятся в каталитическом центре RNН.

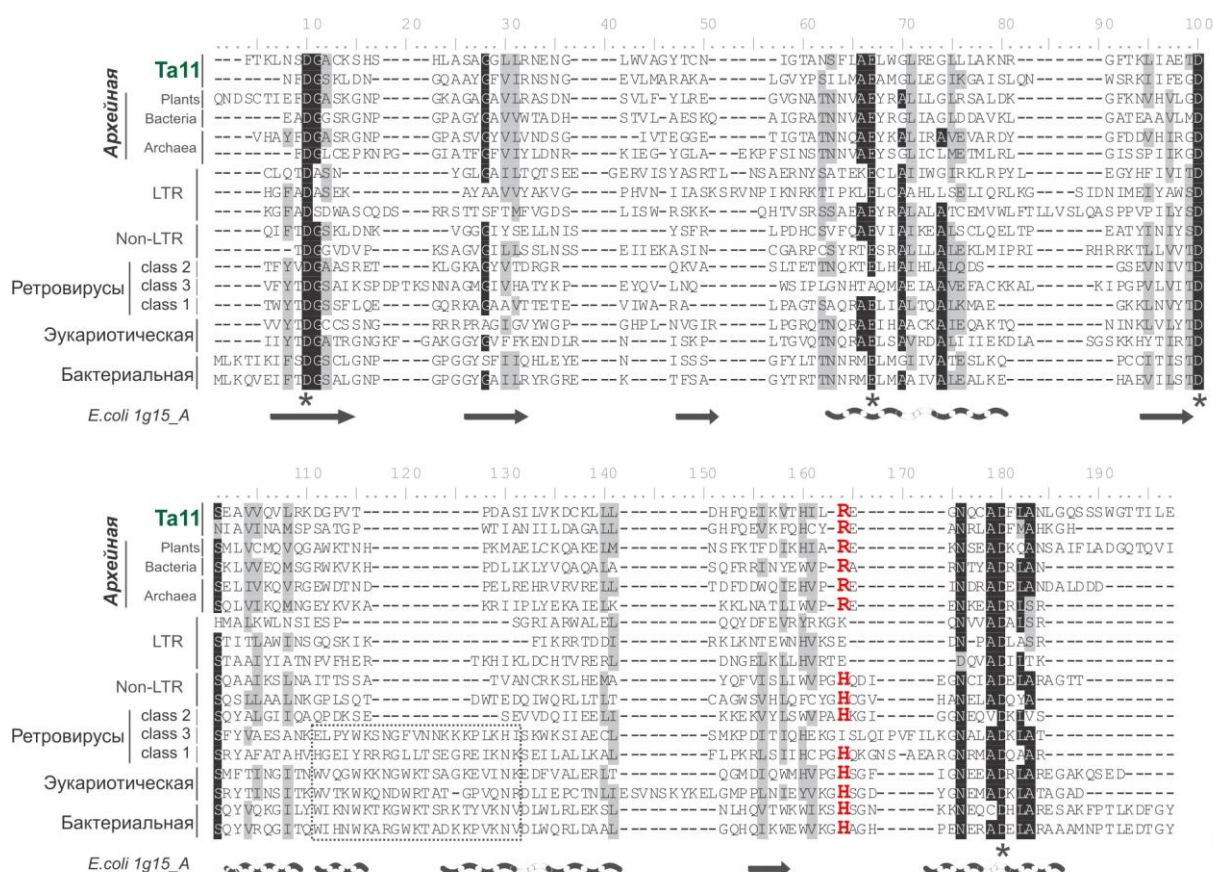


Рис. 16. Выравнивание аминокислотных последовательностей различных типов RNH (Smyshlyaev *et al.*, 2013, с модификациями). Названия трех типов клеточных RNH (не ассоциированных с RT) выделены жирным курсивом. Название *Ta11*-семейства выделено жирным шрифтом и зеленым цветом. “Основной выступ” обозначен пунктирной рамкой. Аминокислоты в консервативной позиции 164 выравнивания, варьирующей между архейной и бактериальной/эукариотической RNH, обозначены увеличенным шрифтом и красным цветом (R - Arg, H - His). Уровень консервативности остальных аминокислот показан черным и серым фоном. Вторичная структура RNH *E. coli* (Protein Data Bank ID код 1G15_A) показана к низу от выравнивания: альфа-спирали в виде волнистых линий и бета цепи в виде стрелок.

Подчеркнем, что RNH LTR-ретротранспозонов не содержит ни His, ни Arg в этой позиции и считается, что из-за этого она каталитически менее активна, чем все другие типы RNH. Предполагается, что преимущество такой менее активной RNH заключается в том, что позволяет сохранить РНК-праймер в районе полипуринового тракта, с которого начинается синтез второй цепи ДНК

ретротранспозона (Malik, Eickbush, 2001). Non-LTR-ретротранспозоны тем не менее полагаются на наличие высокоактивной RNH (будь то клеточная RNH или кодируемая самим ретротранспозоном), поскольку считается, что по крайней мере для некоторых из них их РНК матрица должна быть полностью разрушена прежде, чем начнется синтез второй цепи ДНК в процессе интеграции транспозона (Malik, 2005). Эти различия могут объяснить более серьезное давление отбора на наличие более активной RNH со всеми консервативными позициями аминокислот в RNH non-LTR-ретротранспозонов по сравнению с RNH LTR-ретротранспозонов. С другой стороны, ретровирусы в той же степени, что и LTR-ретротранспозоны полагаются на сохранение праймера на полипуриновом тракте, но, тем не менее, их RNH содержит His в позиции активного центра (Рис. 16). В этом случае предполагается, что так называемый «соединительный домен» (Рис. 1), отсутствующий у LTR-ретротранспозонов, модулирует активность RNH (Malik, 2005).

Таким образом, проведенный сравнительный анализ показал, что RNH *Tall*-элементов больше похожа на архейную RNH, чем на RNH других non-LTR-ретротранспозонов. Интересно, что присутствие гомологов архейной RNH было ранее выявлено в хромосомах растений (Ohtani *et al.*, 2004), а также в их хлоропластных и митохондриальных геномах (Stoppel, Meurer, 2012).

С целью выяснения происхождения архее-подобного домена RNH в *Tall*-ретротранспозонах, мы провели анализ гомологичных последовательностей с использованием архейной RNH в качестве запроса для поиска blastp в базе данных GenBank. В результате мы обнаружили несколько архее-подобных RNH в организмах из различных таксономических групп, начиная от прокариот и заканчивая растениями. Однако, нам не удалось найти гомологов архейной RNH в геномах грибов и животных. Используя аминокислотные последовательности найденных нами и описанных ранее архее-подобных RNH вместе с последовательностями эукариотических, бактериальных и кодируемых ретротранспозонами RNHs, мы реконструировали филогенетическое дерево RNH (Рис. 17).

Проведенный анализ продемонстрировал, что RNH *Tall*-ретротранспозонов и клеточная архейная RNH объединяются в монофилетическую группу. RNHs LTR-

ретротранспозонов формируют прилегающую к ней группу, которая внутри разделяется на ветви, соответствующие основным группам LTR-ретротранспозонов. Из проведенного анализа следует еще один вывод о том, что RNH *Tall*-ретротранспозонов не родственна RNH других non-LTR-ретротранспозонов из всех суперсемейств. Последние филогенетически гораздо ближе к RNH ретровирусов класса 2. RNH ретровирусов классов 1 и 3 филогенетически отделены от RNH других ретровирусов и обладают уникальной особенностью, отсутствующей у других RT-ассоциированных RNHs, – так называемым “основным выступом”, который у присутствует у эукариотических и бактериальных RNH и предположительно ответственен у них за связывание с РНК/ДНК гибридом (Tadokoro, Kanaya, 2009).

Остается неясным, почему “основной выступ” отсутствует у ретротранспозонов и ретровирусов класса 2. Возможно, у этих элементов роль связывания с гибридом выполняют другие структуры полипротеина. Кроме того проведенный анализ выявил потенциальные источники этих RNH: в соответствии с филогенией, ретровирусы класса 2 скорее всего взяли RNH у non-LTR-ретротранспозонов, в то время как ретровирусы классов 1 и 3 независимо от ретровирусов класса 2 “захватили” клеточную эукариотическую или прокариотическую RNH, имеющую “основной выступ”.

Отличие RNH ретровирусов от RNH LTR-ретротранспозонов, которое не вписывалось в строгую вертикальную эволюцию этих структур было выявлено ранее, и, более того, было сформулирована гипотеза о повторном появлении RNH в ретровирусах (Malik, Eickbush, 2001; Malik, 2005). Более того, вариант развития событий, при котором наиболее ранний общий предок приобрел собственный RNH-домен из генома хозяина, предполагает, что этот RNH-домен не имеет никакого отношения к RNH других non-LTR-ретротранспозонов. Этот факт хорошо согласуется с реконструированной нами филогенией RNH.

