

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК»

На правах рукописи

УСТЬЯНЦЕВ КИРИЛЛ ВАЛЕРЬЕВИЧ

**ФУНКЦИОНАЛЬНАЯ И СТРУКТУРНАЯ КОНВЕРГЕНЦИЯ
РЕТРОТРАНСПОЗОНОВ С ДОПОЛНИТЕЛЬНЫМ ДОМЕНОМ
РИБОНУКЛЕАЗЫ N В ГЕНОМАХ РАСТЕНИЙ И ООМИЦЕТОВ**

03.02.07 – генетика

Диссертация на соискание ученой степени

кандидата биологических наук

Научный руководитель:

кандидат биологических наук,

в. н. с. Блинов Александр Геннадьевич

Новосибирск 2017

ОГЛАВЛЕНИЕ

ОГЛАВЛЕНИЕ	2
СПИСОК СОКРАЩЕНИЙ	5
ВВЕДЕНИЕ	7
Актуальность	7
Цели и задачи исследования.....	9
Научная новизна работы.....	9
Теоретическая и практическая значимость исследования	10
Основные положения, выносимые на защиту	11
Вклад автора.....	11
Апробация работы	11
Глава 1. ОБЗОР ЛИТЕРАТУРЫ	12
1.1 Мобильные генетические элементы эукариот	12
1.1.1 Определение	12
1.1.2 Влияние мобильных элементов на геном	12
1.2 Ретротранспозоны.....	14
1.2.1 LTR-ретротранспозоны	16
1.2.1.1 Структура LTR-ретротранспозонов	17
1.2.1.2 Жизненный цикл LTR-ретротранспозонов.....	19
1.2.1.3 Разнообразие и распространение LTR-ретротранспозонов.....	25
1.2.2 non-LTR-ретротранспозоны.....	28
1.2.2.1 Структура non-LTR-ретротранспозонов.....	29
1.2.2.2 Жизненный цикл non-LTR-ретротранспозонов	31
1.2.2.3 Разнообразие и распространение non-LTR-ретротранспозонов	35
1.2.3 Модульная эволюция ретротранспозонов.....	39
1.2.3.1 Домен рибонуклеазы H в эволюции ретротранспозонов	40
1.2.3.2 Модульная эволюция как конвергентная	42
1.3 Ретротранспозоны оомицетов и зеленых растений с доменом «архейной» рибонуклеазы H.....	43

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ.....	47
Базы данных и конкретные геномные последовательности, использованные для анализа.....	47
Компьютерный поиск ретротранспозонов с доменом «архейной» рибонуклеазы N в базе данных повторенных последовательностей Repbase Update	47
Компьютерный поиск последовательностей ретротранспозонов с доменом «архейной» рибонуклеазы N в геномных последовательностях зеленых растений и оомицетов.....	47
Компьютерный поиск последовательностей потенциальных клеточных генов рибонуклеаз N в геномных последовательностях зеленых растений и оомицетов	48
Определение структуры ретротранспозонов с архейной рибонуклеазой N.....	49
Сравнительный и филогенетический анализы	49
ГЛАВА 3. РЕЗУЛЬТАТЫ и ОБСУЖДЕНИЕ.....	51
Разнообразие и распространение ретротранспозонов с доменом «архейной» рибонуклеазы N в геномах растений и оомицетов	51
<i>Анализ разнообразия и распространения ретротранспозонов с доменов «архейной» рибонуклеазы N в базе данных Repbase Update.....</i>	<i>51</i>
<i>Анализ разнообразия и распространения ретротранспозонов с доменом «архейной» рибонуклеазы N в геномных последовательностях растений и оомицетов</i>	<i>52</i>
Структурный и филогенетический анализы ретротранспозонов растений и оомицетов, содержащих домен «архейной» рибонуклеазы N.....	54
<i>Структурный и филогенетический анализы LTR-ретротранспозонов растений и оомицетов.....</i>	<i>54</i>
<i>LTR-ретротранспозоны с доменом «архейной» рибонуклеазы N в геномах зеленых растений</i>	<i>55</i>
<i>LTR-ретротранспозоны с доменом «архейной» рибонуклеазы N в геномах оомицетов</i>	<i>58</i>

<i>Структурный и филогенетический анализы non-LTR-ретротранспозонов растений и оомицетов.....</i>	59
Происхождение домена «архейной» рибонуклеазы H у ретротранспозонов растений и оомицетов	61
<i>Разнообразие и филогенетический анализ последовательностей рибонуклеазы H оомицетов.....</i>	61
<i>Разнообразие и филогенетический анализ последовательностей рибонуклеазы H зеленых растений.....</i>	66
Анализ структуры gRNH-aRNH у LTR-ретротранспозонов Chronos, Archon и Tat IV-VI	67
Конвергенция ретротранспозонов растений и оомицетов	71
ЗАКЛЮЧЕНИЕ.....	76
ВЫВОДЫ	78
СПИСОК ЛИТЕРАТУРЫ.....	80
ПРИЛОЖЕНИЕ	95

СПИСОК СОКРАЩЕНИЙ

- ВПЧ – вирусоподобная частица
- ГП – горизонтальный перенос
- ДНК – дезоксирибонуклеиновая кислота
- кДНК – комплементарная ДНК
- мРНК – матричная (или информационная) РНК
- МКТВ – Международный комитет по таксономии вирусов
- МЭ – мобильный элемент
- НТО – нетранслируемая область
- ОРС – открытая рамка считывания
- РНК – рибонуклеиновая кислота
- рРНК – рибосомная РНК
- тРНК – транспортная РНК
- PR – аспартильная протеаза (Protease)
- ARE – домен апуриновой/апиримидиновой эндонуклеазы
- CHD – хромодомен (CHromo Domain)
- EN – домен эндонуклеазы non-LTR-ретротранспозонов (ENdonuclease)
- Env – ген белка оболочки ретровирусов (Envelope)
- Gag – структурный белок LTR-ретротранспозона
- HMM – скрытая марковская модель (Hidden Markov Model)
- INT – интеграназа (Integrase)
- IRES – внутренний сайт посадки рибосом (Internal Ribosomal Entry Site)
- LTR – длинный концевой повтор (Long Terminal Repeat)
- NLS – сигнал ядерной локализации (Nuclear Localization Signal)
- PBS – сайт связывания затравки/прайма (Primer Binding Site)
- Pol – полипротеин LTR-ретротранспозона (Polyprotein)
- poI – промотор РНК-полимеразы типа I
- poII – промотор РНК-полимеразы типа II
- PPT – полипуриновый тракт (PolyPurine Tract)

RLE – домен рестрикционно-подобной эндонуклеазы (Restriction-like Endonuclease)

RNH – рибонуклеаза H (Ribonuclease H)

aRNH – «архейная» RNH (Archaea-like RNH)

fmRNH – RNH грибов и метазоа (Fungi/Metazoa-like RNH)

gRNH – исходный домен RNH *Ty3/Gypsy* LTR-ретротранспозонов

RNP – рибонуклеопротеиновая частица (Ribonucleoprotein particle)

RT – обратная транскриптаза (Reverse Transcriptase)

RU – база данных повторенных последовательностей Repbase Update.

SAR – филогенетическая клада Stramenopiles-Alveolata-Rhizaria

TPRT – мишень-иницированная обратная транскрипция (Target Primed Reverse Transcription)

TSD – дупликация сайта встраивания (Target Site Duplication)

ВВЕДЕНИЕ

Актуальность

Мобильные генетические элементы (МЭ) являются важными регуляторами структуры и функционирования геномов. Своим перемещением МЭ могут менять как активность отдельных генов, так и приводить к возникновению крупных хромосомных перестроек. МЭ также являются важными агентами горизонтального переноса генетического материала между различными таксонами. Ретротранспозоны – широкий класс МЭ, перемещающихся по механизму «копирования и вставки». При таком перемещении ДНК-копия ретротранспозона синтезируется по матрице его мРНК в процессе обратной транскрипции и затем встраивается в новое место в геноме хозяина. Поэтому, ретротранспозоны могут составлять значительную долю от размера эукариотических геномов, что в особенности оказалось характерным для геномов растений (Vitte et al., 2005; Frost et al., 2005; Gogvadze et al., 2009; Kim et al., 2012).

Единственной общей чертой различных групп ретротранспозонов и их вирусных потомков (ретровирусов позвоночных) является домен обратной транскриптазы (reverse transcriptase – RT) – ключевой фермент обратной транскрипции. Набор же остальных функциональных доменов, входящих в полипротеин ретроэлемента, может существенно варьировать в зависимости от организации жизненного цикла и стратегии встраивания конкретного ретротранспозона (Xiong et al., 1990; Eickbush et al., 2002). Каждый из этих дополнительных доменов ретротранспозонов может иметь эволюционную историю независимую от истории центрального RT-домена, так как на различных этапах эволюции отдельных групп ретроэлементов тот или иной функциональный домен приобретает в виде готового функционального модуля за счет горизонтального переноса (Lerat et al., 1999; Malik et al., 1999b; Malik et al., 2000b; Malik et al., 2001; Smyshlyaev et al., 2013). Данная особенность послужила основой модульного представления об эволюции ретротранспозонов. Конкретный «фенотип», т. е. жизненный цикл отдельного ретротранспозона, будет определяться совокупностью составляющих его функциональных доменов (Lerat et al., 1999).

Следствием модульного строения является возможность возникновения быстрой структурно-функциональной конвергенции между ретротранспозонами, у которых имеется одинаковый набор функциональных белковых доменов. Конвергенция, в широком смысле, подразумевает приобретение структур со сходной формой или функцией эволюционно удаленными генетическими линиями как следствие приспособления к одинаковым условиям существования (Doolittle, 1994; Stern, 2013; Stayton, 2015). Такими удаленными генетическими линиями являются и сами ретротранспозоны, распространенные в эволюционно удаленных таксонах живых организмов и дивергировавшие вместе с геномами своих хозяев. А структурами со сходной формой или функцией являются те самые белок-кодирующие модули, которые независимо приобретают удаленные группы ретротранспозонов. В случае генетически удаленных ретротранспозонов, независимо приобретших сходные функциональные домены в аналогичном положении, речь будет идти о структурно-функциональной конвергенции между ними к единому «фенотипу», а именно к сходной организации их жизненного цикла.

В работах по изучению разнообразия, распространения и эволюции ретротранспозонов из различных таксонов было выявлено, что одним из наиболее часто приобретаемых и теряющихся доменов в эволюции различных групп ретроэлементов является функциональный домен рибонуклеазы H (ribonuclease H, RNH) (Malik et al., 2001; Kojima et al., 2005a; Malik, 2005; Smyshlyaev et al., 2012; Smyshlyaev et al., 2013; Kojima et al., 2015). Два недавних независимых исследования показали, что *Tall* non-LTR-ретротранспозоны зеленых растений (Archaeplastida, Viridiplantae) из группы *L1*, а также non-LTR-ретротранспозоны из геномов паразитических простейших – оомицетов (клада SAR, Stramenopiles) – из группы *Utopia* приобрели нестандартный для остальных ретротранспозонов домен RNH, филогенетически близкий к последовательностям генов RNH архей, бактерий и растений (Archaea-like RNH, «Архейная» RNH, aRNH) (Smyshlyaev et al., 2013; Kojima et al., 2015).

Оомицеты эволюционно значительно удалены от зеленых растений (Beakes et al., 2012). Однако многие виды оомицетов экологически тесно связаны с

зелеными растениями и паразитируют на них, являясь также хозяйственно-важными вредителями, как например, представители рода *Phytophthora* (Bebber et al., 2015). Остается неясным, приобрели ли ретротранспозоны растений и оомицетов домен aRNH независимо друг от друга (наблюдается пример конвергентной эволюции) или же aRNH-домен был приобретен лишь однажды предковым ретротранспозоном, который уже затем был горизонтально перенесен из генома растения в оомицет (или наоборот) за счет тесной связи паразита и хозяина на ранних этапах их коэволюции. При любом из двух перечисленных эволюционных сценариев интересен вопрос о функциональной роли aRNH-домена для несущих его ретротранспозонов. Также неизвестно, ограничивается ли разнообразие ретротранспозонов с aRNH уже описанными представителями *L1* и *Utopia* растений и оомицетов, или же имеет место более обширный тренд в эволюции ретротранспозонов, который определяется особенной функциональной ролью домена aRNH.

Цели и задачи исследования

Целью данной работы является выявление пути эволюции ретротранспозонов с «архейным» доменом рибонуклеазы H из геномов растений и оомицетов с определением функциональной значимости приобретения данного домена для жизненного цикла этих ретротранспозонов.

Задачи:

1. Исследование разнообразия и распространения ретротранспозонов с «архейным» доменом рибонуклеазы H в геномах растений и оомицетов.
2. Реконструкция филогении основных эволюционных групп ретротранспозонов с «архейным» доменом рибонуклеазы H и установление характера их дивергенции/конвергенции.
3. Сравнительный анализ структурных характеристик найденных элементов и поиск источников данного структурного разнообразия.

Научная новизна работы

В данной работе был впервые проведен биоинформационный поиск и детальный анализ ретротранспозонов, содержащих домен «архейной»

рибонуклеазы Н в 65 геномах зеленых растений и 25 геномов оомицетов. По итогам проведенного скрининга данный домен был впервые описан у LTR-ретротранспозонов растений из группы *Ty3/Gypsy* кластера *Tat*, а также у двух ранее неизвестных кластеров *Ty3/Gypsy* LTR-ретротранспозонов оомицетов и кластера *L1 non-LTR*-ретротранспозонов оомицетов.

Результаты сравнительного структурного и филогенетического анализов говорят о конвергентном (независимом) приобретении домена «архейной» рибонуклеазы Н выявленными группами ретротранспозонов растений и оомицетов. Полученные данные отвергают гипотезу одиночного возникновения данного домена и горизонтального переноса предкового ретротранспозона между растениями и оомицетами.

Детальное изучение последовательностей доменов рибонуклеазы Н показало структурное и функциональное конвергентное сходство LTR-ретротранспозонов растений и оомицетов с ретровирусами позвоночных животных. В итоге, по результатам данной работы было расширено представление о роли конвергенции в эволюции ретротранспозонов, а также предположены основные тенденции в эволюции ретротранспозонов из различных групп.

Теоретическая и практическая значимость исследования

Результаты этой работы расширяют представление о роли конвергенции в эволюции ретротранспозонов и создают основу для проведения исследований по выявлению новых групп ретроэлементов с доменными структурами, аналогичными тем, которые были обнаружены в данной работе. С учетом модульности эволюции ретротранспозонов, выявление независимо возникших доменных структур позволяет судить о сходстве жизненных циклов даже у филогенетически значительно удаленных групп ретроэлементов.

Разработанный и успешно примененный в данной работе алгоритм поиска новых групп ретротранспозонов с доменом «архейной» рибонуклеазы Н может быть в дальнейшем использован для идентификации групп ретротранспозонов с другими интересующими доменами и на новых геномных данных.

Основные положения, выносимые на защиту

1. В геномах растений и оомицетов присутствуют LTR- и non-LTR-ретротранспозоны, содержащие домен «архейной» рибонуклеазы H.
2. Домен «архейной» рибонуклеазы H приобретен ретротранспозонами растений и оомицетов независимо как результат конвергенции к единому структурно-функциональному молекулярному фенотипу.
3. LTR-ретротранспозоны растений и оомицетов с доменом «архейной» рибонуклеазы H конвергентно повторяют отдельные этапы эволюционного пути ретровирусов позвоночных животных.

Вклад автора

Все основные научные результаты были получены автором самостоятельно. Материалы для работы были взяты из открытых бесплатных источников. Неопубликованные данные по сборкам геномных последовательностей мхов были получены от к. б. н. Новиковой О. С. (Университет Олбани, США).

Апробация работы

Результаты данной работы были представлены на научных конференциях «EMBO | EMBL Symposium: The Mobile Genome: Genetic and Physiological Impacts of Transposable Elements» (Heidelberg, 2015) и «LI международная научная студенческая конференция» (Новосибирск, 2013).

По теме диссертации были опубликованы две работы в рецензируемых зарубежных журналах:

1. **Ustyantsev, K.**, Novikova, O., Blinov, A., & Smyshlyayev, G. Convergent evolution of ribonuclease H in LTR retrotransposons and retroviruses // *Mol. Biol. Evol.* 2015. Т. 32. № 5. С. 1197–1207.
2. **Ustyantsev K.**, Blinov A., Smyshlyayev G. Convergence of retrotransposons in oomycetes and plants // *Mobile DNA* 2017. Т. 8. №. 1. С. 4.

Структура и объем работы

Диссертация состоит из введения, обзора литературы, материалов и методов, результатов и обсуждений, заключения, выводов и списка литературы. Работа изложена на 105 страницах, содержит 20 рисунков и 4 таблицы.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1 Мобильные генетические элементы эукариот

1.1.1 Определение

В 1950 году своей новаторской работой по изучению «контролирующих элементов» у кукурузы, влияющих на активность соседствующих с ними генов, Барбара Мак-Клинток начала новый виток в истории развития представлений о структуре и функционировании геномов (McClintock, 1950). Геном, как стабильное и постоянное целое, которое нарушается только вследствие случайных мутаций под действием факторов среды, оказался активно изменчивой, пластичной и даже, в какой-то степени, непредсказуемой сущностью (Shapiro, 1995). С открытием в 1960-ых годах инсерционных последовательностей и «подвижных» профагов у бактерий (Taylor, 1963; Fiandt et al., 1972; Hirsch et al., 1972) было окончательно признано, что структура геномов может меняться за счёт внутренних источников рекомбинации – перемещающихся генетических элементов, которые теперь называют мобильными генетическими элементами (МЭ).

МЭ является любая генетическая последовательность, которая способна обеспечить своё независимое перемещение от других частей генома (Frost et al., 2005). По такому широкому критерию к МЭ относят вирусы (фаги), плазмиды, факторы мейотического хромосомного драйва и, собственно, транспозоны (которые и были впервые описаны в работе Мак-Клинток) (Hurst et al., 2001; Frost et al., 2005). Если МЭ самостоятельно кодирует все необходимые факторы для обеспечения своего перемещения, то его называют автономным и поэтому могут активно преобразовывать структуру генома. К числу автономных МЭ относятся вирусы и транспозоны (Hurst et al., 2001; Koonin, 2016).

1.1.2 Влияние мобильных элементов на геном

За способность к независимому перемещению и самостоятельной репликации МЭ в начале 1980-ых годов были названы «эгоистическими» и сугубо «паразитическими» последовательностями геномов (Doolittle et al., 1980; Orgel et al., 1980). В рамках парадигмы «эгоистической» ДНК, огромное количество, по большей части неактивных копий МЭ, обнаруживаемое в некоторых

эукариотических геномах, характеризуется как генетический «груз» или «мусор», накопленный в борьбе за выживание хозяина с геномными «паразитами». «Эгоистичность» МЭ также в значительной мере проявляется в том, что в подавляющем большинстве случаев их перемещение приводит к негативным последствиям для генома-хозяина (Volf, 2006), а сам МЭ не выполняет какой-либо значимой для своего хозяина функции. Встраивание МЭ в насыщенные генами участки генома может напрямую нарушить кодирующие или регуляторные последовательности находящихся там генов. Множественные копии одного и того же МЭ, расположенные на негомологичных участках одной или разных хромосом, уже сами по себе, пассивно, выступают источниками для неравного кроссинговера, что приводит к возникновению крупных геномных хромосомных перестроек: реципрокных транслокаций, делеций, инверсий и дупликаций. Причём, с увеличением числа МЭ в геноме растёт и вероятность хромосомных мутаций (Deininger et al., 1999; Louis et al., 2000; Kidwell et al., 2001; Hughes et al., 2001; Morgante et al., 2005; Chen et al., 2006).

В качестве критики концепции сугубо «паразитических» и «мусорных» элементов, сформулированной еще в 1980-ых годах, ставится то, что она не вполне согласуется с накопленными к настоящему моменту данными о разнообразии и распространении МЭ, пришедшими с началом более детального анализа геномных последовательностей. Уже доказано, что некоторые МЭ были прямо адаптированы своими хозяевами для выполнения клеточно-важных функций (Kidwell et al., 2001; Volf, 2006; Feschotte, 2008; Biémont, 2010). Основные ферменты V(D)J-рекомбинации генов иммуноглобулинов позвоночных (*Rag1* и *Rag2*) напрямую происходят от белка-транспозазы ДНК-транспозона суперсемейства *Transib* (Zhou et al., 2004; Kapitonov et al., 2005; Fugmann, 2010). Non-LTR-ретротранспозоны *HetA* и *TART* у *Drosophila* замещают функцию отсутствующего фермента теломеразы, целенаправленно встраиваясь в район теломерных концов (Levis et al., 1993). Всё еще обсуждается вопрос о происхождении теломеразы из обратной транскриптазы древнего ретротранспозона (Eickbush, 1997; Gladyshev et al., 2011).

Показано, что гены вирусной оболочки *env* (envelope – оболочка) инактивированных накопленными мутациями ретровирусов млекопитающих

(эндогенные ретровирусы) играют важную роль в формировании плаценты (McCooy et al., 2000). Однако, пожалуй, наиболее важна роль МЭ в формировании гибких и расширяющихся генных сетей в геномах за счёт кодируемых ими регуляторных факторов: промоторов, инсуляторов, сайтов импринтига, новых сайтов сплайсинга - что приводит к возможности формирования сложных регуляторных контуров и потенциальному усложнению организации и функционирования организма (Muotri et al., 2007; Feschotte, 2008; Lisch, 2012). Кроме того, клеточные защитные системы (РНК-интерференция и ДНК-метилирование), которые, как считается, развились в результате борьбы с бесконтрольным распространением МЭ (очередной пример усложнения организации генома вследствие МЭ), теперь играют также важную роль в до- и пост-транскрипционной регуляции экспрессии других генов (Slotkin et al., 2007; Slotkin et al., 2009). Еще одной важной эволюционной ролью МЭ является их участие в качестве агентов захвата и горизонтального переноса (ГП) генетического материала между эволюционно удаленными таксонами. За счет ГП может происходить встраивание и приспособление уже готовых эволюционных решений (ферментативных функций) в новую генную сеть, что позволяет значительно ускорить темпы эволюции генома-реципиента (Schaack et al., 2010; Sormacheva et al., 2012; Soucy et al., 2015).

Таким образом, концепция «эгоистической» и «мусорной» ДНК опирается лишь на ближнесрочный (по большей части негативный) эффект от перемещения МЭ для генома-хозяина и не учитывает роль МЭ в качестве факторов быстрой изменчивости геномов и усложнения их структуры, то есть эволюционных драйверов.

1.2 Ретротранспозоны

Вплоть до 27 июня 1970 года считалось, что центральная догма молекулярной биологии, предложенная еще Ф. Криком (Crick, 1970), является незыблемым законом о направлении потока передачи генетической информации в биологических системах (Finnegan, 2012). Однако две опубликованные сразу друг за другом в тот день статьи в *Nature* (Baltimore, 1970; Temin et al., 1970), описывающие открытие фермента, переводящего РНК в ДНК у онкогенных

вирусов, в буквальном смысле перевернули «догму Крика». Этот фермент вскоре стал известен под названием обратная транскриптаза, а онкогенные РНК-вирусы, у которых он был обнаружен, были переименованы в ретровирусы (Finnegan, 2012).

В то время считалось, что наличие обратной транскриптазы – уникальная особенность ретровирусов. Однако после начала реализации проектов по полногеномному секвенированию оказалось, что количество генов обратной транскриптазы в некоторых геномах эукариот может превышать число любых других белок-кодирующих генов. При этом доля генома, образованная в результате обратной транскрипции, может являться самой крупной его компонентой (Havecker et al., 2004; Finnegan, 2012). Ретровирусы – вирусы позвоночных животных, а гены обратной транскриптазы находили в пределах всего надцарства эукариот: в геномах дрозофилы (*Drosophila melanogaster*), нематоды (*Caenorhabditis elegans*), дрожжей (*Saccharomyces cerevisiae*), покрытосеменного растения (*Arabidopsis thaliana*) и малярийного плазмодия (*Plasmodium falciparum*) (Kidwell, 2002). Впоследствии было выяснено, что основными источниками обратной транскриптазы оказались отнюдь не вирусы, а эндогенные МЭ.

Мобильные элементы, которые подобно ретровирусам кодируют обратную транскриптазу и перемещаются через механизм обратной транскрипции, называют ретротранспозонами. Ретротранспозоны также часто называют МЭ класса I, отделяя их от других МЭ эукариот – ДНК-транспозонов (класс II) (Wicker et al., 2007). Разделение на два класса основано на принципиальных различиях в механизмах перемещения и жизненном цикле ДНК-транспозонов и ретротранспозонов. Жизненный цикл ретротранспозона, как и ретровируса, обязательно опосредован участием его РНК-копии, преобразуемой с помощью обратной транскриптазы в его новую ДНК-копию. Полученная ДНК-копия уже затем встраивается в новое место в геноме, при этом исходная последовательность ДНК ретротранспозона остается на своём прежнем месте. Иногда такой способ перемещения ретротранспозонов (ретротранспозиция) еще называют механизмом «копирования и вставки» (Eickbush et al., 2008; Goodier, 2016). В жизненном цикле ДНК-транспозонов отсутствует стадия РНК-копии, и перемещается либо сама исходная ДНК-последовательность элемента через «вырезание» и встраивание на

новое место в геноме, либо же исходный элемент напрямую служит матрицей для синтеза новой ДНК-копии по механизму «катящегося колеса» (Kapitonov et al., 2001).

Вследствие механизма перемещения, количество копий ретротранспозонов постоянно возрастает, что часто делает их основными вкладчиками в повторенную долю эукариотических геномов (SanMiguel, 1998; Kidwell, 2002; Vitte et al., 2005; Jurka et al., 2007; Cordaux et al., 2009). Так, например, в геномах растений доля ретротранспозонов от размера генома может достигать 85%, а в геноме человека ~40%.

Среди автономных ретротранспозонов (кодирующих собственную обратную транскриптазу) выделяют две основные группы, отличающиеся по структуре входящих в них элементов, принципиальной организацией жизненного цикла и по филогении домена обратной транскриптазы (RT). Это ретротранспозоны с длинными концевыми повторами (Long Terminal Repeats – LTR, LTR-ретротранспозоны) и ретротранспозоны без длинных концевых повторов (non-LTR-ретротранспозоны) (Wicker et al., 2007; Eickbush et al., 2008).

1.2.1 LTR-ретротранспозоны

LTR-ретротранспозоны – первый класс ретроэлементов, обнаруженный у эукариот. По своей структуре они были похожи на ретровирусы, но, в отличие от последних, присутствующих только в геномах позвоночных, LTR-ретротранспозоны широко распространены среди эукариот (Eickbush et al., 2002). Самыми первыми изученными LTR-ретротранспозонами стали *Ty1*-элемент дрожжей *S. cerevisiae* (Boeke et al., 1985) и *copia*-элемент *D. melanogaster* (Mount et al., 1985). Похожие элементы вскоре были обнаружены во многих таксонах эукариот - от слизевиков до растений (Xiong et al., 1990). Учитывая чрезвычайное сходство LTR-ретротранспозонов и ретровирусов, было сразу предположено, что они также перемещаются через молекулу РНК-посредника для формирования своей новой ДНК-копии. Прямым доказательством этому послужила экспериментальная демонстрация исчезновения искусственно встроенного интрона

у новых копий *Ty1*-элемента дрожжей (Boeke et al., 1985). Ядерный сплайсосомный интрон был помещен внутрь *Ty1*-элемента, который затем встроили обратно в геном дрожжей. После выявления новых ДНК-копий, встроенных в геноме, оказалось, что большинство из них полностью потеряло интрон, что говорило о синтезе новой копии ретротранспозона по его РНК-матрице.

1.2.1.1 Структура LTR-ретротранспозонов

Схемы устройств различных представителей LTR-ретротранспозонов и их сравнение с ретровирусами позвоночных представлены на рисунке 1 (подробности см. далее по тексту). Как было указано выше, общей классифицирующей структурной особенностью LTR-ретротранспозонов является наличие прямых концевых повторов (LTR-повторов) (Havecker et al., 2004). LTR(ы) фланкируют белок-кодирующую часть элемента. В них содержатся регуляторные промоторные и терминирующие последовательности транскрипции (Kumar et al., 1999). Обычно короткие 5' и 3' нетранслируемые области (НТО) LTR-ретротранспозона (не включая LTR-повторы) содержат последовательности, с которых происходит начало инициации синтеза минус- и плюс-цепей кДНК. Это праймер-связывающий сайт (PBS – primer binding site) – место посадки тРНК-затравки и полипуриновый тракт (пурин-богатый мотив, устойчивый к деградации от рибонуклеазы H, PPT – polypurine tract) в 5' НТО и 3' НТО, соответственно (Kumar et al., 1999).

Белок-кодирующая часть автономного LTR-ретротранспозона представлена двумя генами – *gag* и *pol*, – которые обычно находятся в составе двух открытых рамок считывания (ОРС). Однако также часто встречается слияние двух ОРС в одну (Kumar et al., 1999; Havecker et al., 2004; Sabot et al., 2006). Первая ОРС (*gag*) кодирует структурный белок Gag, участвующий в формировании нуклеокапсидной капсулы, где проходит обратная транскрипция LTR-ретротранспозона. Вторая ОРС (*pol*) кодирует полипротеин Pol, включающий следующие белковые домены: аспартильную протеазу (PR), обратную транскриптазу (RT), рибонуклеазу H (RNH) и интегразу (INT). PR осуществляет пост-трансляционный процессинг полипротеина Pol, разбивая его на отдельные функциональные белки: собственно PR, двухдоменный фермент с доменами RT и RNH, который производит обратную транскрипцию, и INT, встраивающую полученную копию LTR-ретротранспозона в

новое место в геноме. Также в состав Pol отдельного кластера LTR-ретротранспозонов (хромовирусов) входит дополнительный домен - хромодомен (CHD), слитый в единый белок с доменом INT (Malik et al., 1999b; Novikova, 2009). CHD направляет встраивание ретротранспозона преимущественно в гетерохроматиновые районы генома (Gao et al., 2008).

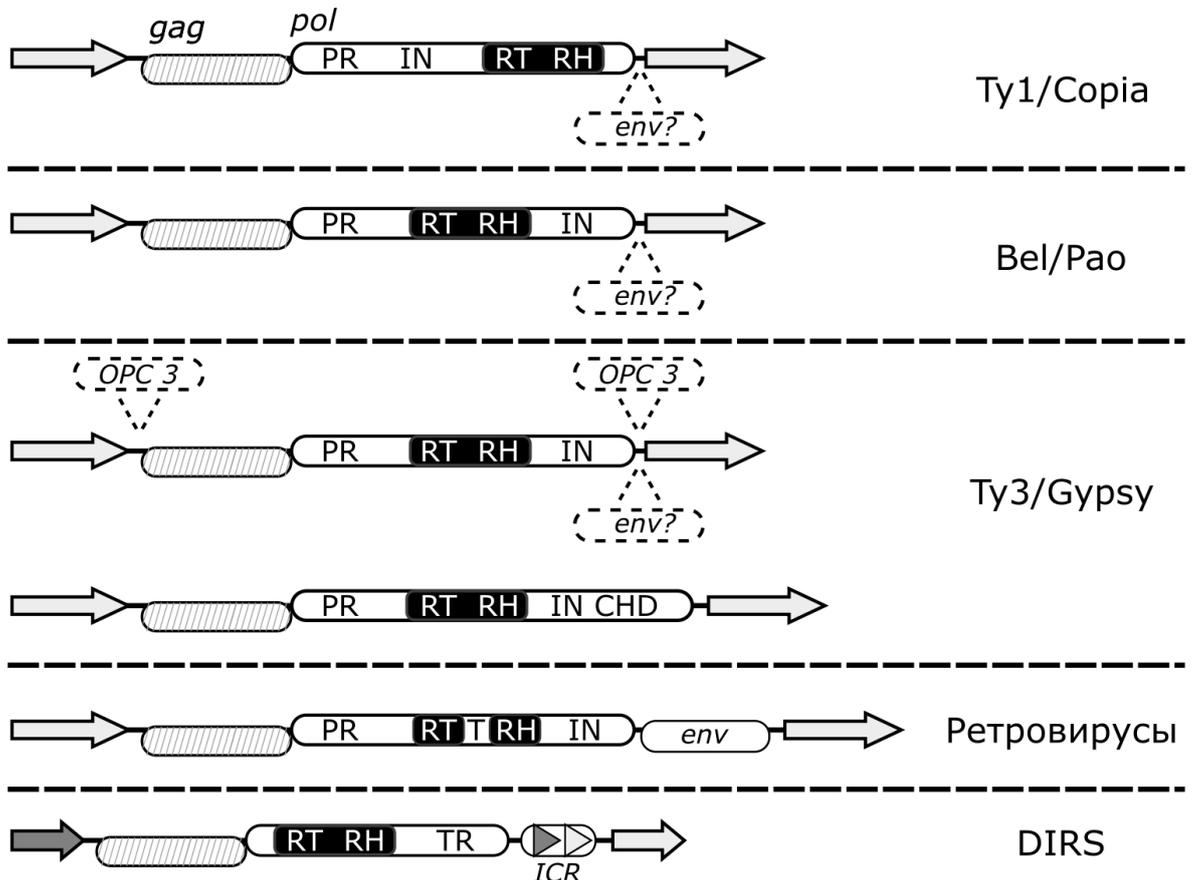


Рисунок 1. Разнообразие структурных вариантов LTR-ретротранспозонов (Eickbush et al., 2008; Steinbauerová et al., 2011). PR – протеаза, RT – обратная транскриптаза, RH – рибонуклеаза H, IN – интеграна, CHD – хромодомен, T – связующий (tether) домен ретровирусов, ICR – внутренние комплементарные повторы, TR – тирозиновая рекомбиназа. LTR(ы) обозначены стрелками. Овалами обозначены открытые рамки считывания. Дополнительные открытые рамки считывания, которых нет у большинства представителей группы, выделены пунктиром.

Кроме двух основных OPC, некоторые LTR-ретротранспозоны также содержат дополнительные кодирующие последовательности. Например, открытые рамки считывания, подобные гену *env* ретровирусов (OPC 3). Аминокислотные последовательности OPC 3 сильно различаются между различными LTR-

ретротранспозонами и, тем более, ретровирусами. Сходство наблюдается в присутствии у таких ОРС мотивов потенциальных трансмембранных доменов и доменов типа coiled-coil, характерных для *env* ОРС настоящих ретровирусов (Havecker et al., 2004). Нет прямых доказательств того, что ОРС 3, обнаруживаемые у отдельных групп LTR-ретротранспозонов, выполняют ту же роль, что и ген *env* ретровирусов. Но всё же принято считать, что функция этих дополнительных ОРС аналогична таковой у ретровирусов, предполагая конвергентное приобретение вирулентных свойств различными группами LTR-ретротранспозонов (Malik et al., 2000b).

Помимо *env*-подобных ОРС, у представителей *Ty3/Gypsy* LTR-ретротранспозонов из кластера *Tat* зеленых растений обнаруживают дополнительные ОРС с неизвестными функциями (Steinbauerová et al., 2011). Такие ОРС могут быть расположены как перед геном *gag*, так и в бессмысловом направлении по отношению к кодирующей последовательности РНК после гена *pol* (Havecker et al., 2004).

1.2.1.2 Жизненный цикл LTR-ретротранспозонов

Данные о полном жизненном цикле для всех LTR-ретротранспозонов во многом получены путем экстраполяции информации по жизненному циклу элементов дрожжей и ретровирусов позвоночных (Boeke et al., 1989; Sabot et al., 2006; Hu et al., 2012).

В жизненном цикле автономных LTR-ретротранспозонов можно выделить следующие фазы (Sabot et al., 2006): (1) транскрипция, (2) трансляция, (3) димеризация и упаковка, (4) обратная транскрипция и (5) встраивание новой копии (Рисунок 2).

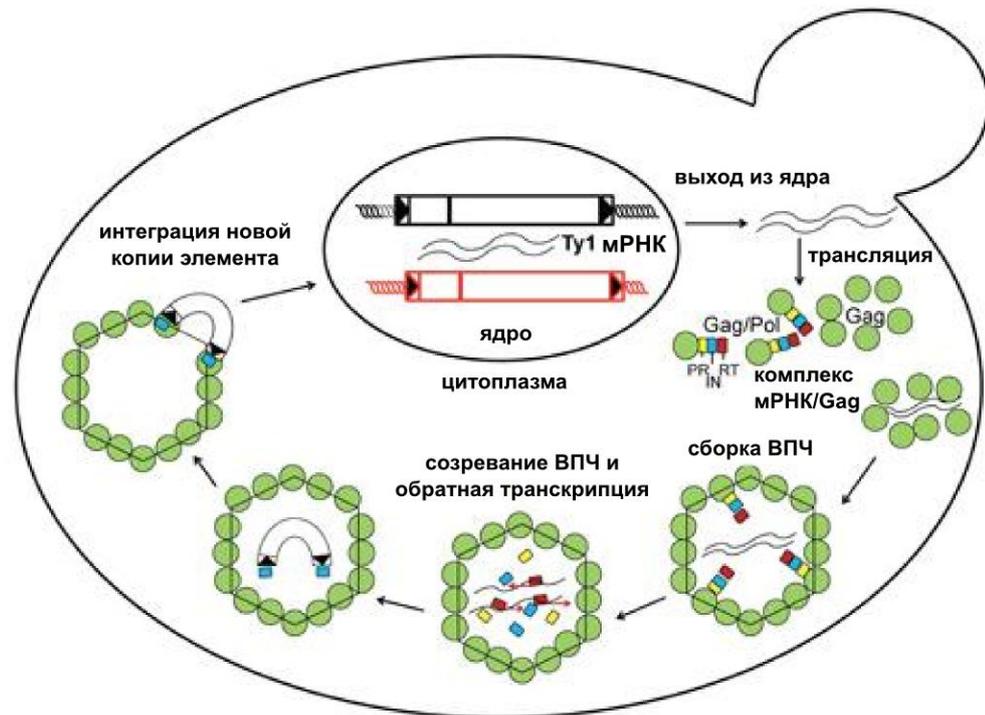


Рисунок 2. Общая схема жизненного цикла LTR-ретротранспозонов на примере *Ty1*-элемента дрожжей (<http://www.bmb.uga.edu/labs/garfinkel>, с модификациям).