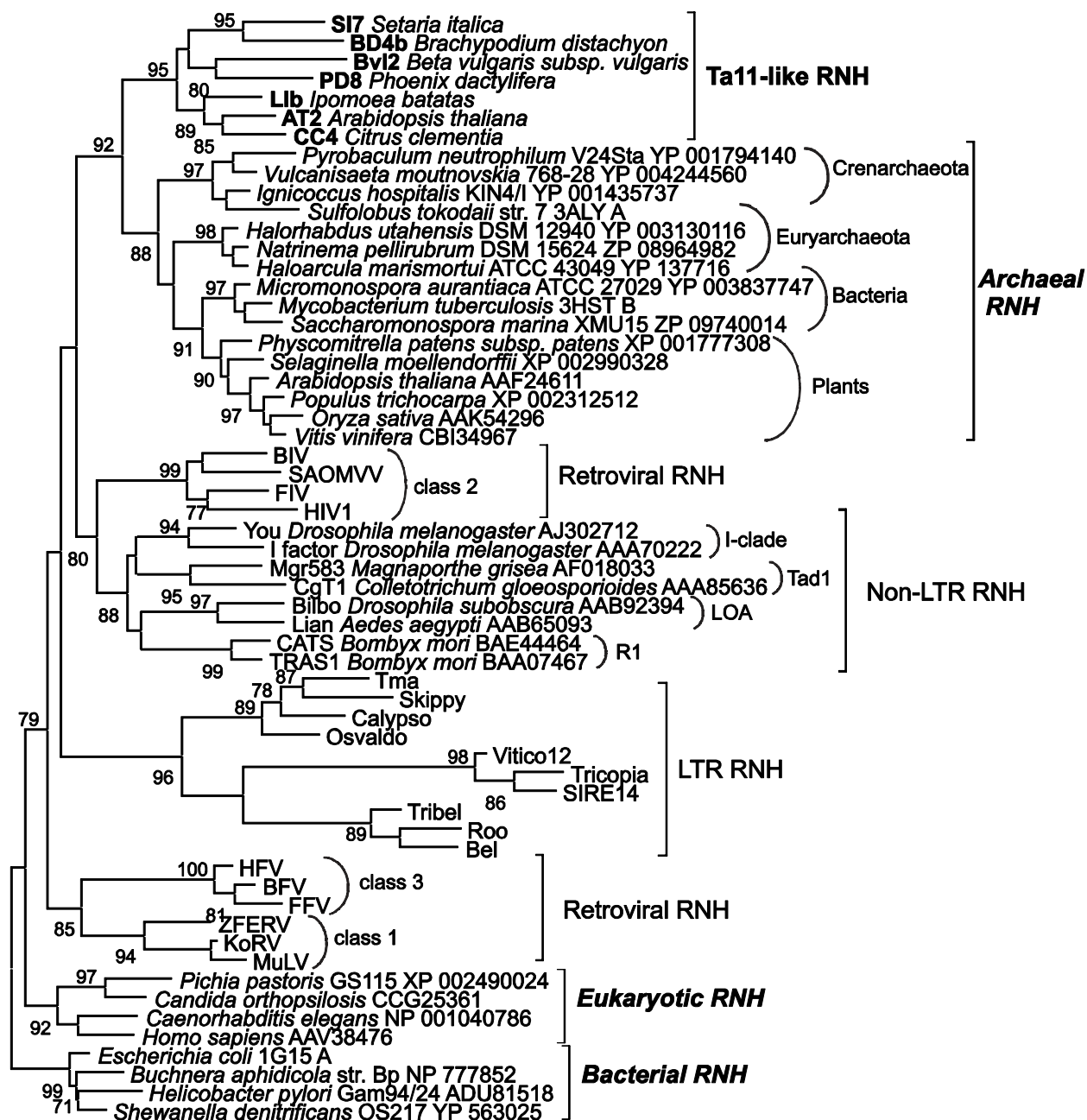


Рис. 17. Филогенетическое дерево различных типов RNH, построенное методом максимального правдоподобия (Smyshlyaev *et al.*, 2013). На дереве присутствует информация о номерах последовательностей, использованных для анализа, в базе данных GenBank. Названия трех типов клеточных RNH (не ассоциированных с RT) выделены жирным курсивом.

Еще одним важным результатом нашего филогенетического и сравнительного анализов является обнаружение высокой степени гомологии между 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, 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*. Гены, которые подвергаются ГП между видами, могут принести пользу организму, в который они переносятся, только в том случае, если их активность сохраняется после переноса. 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 ORF2p, 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.

3.7. 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 (LlbRNH) из генома *Ipomoea batatas*. Этот ретротранспозон был выбран потому, что для него было показано активное перемещение в геноме (Yamashita, Tahara, 2006). Это, скорее всего, означает, что все структурные домены кодируемых белков *Llb*-элемента функционально активны. Для проведения этого анализа рекомбинантный полипротеин LlbRNH был экспрессирован в *E. coli*.

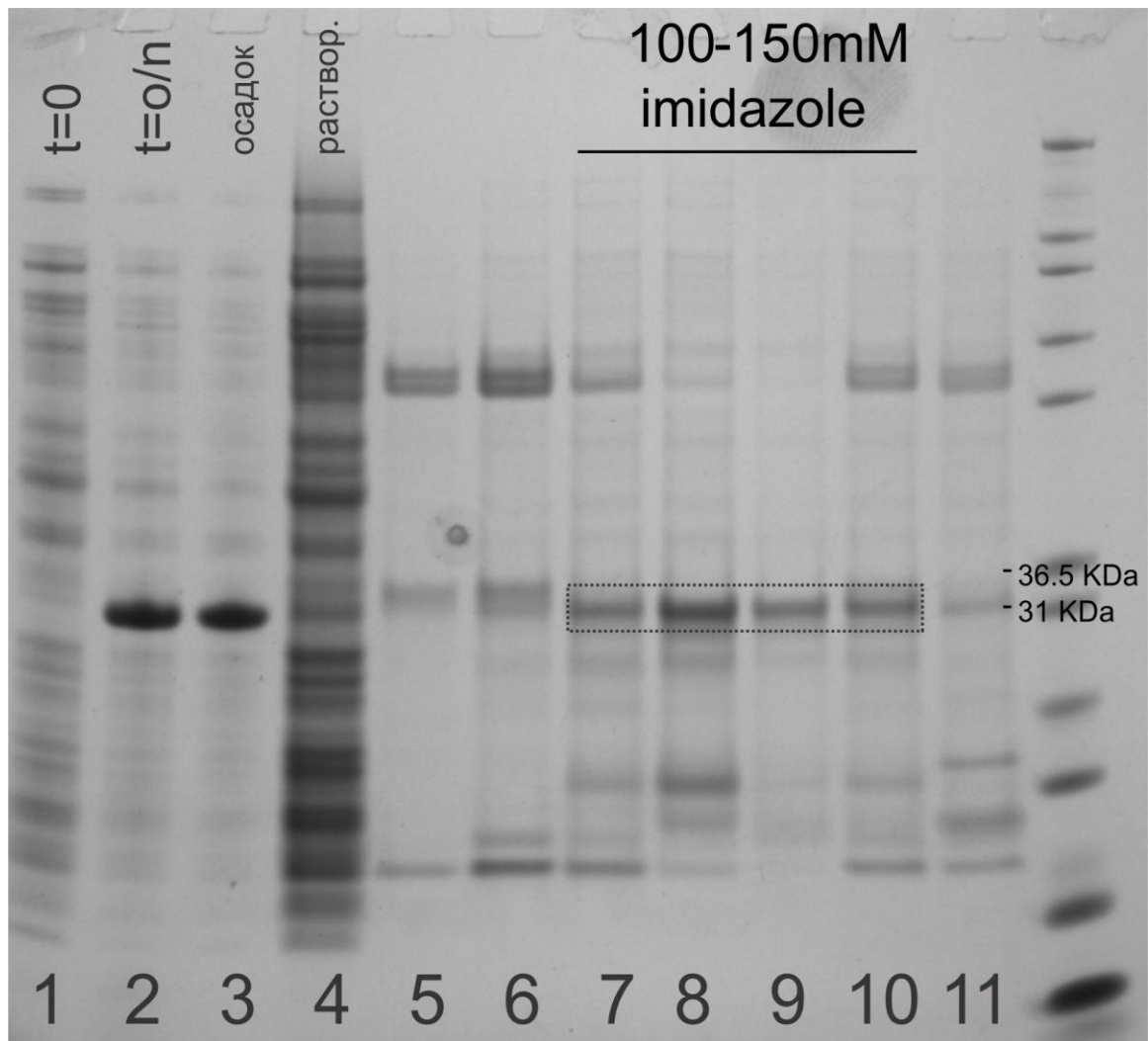


Рис. 18. SDS/PAGE различных стадий экспрессии рекомбинантного белка LlbRNH. Ряды 1 и 2 – экстракт *E. coli* без экспрессии и с экспрессией белка (o/n - overnight) соответственно. Ряды 3 и 4 – белки *E. coli* после лизирования, остающиеся в осадке и в растворе соответственно. Ряды 5-11 – образцы, взятые при различных концентрациях имидазола в качестве элюента. Наибольшая часть экспрессируемого белка элюировалась при концентрации имидазола 100-150 mM. Последний ряд – маркер длины.