(1) Транскрипция

Принято считать, что экспрессия LTR-ретротранспозонов осуществляется по классическому механизму для промоторов РНК-полимеразы II (*polII*). Транскрипция начинается в 5' LTR элемента, содержащем последовательность ТАТА-бокса сразу перед 5' R участком, и продолжается вплоть до 3' R участка 3' LTR (Kumar et al., 1999). В результате образуется бицистронная мРНК, кодирующая как минимум две ОРС – *gag* и *pol*. Ретротранспозоны не содержат интронов, и RT действует на зрелую, сплайсированную, мРНК. Таким образом, транскрипция LTR-ретротранспозона, как и любых других генов, регулируемых *polII*-промотором, зависит от факторов транскрипции, предоставляемых клеткой-хозяином. С данной точки зрения ретротранспозон паразитирует на транскрипционной машине клетки. Образованная мРНК ретротранспозона полиаденилируется, после чего доставляется в цитоплазму, как и любые другие матрицы, полученные с помощью *polII* (Sabot et al., 2006).

(2) Трансляция

С соответствующих ОРС, *gag* и *pol*, с бидистронной мРНК автономных LTR-ретротранспозонов транслируются белки Gag и Pol. Для сборки упаковочной вирусоподобной частицы (ВПЧ) необходимо большее количество структурного белка Gag в стехиометрическом отношении к функциональному белку Pol (Sabot et al., 2006). Такой трансляционный сдвиг может быть достигнут различными путями. Например, при наличии обычного сайта посадки рибосомы в 5' конце от *gag* ОРС и менее эффективного сайта внутренней посадки рибосомы перед *pol* ОРС (Internal Ribosome Entry Site, IRES) (Meignin et al., 2003). У ретровирусов, у которых гены *gag* и *pol* расположены на разных перекрывающихся рамках считывания, достижение необходимого стехиометрического отношения достигается за счет случайного смещения рибосом на вторую рамку во время трансляции (Jacks et al., 1988). Также известны другие способы: наличие короткого повторенного мотива (например, ААААА) между *gag* и *pol*, который приводит к случайному проскальзыванию рибосомы (Jin et al., 1989), а также накопление редких кодонов тРНК в 3' *gag*, приводящих к задержке рибосом и их дальнейшему проскальзыванию на матрицу *pol* (Hull et al., 1995).

Несмотря на очевидную логику наличия двух ОРС для обеспечения необходимой стехиометрии между структурной и функциональной компонентами, некоторые LTR-ретротранспозоны используют более «расточительный» подход при наличии единственной слитой *gag-pol* ОРС. Это приводит к синтезу длинного полипротеина Gag-Pol, который уже впоследствии «нарезается» на функциональные единицы (Sabot et al., 2006). После трансляции Pol у всех LTR-ретротранспозонов происходит его эндопротеолитическое расщепление на функциональные домены с помощью домена PR (Jaaskelainen et al., 1999).

(3) Образование нуклеокапсида, димеризация и упаковка

Gag-белок ретровирусов, по своей сути, – тоже полипротеин, состоящий из трех функциональных доменов, которые впоследствии «нарезаются» из Gag-предшественника. Это капсид (полимеризующий домен), собственно нуклеокапсид (содержит структуры цинковых пальцев и основные аминокислотные остатки для взаимодействия с РНК) и матричный домен (ответственен за связь с белком оболочки ретровирусов). В Gag-белках LTR-ретротранспозонов однозначно

обнаруживают капсид и нуклеокапсид, в то время как имеется лишь очень слабое сходство с последовательностью матричного домена (Jaaskelainen et al., 1999; Sabot et al., 2006).

ВПЧ необходима для формирования среды проведения обратной транскрипции (теоретически) лишь собственной РНК ретротранспозона. ВПЧ образуется вследствие полимеризации Gag, опосредованной его капсидным доменом. Параллельно должна происходить упаковка РНК внутрь ВПЧ. Специфичность упаковки РНК у ретровирусов обеспечивается наличием у неё вторичной структуры на 5' конце перед старт-кодоном Gag (сигнал упаковки). У ретровирусов данная вторичная структура РНК специфически узнается нуклеокапсидным доменом Gag. У LTR-ретротранспозонов наличие сигнала упаковки еще не было экспериментально верифицировано, однако имеется высокий уровень консерватизма вторичной структуры 5' концевого участка их РНК, что предполагает наличие механизма сходного с ретровирусами (Sabot et al., 2006).

В основном РНК ретровирусов представляет собой гомодимер внутри ВПЧ (Brunel et al., 2002). Димеризация происходит либо перед, либо непосредственно во время упаковки в ВПЧ. Для LTR-ретротранспозонов только на *Ty1*-элементе дрожжей было показано образование РНК-димера (Feng et al., 2000). Имеются также косвенные свидетельства димеризации у других LTR-ретротранспозонов, основанные на обнаружении химерных LTR-элементов за счёт процесса смены матрицы во время обратной транскрипции (Vicient et al., 2005; Sabot et al., 2005). Димеризация обусловлена наличием структурного сигнала инициации димеризации, находящимся рядом с сигналом упаковки. Два сигнала димеризации образуют нековалентные связи с формированием так называемой структуры «петли-поцелуя». Также считается, что сигнал инициации димеризации, как и сигнал упаковки, обладает специфичностью к спариванию с соответствующим ему сигналом (Paillart et al., 2004).

(4) Обратная транскрипция и образование комплементарной ДНК

Вся информация по механизму обратной транскрипции взята из обзорной работы Sabot и Schulman 2006 (Sabot et al., 2006). Принципиальная схема процесса изображена на рисунке 3. После упаковки молекул РНК в ВПЧ, они используются

в качестве матрицы для обратной транскрипции. В ВПЧ помимо РНК также должны попасть каталитически-активные белки LTR-ретротранспозона (двухдоменный белок с RT и RNH, а также INT) и молекулы тРНК, которые служат в качестве затравки для начала синтеза минус-цепи комплементарной ДНК (кДНК). Молекула тРНК связывается с PBS, который расположен между концом 5' LTR и началом *gag* и содержит последовательность комплементарную тРНК-затравке (Рисунок 3.1). С затравки синтез кДНК осуществляется в сторону 5' LTR до его R-домена, образуя так называемую strong-stop минус-цепь кДНК (Рисунок 3.2). Параллельно с синтезом минус-цепи домен RNH осуществляет деградацию РНК-матрицы, спаренной с кДНК. Это освобождает кДНК и позволяет произойти первой смене матрицы синтеза – переброске открытого 3' конца кДНК на комплементарную последовательность R-домена 3' LTR (Рисунок 3.3). После переброса, синтез минус-цепи идет до самого конца оставшейся РНК-матрицы - вплоть до участка PBS. Параллельно идет деградация всей РНК-матрицы в дуплексе с кДНК, кроме РРТ, расположенного сразу перед 3' LTR. Свободный 3' конец РРТ служит затравкой для начала синтеза плюс-цепи кДНК по матрице минус-цепи кДНК, продолжающегося вплоть до бывшего PBS – strong-stop плюс-цепь кДНК (Рисунок 3.4). После того как RNH деградирует тРНК-затравку, происходит вторая смена матрицы – переброска открытого 3' конца strong-stop плюс-цепи кДНК на комплементарную ему последовательность PBS на другом конце минус-цепи кДНК. По завершению синтеза плюс- и минус-цепей кДНК и полной деградации РНК-матрицы, включая РРТ, образуется новая ДНК-копия LTR-ретротранспозона с идентичными концевыми LTR (Рисунок 3).

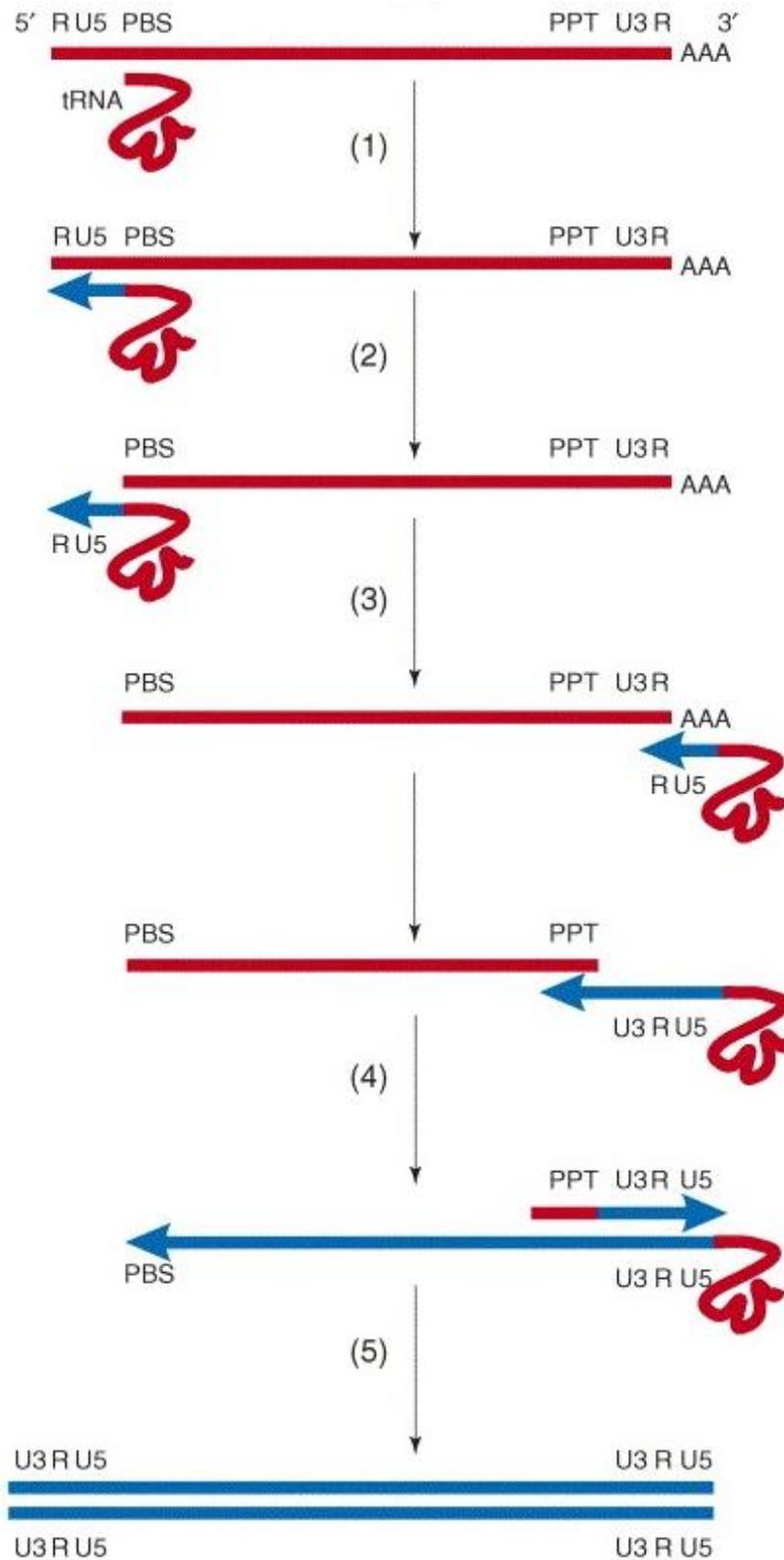


Рисунок 3. Схема механизма обратной транскрипции LTR-ретротранспозона (Holmes et al., 2007) с модификациями. мРНК – красная линия; ДНК – синяя линия; tRNA – тРНК; U3 и U5 уникальные (unique) участки на 3' и 5' концах мРНК; R – повторенные участки на 3' и 5' концах мРНК.

(5) Встраивание новой копии

ВПЧ целиком импортируется в ядро с помощью клеточных систем ядерного транспорта. Импорт ВПЧ внутрь ядра осуществляется за счет наличия сигналов ядерной локализации (nuclear localization signals, NLS) на поверхности Gag-белков, составляющих оболочку ВПЧ (Dang et al., 2000), после чего происходит внутриядерная разборка ВПЧ и кДНК LTR-ретротранспозона, ассоциированная с INT направляется к месту встройки в геном. INT специфично связывается с последовательностями концов LTR. Специфичность определяется наличием мотивов узнавания, расположенных на этих концах. Далее, INT, подобно бактериальным транспозазам, производит асимметричный двунитевой разрыв на геномной ДНК в месте встраивания, который обычно занимает 2-16 п. о. Место разрыва затем лигируется клеточной системой ДНК-репарации, «вшивая» ретротранспозон в геномную ДНК. При этом образуется характерная дупликация сайта встраивания по обеим сторонам от места встройки ретротранспозона (Sabot et al., 2006).

1.2.1.3 Разнообразие и распространение LTR-ретротранспозонов

На основании анализа филогении, реконструированной на основе аминокислотных последовательностей RT-доменов, можно выделить четыре отдельных группы LTR-ретротранспозонов (*Ty3/Gypsy*, *Ty1/Copia*, *BEL* и *DIRS*), а также три родственных им группы вирусов (ретровирусы, гепаднавирусы и каулимовирусы) (Eickbush et al., 2002). Две группы, *Ty3/Gypsy* и *Ty1/Copia* являются наиболее широко распространенными и детально исследованными из четырех групп LTR-ретротранспозонов.

(1) *Ty1/Copia*

Данная группа названа по первым ретротранспозонам, обнаруженным в геномах *S. cerevisiae* и *D. melanogaster* (Boeke et al., 1985; Mount et al., 1985). *Ty1/Copia* представляют наиболее древнюю группу LTR-ретротранспозонов на основании дивергенции последовательности их RT (Рисунок 4) (Eickbush et al., 2002). Основным структурным отличием *Ty1/Copia* от остальных групп LTR-ретротранспозонов и ретровирусов является положение домена INT в

(2) *Ty3/Gypsy*

Эта вторая обширная группа LTR-ретротранспозонов была также названа по соответствующим первым описанным элементам дрожжей и дрозофилы (Marlor et al., 1986; Hansen et al., 1988). Исходя из анализа RT-филогении, *Ty3/Gypsy* значительно более близки к ретровирусам и каулимовирусам, чем представители *Ty1/Copia* (Рисунок 4) (Eickbush et al., 2002). За исключением одной подгруппы (хромовирусов, у которых имеется дополнительный белковый домен в гене полипротеина – CHD) (Novikova, 2009), все элементы группы *Ty3/Gypsy* имеют одинаковую доменную организацию генов *gag* и *pol* с ретровирусами (Рисунок 1) (Havecker et al., 2004). Как и в случае с *Ty1/Copia*, имеется название представленное МКТВ для *Ty3/Gypsy* – Metaviridae. Основываясь на количестве известных на данный момент последовательностей, *Ty3/Gypsy* является самой разнообразной и широко распространенной группой LTR-ретротранспозонов эукариот. *Ty3/Gypsy* находят в геномах различных простейших, животных, грибов и растений (Eickbush et al., 2002; Havecker et al., 2004).

(3) *Bel/Pao*

Название группы происходит от имен первых найденных её представителей: ретротранспозона *Pao* из генома *Bombyx mori* и *Bel* из генома *D. melanogaster* (Xiong et al., 1993a; Davis et al., 1995). До настоящего времени всё известное разнообразие данных ретротранспозонов ограничивается только геномами метазоа. С точки зрения RT-филогении *Bel/Pao* дивергируют первыми от ближайшего общего предка с *Ty1/Copia* (Рисунок 4). Структура известных элементов *Bel/Pao* в общем повторяет структуру *Ty3/Gypsy* и ретровирусов, создавая основу для предположения, что они являются потенциальными предками *Ty3/Gypsy* и ретровирусов (Рисунок 1) (Eickbush et al., 2002). По классификации МКТВ *Bel/Pao* присвоено название Semotiviridae (Havecker et al., 2004).

(4) *DIRS*

Среди всех групп LTR-ретротранспозонов, *DIRS*, пожалуй, является самой необычной. Впервые открытые в геноме слизевика *Dictyostelium discoideum* (Carpello et al., 1985), представители *DIRS* не содержат INT, а вместо этого

кодируют тирозиновую рекомбиназу. Структура их концевых повторов существенно отличается от стандартной структуры остальных LTR-ретротранспозонов и содержит либо внутренние комплементарные прямые повторы, либо инвертированные повторы (Рисунок 1). Несмотря на эти различия, филогенетический анализ RT-доменов помещает группу *DIRS* внутрь LTR-ретротранспозонов, что говорит об уникальном пути эволюции данной группы, в результате которого произошла замена INT доменом тирозиновой рекомбиназы (Рисунок 4) (Eickbush et al., 2008).

Наличие тирозиновой рекомбиназы и то, что многие *DIRS* встраиваются внутрь уже имеющихся геномных копий *DIRS*, легли в основу предположения о рекомбинативном, а не интегративном, механизме встраивания этих ретроэлементов. Представители *DIRS* были обнаружены во многих исследованных геномах эукариот: от простейших до зеленых водорослей, грибов и животных (Eickbush et al., 2008).

1.2.2 non-LTR-ретротранспозоны

Второй обширной группой ретроэлементов, значительно отличающейся от LTR-ретротранспозонов, являются non-LTR-ретротранспозоны. Non-LTR-ретротранспозоны также известны как LINE-элементы (Long Interspersed Elements – длинные диспергированные повторы – так был назван первый открытый non-LTR-элемент человека), ретропозоны и поли(A)-ретротранспозоны (Eickbush et al., 2002). Несмотря на то что эти элементы широко распространены в геномах эукариот, их долгое время не выделяли в собственную группу МЭ. Такое позднее выявление связывают с полным отсутствием non-LTR-ретротранспозонов в геноме дрожжей *S. cerevisiae* (пример редких геномов без non-LTR-ретротранспозонов), на которых до этого были очень детально изучены LTR-ретротранспозоны. Однако наличие поли(A)-хвоста на 3'-конце многих ДНК-копий этих элементов послужило ясным свидетельством в пользу того, что они произошли в результате обратной транскрипции исходной РНК-матрицы. Впоследствии у LINE-элементов приматов были обнаружены ОРС, кодирующие последовательность RT, что послужило первой основой для их выделения в новую группу автономных МЭ. Обнаружение гомологичных последовательностей LINE в геномах насекомых, растений и

трипаносом показало широкую распространенность данных ретроэлементов (Eickbush et al., 2002).

Используя аналогичный подход, который был применен для доказательства ретротранспозиции *Ty1*-элемента дрожжей (Boeke et al., 1985), было продемонстрировано, что встраивание новых копий *I*-элемента *D. melanogaster* и *LINE1* (или, более коротко, *LI*) млекопитающих сопровождается точным исчезновением встроенного интрона (Jensen et al., 1991). После данного открытия практически не осталось сомнений, что non-LTR-ретротранспозоны представляют собой группу автономных МЭ, перемещающихся посредством ретротранспозиции.

1.2.2.1 Структура non-LTR-ретротранспозонов

Схемы структурного устройства различных представителей non-LTR-ретротранспозонов представлены на рисунке 5. Полноразмерный и автономный non-LTR-ретротранспозон обычно содержит одну или две ОРС. Обязательным фактором для обратной транскрипции является обратная транскриптаза, поэтому все автономные (как двух-, так и однорамочные) non-LTR-ретротранспозоны кодируют RT-домен. Так как у двухрамочных элементов RT-домен расположен во второй, начиная с 5'-конца ОРС (как и у LTR-ретротранспозонов), то её принято называть ОРС 2 у всех non-LTR-ретротранспозонов (Han, 2010). Также практически все non-LTR-ретротранспозоны кодируют домен с эндонуклеазной активностью (EN – endonuclease), хотя наличие EN-домена не является обязательным для осуществления полного жизненного цикла non-LTR-ретротранспозона (Malik et al., 1999a). Также ОРС 2 некоторых non-LTR-ретротранспозонов содержит домен RNH (Malik, 2005). Вторая ОРС (ОРС 1), при её наличии, содержит домен с РНК-связывающей активностью, а также шаперонной активностью для нуклеиновых кислот (Hohjoh et al., 1997; Martin et al., 2001; Kolosha et al., 2003). Поэтому, считается, что ОРС 1 играет сходную роль с *gag* ОРС LTR-ретротранспозонов и ретровирусов. 5' и 3' НТО non-LTR-ретротранспозонов достаточно разнообразны. 5' НТО большинства активных элементов содержит промоторную последовательность, однако, по-видимому, имеет место частая замена последовательности промотора у схожих во всем остальном non-LTR-ретротранспозонов, но обитающих в разных видах, из-за чего часто наблюдается

очень низкий консерватизм 5' НТО как по длине, так и по самой последовательности, у non-LTR-ретротранспозонов из разных видов (Schumann et al., 1994; Contursi et al., 1995; DeBerardinis et al., 1999). Среди non-LTR-ретротранспозонов существуют элементы без собственного эндогенного промотора, такие как *R2*, которые находят в генах рибосомной РНК (рРНК). Считается, что *R2*-элементы транскрибируются совместно с повторами рРНК (George et al., 1999). 3' НТО обычно содержит специфические последовательности, структура которых узнается продуктами OPC 2 non-LTR-ретротранспозона (Osanaï et al., 2004; Anzai et al., 2005), хотя в случае элемента *L1* человека наличие 3' НТО необязательно (Moran et al., 1996). 3'-конец non-LTR-ретротранспозона может быть представлен поли(А)-последовательностью или короткими повторами (Chambeyron et al., 2002; Kajikawa et al., 2002). Большинство геномных копий non-LTR-ретротранспозонов укорочены на разную длину с 5' конца, что является следствием не полностью прошедшей обратной транскрипции (Szak et al., 2002).

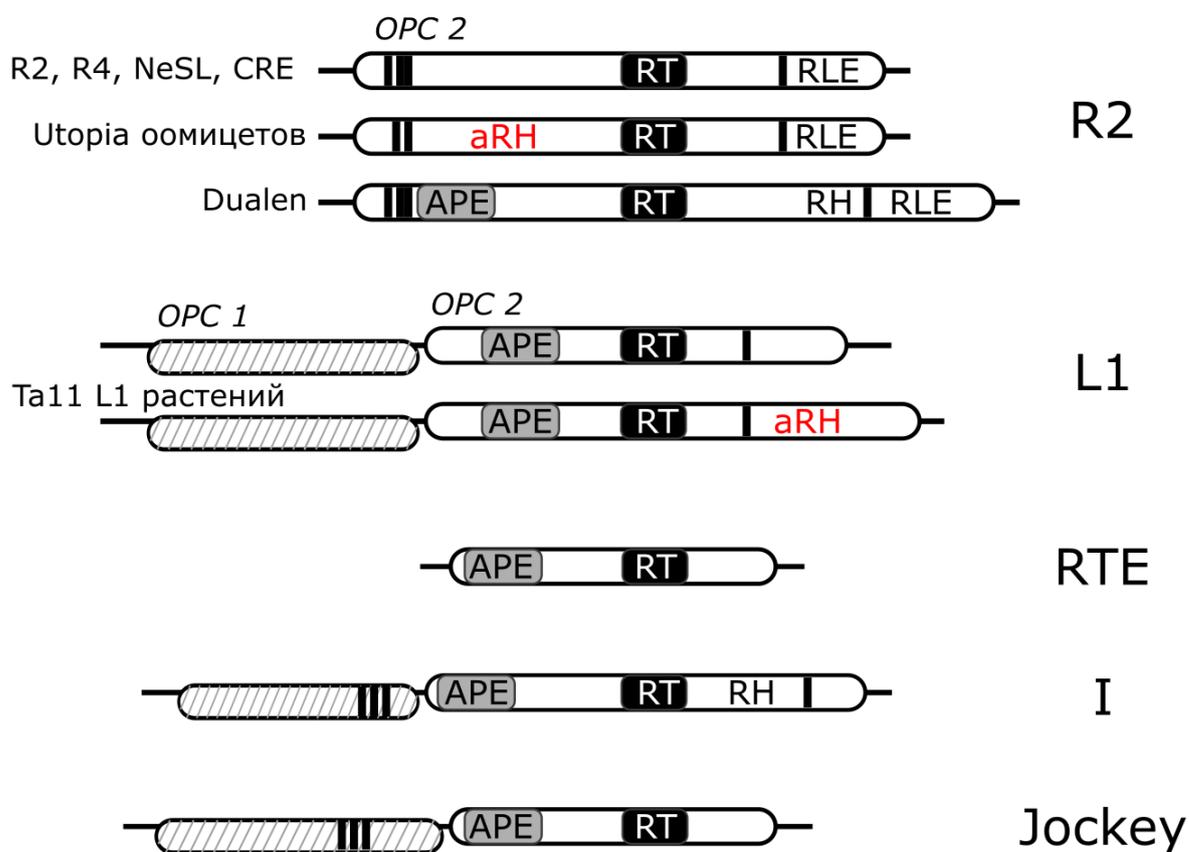


Рисунок 5. Разнообразие структурных вариантов non-LTR-ретротранспозонов (Eickbush et al., 2002; Kojima et al., 2005a; Smyshlyayev et al., 2013; Kojima et al., 2015). RH – рибонуклеаза H; aRH – «археная» рибонуклеаза H (выделено красным);

черные вертикальные полосы – цистеин-гистидиновый ССНС ДНК-связывающий домен. Овалами обозначены открытые рамки считывания.

1.2.2.2 Жизненный цикл non-LTR-ретротранспозонов

Принципиальная организация жизненного цикла non-LTR-ретротранспозонов изображена на рисунке 6.

(1) Транскрипция

Транскрипция большинства non-LTR-ретротранспозонов контролируется эндогенным промотором *polII*, который инициируется синтез мРНК выше 5'-конца элемента (Mizrokhi et al., 1988; Swergold, 1990). Некоторые элементы не имеют собственного промотора, и поэтому возможность их транскрипции случайным образом зависит от того, встроился ли такой ретроэлемент под контроль промотора другого гена, чтобы быть транскрибированным вместе с ним, или нет (Eickbush, 1992). Так, например, экспрессия элементов *R1* и *R2* контролируется работой промотора РНК-полимеразы *I* (*polI*) (George et al., 1999). Нарботанная мРНК кэпируется и полиаденилируется (в случае экспрессии с *polII*-промотора), после чего экспортируется из ядра в цитоплазму (Han, 2010).

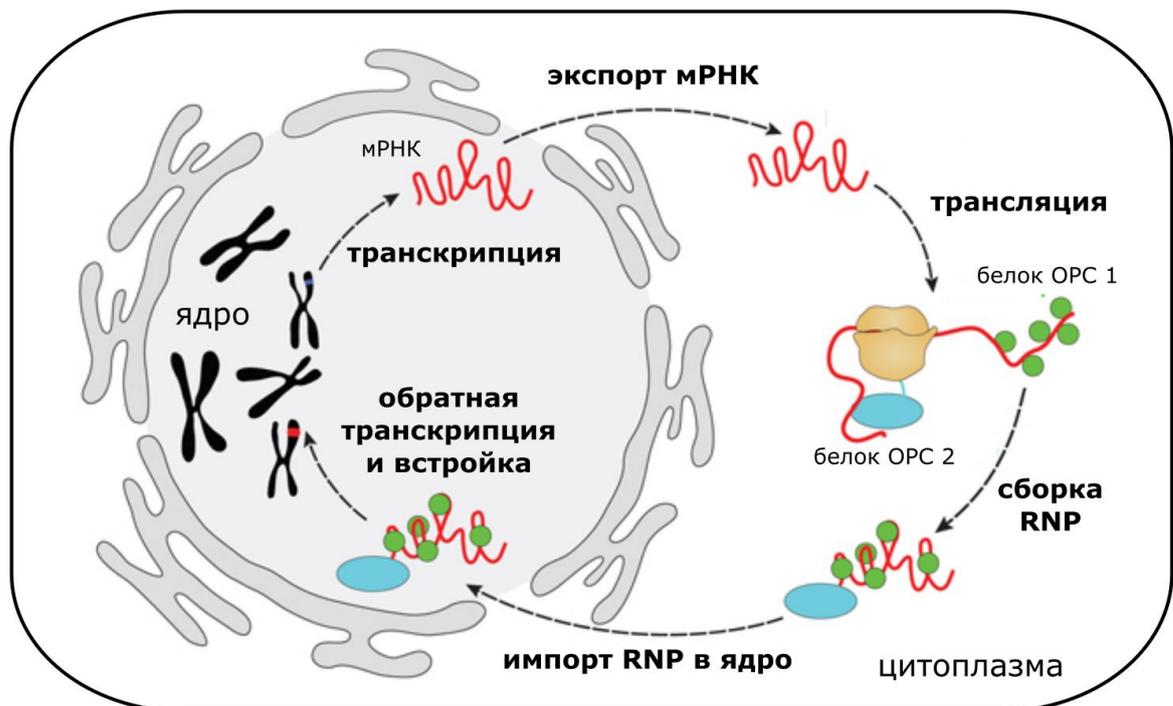


Рисунок 6. Общая схема жизненного цикла non-LTR-ретротранспозонов на примере *L1*-элемента человека (Vincent Dollard, 2010) с модификациями.

(2) Трансляция

Трансляция для *polIII*-транскрибируемых элементов происходит по классической модели с узнаванием кэпа и последующего рибосомного сканирования. Данная модель трансляции (в противоположность модели с IRES) поддерживается результатами экспериментов на млекопитающих, которые демонстрируют, что 5' НТО, 3' НТО и промежуточная последовательность между двумя ОРС *L1*-ретротранспозона не обязательны для полноценного прохождения ретроинверсии элемента (Moran et al., 1996; Han, 2010). Также показано, что кодирующая последовательность ОРС 1 может содержать множественные миссенс и нонсенс мутации, которые не нарушают ход ретроинверсии. При этом была показана зависимость инициации трансляции *L1*-ретротранспозона человека от наличия кэпа перед 5' НТО (Han et al., 2004; Dmitriev et al., 2007). *PolIII*-транскрибируемые *R1*- и *R2*-элементы, по-видимому, транслируются через наличие IRES в 5' НТО (George et al., 1999).

Трансляция ОРС 2 в бицистронных транскриптах проходит за счет нестандартных сигналов инициации. Исследования элемента *SART1* тутового шелкопряда и *L1*-элемента человека предполагают сходные механизмы трансляции, когда трансляция ОРС 2 зависит от трансляции ОРС 1 (Kojima et al., 2005b; Alisch et al., 2006). Рибосома, которая транслирует ОРС 1 может начать транслировать ОРС 2 или же «подозвать» другую рибосому для выполнения данной функции. При этом, при нарушении прохождения трансляции ОРС 1 значительно снижается трансляция ОРС 2 и падает уровень ретроинверсии (Alisch et al., 2006).

Продукты трансляции ОРС 1 и ОРС 2 связываются со своей мРНК, образуя рибонуклеопротеиновую частицу (RNP – ribonucleoprotein particle) (Han, 2010). Показано, что RNP локализуется преимущественно в цитоплазме (Hohjoh et al., 1997; Martin et al., 2001; Han, 2010), хотя известен и яркий обратный пример – RNP теломер-специфичных non-LTR-ретротранспозонов дрозофилид содержатся в основном в ядре (Pardue et al., 2005). В цитоплазме RNP non-LTR-ретротранспозонов обнаруживают в составе стресс-гранул, которые были исходно описаны как хранилище нетранслируемых мРНК (Goodier et al., 2007; Balagopal et

al., 2009). Является ли попадание в стресс-гранулы путем деградации RNP или же это важный этап их жизненного цикла еще не ясно (Han, 2010). Однако определено, что впоследствии RNP non-LTR-ретротранспозона должна попасть обратно в ядро, где происходит обратная транскрипция.

(3) Обратная транскрипция и встраивание non-LTR-ретротранспозона

Для встройки в геном, non-LTR-ретротранспозон начинает копирование своей ДНК в целевой локус по матрице своей мРНК. Механизм, по которому обратная транскрипция происходит напрямую в месте встраивания, называется мишень-инициируемой обратной транскрипцией (target-primed reverse transcription, TPRT) (Рисунок 7) (Eickbush et al., 2002; Han, 2010). Начальные стадии этого процесса были убедительно продемонстрированы в *in vitro* исследованиях R2-элемента тутового шелкопряда (*Bombyx mori*), а более общее описание остальных стадий TPRT было показано в *in vivo* экспериментах на элементах L1 и I человека и дрозофилы, соответственно (Eickbush et al., 2002).

В самом начале домен EN OPC 2 производит однонитевой разрыв в целевом сайте, что приводит к высвобождению 3' гидроксильной группы ДНК, которая затем служит затравкой для синтеза минус-цепи кДНК по матрице мРНК в процессе обратной транскрипции (Рисунок 7.1 и Рисунок 7.2) (Han, 2010). Последовательность сайта встраивания зависит от типа кодируемого EN-домена ретротранспозона. Более «древние» кластеры non-LTR-ретротранспозонов кодируют рестрикционно-подобную эндонуклеазу (Restriction-Like Endonuclease – RLE), которая узнает специфичные последовательности. У более «молодых» кластеров non-LTR-ретротранспозонов RLE заменяется на менее специфичную апуриновую/апириимидиновую эндонуклеазу (Apurinic/apyrimidinic endonuclease-like endonuclease - APE) (Malik et al., 1999a). Несмотря на несомненно важную приспособительную роль EN в формировании 3'-затравки в результате однонитевого разрыва хромосомной ДНК, инициация TPRT может проходить и без участия данного домена (Morrish et al., 2002). Non-LTR-ретротранспозоны *TART*, *HeT-A* и *TANRE*, которые замещают функцию теломеразы у дрозофилы, лишены EN (Pardue et al., 2005). Также было показано, что L1 ретротранспозон человека с делецией EN-домена впоследствии обнаруживается встроенным в концевых

участках хромосом (Morrish et al., 2007). Таким образом, вероятно, что существовала «эпоха», когда non-LTR-ретротранспозоны были лишены EN и встраивались в места спонтанного возникновения разрывов в ДНК (Han, 2010).

Дальнейшие детали механизма TPRT не являются столь определенными. Сразу после, или во время синтеза, минус-цепи кДНК происходит разрыв на второй

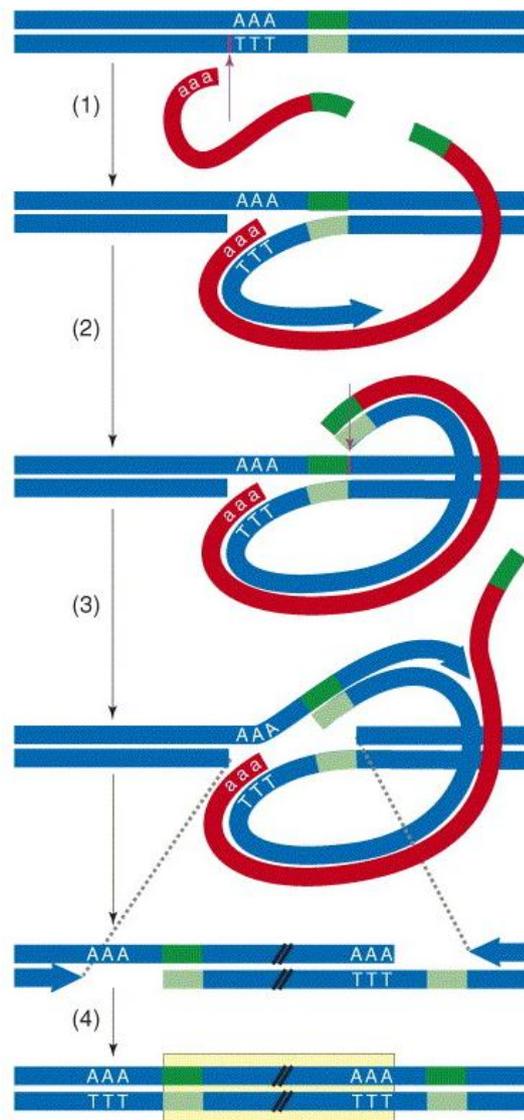


Рисунок 7. Механизм обратной транскрипции non-LTR-ретротранспозонов (Holmes et al., 2007) с модификациями. ДНК – синяя линия; мРНК – красная линия; Зеленым цветом обозначен поли(А)-конец мРНК и соответствующий ему сайт на ДНК.

цепи геномной ДНК выше, ниже или комплементарно месту первого разрыва, что впоследствии приводит либо к дупликации сайта встраивания (target site duplication, TSD), делеции промежутка сайта встраивания или чистой встройке ретротранспозона. Разрыв на второй цепи также служит праймером, но уже для

синтеза плюс-цепи кДНК по матрице минус-цепи кДНК. Считается, что каким-то образом происходит переброс минус-цепи кДНК к месту разрыва, что приводит к началу синтеза плюс-цепи кДНК (Рисунок 7.3). Также неясно, деградирует ли РНК-матрица в процессе замещения её на кДНК плюс-цепи, или же её удаление происходит активно за счет активности RNH, кодируемой геномом хозяина (Han, 2010). Несмотря на это, встречаются non-LTR-ретротранспозоны, которые кодируют собственные домены RNH (Malik et al., 1999a; Eickbush et al., 2002; Kojima et al., 2005a; Smyshlyaev et al., 2013; Kojima et al., 2015). Однако до сих пор не совсем понятно, синтезируется ли плюс-цепь кДНК RT-доменом самого non-LTR-ретротранспозона, или же это происходит за счет работы клеточной системы ДНК-репарации.

По завершению обратной транскрипции разрывы в геномной ДНК лигируются системой ДНК-репарации клетки (Рисунок 7.4) и, в зависимости от положения второго разрыва в геномной ДНК (как уже упоминалось выше), происходят либо TSD, либо делеция участка встраивания (Han, 2010).

1.2.2.3 Разнообразие и распространение non-LTR-ретротранспозонов

На основании анализа филогении, реконструированной на основе аминокислотных последовательностей RT-доменов, можно выделить как минимум 16 кластеров non-LTR-ретротранспозонов, названных по наиболее исследованным их представителям (Рисунок 8) (Eickbush et al., 2002; Kojima et al., 2005a; Kapitonov et al., 2009; Kojima et al., 2015). Однако, данное разделение не столь удобно, т.к. не отражает более глубокой структурной общности между различными кластерами (Eickbush et al., 2002). Поэтому была предложена дополнительная классификация non-LTR-ретротранспозонов, объединяющая различные кластеры как по их эволюционной близости по последовательности RT, так и по наличию/отсутствию и расположению кодируемых функциональных белковых доменов. На основе данного критерия можно разбить все non-LTR-ретротранспозоны на пять широких групп: *R2*, *L1*, *RTE*, *I* и *Jockey* (Eickbush et al., 2002).