На Рис. 18 приведено изображение SDS/PAGE различных стадий экспрессии белка, из которого видно, что белок экспрессировался (ряд 2, Рис. 18), но его большая часть шла в осадок при выделении (ряд 3, Рис. 18). Тем не менее, часть белка находилась в растворимой фазе в 100-150mM градиенте имидазола и далее очищалась в соответствии с протоколом, указанным в разделе 2.2.1. материалов и методов. Гель-фильтрация и SDS/PAGE очищенного белка подтвердили его

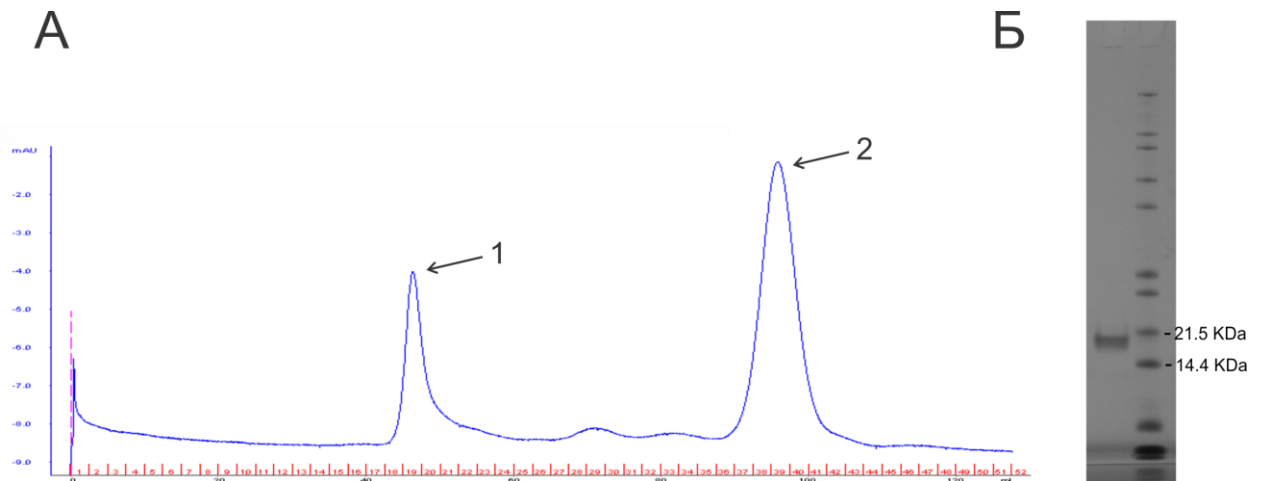


Рис. 19. Экспрессия рекомбинантного белка LlbRNH. **А.** Гель фильтрация очищенного белка. Пик 1 – агрегированный белок; пик 2 – мономер LlbRNH. **Б.** SDS-PAGE очищенного белка, взятого после гель-фильтрации из фракций соответствующим пику 2.

чистоту (Рис. 19А и Б) и малое количество агрегации (пик 1, Рис. 19А), которая могла бы нарушить нативную конформацию белка. Только те фракции LlbRNH, которые элюировались как мономер при гель-фильтрации (пик 2, Рис. 19А), использовались для последующего анализа активности. Точно таким же способом проводилась экспрессия и очистка мутантного белка LlbRNHD1326N.

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^{2+} или Mn^{2+} , причем каталитический мутант LlbRNHD1326N не проявлял подобной активности (Рис. 20). Более того, обнаруженная активность оказалось специфической для РНК/ДНК-гибрида: инкубация LlbRNH с одноцепочечной РНК не приводила к гидролизу субстрата.

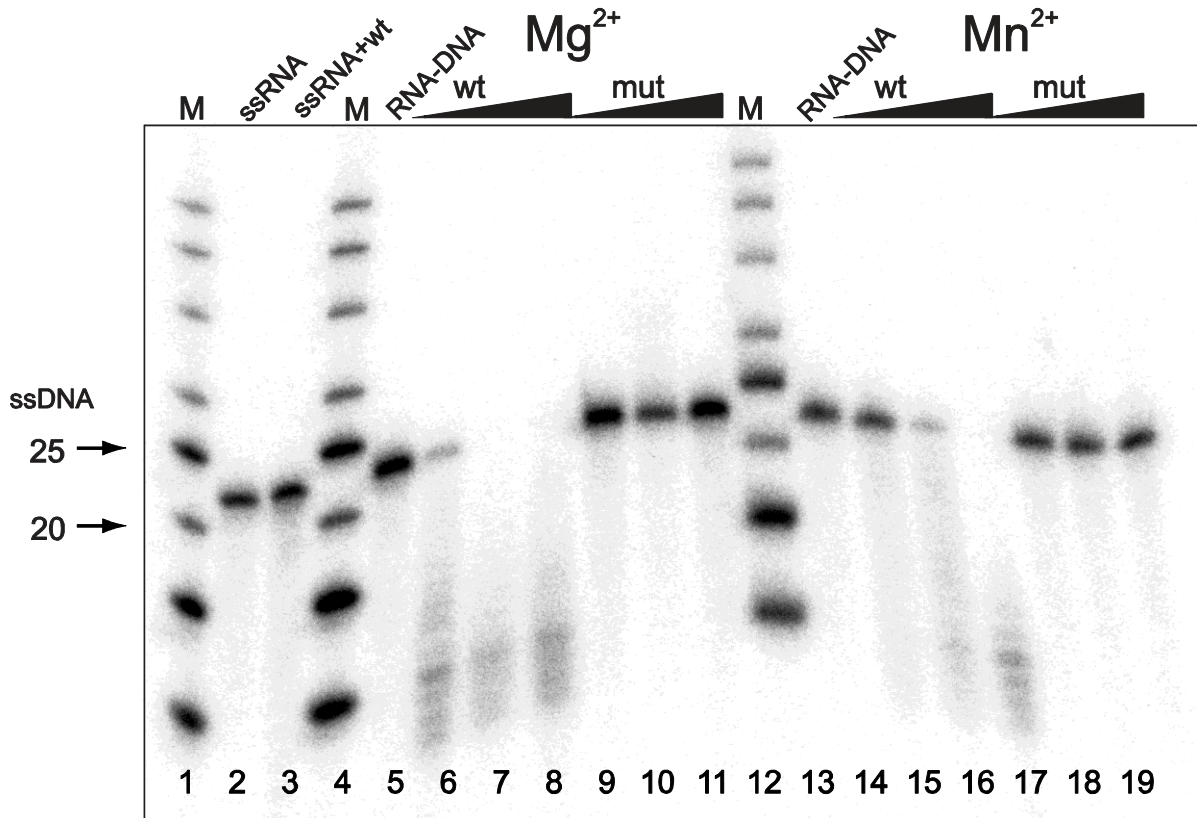


Рис. 20. Ферментативная активность домена RNH *Llb*-ретротранспозона (Smyshlyayev *et al.*, 2013); ‘wt’ обозначает LlbRNH дикого типа; ‘mut’ обозначает мутантный LlbRNHD1326N белок, содержащий замену в консервативной позиции RNH. Образцы, содержащие одноцепочечную РНК (ssRNA) и ssRNA вместе с LlbRNH в присутствии Mg^{2+} , 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^{2+} или Mn^{2+} соответственно). Отсутствие расщепления наблюдается при добавлении мутантного белка LlbRNHD1326N (ряды 9-11 и 17-19 в присутствии Mg^{2+} или Mn^{2+} соответственно).

Глава 4. Обсуждение. Модульная эволюция *L1-non-LTR*-ретротранспозонов растений

Теория модульной эволюции для 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 вирусо-подобных систем. Считается, что подобным свойством обладают и МЭ (Lerat *et al.*, 1999), а значит, и их эволюция может характеризоваться как модульная эволюция.

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. 21).