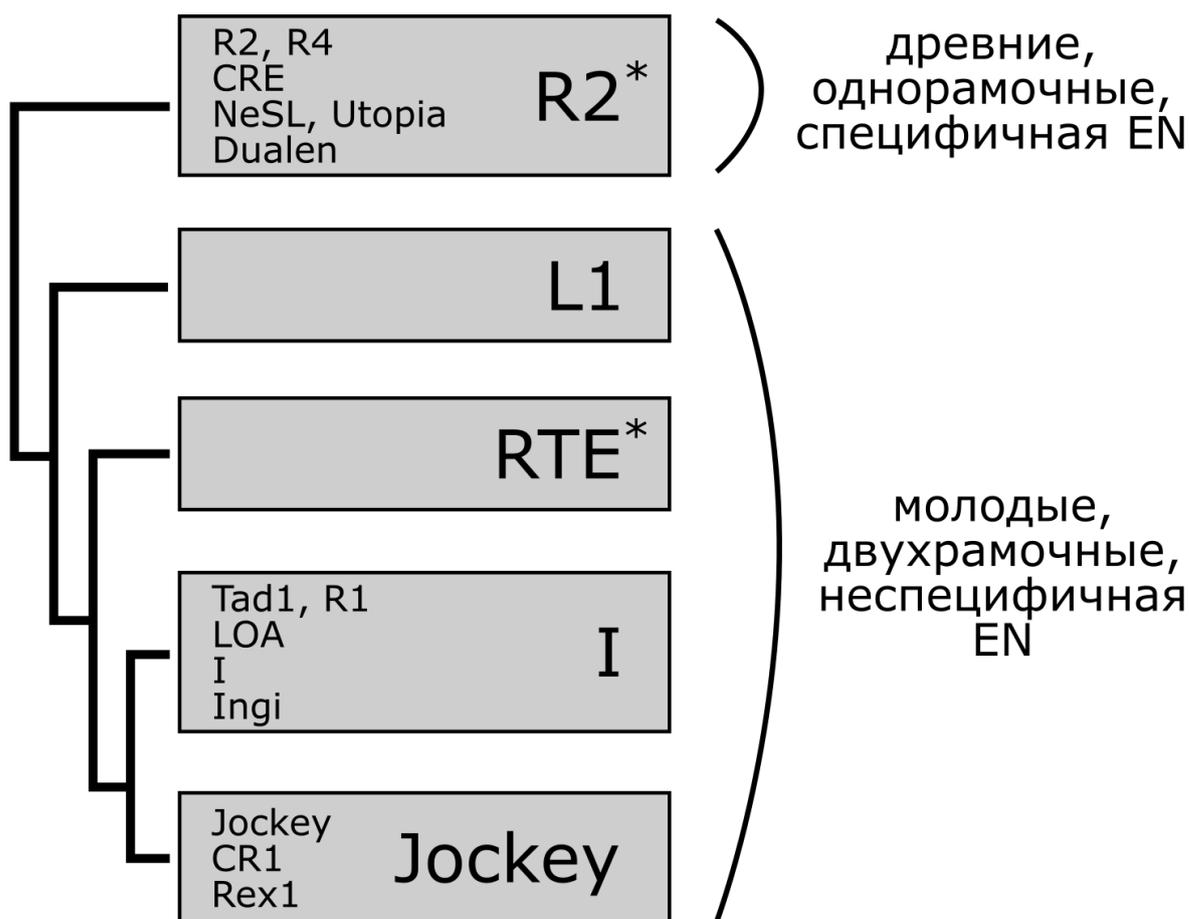


Рисунок 8. Филогенетические взаимоотношения между основными группами non-LTR-ретротранспозонов на основе аминокислотных последовательностей их RT-доменов (Eickbush et al., 2002; Kojima et al., 2005a; Kojima et al., 2015). Структуры элементов для каждой группы изображены на рисунке 5. *R2** - элементы *Dualen* группы *R2* кодируют неспецифическую APE-эндонуклеазу; *RTE** - элементы *RTE* однорамочные, у них нет OPC 1.

(1) *R2*

Группа *R2* объединяет наиболее древние кластеры non-LTR-ретротранспозонов, *R2*, *CRE*, *NeSL*, *R4*, и относительно недавно описанные *Dualen* и *Utopia* (Рисунок 8) (Eickbush et al., 2002; Kojima et al., 2005a; Kojima et al., 2015). Представители этой группы наиболее характерны для геномов беспозвоночных и простейших. Все элементы группы *R2* обладают одной OPC, которая, помимо центрального RT-домена, кодирует С-концевой домен сайт-специфичной эндонуклеазы RLE и ДНК-связывающий домен типа ССНС на N-конце (Рисунок 5) (Yang et al., 1999). Особенной структурой обладают элементы кластеров *Dualen* и *Utopia*, которые

характеризуются дополнительными функциональными доменами (Kojima et al., 2005a; Kojima et al., 2015). Так *Dualen* элементы, помимо RLE, кодируют APE-эндонуклеазу на N-конце, а также домен RNH на N-конце, перед RLE (Kojima et al., 2005a). Отдельные представители *Utopia* из геномов оомицетов кодируют RNH-домен на N-конце OPC (Kojima et al., 2015).

Наличие сайт-специфичной RLE-эндонуклеазы отражает главную особенность характерную для *R2*: элементы этой группы встраиваются в специфические участки генома (Eickbush et al., 2002). Элементы кластеров *CRE* и *NeSL* встраиваются в повторенные гены сплайс-лидерной РНК (Spliced leader RNA, участвует в транссплайсинге) (Blumenthal, 1998), которые найдены у трипаносом, нематод и плоских червей (Gabriel et al., 1990; Aksoy et al., 1990; Malik et al., 2000a). Представители кластеров *R2* и *R4* встраиваются в гены 28S рРНК у членистоногих и нематод (Yang et al., 1999). Отдельные представители *R4* встраиваются в тандемные ТАА-повторы (Xiong et al., 1993b). Недавно открытый кластер non-LTR-ретротранспозонов, *Utopia*, характеризуется преимущественным встраиванием его элементов внутрь генов U2 малых ядерных РНК (Kojima et al., 2015). Из-за наличия неспецифичной эндонуклеазы APE, элементы *Dualen* не имеют преимущественного места встраивания (Kojima et al., 2005a).

Встройка в активно экспрессирующиеся гены отражается в отсутствии промоторов у представителей группы *R2*. Высокая повторенность и дисперсность расположения генов рРНК и сплайс-лидерной РНК должны минимизировать вероятность летального эффекта от встраивания этих non-LTR-ретротранспозонов (Eickbush et al., 2002).

(2) *L1*

Представители группы *L1* широко распространены в геномах эукариот. Их можно обнаружить в геномах различных видов простейших, зеленых водорослей и растений, грибов и животных (включая позвоночных). Все элементы *L1* несут две OPC, где OPC 2 кодирует N-концевую APE-эндонуклеазу, центральный RT-домен и ССНС ДНК-связывающий домен на С-конце (Рисунок 5, Рисунок 8) (Eickbush et al., 2002). У представителей кластера *Ta11* зеленых растений показано наличие С-

концевого домена RNH, который расположен сразу за ССНС-доменом (Wenke et al., 2009; Smyshlyaev et al., 2013).

Ввиду наличия APE, большинство *LI*-элементов не имеют предпочтения по встраиванию в специфические последовательности генома. Однако имеются отдельные примеры, когда такая сайт-специфичность наблюдается, но, скорее всего, вследствие наличия каких-либо иных факторов (Eickbush et al., 2002). Так, элемент *Tx1* гладкой шпорцевой лягушки *Xenopus laevis* встраивается исключительно в многочисленные последовательности ДНК-транспозона *Tx1d* (Garrett et al., 1989). Элементы *DRE* и *Tdd-3* слизевика *D. discoideum* встраиваются рядом с генами тРНК (Szafranski et al., 1999). Элемент *Zepp* зеленой водоросли *Chlorella vulgaris* обнаруживают рядом с теломерными участками хромосом, в результате чего было предположено, что он играет частичную роль в поддержании длины теломер для клетки (Higashiyama et al., 1997).

(3) *RTE*

Представители группы *RTE* также широко распространены среди эукариот (Eickbush et al., 2002). Их обнаруживают в геномах грибов, растений, а также беспозвоночных и позвоночных животных. Характерной особенностью *RTE* является отсутствие OPC 1 и С-концевого ССНС-домена OPC 2 (Malik et al., 1998). В остальном структура OPC 2 соответствует таковой у *LI* (Рисунок 5, Рисунок 8). Для представителей *RTE* не было показано специфичности встраивания.

(4) *I*

Представителей данной группы двухрамочных элементов можно обнаружить в геномах простейших, грибов и беспозвоночных животных. Кластеры *I* филогенетически хорошо обособляются от групп *R2*, *LI* и *RTE*, но не от группы *Jockey* (Рисунок 8) (Eickbush et al., 2002). Выделение пяти кластеров (*Tad*, *I*, *Loa*, *R1* и *Ingi*) в отдельную группу *I* от *Jockey* группы, связано с наличием у них С-концевого домена RNH в OPC 2, филогения которого совпадает с филогенией их RT-домена, что говорит о приобретении RNH-домена общим предком данной группы (Malik et al., 1999a; Eickbush et al., 2002). Отдельные элементы кластеров *R1* и *Tad*, однако, теряют RNH (Malik et al., 1999a). В остальном, структура элементов

группы *I* совпадает с таковой у представителей *L1* (Рисунок 5). Стоит отдельно выделить кластер *R1*, представители которого обладают преимущественной специфичностью встраивания. Так, собственно *R1*-элементы встраиваются в гены 28S рРНК, а элементы *TRAS* и *SART B. mori* встраиваются в теломерные повторы (Besansky, 1990; Okazaki et al., 1995). Для APE-доменов элементов *R1* и *TRAS* была экспериментально показана специфичность к их целевым сайтам (Feng et al., 1998; Anzai et al., 2001).

(5) *Jockey*

На данный момент элементы данной группы (кластеры *Jockey*, *CRI* и *REXI*) были обнаружены в геномах нематод, плоских червей, насекомых и позвоночных. *Jockey*-элементы несут две ОРС, но не имеют С-концевого ССНС-домена в ОРС 2 (Eickbush et al., 2002). В остальном структура этих элементов напоминает структуру представителей *L1* (Рисунок 5, Рисунок 8). Единственные представители *Jockey*, обладающие характерной сайт-специфичностью – элементы дрозофилид *TART* и *HeT-A*. Данные non-LTR-ретротранспозоны замещают функцию фермента теломеразы, отсутствующего у дрозофил, встраиваясь исключительно в теломерные концы хромосом (Levis et al., 1993).

1.2.3 Модульная эволюция ретротранспозонов

Теория модульной эволюции была предложена еще в 1980 г. Дэвидом Ботштейном для вирусов (Botstein, 1980). В ней он предположил, что продуктом эволюции является не отдельный вирус, а совокупность взаимозаменяемых генетических элементов (модулей), каждый из которых несет определенную биологическую функцию. По данной причине многие значительно удаленные по своей морфологии и жизненному циклу вирусы могут успешно рекомбинировать друг с другом, образуя гибриды, составленные из новых комбинаций модулей, которые представляют уже новый вирус.

Ретротранспозоны и другие МЭ также обладают модульным строением (Lerat et al., 1999; Toussaint et al., 2002). Исследование разнообразия и эволюции как LTR-, так и non-LTR-ретротранспозонов говорит о том, что основным двигателем их эволюции является либо замена, либо приобретение новых готовых отдельных

функциональных белковых доменов (модулей). До попадания в состав ретротранспозона эти домены эволюционируют независимо, и их эволюционная история может значительно отличаться от таковой других доменов ретроэлемента, и в частности центрального RT-домена (Lerat et al., 1999; Malik et al., 1999b; Malik et al., 1999a; Malik et al., 2000b; Smyshlyaev et al., 2013).

Конкретный фенотип, т. е. жизненный цикл отдельного ретротранспозона, будет определяться совокупностью составляющих его функциональных доменов. Так, ретровирусы позвоночных приобретают вирулентность за счет приобретения дополнительной ОРС, кодирующей трансмембранный рецептор-связывающий белок оболочки (*env*), хотя во всем остальном (структуре и деталях обратной транскрипции) они практически идентичны *Ty3/Gypsy* LTR-ретротранспозонам (Malik et al., 2000b). Повторное и независимое приобретение генов, подобных *env* ретровирусов, наблюдается в отдельных кластерах *Ty3/Gypsy*, *Ty1/Copia* и *Bel/Pao*, хотя прямых свидетельств их вирулентности так и не было представлено (Eickbush et al., 2002). Подгруппы *Ty3/Gypsy* и *Ty1/Copia* LTR-ретротранспозонов хромовирусы и CoDi-I, соответственно, независимо приобретают CHD-домен, который располагается сразу за доменом INT (Malik et al., 1999b; Novikova, 2009; Llorens et al., 2009). Примечательно, что INT в случае *Ty3/Gypsy* и *Ty1/Copia* расположен в разных местах гена *pol* – после домена RNH и до домена RT, соответственно (Рисунок 1). Важную роль для изменения стратегии встраивания сыграла замена сайт-специфичного RLE-домена эндонуклеазы non-LTR-ретротранспозонов на домен неспецифичной APE-эндонуклеазы (Malik et al., 1999a). Среди прочих особенно выделяется домен RNH, который был независимо и неоднократно приобретен в эволюции различных представителей как non-LTR-, так и LTR-ретротранспозонов (Malik et al., 1999a; Kojima et al., 2005a; Malik, 2005; Wenke et al., 2009; Kojima et al., 2015).

1.2.3.1 Домен рибонуклеазы *H* в эволюции ретротранспозонов

Активность RNH необходима для удаления РНК-матрицы из кДНК/РНК-гибрида, образующегося в процессе обратной транскрипции (см. выше). Филогенетически RNH ретротранспозонов относится к группе RNH типа I. RNH типа I, в свою очередь, включает в себя три крупных группы: RNH грибов и

метазоа (Fungi/Metazoa RNH, fmRNH), RNH LTR-ретротранспозонов и «архейную» RNH (Archaeal-like RNH, aRNH) (Malik et al., 2001; Malik, 2005; Smyshlyaev et al., 2013). Для fmRNH и aRNH характерно наличие аминокислот гистидина (H) или аргинина (R) в одной из консервативных позиций активного центра фермента, однако данные аминокислоты отсутствуют в аналогичной позиции у RNH LTR-ретротранспозонов. aRNH были исходно описаны в геномах архей (Ohtani et al., 2004), а в дальнейшем они были обнаружены в геномах зеленых растений и некоторых бактерий (Wenke et al., 2009; Smyshlyaev et al., 2012). В 2015 году, Kojima и Jurka обнаружили новую подгруппу non-LTR-ретротранспозонов (*Utopia*) в геномах грибоподобных паразитических простейших – оомицетов (Stramenopiles, Oomycetes), – которые также содержали домен aRNH (Kojima et al., 2015).

Ретротранспозоны используют либо фермент RNH, кодируемый геномом хозяина, либо собственные домены RNH (Malik, 2005). К примеру, non-LTR-ретротранспозоны часто опираются на RNH, кодируемую геномом хозяина, так как их обратная транскрипция происходит напрямую в ядре, где уже находится клеточный фермент RNH. Тем не менее, некоторые non-LTR-ретротранспозоны кодируют свою собственную RNH. Эволюция RNH разных кластеров non-LTR-ретротранспозонов группы *I* (см. выше) повторяет эволюцию их RT-домена, что говорит об единичном приобретении общим предком этих групп (Malik et al., 1999a). При этом филогенетически RNH non-LTR-элементов наиболее близка к генам RNH эукариот из группы fmRNH, что согласуется с гипотезой её приобретения из генома хозяина (Malik, 2005). Элементы кластера *Dualen* группы *R2* удалены по филогенетическому положению последовательности RT-домена от элементов группы *I* с RNH (Рисунок 8). Однако последовательность RNH *Dualen* монофилетична с последовательностям доменов RNH остальных групп, что свидетельствует либо о повторном приобретении одного и того же домена RNH, либо о рекомбинации с уже несущими её элементами (Kojima et al., 2005a). Условность наличия RNH у non-LTR-ретротранспозонов проявляется в том, что отдельные представители non-LTR-элементов кластера *R1* из группы *I* впоследствии теряют RNH-домен (Malik et al., 1999a). Однако, отдельные кластеры non-LTR-ретротранспозонов оомицетов (*Utopia*, группа *R2*) и высших растений

(Tall, группа L1) независимо приобрели RNH из группы aRNH (Smyshlyayev et al., 2013; Kojima et al., 2015).

LTR-ретротранспозонам необходимо кодировать собственную RNH, т. к. их обратная транскрипция проходит в цитоплазме, внутри ВПЧ, где отсутствует белок RNH хозяина. Соответственно, домен RNH обнаруживают у всех известных LTR-ретротранспозонов и его эволюция повторяет эволюцию RT-домена (Eickbush et al., 2002; Malik, 2005). Однако некоторые LTR-ретроэлементы, такие как ретровирусы, приобрели дополнительный домен RNH, который функционально заместил исходный ретротранспозонный домен (Malik, 2005). RNH ретровирусов наиболее филогенетически близка к генам RNH эукариот (fmRNH), а не к доменам RNH LTR-ретротранспозонов, что и подтверждает теорию его приобретения.

Множественные независимые приобретения и даже потери домена RNH «делают» его типичным эволюционным функциональным модулем в понимании Ботштейна применимо к ретротранспозонам.

1.2.3.2 Модульная эволюция как конвергентная

Конвергенция в целом подразумевает независимое (не от ближайшего общего предка) приобретение (или развитие) сходных адаптивных изменений у двух генетически удаленных линий за счет которых эти генетические линии могут иметь сходный фенотип (Doolittle, 1994; Stern, 2013; Stayton, 2015). Дальнейшая и более подробная интерпретация термина «конвергенция» зависит от рассматриваемого уровня организации (молекулярного, генового, геномного, клеточного, органно-тканевого, организменного, популяционного и т. п.) и взгляда на это сходство (структурное, функциональное, механистическое) (Doolittle, 1994). Таким образом, может идти речь как о конвергенции белковых молекул, приводящей к возникновению у них сходной укладки (структурная молекулярная конвергенция) и/или к сходству выполняемых ими функций (функциональная конвергенция), так и структурно-функциональной конвергенции организмов (возникновения летательных органов типа крыла у насекомых, птиц и летучих мышей) (Doolittle, 1994; Stern, 2013).

В случае генетически удаленных ретротранспозонов, независимо приобретших сходные функциональные домены в аналогичном положении, речь идет о структурно-функциональной конвергенции между ними. Согласно теории модульной эволюции, «фенотип» ретротранспозона или вируса определяется совокупностью составляющих его функциональных модулей (Botstein, 1980; Lerat et al., 1999). Поэтому, повторное возникновение этой совокупности у удаленных линий ретротранспозонов, будет обозначать возникновение сходного «фенотипа». В случае ретротранспозонов «фенотип» – это организация жизненного цикла. Значит, выявление структурно-функциональной конвергенции между ретротранспозонами может дать основу для выявления факторов давления отбора, приводящих к формированию сходных адаптаций (Stern, 2013) жизненного цикла у различных групп ретроэлементов.