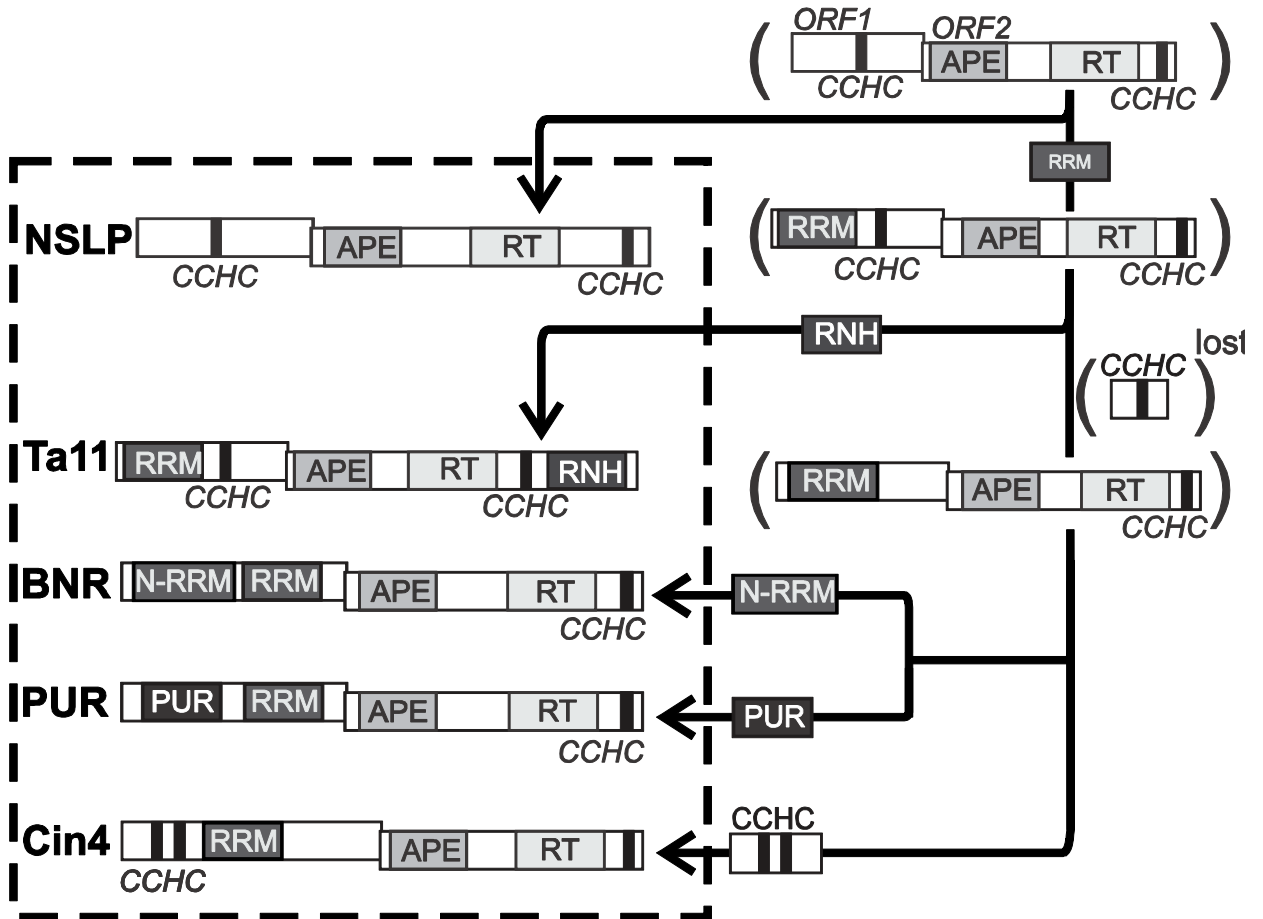


Рис. 21. Модульная эволюция *LI*-non-LTR-ретротранспозонов растений (Smyshlyaev *et al.*, 2013). Схемы структур современных растительных *LI*-элементов показаны в пунктирной рамке. События элиминации или приобретения функциональных доменов показаны снаружи рамки. Предполагаемые схемы структур элементов-предшественников 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.

За основу эволюции *LI*-non-LTR-ретротранспозонов растений была взята реконструированная с помощью домена RT филогения этих элементов (Рис. 7). В соответствии с этой филогенией, наиболее древней из выявленных групп *LI*-элементов растений, является группа *NSLP*, специфичная для мхов и плаунов. Как было сказано ранее, в ORF1p этих элементов не было идентифицировано никаких РНК-связывающих доменов за исключением CCHC-мотива (Рис. 8). *LI*-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 HNPRED-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; Tajrishii *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. 21).

Следующим эволюционным шагом к появлению современных форм *L1*-ретротранспозонов, возможно, стала потеря у некоторых из них С-концевого ССНС-мотива в ORF1p 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. 21). 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. 21). 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 ORF1p, 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 ORF1p.

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 5 основных групп. Кроме того была предпринята попытка реконструировать 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 архееподобной 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 взятых из баз данных 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 *Tall*, 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 используемой литературы:

1. Капелинская Т. В., Каграманова А. С., Королев А. Л., Муха Д. В. Белок первой открытой рамки считывания (ORF1p) *RI*-ретропозона *Blattella germanica* и филогенетически близкие ему GAG-подобные белки насекомых и грибов содержат RRM домены // Генетика. – 2011. - Т. 47. - № 2. - С. 149-158.
2. Новикова О. С., Блинов А. Г. Возникновение, эволюция и распространение различных групп non-LTR-ретротранспозонов среди эукариотических организмов // Генетика. – 2009. - Т. 45. - № 2. С. 149-159
3. Смышляев Г. А., Блинов А. Г. Эволюция и разнообразие *LI*-ретротранспозонов в геномах покрытосеменных растений // Вавиловский журнал генетики и селекции. – 2011. – Т. 15. - № 3. – С. 72-78.
4. Alisch R. S., Garcia-Perez J. L., Muotri A. R. *et al.* Unconventional translation of mammalian *LINE-1* retrotransposons // Genes Dev. - 2006. - V. 20. - No. 2. - P. 210–224.
5. Alzohairy A. M., Gyulai G., Jansen R. K., Bahieldin A. Transposable elements domesticated and neofunctionalized by eukaryotic genomes // Plasmid. - 2013. - V. 69. - No. 1. - P. 1–15.
6. An W., Dai L., Niewiadowska A. M., Yetil A. *et al.* Characterization of a synthetic human *LINE-1* retrotransposon *ORFeus-Hs* // Mob. DNA. - 2011. - V. 2. - No. 1. - P. 2.
7. Anisimova M., Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative // Syst. Biol. - 2006. - V. 55. - No. 4. - P. 539–552.
8. Anzai T., Osanai M., Hamada M., Fujiwara H. Functional roles of 3'-terminal structures of template RNA during *in vivo* retrotransposition of non-LTR retrotransposon, *R1Bm* // Nucleic Acids Res. - 2005. - V. 33. - No. 6. - P. 1993–2002.
9. Aravin A. A., Lagos-Quintana M., Yalcin A. *et al.* The small RNA profile during *Drosophila melanogaster* development // Dev. Cell. - 2003. - V. 5. - No. 2. - P. 337–350.
10. Aravin A. A., Sachidanandam R., Bourc'his D. *et al.* Developmentally regulated piRNA clusters implicate MILI in transposon control // Science. - 2007. - V. 316. - No. 5825. - P. 744–747.

11. Aravin A. A., Sachidanandam R., Girard A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice // *Mol. Cell.* - 2008. - V. 31. - No. 6. - P. 785–799.
12. Arkhipova I. R., Morrison H. G. Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead // *Proc. Natl Acad. Sci. U. S. A.* - 2001. - V. 98. - No. 25. - P. 14497–14502.
13. Aye M., Irwin B., Beliakova-Bethell N. *et al.* Host factors that affect *Ty3* retrotransposition in *Saccharomyces cerevisiae* // *Genetics.* - 2004. - V. 168. - No. 3. - P. 1159–1176.
14. Balagopal V., Parker R. Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs // *Curr. Opin. Cell Biol.* - 2009. - V. 21. - No. 3. - P. 403–408.
15. Beauregard A., Curcio M. J., Belfort M. The take and give between retrotransposable elements and their hosts // *Annu. Rev. Genet.* - 2008. - V. 42. - P. 587–617.
16. Bennett P. Genome Plasticity: Insertion Sequence Elements, Transposons and Integrons, and DNA Rearrangement // *Genomics, Proteomics, and Clinical Bacteriology: Methods and Reviews* / ed. by N. Woodford, A. Johnson, 2004. - P. 71–113.
17. Bhattacharya S., Bakre A., Bhattacharya A. Mobile genetic elements in protozoan parasites // *J. Genet.* - 2002. - V. 81. - No. 2. - P. 73–86.
18. Biedler J., Tu Z. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity // *Mol. Biol. Evol.* - 2003. - V. 20. - No. 11. - P. 1811–1825.
19. Biémont C., Vieira C. Genetics: junk DNA as an evolutionary force // *Nature.* - 2006. - V. 443. - No. 7111. - P. 521–524.
20. Biessmann H., Valgeirsdottir K., Lofsky A. *et al.* *HeT-A*, a transposable element specifically involved in “healing” broken chromosome ends in *Drosophila melanogaster* // *Mol. Cell. Biol.* - 1992. - V. 12. - No. 9. - P. 3910–3918.
21. Blesa D., Martínez-Sebastián M. J. *Bilbo*, a non-LTR retrotransposon of *Drosophila subobscura*: a clue to the evolution of LINE-like elements in *Drosophila* // *Mol. Biol. Evol.* - 1997. - V. 14. - No. 11. - P. 1145–1153.