1.3 Ретротранспозоны оомицетов и зеленых растений с доменом «архейной» рибонуклеазы H

Как уже упоминалось выше, группы *I* и *R2* (кластер *Dualen*) non-LTR-ретротранспозонов несут монофилетичный домен RNH, который филогенетически близок к группе fmRNH. В работах по изучению разнообразия и распространения ретротранспозонов в геномах паразитических простейших – оомицетов (*Stramenopiles*) – и зеленых растений (*Viridiplantae*) были найдены элементы с доменом RNH, наиболее близким к группе aRNH. У оомицетов это подгруппа однорамочных «древних» элементов с сайт-специфичной RLE-эндонуклеазой группы *R2* кластера *Utopia* (Kojima et al., 2015). У высших зеленых растений это кластер *Ta11* элементов из группы двухрамочных non-LTR-ретротранспозонов *L1* (Smyshlyaev et al., 2013). Для остальных представителей *Utopia* и *L1* наличие какого-либо домена RNH показано не было. Элементы *Utopia* оомицетов характеризуются наличием aRNH на N-конце ОРС, в то время как *Ta11 L1*-элементы зеленых растений несут aRNH на C-концевой части белка, как и другие non-LTR-ретротранспозоны с RNH (Рисунок 5, Рисунок 8). Очевидно раздельное положение доменов aRNH и fmRNH на филогенетическом древе исключает

возможность приобретения домена за счет рекомбинации с другими элементами, содержащими RNH (Smyshlyaev et al., 2013; Kojima et al., 2015).

Non-LTR-ретротранспозоны с монофилетичным доменом RNH из филогенетической группы fmRNH (группа *R2 (Dualen)* и группа *I (I, Ingi, R1, Tad, LOA)*) распространены в геномах простейших, членистоногих, моллюсков и грибов (Eickbush et al., 2002). Таким образом, распространение non-LTR-элементов с aRNH не перекрывается с распространением других известных non-LTR-ретротранспозонов с доменом RNH. Филогенетический анализ показал, что домены aRNH *Utopia* оомицетов при этом оказались достаточно далеки по своим последовательностям от aRNH *Tall L1*-элементов зеленых растений (Kojima et al., 2015). Однако разрешающая способность филогенетического анализа доменов RNH остается достаточно низкой, чтобы точно опровергнуть возможную связь между соответствующими доменами aRNH двух удаленных групп non-LTR-ретротранспозонов. В случае *Tall L1*-элементов зеленых растений достаточно уверенно можно предположить, что домен aRNH был приобретен из генома своих хозяев, т. к. гены RNH зеленых растений также принадлежат к «архейному» кластеру (Smyshlyaev et al., 2013). В случае с оомицетами подобной проверки на филогенетическую принадлежность их генов RNH проведено не было, однако авторы, открывшие *Utopia*, также склоняются в сторону гипотезы приобретения домена aRNH из генома хозяина (Kojima et al., 2015). Однако, низкое разрешение филогении по aRNH опять же не позволяет окончательно определить источник появления домена aRNH даже в случае *Tall*-элементов растений. Филогенетические кластеры доменов aRNH *L1* ретротранспозонов не показали значимой кластеризации с последовательностями aRNH генов растений (Smyshlyaev et al., 2013).

Обнаружение aRNH в паре таксонов, как растения и оомицеты, представляет отдельный интерес. Раньше оомицетов на основании сходства морфологии относили к грибам за счет наличия характерного мицелиального строения, осмотрфного типа питания, присутствия клеточной стенки и сходства экологических ниш (Beakes et al., 2012). Однако это было неверное предположение. Во-первых, биохимический и цитологический анализы показали, что в отличие от

грибов, оомицеты обладают целлюлозной, а не хитиновой клеточной стенкой и их гаметы образуются напрямую через мейоз диплоидной стадии, а митохондрии оомицетов содержат трубчатые кристы. Во-вторых, проведенный филогенетический анализ на основе генов рибосомной РНК показал, что оомицеты относятся абсолютно к другому надцарству эукариот – так называемых хромальвеолят (*Chromalveolata*) или, иначе, SAR (*Stramenopiles*, *Alveolata* и *Rhizaria*) (Cavalier-Smith et al., 2006; Beakes et al., 2012). Группа *Stramenopiles*, куда относятся оомицеты, объединяет «простейших» гетеро- и автотрофных одноклеточных организмов, а также единственную группу с многоклеточными представителями за пределами царств грибов, животных и растений – бурых водорослей. Считается, что вся группа *Stramenopiles* происходит от одноклеточного фотосинтезирующего предка с вторичными пластидами, полученными при неполном поглощении древних красных водорослей (вторичный эндосимбиоз) (Harper et al., 2005; Cavalier-Smith et al., 2006). Все оомицеты – гетеротрофы, у них отсутствуют следы пластид, однако сравнительный геномный анализ с диатомовыми водорослями показал, что они еще сохраняют общее множество генов, ответственных за фотосинтез (Tyler et al., 2006; Lamour et al., 2007). Многие оомицеты занимают экологическую нишу паразитов, хотя имеются и свободноживущие сапротрофные и бактериотрофные представители. Различные группы оомицетов паразитируют на красных и бурых водорослях, нематодах, беспозвоночных и позвоночных животных и наземных зеленых растениях (Beakes et al., 2012). В качестве паразитов последних оомицеты выступают в роли важных сельскохозяйственных вредителей, наносящих ущерб множеству выращиваемых человеком растительных культур (Webber et al., 2015).

Паразитизм является единственным прямо связующим звеном между отдельными таксонами оомицетов и растениями (Beakes et al., 2012). Постоянное и тесное взаимодействие паразита с его хозяином может служить основой для ГП генетического материала из одного организм в другой на одном из этапов их коэволюции (Wijayawardena et al., 2013; Soanes et al., 2014). Не исключено, что либо ген *aRNH*, либо уже какой-то несущий его предковый ретротранспозон, были горизонтально перенесены из растения в оомицет (или наоборот). Также

неизвестно, ограничивается ли разнообразие ретротранспозонов с aRNH в геномах растений и оомицетов лишь уже описанными кластерами элементов *Tall L1*, и *Utopia* (Smyshlyaev et al., 2013; Kojima et al., 2015). Для поиска ответа на вопрос происхождения домена aRNH у ретротранспозонов растений и оомицетов (ГП или вертикальная эволюция) требуется проанализировать разнообразие, распространение и эволюцию ретротранспозонов в этих геномах. Полученные данные должны помочь понять роль домена aRNH как функционального модуля для эволюции ретротранспозонов.

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ

Базы данных и конкретные геномные последовательности, использованные для анализа

Геномные последовательности зеленых растений и оомицетов были получены из открытых баз данных, указанных в приложении 1. Последовательности известных аннотированных ретротранспозонов были взяты из базы данных повторенных последовательностей Repbase Update v. 20.08 (RU) (Bao et al., 2015).

Компьютерный поиск ретротранспозонов с доменом «архейной» рибонуклеазы H в базе данных повторенных последовательностей Repbase Update

Полная база данных RU была загружена и проанализирована на наличие в ней известных ретротранспозонов, содержащих домен aRNH. Сигналы aRNH искали с помощью специфичного профиля скрытой марковской модели (HMM, Hidden Markov Model) в приложении hmmsearch из пакета HMMER (Eddy, 2011) в предварительно протранслированных по трем рамкам считывания в прямом и обратно-комплементарном направлении последовательностях RU. Профиль aRNH был сконструирован на основе множественного выравнивания аминокислотных последовательностей aRNH, взятого из работы Smyshlyayev et al. (Smyshlyayev et al., 2013). Повторы, в последовательностях которых программой hmmersearch не было обнаружено значимого ($e\text{-value} < 1e\text{-5}$) сходства с HMM-профилем aRNH, отфильтровывали. Оставшиеся RU-повторы были сначала сгруппированы по таксономическому положению организма, из которого они происходят, а затем по типу повторов согласно классификации RU.

Компьютерный поиск последовательностей ретротранспозонов с доменом «архейной» рибонуклеазы H в геномных последовательностях зеленых растений и оомицетов

Для поиска всех возможных ретротранспозонов, содержащих домен aRNH, в геномах растений и оомицетов был использован следующий алгоритм поиска. Во-первых, основываясь на HMM-профиле aRNH, в программе hmmsearch пакета

HMMER (Eddy, 2011), искали сигналы aRNH в предварительно протранслированных по трем рамкам считывания в прямом и обратнокопплементарном направлении геномных последовательностях. Во-вторых, после обнаружения значимого сходства с HMM-профилем aRNH ($e\text{-value} < 1e\text{-}5$), если такое было возможно, искали сигналы RT-домена с помощью соответствующих HMM-профилей RT LTR- и non-LTR-ретротранспозонов во фланкирующих последовательностях на расстоянии до 10000 пар нуклеотидов от локализации сигнала aRNH. HMM-профиль, специфичный к доменам RT non-LTR-ретротранспозонов, был сгенерирован на основе множественного выравнивания аминокислотных последовательностей RT доменов, которое получается на выходе программы RTclass1 (Kapitonov et al., 2009). HMM-профиль, специфичный к доменам RT LTR-ретротранспозонов, был сконструирован на основе множественного выравнивания аминокислотных последовательностей RT, доступных на сайте базы данных Gypsy Database (GyDB) (Llorens et al., 2011). В-третьих, производили разделение последовательностей со значимым сходством к одному из двух RT HMM-профилей ($e\text{-value} < 1e\text{-}5$) на две группы, соответствующие LTR и non-LTR-ретротранспозонам. Оставшиеся последовательности aRNH были отфильтрованы и сохранены для отдельного анализа, но уже как индивидуальные (не из ретротранспозона) домены aRNH. В-пятых, первично кластеризованные последовательности ретротранспозонов были классифицированы по родству к уже известным последовательностям ретротранспозонов, доступных в RU с помощью программы tBLASTx. На последнем этапе для каждой группы гомологии (как *Ty3/Gypsy*, *L1* или *Utopia*) отбирали последовательность-представитель для дальнейшего анализа её структуры и сравнительного и филогенетического анализов.

Компьютерный поиск последовательностей потенциальных клеточных генов рибонуклеаз H в геномных последовательностях зеленых растений и оомицетов

Помимо HMM-профиля, специфичного к последовательностям aRNH, был сконструирован HMM-профиль для генов RNH из филогенетической группы грибов и метазоа (fmRNH), так как заранее не было известно – относятся ли гены

RNH оомицетов к группе aRNH или fmRNH. Профиль fmRNH был сконструирован на основе множественного выравнивания аминокислотных последовательностей fmRNH из работы Smyshlyayev et al. (Smyshlyayev et al., 2013). Данный HMM-профиль был использован для поиска сигналов fmRNH в транслированных геномных последовательностях в программе hmmsearch (Eddy, 2011). Найденные последовательности со значимым сходством к fmRNH ($e\text{-value} < 1e\text{-}5$) отбирались для дальнейшего сравнительного и филогенетического анализов генов и доменов RNH.

Определение структуры ретротранспозонов с архейной рибонуклеазой H

Для каждой отобранной последовательности-представителя ретротранспозона был проведен детальный анализ его структуры. Программа NCBI ORFfinder (NCBI Open Reading Frame finder,) была использована для поиска OPC, а программы NCBI CD-search (Marchler-Bauer et al., 2011) и HHpred (Söding et al., 2005) применяли для поиска гомологии к известным консервативным белковым доменам. Для представителей LTR-ретротранспозонов были дополнительно определены последовательности их LTR с помощью выравнивания 5' и 3' областей, фланкирующих OPC 1 и OPC 2 в программе BLAST (Altschul et al., 1990). Предсказание вторичной структуры для последовательностей aRNH производили в программе Quick2D из биоинформационного пакета программ MPI (Alva et al., 2016).

Сравнительный и филогенетический анализы

Аминокислотные последовательности RT-доменов отобранных представителей LTR- и non-LTR-ретротранспозонов растений и оомицетов, содержащих aRNH, были выровнены на соответствующие HMM-профили с помощью программы hmalign пакета HMMER (Eddy, 2011). Существенно менее консервативные, чем последовательности RT, аминокислотные последовательности RNH были выровнены с помощью программы PROMALS3D, которая использует данные о предсказании вторичной структуры для оптимизации локальных участков выравнивания (Pei et al., 2008). На основе полученных выравниваний были реконструированы соответствующие филогенетические деревья LTR- и non-LTR-ретротранспозонов, а также генов и доменов RNH.

Реконструкцию филогении проводили методом максимального правдоподобия и байесовской реконструкцией в программах PhyML (Guindon et al., 2010) и MrBayes (Ronquist et al., 2012), соответственно. Наиболее оптимальная эволюционная модель для реконструкции филогении, LG + G, была определена на основе Информационного критерия Акаике (AIC) и Байесовского информационного критерия (BIC) в программе ProtTest (Darriba et al., 2011) для каждого из полученных аминокислотных выравниваний. В программе PhyML наиболее оптимальная топология филогенетического дерева была выбрана из 100 случайных стартовых деревьев с применением алгоритма SPR. Из полученных 100 реконструкций выбиралась одна с максимальным значением логарифма максимального правдоподобия, а статистическая значимость реконструкции была определена с помощью Байесовской трансформации теста aLRT (aBayes transformation) в программе PhyML (Anisimova et al., 2011). В программе MrBayes 10 отдельных марковских цепей Монте Карло (MCMC) были запущены на 2500000 поколений с отбором каждые 250 поколений. Перед реконструкцией финального консенсусного дерева первые 5000 отобранных образцов были отфильтрованы.

ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Разнообразие и распространение ретротранспозонов с доменом «архейной» рибонуклеазы H в геномах растений и оомицетов

Анализ разнообразия и распространения ретротранспозонов с доменов «архейной» рибонуклеазы H в базе данных Repbase Update

Домен aRNH известен у non-LTR-ретротранспозонов растений (*Ta11*, группа *L1*) (Smyshlyaev et al., 2013) и оомицетов (*Utopia*, группа *R2*) (Kojima et al., 2015). Для того чтобы проверить наличие домена aRNH у других мобильных элементов, мы проанализировали базу данных повторенных последовательностей RU (см. главу 2. Материалы и методы).

Все значимые сигналы aRNH относились к последовательностям ретротранспозонов из RU, аннотированных в геномах 42 видов зеленых растений и пяти видов оомицетов из рода *Phytophthora* (Oomycetes, Peronosporales) (Приложение 2). Согласно классификации, используемой в базе данных RU, ретротранспозоны с aRNH встречаются в следующих группах: “LTR/Gypsy”, “Non-LTR/Utopia” и “Non-LTR/L1”. То есть помимо non-LTR-ретротранспозонов, aRNH встречается в составе LTR-ретроэлементов из группы *Ty3/Gypsy*. Причем, LTR-ретротранспозоны с aRNH были обнаружены как у растений, так и у оомицетов. Неожиданным также оказалось обнаружение non-LTR-ретротранспозонов из группы *L1*, содержащих домен aRNH, не только у растений, но и у оомицетов. Non-LTR-элементы *Utopia* были найдены только у оомицетов (Приложение 2). Таким образом, в базе данных RU ретротранспозоны с доменом aRNH представлены LTR-элементами из группы *Ty3/Gypsy* и non-LTR-элементами из группы *L1* у зеленых растений и LTR-элементами *Ty3/Gypsy* и non-LTR-элементами из групп *L1* и *R2* (кластер *Utopia*) у оомицетов.

Ретротранспозоны с aRNH в RU появляются у зеленых растений, начиная с отдела плауновидных (Lycopodiophyta), с единственным аннотированным представителем – *Ty3/Gypsy* LTR-ретротранспозоном из генома *Selaginella moellendorffii*. Далее, *Ty3/Gypsy* LTR-элементы и *L1* non-LTR-элементы встречаются у представителей хвойных голосеменных (Gymnospermae, Pinacea) и

цветковых растений (Angiospermae), причем как у однодольных (Liliopsida), так и у двудольных (Magnoliopsida) (Приложение 2).

Как уже упоминалось выше, всё разнообразие оомицетов в базе данных RU было представлено элементами из геномов рода *Phytophthora* – облигатные паразиты зеленых растений. Соответственно, всё найденное разнообразие ретротранспозонов с aRNH в RU ограничено этим родом (Приложение 2).

Анализ разнообразия и распространения ретротранспозонов с доменом «архейной» рибонуклеазы H в геномных последовательностях растений и оомицетов

Для получения более широкой и детальной картины о разнообразии и распространении ретротранспозонов с aRNH был проведен биоинформационный поиск таких элементов в геномных последовательностях из максимально доступного на данный момент разнообразия таксонов зеленых растений и оомицетов (Приложение 1). Так как разнообразие и распространение *L1*-элементов растений с aRNH уже было детально изучено (Smyshlyaev et al., 2012; Smyshlyaev et al., 2013), в настоящей работе внимание было уделено именно LTR-ретротранспозонам с доменом aRNH из геномов зеленых растений. В работе Коджима и Юрка (Kojima et al., 2015), в которой были открыты элементы *Utopia*, было проанализировано лишь семь геномов оомицетов, пять из которых принадлежит к одному роду (*Phytophthora*). Поэтому в нашей работе, ввиду значительно большего разнообразия проанализированных геномов, уделено внимание разнообразию и распространению всех групп ретротранспозонов с aRNH у оомицетов.

Всего было проанализировано 65 геномов зеленых растений и 25 геномов оомицетов, соответственно (Приложение 1). Присутствие LTR-ретротранспозонов с aRNH было обнаружено для 43 из 65 исследованных геномов растений, а представители хотя бы одной из трех групп ретротранспозонов с aRNH (*Ty3/Gypsy*, *L1* или *Utopia*) были найдены в 21 из 25 исследованных геномов оомицетов (Рисунок 9).

LTR-элементы с aRNH не были обнаружены в исследованных геномах зеленых водорослей и мхов и, как и в случае с анализом базы данных RU, они появляются, начиная с плауновидных растений, встречаясь далее у представителей всех крупных таксонов зеленых растений (Рисунок 9). В тоже время *LI*-элементы с aRNH начинают встречаться, начиная с таксонов голосеменных хвойных растений.