22. Bogerd H. P., Wiegand H. L., Hulme A. E. *et al.* Cellular inhibitors of *long interspersed element 1* and *Alu* retrotransposition // Proc. Natl Acad. Sci. U. S. A. - 2006. - V. 103. - No. 23. - P. 8780–8785.
23. Botstein D. A theory of modular evolution for bacteriophages // Ann. N. Y. Acad. Sci. - 1980. - V. 354. - P. 484–490.
24. Brennecke J., Aravin A. A., Stark A. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila* // Cell. - 2007. - V. 128. - No. 6. - P. 1089–1103.
25. Bringaud F., Biteau N., Zuiderwijk E. *et al.* The *ingi* and *RIME* non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei* // Mol. Biol. Evol. - 2004. - V. 21. - No. 3. - P. 520–528.
26. Brouha B., Schustak J., Badge R. M. *et al.* Hot *L1s* account for the bulk of retrotransposition in the human population // Proc. Natl Acad. Sci. U. S. A. - 2003. - V. 100. - No. 9. - P. 5280–5285.
27. Bucheton A., Vaury C., Chaboissier M. C. *et al.* *I* elements and the *Drosophila* genome // Genetica. - 1992. - V. 86. - No. 1-3. - P. 175–190.
28. Burke W. D., Eickbush D. G., Xiong Y. *et al.* Sequence relationship of retrotransposable elements *R1* and *R2* within and between divergent insect species // Mol. Biol. Evol. - 1993. - V. 10. - No. 1. - P. 163–185.
29. Burke W. D., Malik H. S., Jones J. P., Eickbush T. H. The domain structure and retrotransposition mechanism of *R2* elements are conserved throughout arthropods // Mol. Biol. Evol. - 1999. - V. 16. - No. 4. - P. 502–511.
30. Burke W. D., Müller F., Eickbush T. H. *R4*, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes // Nucleic Acids Res. - 1995. - V. 23. - No. 22. - P. 4628–4634.
31. Burke W. D., Malik H. S., Rich S. M., Eickbush T. H. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia* // Mol. Biol. Evol. - 2002. - V. 19. - No. 5. - P. 619–630.
32. Capy P., Gasperi G., Biéumont C., Bazin C. Stress and transposable elements: co-evolution or useful parasites? // Heredity (Edinb). - 2000. - V. 85. - No. 2. - P. 101–106.

33. Carmell M. A., Girard A., van de Kant H. J. G. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline // *Dev. Cell.* - 2007. - V. 12. - No. 4. - P. 503–514.
34. Casaregola S., Neuvéglise C., Bon E., Gaillardin C. *Ylli*, a non-LTR retrotransposon *L1* family in the dimorphic yeast *Yarrowia lipolytica* // *Mol. Biol. Evol.* - 2002. - V. 19. - No. 5. - P. 664–677.
35. Cerritelli S. M., Crouch R. J. Ribonuclease H: the enzymes in eukaryotes // *FEBS J.* - 2009. - V. 276. - No. 6. - P. 1494–1505.
36. Chambeyron S., Bucheton A., Busseau I. Tandem UAA repeats at the 3'-end of the transcript are essential for the precise initiation of reverse transcription of the I factor in *Drosophila melanogaster* // *J. Biol. Chem.* - 2002. - V. 277. - No. 20. - P. 17877–17882.
37. Chaw S. M., Chang C. C., Chen H. L., Li W. H. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes // *J. Mol. Evol.* - 2004. - V. 58. - No. 4. - P. 424–441.
38. Checkley M. A., Nagashima K., Lockett S. J. *et al.* P-body components are required for *Ty1* retrotransposition during assembly of retrotransposition-competent virus-like particles // *Mol. Cell. Biol.* - 2010. - V. 30. - No. 2. - P. 382–398.
39. Christensen S. M., Bibillo A., Eickbush T. H. Role of the *Bombyx mori* *R2* element N-terminal domain in the target-primed reverse transcription (TPRT) reaction // *Nucleic Acids Res.* - 2005. - V. 33. - No. 20. - P. 6461–6468.
40. Christensen S. M., Ye J., Eickbush T. H. RNA from the 5' end of the *R2* retrotransposon controls *R2* protein binding to and cleavage of its DNA target site // *Proc. Natl Acad. Sci. U. S. A.* - 2006. - V. 103. - No. 47. - P. 17602–17607.
41. Daniels S. B., Peterson K. R., Strausbaugh L. D. *et al.* Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species // *Genetics.* - 1990. - V. 124. - No. 2. - P. 339–355.
42. Davies J. F., Hostomska Z., Hostomsky Z. *et al.* Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase // *Science.* - 1991. - V. 252. - No. 5002. - P. 88–95.

43. De Koning A.P.J., Gu W., Castoe T. A. *et al.* Repetitive elements may comprise over two-thirds of the human genome // *PLoS Genet.* - 2011. - V. 7. - No. 12. - P. e1002384.
44. De la Chaux N., Tsuchimatsu T., Shimizu K. K., Wagner A. The predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata* // *Mob. DNA.* - 2012. - V. 3. - No. 1. - P. 2.
45. Di Giacomo M., Comazzetto S., Saini H. *et al.* Multiple epigenetic mechanisms and the piRNA pathway enforce *LINE1* silencing during adult spermatogenesis // *Mol. Cell.* - 2013. - V. 50. - No. 4. - P. 601–608.
46. Dmitriev S. E., Andreev D. E., Terenin I. M. *et al.* Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon *LINE-1* mRNA is strictly cap dependent rather than internal ribosome entry site mediated // *Mol. Cell. Biol.* - 2007. - V. 27. - No. 13. - P. 4685–4697.
47. Domon C., Lorković Z. J., Valcárcel J., Filipowicz W. Multiple forms of the U2 small nuclear ribonucleoprotein auxiliary factor U2AF subunits expressed in higher plants // *J. Biol. Chem.* - 1998. - V. 273. - No. 51. - P. 34603–34610.
48. Doolittle W. F. Is junk DNA bunk? A critique of ENCODE // *Proc. Natl Acad. Sci. U. S. A.* - 2013. - V. 110. - No. 14. - P. 5294–5300.
49. Doolittle W. F., Sapienza C. Selfish genes, the phenotype paradigm and genome evolution // *Nature.* - 1980. - V. 284. - No. 5757. - P. 601–603.
50. Doucet A. J., Hulme A. E., Sahinovic E. *et al.* Characterization of *LINE-1* ribonucleoprotein particles // *PLoS Genet.* - 2010. - V. 6. - No. 10.
51. Dutko J. A., Schäfer A., Kenny A. E. *et al.* Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases // *Curr. Biol.* - 2005. - V. 15. - No. 7. - P. 661–666.
52. Eddy S. R. Profile hidden Markov models // *Bioinformatics.* - 1998. - V. 14. - No. 9. - P. 755–763.
53. Eddy S. R. The C-value paradox, junk DNA and ENCODE // *Curr. Biol.* - 2012. - V. 22. - No. 21. - P. R898–899.
54. Edgar R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity // *BMC Bioinformatics.* - 2004. - V. 5. - P. 113.

55. Eickbush D. G., Eickbush T. H. *R2* retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript // *Mol. Cell. Biol.* - 2010. - V. 30. - No. 13. - P. 3142–3150.
56. Eickbush T. H., Burke W. D., Eickbush D. G., Lathe W. C. Evolution of *R1* and *R2* in the rDNA units of the genus *Drosophila* // *Genetica.* - 1997. - V. 100. - No. 1-3. - P. 49–61.
57. Eickbush T. H., Jamburuthugoda V. K. The diversity of retrotransposons and the properties of their reverse transcriptases // *Virus Res.* - 2008. - V. 134. - No. 1-2. - P. 221–234.
58. Evgen'ev M. B., Arkhipova I. R. *Penelope*-like elements – a new class of retroelements: distribution, function and possible evolutionary significance // *Cytogenet. Genome Res.* - 2005. - V. 110. - No. 1-4. - P. 510–521.
59. Fedoroff N. - V. Presidential address. Transposable elements, epigenetics, and genome evolution // *Science.* - 2012. - V. 338. - No. 6108. - P. 758–767.
60. Feng Q., Moran J. V., Kazazian H. H., Boeke J. D. Human *L1* retrotransposon encodes a conserved endonuclease required for retrotransposition // *Cell.* - 1996. - V. 87. - No. 5. - P. 905–916.
61. Ferrandon D., Elphick L., Nüsslein-Volhard C., St Johnston D. Staufén protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner // *Cell.* - 1994. - V. 79. - No. 7. - P. 1221–1232.
62. Fire A., Xu S., Montgomery M. K., Kostas S. A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans* // *Nature.* - 1998. - V. 391. - No. 6669. - P. 806–811.
63. Fisher C. L., Pei G. K. Modification of a PCR-based site-directed mutagenesis method // *Biotechniques.* - 1997. - V. 23. - No. 4. - P. 570–574.
64. Fujiwara H., Osanai M., Matsumoto T., Kojima K. K. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori* // *Chromosome Res.* - 2005. - V. 13. - No. 5. - P. 455–467.
65. Gall J.G. Chromosome structure and the C-value paradox // *J. Cell Biol.* - 1981. - V. 91. - No. 3 Pt 2. - P. 3s–14s.

66. Gallois-Montbrun S., Kramer B., Swanson C. M. *et al.* Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules // *J. Virol.* - 2007. - V. 81. - No. 5. - P. 2165–2178.
67. Gasior S. L., Wakeman T. P., Xu B., Deininger P. L. The human *LINE-1* retrotransposon creates DNA double-strand breaks // *J. Mol. Biol.* - 2006. - V. 357. - No. 5. - P. 1383–1393.
68. George J. A., Eickbush T. H. Conserved features at the 5' end of *Drosophila R2* retrotransposable elements: implications for transcription and translation // *Insect Mol. Biol.* - 1999. - V. 8. - No. 1. - P. 3–10.
69. Ghildiyal M., Seitz H., Horwich M. D. *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells // *Science.* - 2008. - V. 320. - No. 5879. - P. 1077–1081.
70. Goodier J. L., Zhang L., Vetter M. R., Kazazian H. H. *LINE-1* ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex // *Mol. Cell. Biol.* - 2007. - V. 27. - No. 18. - P. 6469–6483.
71. Goodwin T. J., Ormandy J. E., Poulter R. T. *LI*-like non-LTR retrotransposons in the yeast *Candida albicans* // *Curr. Genet.* - 2001. - V. 39. - No. 2. - P. 83–91.
72. Goodwin T. J., Poulter R. T. The diversity of retrotransposons in the yeast *Cryptococcus neoformans* // *Yeast.* - 2001. - V. 18. - No. 9. - P. 865–880.
73. Grewal S. I. S., Elgin S. C. R. Transcription and RNA interference in the formation of heterochromatin // *Nature.* - 2007. - V. 447. - No. 7143. - P. 399–406.
74. Guindon S., Dufayard J. F., Lefort V. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0 // *Syst. Biol.* - 2010. - V. 59. - No. 3. - P. 307–321.
75. Haas N. B., Grabowski J. M., North J. *et al.* Subfamilies of *CR1* non-LTR retrotransposons have different 5'UTR sequences but are otherwise conserved // *Gene.* - 2001. - V. 265. - No. 1-2. - P. 175–183.
76. Han J. S. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions // *Mob. DNA.* - 2010. - V. 1. - No. 1. - P. 15.

77. Han J. S., Boeke J. D. A highly active synthetic mammalian retrotransposon // *Nature*. - 2004. - V. 429. - No. 6989. - P. 314–318.
78. Heitkam T., Schmidt T. BNR - a LINE family from *Beta vulgaris* - contains a RRM domain in open reading frame 1 and defines a *L1* sub-clade present in diverse plant genomes // *Plant J.* - 2009. - V. 59. - No. 6. - P. 872–882.
79. Heras S. R., López M. C., García-Pérez J. L. *et al.* The *LITc* C-terminal domain from *Trypanosoma cruzi* non-long terminal repeat retrotransposon codes for a protein that bears two C2H2 zinc finger motifs and is endowed with nucleic acid chaperone activity // *Mol. Cell. Biol.* - 2005. - V. 25. - No. 21. - P. 9209–9220.
80. Higashiyama T., Noutoshi Y., Fujie M., Yamada T. *Zepp*, a LINE-like retrotransposon accumulated in the *Chlorella* telomeric region // *The EMBO J.* - 1997. - V. 16. - No. 12. - P. 3715–3723.
81. Hohjoh H., Singer M. F. Cytoplasmic ribonucleoprotein complexes containing human *LINE-1* protein and RNA // *The EMBO J.* - 1996. - V. 15. - No. 3. - P. 630–639.
82. Hohjoh H., Singer M. F. Sequence-specific single-strand RNA binding protein encoded by the human *LINE-1* retrotransposon // *The EMBO J.* - 1997. - V. 16. - No. 19. - P. 6034–6043.
83. Horn A. V., Klawitter S., Held U. *et al.* Human *LINE-1* restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity // *Nucleic Acids Res.* - 2014. - V. 42. - No. 1. - P. 396–416.
84. Houwing S., Kamminga L. M., Berezikov E. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish // *Cell*. - 2007. - V. 129. - No. 1. - P. 69–82.
85. Hutvagner G., Simard M.J. Argonaute proteins: key players in RNA silencing // *Nat. Rev. Mol. Cell Biol.* - 2008. - V. 9. - No. 1. - P. 22–32.
86. Ibarra-Laclette E., Lyons E., Hernández-Guzmán G. *et al.* Architecture and evolution of a minute plant genome // *Nature*. - 2013. - V. 498. - No. 7452. - P. 94–98.
87. Ichiyanagi K., Nakajima R., Kajikawa M., Okada N. Novel retrotransposon analysis reveals multiple mobility pathways dictated by hosts // *Genome Res.* - 2007. - V. 17. - No. 1. - P. 33–41.

88. Jakymiw A., Lian S., Eystathioy T. *et al.* Disruption of GW bodies impairs mammalian RNA interference // *Nat. Cell Biol.* - 2005. - V. 7. - P. 1267–1274.
89. Janicki M., Rooke R., Yang G. Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes // *Chromosome Res.* - 2011. - V. 19. - No. 6. - P. 787–808.
90. Jurka J., Kapitonov V. V., Pavlicek A. *et al.* Repbase Update, a database of eukaryotic repetitive elements // *Cytogenet. Genome Res.* - 2005. - V. 110. - No. 1-4. - P. 462–467.
91. Kajikawa M., Okada N. LINEs mobilize SINEs in the eel through a shared 3' sequence // *Cell.* - 2002. - V. 111. - No. 3. - P. 433–444.
92. Kapitonov V. V., Jurka J. The esterase and PHD domains in *CRI*-like non-LTR retrotransposons // *Mol. Biol. Evol.* - 2003. - V. 20. - No. 1. - P. 38–46.
93. Kapitonov V. V., Tempel S., Jurka J. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences // *Gene.* - 2009. - V. 448. - No. 2. - P. 207–213.
94. Ketting R. F., Haverkamp T. H., van Luenen H. G., Plasterk R. H. Mut-7 of *Caenorhabditis elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD // *Cell.* - 1999. - V. 99. - No. 2. - P. 133–141.
95. Khazina E., Truffault V., Büttner R. *et al.* Trimeric structure and flexibility of the L1ORF1 protein in human *L1* retrotransposition // *Nat. Struct. Mol. Biol.* - 2011. - V. 18. - No. 9. - P. 1006–1014.
96. Khazina E., Weichenrieder O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame // *Proc. Natl Acad. Sci. U. S. A.* - 2009. - V. 106. - No. 3. - P. 731–736.
97. Kidwell M. G., Kidwell J. F., Sved J. A. Hybrid Dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination // *Genetics.* - 1977. - V. 86. - No. 4. - P. 813–833.
98. Kinomoto M., Kanno T., Shimura M. *et al.* All APOBEC3 family proteins differentially inhibit *LINE-1* retrotransposition // *Nucleic Acids Res.* - 2007. - V. 35. - No. 9. - P. 2955–2964.

99. Kirilyuk A., Tolstonog G. V., Damert A. *et al.* Functional endogenous *LINE-1* retrotransposons are expressed and mobilized in rat chloroleukemia cells // *Nucleic Acids Res.* - 2008. - V. 36. - No. 2. - P. 648–665.
100. Kojima K. K., Fujiwara H. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets // *Mol. Biol. Evol.* - 2004. - V. 21. - No. 2. - P. 207–217.
101. Kojima K. K., Fujiwara H. An extraordinary retrotransposon family encoding dual endonucleases // *Genome Res.* - 2005. - V. 15. - No. 8. - P. 1106–1117.
102. Kojima K. K., Matsumoto T., Fujiwara H. Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon *SART1* // *Mol. Cell. Biol.* - 2005. - V. 25. - No. 17. - P. 7675–7686.
103. Komatsu M., Shimamoto K., Kyojuka J. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon *Karma* // *Plant Cell.* - 2003. - V. 15. - No. 8. - P. 1934–1944.
104. Koonin E. V, Makarova K. S., Aravind L. Horizontal gene transfer in prokaryotes: quantification and classification // *Annu. Rev. Microbiol.* - 2001. - V. 55. - P. 709–742.
105. Kordis D., Gubensek F. Horizontal transfer of non-LTR retrotransposons in vertebrates // *Genetica.* - 1999. - V. 107. - No. 1-3. - P. 121–128.
106. Kramerov D. A., Vassetzky N. S. Origin and evolution of SINEs in eukaryotic genomes // *Heredity (Edinb).* - 2011. - V. 107. - No. 6. - P. 487–495.
107. Kumar A., Bennetzen J. L. Plant retrotransposons // *Annu. Rev. Genet.* - 1999. - V. 33. - P. 479–532.
108. Kuramochi-Miyagawa S., Watanabe T., Gotoh K. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes // *Genes Dev.* - 2008. - V. 22. - No. 7. - P. 908–917.
109. Kurzynska-Kokorniak A., Jamburuthugoda V. K., Bibillo A., Eickbush T. H. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the *R2* retrotransposon // *J. Mol. Biol.* - 2007. - V. 374. - No. 2. - P. 322–333.