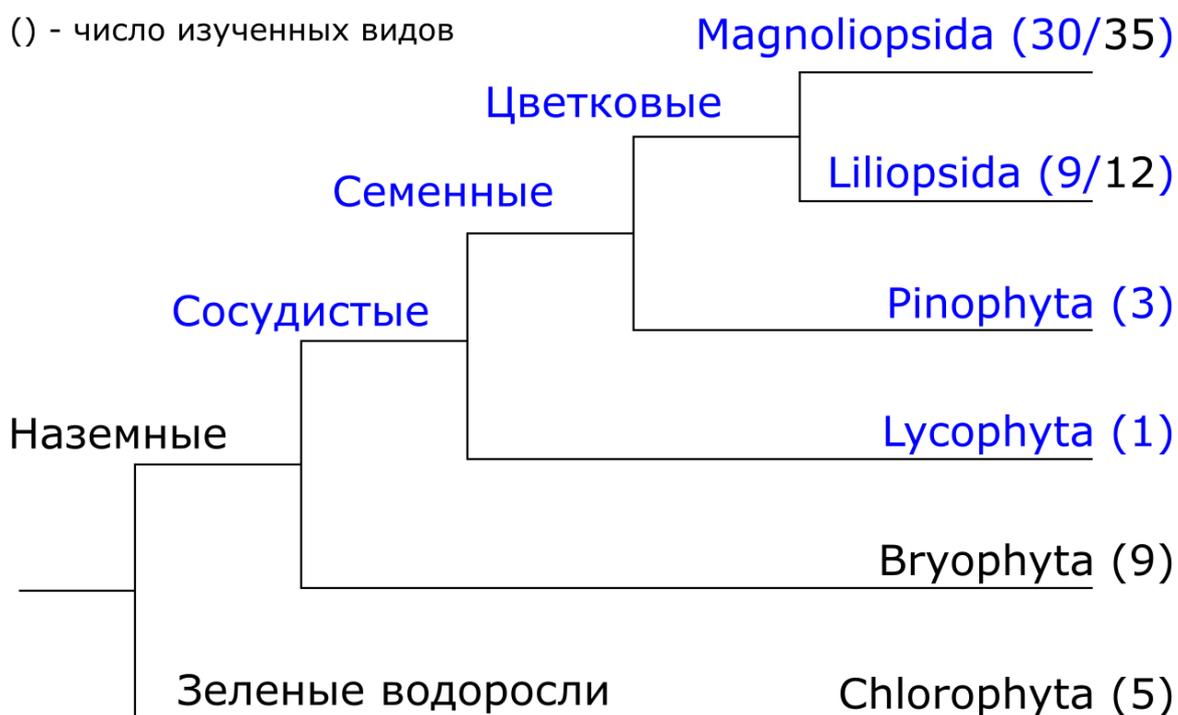


Рисунок 9. Схема филогенетических взаимоотношений и разнообразие исследованных геномов зеленых растений. Синим цветом выделены названия таксонов, в которых были найдены ретротранспозоны с aRNH.

В базах данных были доступны геномные последовательности четырех отрядов оомицетов, занимающих три различных экологических ниши (Beakes et al., 2012): Albuginales, Peronosporales, Pythiales и Saprolegniales (Рисунок 10, Приложение 1). В геномах оомицетов ретротранспозоны с aRNH встречаются уже в наиболее рано дивергировавшем отряде Saprolegniales. Хотя и в штучном количестве, у Saprolegniales были найдены как *Ty3/Gypsy* LTR-элементы с aRNH, так и non-LTR-элементы *Utopia*. У представителей Albuginales не удалось обнаружить каких-либо ретротранспозонов с aRNH, а уже в геномах отрядов Pythiales и Peronosporales встречаются все три группы ретроэлементов с aRNH,

исходно идентифицированные при анализе базы данных RU: *Ty3/Gypsy* LTR-элементы и *L1* и *Utopia* non-LTR-ретротранспозоны.

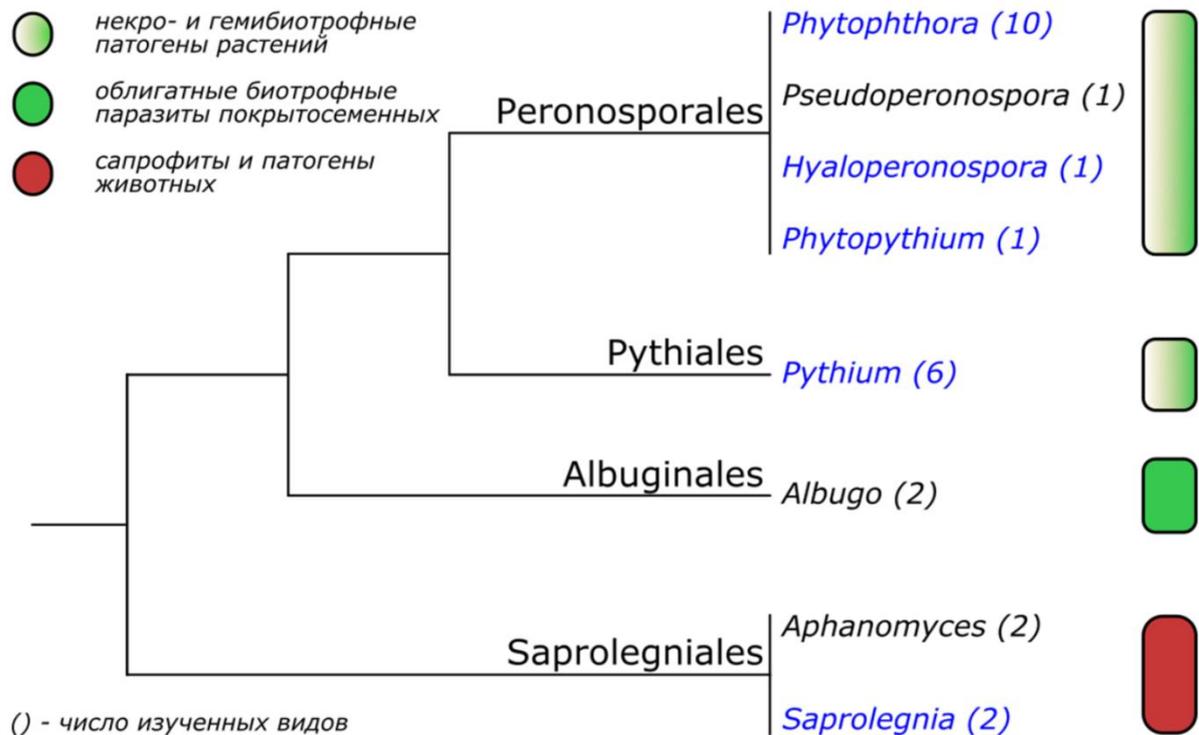


Рисунок 10. Схема филогенетических взаимоотношений и разнообразие исследованных геномов оомицетов из различных экологических ниш. Синим цветом выделены названия родов, в которых были найдены ретротранспозоны с aRNH.

Структурный и филогенетический анализы ретротранспозонов растений и оомицетов, содержащих домен «архейной» рибонуклеазы H

Из каждого генома было отобрано по одному представителю из каждой найденной у него группы ретротранспозонов с aRNH: *Ty3/Gypsy*, *L1* и *Utopia* (см. Материалы и методы). Далее, для каждого отобранного элемента была определена доменная структура и на основе аминокислотных последовательностей RT-доменов был проведен анализ их филогении.

Структурный и филогенетический анализы LTR-ретротранспозонов растений и оомицетов

Результаты филогенетического и структурного анализа найденных LTR-ретротранспозонов растений и оомицетов представлены на рисунке 11. Всего был

выявлен один эволюционный кластер LTR-ретротранспозонов растений и два кластера LTR-элементов у оомицетов.

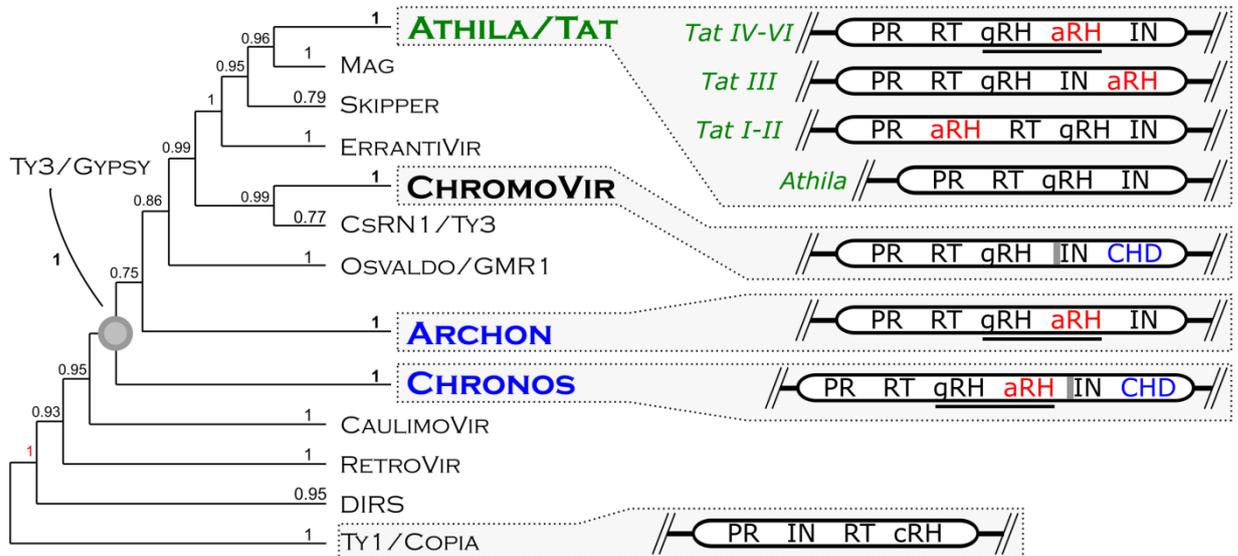


Рисунок 11. Консенсусное дерево филогенетических отношений LTR-ретротранспозонов реконструированное по аминокислотным последовательностям их RT-доменов. Кластеры *Tat* (зеленый) и *Archon* и *Chronos* (синий) принадлежат геномам растений и оомицетов, соответственно. В узлах дерева и на его ветвях обозначена статистическая aLRT-поддержка топологии кластера в реконструкции методом максимального правдоподобия (доли единицы). Справа от ветвей изображена схема консенсусной структуры гена *pol* (ОПС 2) элементов соответствующего кластера. aRH – aRNH, gRH – gRNH, IN – INT. Vir – сокращение для "viridae". Подробное дерево со всеми коэффициентами см. в приложении 3.

LTR-ретротранспозоны с доменом «архейной» рибонуклеазы H в геномах зеленых растений

Филогенетический анализ RT-доменов показал, что все найденные LTR-элементы растений, содержащие aRNH, принадлежат к уже известному монофилетичному кластеру *Ty3/Gypsy* LTR-ретротранспозонов, специфичных для геномов растений, – *Tat* (Wright et al., 1998; Llorens et al., 2011). До сих пор для элементов данной группы не было показано наличие домена aRNH. Ближайшим родственным кластером для *Tat* является кластер *Athila*. Элементы данного кластера, для которых было показано наличие ОПС 3 похожей на *env*-ген ретровирусов, также обнаруживаются лишь в геномах растений (Vicent et al.,

2001). Как и *Athila*, *Tat* LTR-ретротранспозоны характеризуются наличием дополнительных ОРС (ОРС 3, функция которых неизвестна), которые могут иметь различное положение относительно гена *pol* (Steinbauerová et al., 2011).

Нам удалось выявить шесть филогенетических кластеров *Tat* LTR-ретротранспозонов, характеризующихся различной структурой гена *pol*, а также типом и положением ОРС 3 (Рисунки 11 и 12). Кластер I представлен единственным элементом, найденным в геноме плауна *S. moellendorffii*. Элементы, формирующие кластеры II и III, распространены в геномах голосеменных растений, а представители кластеров IV-VI представлены в геномах цветковых растений (Рисунок 12А). Отсутствие каких-либо следов *Tat* LTR-ретротранспозонов с aRNH в геномах мхов, говорит о том, что домен aRNH был приобретен *Tat* LTR-ретротранспозонами уже после отделения мхов от общего эволюционного древа растений.

Наиболее интересной особенностью обнаруженных *Tat* LTR-ретротранспозонов является одновременное присутствие двух доменов RNH: aRNH и исходного *Ty3/Gypsy* домена RNH (gRNH) в белке Pol. Особенностью домена aRNH по сравнению с остальными доменами, образующими ретротранспозон, является то, что в разных кластерах он имеет различное положение в ОРС гена *pol*. В элементах кластеров I и II домен aRNH расположен между доменами PR и RT, в кластере III – на 3' конце гена *pol*, а в кластерах IV-VI – сразу после исходного gRNH-домена.

Дополнительные ОРС – другая интересная особенность *Tat* LTR-ретротранспозонов, которая также коррелирует с разделением на описанные шесть кластеров. Элементы кластеров I и II не обладают дополнительной ОРС. Впервые ОРС 3, расположенная в 3' конце элемента и в антисмысловом направлении, возникает у элементов кластера III, т.е. у голосеменных растений и далее сохраняется во всех последующих дивергировавших кластерах *Tat* LTR-ретротранспозонов. Аналогичным расположением ОРС 3 обладают элементы кластеров V-VI, принадлежащие цветковым растениям, а у наиболее рано обособленного кластера IV происходит смена её расположения и ориентации – она

расположена на 5' конце элемента и в направлении основной кодирующей цепи (Рисунок 12В).

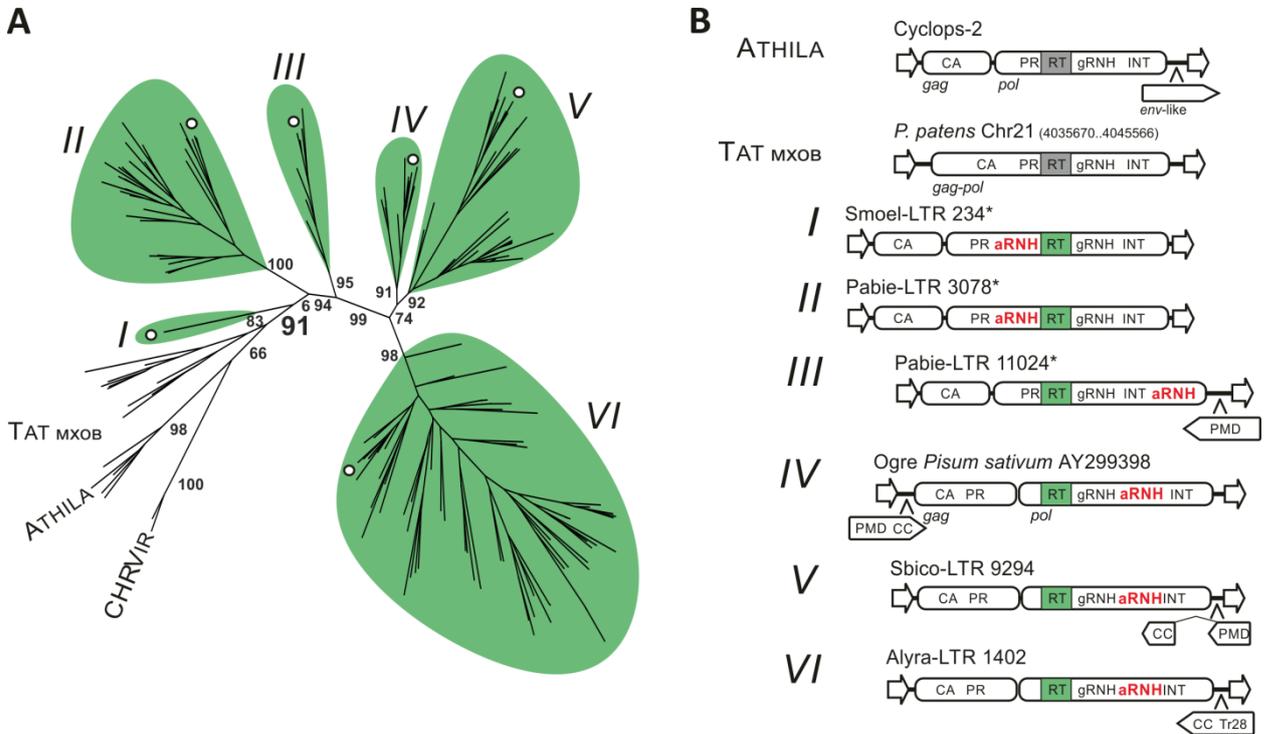


Рисунок 12. Структурное и филогенетическое разнообразие *Tat* LTR-ретротранспозонов. (А) Древо филогенетических взаимоотношений, реконструированное по аминокислотным последовательностям домена RT. Кластеры *Tat* с aRNH выделены зелеными овалами и обозначены латинскими цифрами I-VI. ChrVir – внешняя группа, хромовирусы. Ветви с выбранными элементами представителями отмечены белым кружком. Статистическая поддержка кластеров представлена в виде значений aLRT, которые обозначены в узлах ветвей. (В) Структурная организация элементов-представителей из кластеров, обозначенных на древе. Основные ОРС обозначены овалами, дополнительные ОРС – пятиугольниками. СА – капсидный домен Gag. СС – coiled coil. PMD и Tr28 – консервативные домены с неизвестной функцией в ОРС 3.

Еще одной уникальной структурной особенностью обладают *Tat* LTR-ретротранспозоны из кластеров IV-VI. Домен PR у них оказывается слитым с ОРС белка Gag, а не Pol, как у остальных двухрамочных LTR-ретротранспозонов

(Рисунок 12В). У более ранних кластеров *Tat* такого слияния не наблюдается. Можно предположить, что слияние PR с Gag является эволюционным новшеством, приобретенным элементами из более молодых кластеров *Tat*.

Различное положение aRNH в гене *pol Tat* LTR-ретротранспозонов косвенно может свидетельствовать в пользу того, что данный домен был приобретен независимо несколько раз элементами из разных кластеров *Tat* в течение их эволюции. Альтернативной гипотезой может являться приобретение домена aRNH единожды предковым *Tat*-элементом до отделения плаунов, а затем смена положения aRNH за счёт рекомбинации. Предпочтения любому из этих сценариев могут быть отданы лишь после анализа филогении доменов RNH (см. ниже).

LTR-ретротранспозоны с доменом «архейной» рибонуклеазы H в геномах оомицетов

Основываясь на филогении доменов RT и структурном анализе последовательностей ретротранспозонов, мы обнаружили два кластера среди выявленных *Ty3/Gypsy* LTR-ретротранспозонов с aRNH у оомицетов (Рисунок 11). Обе группы не были описаны ранее и впервые характеризуются в данной работе. Первый кластер, названный нами *Archon*, является специфичным для геномов Saprolegniales. У представителей *Archon* aRNH находится сразу же за исходным gRNH-доменом. Такое сочленение gRNH-aRNH напоминает аналогичную структуру, выявленную ранее для *Tat* LTR-элементов цветковых растений из кластеров *Tat* IV-VI (Рисунки 11 и 12В). Второй кластер, названный *Chronos*, по большей части представлен элементами из геномов отрядов Peronosporales и Pythiales (Рисунок 10) и единственной копией в геноме *Aphanomyces invadans* (Saprolegniales). Элементы *Chronos* также характеризуются наличием структуры gRNH-aRNH, но дополнительно к этому, несут CHD на 3' конце гена *pol*, сразу за доменом INT (Рисунок 11). Ранее наличие CHD было показано лишь для двух кластеров LTR-ретротранспозонов: хромовирусов (группа *Ty3/Gypsy*) (Gorinsek et al., 2004), распространенных в геномах растений, грибов и позвоночных, и CoDi-I-элементов (группа *Ty1/Copia*) из генома диатомовой водоросли *Phaedactylum tricornutum* (надцарство SAR, Stramenopiles) (Llorens et al., 2009).