110. Law J. A., Jacobsen S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals // *Nat. Rev. Genet.* - 2010. - V. 11. - No. 3. - P. 204–220.
111. Lerat E., Brunet F., Bazin C., Capy P. Is the evolution of transposable elements modular? // *Genetica.* - 1999. - V. 107. - No. 1-3. - P. 15–25.
112. Levis R. W., Ganesan R., Houtchens K. *et al.* Transposons in place of telomeric repeats at a *Drosophila* telomere // *Cell.* - 1993. - V. 75. - No. 6. - P. 1083–1093.
113. Lisch D. Regulation of transposable elements in maize // *Curr. Opin. Plant Biol.* - 2012. - V. 15. - No. 5. - P. 511–516.
114. Luchetti A., Mantovani B. Non-LTR *R2* element evolutionary patterns: phylogenetic incongruences, rapid radiation and the maintenance of multiple lineages // *PLoS One.* - 2013. - V. 8. - No. 2. - P. e57076.
115. Malik H. S. Ribonuclease H evolution in retrotransposable elements // *Cytogenet. Genome Res.* - 2005. - V. 110. - No. 1-4. - P. 392–401.
116. Malik H. S., Burke W. D., Eickbush T. H. The age and evolution of non-LTR retrotransposable elements // *Mol. Biol. Evol.* - 1999. - V. 16. - No. 6. - P. 793–805.
117. Malik H. S., Eickbush T. H. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons // *J. Virol.* - 1999. - V. 73. - No. 6. - P. 5186–5190.
118. Malik H. S., Eickbush T. H. *NeSL-1*, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans* // *Genetics.* - 2000. - V. 154. - No. 1. - P. 193–203.
119. Malik H. S., Eickbush T. H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses // *Genome Res.* - 2001. - V. 11. - No. 7. - P. 1187–1197.
120. Mandal P. K., Bagchi A., Bhattacharya A., Bhattacharya S. An *Entamoeba histolytica* LINE/SINE pair inserts at common target sites cleaved by the restriction enzyme-like LINE-encoded endonuclease // *Eukaryot. Cell.* - 2004. - V. 3. - No. 1. - P. 170–179.
121. Mandal P. K., Ewing A. D., Hancks D. C., Kazazian H. H. Enrichment of processed pseudogene transcripts in *L1*-ribonucleoprotein particles // *Hum. Mol. Genet.* - 2013. - V. 22. - No. 18. - P. 3730–3748.

122. Mangeat B., Turelli P., Caron G. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts // *Nature*. - 2003. - V. 424. - No. 6944. - P. 99–103.
123. Marchler-Bauer A., Lu S., Anderson J. B. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins // *Nucleic Acids Res.* - 2011. - V. 39. - No. Database issue. - P. D225–229.
124. Martin S. L., Branciforte D., Keller D., Bain D. L. Trimeric structure for an essential protein in *LI* retrotransposition // *Proc. Natl Acad. Sci. U. S. A.* - 2003. - V. 100. - No. 24. - P. 13815–13820.
125. Martin S. L. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, *LINE-1* // *RNA Biol.* - 2010. - V. 7. - No. 6. - P. 706–711.
126. Matzke M., Kanno T., Daxinger L. *et al.* RNA-mediated chromatin-based silencing in plants // *Curr. Opin. Cell Biol.* - 2009. - V. 21. - No. 3. - P. 367–376.
127. McClintock B. The origin and behavior of mutable loci in maize // *Proc. Natl Acad. Sci. U. S. A.* - 1950. - V. 36. - No. 6. - P. 344–355.
128. Meister G. Argonaute proteins: functional insights and emerging roles // *Nat. Rev. Genet.* - 2013. - V. 14. - No. 7. - P. 447–459.
129. Mizrokhi L. J., Georgieva S. G., Ilyin Y. V. *Jockey*, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II // *Cell*. - 1988. - V. 54. - No. 5. - P. 685–691.
130. Moran J. V., Holmes S. E., Naas T. P. *et al.* High frequency retrotransposition in cultured mammalian cells // *Cell*. - 1996. - V. 87. - No. 5. - P. 917–927.
131. Morrish T. A., Gilbert N., Myers J. S. *et al.* DNA repair mediated by endonuclease-independent *LINE-1* retrotransposition // *Nat. Genet.* - 2002. - V. 31. - No. 2. - P. 159–165.
132. Morrish T. A., Garcia-Perez J. L., Stamato T. D. *et al.* Endonuclease-independent *LINE-1* retrotransposition at mammalian telomeres // *Nature*. - 2007. - V. 446. - No. 7132. - P. 208–212.
133. Muckenfuss H., Hamdorf M., Held U. *et al.* APOBEC3 proteins inhibit human *LINE-1* retrotransposition // *J. Biol. Chem.* - 2006. - V. 281. - No. 31. - P. 22161–2272.

134. Noma K., Ohtsubo E., Ohtsubo H. Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes // *Mol. Gen. Genet.* - 1999. - V. 261. - No. 1. - P. 71–79.
135. Novick P. A., Basta H., Floumanhaft M. *et al.* The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals // *Mol. Biol. Evol.* - 2009. - V. 26. - No. 8. - P. 1811–1122.
136. Novikova O., Fet V., Blinov A. Non-LTR retrotransposons in fungi // *Funct. Integr. Genomics.* - 2009. - V. 9. - No. 1. - P. 27–42.
137. Odon V., Luke G. A., Roulston C. *et al.* APE-type non-LTR retrotransposons of multicellular organisms encode virus-like 2A oligopeptide sequences, which mediate translational recoding during protein synthesis // *Mol. Biol. Evol.* - 2013. - V. 30. - No. 8. - P. 1955–1965.
138. Ohshima K. Parallel relaxation of stringent RNA recognition in plant and mammalian *L1* retrotransposons // *Mol. Biol. Evol.* - 2012. - V. 29. - No. 11. - P. 3255–3259.
139. Ohtani N., Yanagawa H., Tomita M., Itaya M. Identification of the first archaeal Type 1 RNase H gene from *Halobacterium sp. NRC-1*: archaeal RNase HI can cleave an RNA-DNA junction // *Biochem. J.* - 2004. - V. 381. - No. Pt 3. - P. 795–802.
140. Okazaki S., Ishikawa H., Fujiwara H. Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori* // *Mol. Cell. Biol.* - 1995. - V. 15. - No. 8. - P. 4545–4552.
141. Orgel L. E., Crick F. H., Sapienza C. Selfish DNA // *Nature.* - 1980. - V. 288. - No. 5792. - P. 645–646.
142. Ostertag E. M., Kazazian H. H. Biology of mammalian *L1* retrotransposons // *Annu. Rev. Genet.* - 2001. - V. 35. - P. 501–538.
143. Pardue M. L., Rashkova S., Casacuberta E. *et al.* Two retrotransposons maintain telomeres in *Drosophila* // *Chromosome Res.* - 2005. - V. 13. - No. 5. - P. 443–453.
144. Pflingsten J. S., Kieft J. S. RNA structure-based ribosome recruitment: lessons from the Dicistroviridae intergenic region IRESes // *RNA.* - 2008. - V. 14. - No. 7. - P. 1255–1263.

145. Piskurek O., Okada N. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals // Proc. Natl Acad. Sci. U. S. A. - 2007. - V. 104. - No. 29. - P. 12046–12051.
146. Poulter R. T. M., Goodwin T. J. D. DIRS-1 and the other tyrosine recombinase retrotransposons // Cytogenet. Genome Res. - 2005. - V. 110. - No. 1-4. - P. 575–588.
147. Rho M., Tang H. MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes // Nucleic Acids Res. - 2009. - V. 37. - No. 21. - P. e143.
148. Roberts A. P., Mullany P. A modular master on the move: the *Tn916* family of mobile genetic elements // Trends Microbiol. - 2009. - V. 17. - No. 6. - P. 251–258.
149. Sánchez-Gracia A., Maside X., Charlesworth B. High rate of horizontal transfer of transposable elements in *Drosophila* // Trends Genet. - 2005. - V. 21. - No. 4. - P. 200–203.
150. Schaack S., Gilbert C., Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution // Trends Ecol. Evol. - 2010. - V. 25. - No. 9. - P. 537–546.
151. Schnable P. S., Ware D., Fulton R. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics // Science. - 2009. - V. 326. - No. 5956. - P. 1112–1115.
152. Schneider A. M., Schmidt S., Jonas S. *et al.* Structure and properties of the esterase from non-LTR retrotransposons suggest a role for lipids in retrotransposition // Nucleic Acids Res. - 2013. - V. 41. - No. 22. - P. 10563–10572.
153. Schumann G., Zündorf I., Hofmann J. *et al.* Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the Dictyostelium repetitive element, from *Dictyostelium discoideum* // Mol. Cell. Biol. - 1994. - V. 14. - No. 5. - P. 3074–3084.
154. Schwarz-Sommer Z., Leclercq L., Göbel E., Saedler H. *Cin4*, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons // The EMBO J. - 1987. - V. 6. - No. 13. - P. 3873–3880.
155. Seleme M. C., Disson O., Robin S. *et al.* *In vivo* RNA localization of *I factor*, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein // Nucleic Acids Res. - 2005. - V. 33. - No. 2. - P. 776–785.

156. Seleme M. C., Busseau I., Malinsky S. *et al.* High-frequency retrotransposition of a marked *I factor* in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes // *Genetics*. - 1999. - V. 151. - No. 2. - P. 761–771.
157. Sheen F. M., Levis R. W. Transposition of the LINE-like retrotransposon *TART* to *Drosophila* chromosome termini // *Proc. Natl Acad. Sci. U. S. A.* - 1994. - V. 91. - No. 26. - P. 12510–12514.
158. Siefert J. L. Defining the mobilome // *Methods Mol. Biol.* - 2009. - V. 532. - P. 13–27.
159. Silva J. C., Loreto E. L., Clark J. B. Factors that affect the horizontal transfer of transposable elements // *Curr. Issues Mol. Biol.* - 2004. - V. 6. - No. 1. - P. 57–71.
160. Slotkin R. K. The epigenetic control of the *Athila* family of retrotransposons in *Arabidopsis* // *Epigenetics*. - 2010. - V. 5. - No. 6. - P. 483–490.
161. Smith D., Zhong J., Matsuura M. *et al.* Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming // *Genes Dev.* - 2005. - V. 19. - No. 20. - P. 2477–2487.
162. Smyshlyaev G., Voigt F., Blinov A. *et al.* Acquisition of an Archaea-like ribonuclease H domain by plant *LI* retrotransposons supports modular evolution // *Proc. Natl Acad. Sci. U. S. A.* - 2013. - V. 110. - No. 50. - P. 20140–20145.
163. Söding J., Biegert A., Lupas A. N. The HHpred interactive server for protein homology detection and structure prediction // *Nucleic Acids Res.* - 2005. - V. 33. - No. Web Server issue. - P. W244–248.
164. Sormacheva I., Smyshlyaev G., Mayorov V. *et al.* Vertical evolution and horizontal transfer of *CRI* non-LTR retrotransposons and *Tc1/mariner* DNA transposons in *Lepidoptera* species // *Mol. Biol. Evol.* - 2012. - V. 29. - No. 12. - P. 3685–3702.
165. Stenglein M. D., Harris R. S. APOBEC3B and APOBEC3F inhibit *LI* retrotransposition by a DNA deamination-independent mechanism // *J. Biol. Chem.* - 2006. - V. 281. - No. 25. - P. 16837–16841.
166. Stoppel R., Meurer J. The cutting crew - ribonucleases are key players in the control of plastid gene expression // *J. Exp. Bot.* - 2012. - V. 63. - No. 4. - P. 1663–173.

167. Suzuki J, Yamaguchi K, Kajikawa M. *et al.* Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition // PLoS Genet. - 2009. - V. 5. - No. 4. - P. e1000461.
168. Symer D. E., Connelly C., Szak S. T. *et al.* Human *L1* retrotransposition is associated with genetic instability *in vivo* // Cell. - 2002. - V. 110. - No. 3. - P. 327–338.
169. Szak S. T., Pickeral O. K., Makalowski W. *et al.* Molecular archeology of *L1* insertions in the human genome // Genome Biol. - 2002. - V. 3. - No. 10. - P. research0052.
170. Tabara H., Sarkissian M., Kelly W. G. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *Caenorhabditis elegans* // Cell. - 1999. - V. 99. - No. 2. - P. 123–132.
171. Tadokoro T., Kanaya S. Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes // FEBS J. - 2009. - V. 276. - No. 6. - P. 1482–1493.
172. Tajrishli M. M., Tuteja R., Tuteja N. Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus // Commun. Integr. Biol. - 2011. - V. 4. - No. 3. - P. 267–275.
173. Takahashi H., Fujiwara H. Transcription analysis of the telomeric repeat-specific retrotransposons *TRAS1* and *SART1* of the silkworm *Bombyx mori* // Nucleic Acids Res. - 1999. - V. 27. - No. 9. - P. 2015–2021.
174. Tamura K., Peterson D., Peterson N. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods // Mol. Biol. Evol. - 2011. - V. 28. - No. 10. - P. 2731–2739.
175. Teng S. C., Wang S. X., Gabriel A. A new non-LTR retrotransposon provides evidence for multiple distinct site-specific elements in *Crithidia fasciculata* miniexon arrays // Nucleic Acids Res. - 1995. - V. 23. - No. 15. - P. 2929–2936.
176. Trelogan S. A., Martin S. L. Tightly regulated, developmentally specific expression of the first open reading frame from *LINE-1* during mouse embryogenesis // Proc. Natl Acad. Sci. U. S. A. - 1995. - V. 92. - No. 5. - P. 1520–1524.

177. Voigt F., Reuter M., Kasaruho A. *et al.* Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc // RNA. - 2012. - V. 18. - No. 12. - P. 2128–2134.
178. Volff J. N., Körting C., Froschauer A. *et al.* Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates // J. Mol. Evol. - 2001. - V. 52. - No. 4. - P. 351–360.
179. Volff J. N., Körting C., Scharl M. Multiple lineages of the non-LTR retrotransposon *Rex1* with varying success in invading fish genomes // Mol. Biol. Evol. - 2000. - V. 17. - No. 11. - P. 1673–1684.
180. Wassenegger M., Heimes S., Riedel L., Sängler H. L. RNA-directed *de novo* methylation of genomic sequences in plants // Cell. - 1994. - V. 76. - No. 3. - P. 567–576.
181. Wenke T., Holtgräwe D., Horn A. V. *et al.* An abundant and heavily truncated non-LTR retrotransposon (LINE) family in *Beta vulgaris* // Plant Mol. Biol. - 2009. - V. 71. - No. 6. - P. 585–597.
182. White M. K., Johnson E. M., Khalili K. Multiple roles for Puralpha in cellular and viral regulation // Cell Cycle. - 2009. - V. 8. - No. 3. - P. 1–7.
183. Wicker T., Sabot F., Hua-Van A. *et al.* A unified classification system for eukaryotic transposable elements // Nat. Rev. Genet. - 2007. - V. 8. - No. 12. - P. 973–982.
184. Wright D. A., Ke N., Smalle J. *et al.* Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana* // Genetics. - 1996. - V. 142. - No. 2. - P. 569–578.
185. Xiong Y., Eickbush T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences // The EMBO J. - 1990. - V. 9. - No. 10. - P. 3353–3362.
186. Yamashita H., Tahara M. A LINE-type retrotransposon active in meristem stem cells causes heritable transpositions in the sweet potato genome // Plant Mol. Biol. - 2006. - V. 61. - No. 1-2. - P. 79–94.
187. Yang J., Eickbush T. H. RNA-induced changes in the activity of the endonuclease encoded by the *R2* retrotransposable element // Mol. Cell. Biol. - 1998. - V. 18. - No. 6. - P. 3455–3465.

188. Yang J., Malik H. S., Eickbush T. H. Identification of the endonuclease domain encoded by *R2* and other site-specific, non-long terminal repeat retrotransposable elements // *Proc. Natl Acad. Sci. U. S. A.* - 1999. - V. 96. - No. 14. - P. 7847–7852.
189. Yoder J. A., Walsh C. P., Bestor T. H. Cytosine methylation and the ecology of intragenomic parasites // *Trends Genet.* - 1997. - V. 13. - No. 8. - P. 335–340.
190. Zingler N., Willhoeft U., Brose H. P. *et al.* Analysis of 5' junctions of human *LINE-1* and *Alu* retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining // *Genome Res.* - 2005. - V. 15. - No. 6. - P. 780–789.
191. Zingler N., Weichenrieder O., Schumann G. G. APE-type non-LTR retrotransposons: determinants involved in target site recognition // *Cytogenet. Genome Res.* - 2005. - V. 110. - No. 1-4. - P. 250–268.
192. Zuckerkandl E., Cavalli G. Combinatorial epigenetics, “junk DNA”, and the evolution of complex organisms // *Gene.* - 2007. - V. 390. - No. 1-2. - P. 232–242.
193. Zupunski V., Gubensek F., Kordis D. Evolutionary dynamics and evolutionary history in the *RTE* clade of non-LTR retrotransposons // *Mol. Biol. Evol.* - 2001. - V. 18. - No. 10. - P. 1849–1863.

Приложение 1. Филогенетическое дерево non-LTR-ретротранспозонов, реконструированное с помощью метода максимального правдоподобия на основе аминокислотных последовательностей их обратных транскриптаз. Серым цветом, обозначены элементы растений, исследованные в данной работе. Для обозначения элементов найденных в геноме использовались инициалы вида, в геноме которого был найден элемент и номер элемента, например, *OS1a* – это элемент из *O. sativa* под номером 1a. Для обозначения элементов, взятых из базы данных RepBase, использовалось исходное название в базе данных, по которому элемент может быть в ней найден, например SHALINE12_MT_L1_Medicago_truncatula. Названия всех остальных элементов на дереве образованы названием элемента, используемым в литературе, названием вида и номером доступа в GenBank, например, Karma_Oryza_sativa_BAC15618. В качестве корня использовались последовательности *CRE*-элементов, а также *L1*-элементы водорослей. Слева от некоторых узлов дерева указана их статистическая достоверность согласно aLRT.