Несмотря на то что *Archon* и *Chronos* обладают сходным положением домена aRNH и принадлежат к *Ty3/Gypsy*, как и *Tat* LTR-элементы растений (кластеры IV-VI), они очевидно являются эволюционно давно разошедшимися кластерами (Рисунок 11). Такое же очевидное эволюционное расхождение наблюдается между кластерами *Chronos* оомицетов и хромовирусами, несущими CHD. Таким образом, наиболее вероятным объяснением возникновения схожих структур между элементами из этих кластеров LTR-ретротранспозонов может быть их конвергентная эволюция к данному «фенотипу» за счёт независимого приобретения домена aRNH в результате воздействия одинаковых условий отбора на эти элементы (см. пояснение ниже в главе «Конвергентная эволюция ретротранспозонов растений и оомицетов»).

Структурный и филогенетический анализы non-LTR-ретротранспозонов растений и оомицетов

Результаты филогенетического и структурного анализов выявленных non-LTR-ретротранспозонов растений и оомицетов представлены на рисунке 13.

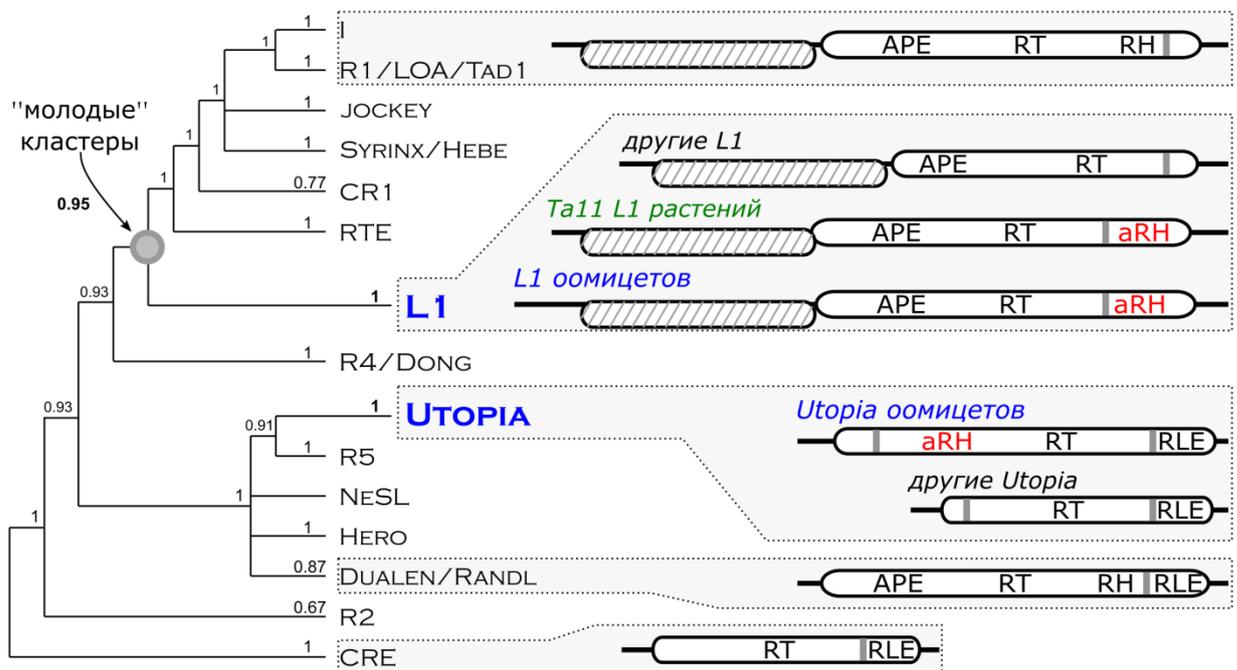


Рисунок 13. Консенсусное дерево филогенетических отношений non-LTR-ретротранспозонов, реконструированное по аминокислотным последовательностям их RT-доменов. В узлах дерева и на его ветвях обозначена статистическая aLRT-поддержка топологии кластера в реконструкции методом максимального

правдоподобия (доли единицы). Справа от ветвей изображена схема консенсусной структуры OPC 2 элементов соответствующего кластера. OPC обозначены овалами, OPC 1 заштрихована, RN – RNH non-LTR-элементов, aRN – aRNH. Подробное дерево со всеми коэффициентами см. в приложении 4.

Non-LTR-ретротранспозоны с aRNH из группы *L1* были обнаружены в отрядах *Peronosporales* и *Pythiales* оомицетов. По своей общей структурной организации они очень схожи с aRNH-содержащими *Tall L1*-элементами из геномов зеленых растений (Рисунок 13). В обеих группах домен aRNH расположен на 3' конце OPC 2. Более того, у обеих групп ССНС цистеиновый мотив расположен перед aRNH, в то время как во всех остальных кластерах non-LTR-ретротранспозонов с доменом RNH, ССНС расположен после RNH (Smyshlyayev et al., 2012). Однако, несмотря на сходства в общей структурной организации, aRNH-содержащие элементы зеленых растений и оомицетов не образуют единого монофилетического кластера внутри группы *L1* (Приложение 4). Отсутствие монофилии может быть объяснено независимым приобретением домена aRNH этими *L1*-элементами. Однако для полного подтверждения этого предположения необходимо определить источник происхождения домена aRNH в случае элементов оомицетов и растений (см. ниже).

Non-LTR-элементы группы *Utopia* были найдены в большинстве геномов отрядов *Peronosporales* и *Pythiales*, в то время как лишь одна копия была обнаружена в геноме *Saprolegniales* – *Saprolegnia diclina* (Рисунок 10, Приложение 2). *Utopia* относятся к древней группе non-LTR-ретротранспозонов – *R2*. Все *R2*-элементы характеризуются наличием RLE-эндонуклеазы на 3' конце OPC 2 (Рисунок 13). Сайт-специфичность RLE *Utopia* направляет их встраивание преимущественно в гены U2 малых ядерных РНК. Элементы *Utopia*, обнаруженные в данной работе, не отличались по своей организации от исходно найденных *Utopia* из работы Коджима и Юрка (Kojima et al., 2015). Очевидно, что раздельное и удаленное друг от друга положение *Utopia* на эволюционном древе non-LTR-ретротранспозонов от элементов группы *L1* однозначно говорит о независимом приобретении домена aRNH элементами данных групп.

Происхождение домена «архейной» рибонуклеазы H у ретротранспозонов растений и оомицетов

После проведения скрининга геномных последовательностей оомицетов и зеленых растений мы обнаружили, что последовательности aRNH не всегда были расположены совместно или в непосредственной близости от RT или других доменов, характерных для ретротранспозонов. Для удобства, далее в тексте такие последовательности aRNH будут обозначаться как индивидуальные aRNH. В дополнение, мы произвели поиск последовательностей, соответствующих другому, наиболее распространенному кластеру эукариотических RNH – fmRNH, – для обнаружения потенциальных последовательностей клеточных генов RNH.

Разнообразие и филогенетический анализ последовательностей рибонуклеазы H оомицетов

В таблице 1 обобщены результаты анализа по распределению индивидуальных aRNH и fmRNH в геномах оомицетов, а также проведено их сравнение с распределением доменов aRNH, найденных у ретротранспозонов. Индивидуальные aRNH были найдены в 21 из 25 исследованных геномах оомицетов. Именно в этих 21 геномах были обнаружены ретротранспозоны с aRNH. Последовательности fmRNH были найдены во всех исследованных геномах оомицетов. Для большинства геномов индивидуальные aRNH были обнаружены в одной единственной копии, в то время как в остальных геномах число копий индивидуальных aRNH достигало 11. Количество копий fmRNH также было относительно низким, варьируя от одной до семи на геном (Таблица 1). Ввиду 100% распространенности и низкого числа копий на геном, fmRNH выглядит как наиболее правдоподобный кандидат на роль клеточного гена RNH у оомицетов. Однако, тогда происхождение и функция индивидуальных aRNH остается неясной.

Для того чтобы разобраться с происхождением доменов aRNH ретротранспозонов и индивидуальных aRNH у оомицетов, был произведен филогенетический анализ различных генов и доменов RNH из широкого круга таксонов (Рисунок 14). Домены aRNH ретротранспозонов оомицетов из кластеров *L1*, *Utopia*, *Archon* и *Chronos* и домены aRNH растительных ретротранспозонов (*Tat* и *L1*) занимают отдельные позиции на филогенетическом древе, т. е. не образуют

монофилетического кластера (см. ниже). Индивидуальные aRNH оомицетов разделяются на три кластера: aRNH 1, aRNH 2 и aRNH 3. Кластеры aRNH 1 и aRNH 3 группируются совместно с доменами aRNH ретротранспозонов *Archon* и *L1*, соответственно. Кластер aRNH 2 не показал значимой кластеризации ни с одним из доменов aRNH ретротранспозонов (Рисунок 14). Стоит отметить, что в отличие от aRNH 2, последовательности кластеров aRNH 1 и aRNH 3 были найдены в большем количестве копий (Таблица 1). Учитывая это, а также совместную кластеризацию aRNH 1 и aRNH 3 с доменами aRNH

Таблица 1. Разнообразие, распространение и количество доменов aRNH и fmRNH в исследованных видах оомицетов

Таксономическое положение по NCBI Таксоному			Число ретротранспозонных aRNH				Число индивидуальных RNH			
Порядок	Род	Вид	Chronos	Archon	L1	Utopia	aRNH			fmRNH
							aRNH 1	aRNH 2	aRNH 3	
Albuginales	<i>Albugo</i>	<i>A. candida</i>	-	-	-	-	-	-	-	2
		<i>A. laibachii</i>	-	-	-	-	-	-	-	1
Peronosporales	<i>Hyaloperonospora</i>	<i>H. arabidopsidis</i>	57	-	88	-	-	1	-	1
		<i>Phytophthora</i>	<i>P. alni</i>	106	-	3	2	4	-	7
	<i>P. capsici</i>	64	-	14	16	1	1	-	1	
	<i>P. cinnamomi</i>	92	-	3	5	-	1	-	1	
	<i>P. infestans</i> T30-4	1555	-	43	25	-	1	-	2	
	<i>P. kernovia</i> 00238/432	3	-	-	1	-	1	-	1	
	<i>P. lateralis</i> MPF4	30	-	7	5	-	1	-	2	
	<i>P. parasitica</i> P1569	11	-	-	1	1	1	-	4	
	<i>P. pinifolia</i> CBS	75	-	26	2	8	2	-	2	
	<i>P. ramorum</i>	120	-	11	14	-	1	-	3	
	<i>P. sojae</i>	271	-	35	31	2	1	-	4	
		<i>Среднее</i>	233	-	18	10	3	1	-	3
	<i>Phytophthora</i>	<i>P. vexans</i>	1	-	-	1	-	1	-	2
	<i>Pseudoperonospora</i>	<i>P. cubensis</i> MSU-1	-	-	-	-	-	-	-	1
Pythiales	<i>Pythium</i>	<i>P. aphanidermatum</i>	5	-	3	-	-	1	-	1
		<i>P. arrhenomanes</i>	5	-	5	2	-	1	-	1
		<i>P. insidiosum</i>	33	-	43	2	-	1	-	1
		<i>P. irregulare</i>	1	-	-	1	-	1	-	2
		<i>P. iwayamai</i>	1	-	3	-	-	1	-	2
		<i>P. ultimum</i>	-	-	74	7	-	1	-	2
		<i>Среднее</i>	9	-	26	3	-	1	-	2
Saprolegniales	<i>Aphanomyces</i>	<i>A. astaci</i> APO3.2	1	-	-	-	-	-	3	1
		<i>A. invadans</i> 9901.2	-	-	-	-	-	-	-	1
			<i>Среднее</i>	1	-	-	-	-	3	1
	<i>Saprolegnia</i>	<i>S. diclina</i> VS20	-	3	-	1	-	1	10	1
		<i>S. parasitica</i> CBS	-	2	-	-	-	1	4	1
		<i>Среднее</i>	-	3	-	1	-	1	7	1

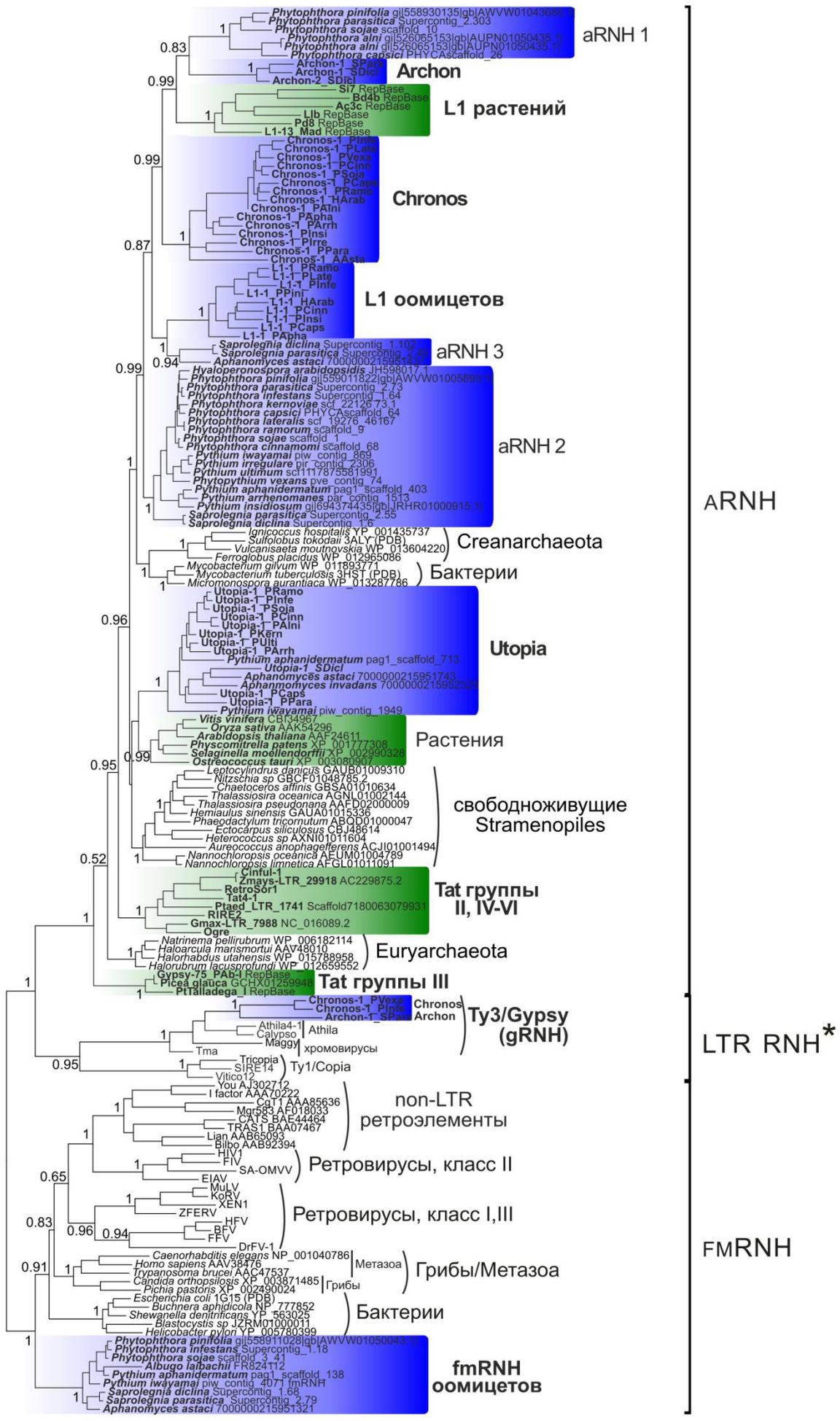


Рисунок 14. Древо филогенетических взаимоотношений RNH типа I из различных источников, реконструированное методом максимального правдоподобия. В узлах дерева и на его ветвях обозначена статистическая поддержка топологии кластера aLRT (доли единицы). Кластеры, соответствующие последовательностям RNH из геномов оомицетов и растений, выделены синим и зеленым, соответственно. Полное дерево со всеми коэффициентами, а также сравнение с деревом, построенным байесовским методом, см. в приложении 5.

ретротранспозонов оомицетов, можно предположить, что они могут представлять собой остатки элементов *Archon* и *L1*, соответственно. Учитывая низкую копияность, а также относительно широкое распределение домена aRNH 2 среди исследованных геномов оомицетов (Таблица 1), можно также предложить aRNH 2 в качестве кандидата на роль клеточного гена RNH у оомицетов.

Для того чтобы пролить больше света на эволюцию aRNH и fmRNH у оомицетов, мы проанализировали геномные и транскриптомные последовательности свободноживущих представителей Stramenopiles – эволюционно ближайших таксонов к оомицетам, доступных в базах данных (Таблица 2). Для этого был использован поиск в программе tBLASTn в базах данных NCBI WGS и NCBI TSA, используя обнаруженные в данной работе последовательности aRNH и fmRNH оомицетов в качестве поисковых запросов. Индивидуальные aRNH были обнаружены во всех геномах свободноживущих Stramenopiles, в то время как сигнал fmRNH полностью отсутствовал. Индивидуальные aRNH свободноживущих Stramenopiles образуют монофилетичный кластер на древе, реконструированном методом максимального правдоподобия, который распадался (парафилетичен) в реконструкции байесовским методом (Приложение 5). При этом их последовательности не показали значимой кластеризации ни с одним из других исследованных кластеров aRNH (Рисунок 14).

В поиске гомологов aRNH и fmRNH в геномах оомицетов, мы обнаружили aRNH одновременно как у оомицетов, так и у их свободноживущих «родственников». В тоже время, fmRNH была обнаружена лишь в геномах оомицетов (Таблицы 1 и 2). При этом aRNH отсутствует в некоторых

исследованных таксонах оомицетов, а в тех таксонах, где она есть, не показывает никакой значимой кластеризации с последовательностями свободноживущих *Stramenopiles* (Таблица 1, Рисунок 14). С одной стороны, можно предположить, что aRNH присутствовала у общего предка всех *Stramenopiles*, а затем могла быть вертикально унаследована оомицетами. С другой стороны, aRNH могла попасть в оомицеты горизонтально от зеленых растений, на которых паразитируют большинство из исследованных в данной работе таксонов оомицетов (Beakes et al., 2012). При любом из сценариев, полная потеря aRNH некоторыми таксонами оомицетов может быть объяснена избыточностью одновременного наличия aRNH и fmRNH, выполняющих сходную функцию.

Отсутствие fmRNH у свободноживущих *Stramenopiles* наиболее вероятно отражает то, что оомицеты приобрели данный ген уже после отделения от общего корня *Stramenopiles*. Одним из потенциальных источников домена fmRNH у оомицетов могут являться грибы. ГП генов из грибов в оомицеты как адаптация к паразитизму на водорослях и растениях был уже неоднократно показан (Richards et al., 2011; Soanes et al., 2014). Геномы грибов кодируют fmRNH, функция которой заключается в удалении РНК-праймеров фрагментов Оказаки во время репликации ДНК, а также играет важную роль в поддержании стабильности генома (Qiu et al., 1999; Cerritelli et al., 2009). Таким образом, можно предположить, что fmRNH была перенесена в оомицеты с ГП других генов от древней группы грибов.

Разнообразие и филогенетический анализ последовательностей рибонуклеазы H зеленых растений

Индивидуальные aRNH были обнаружены во всех геномах зеленых растений, включая зеленые водоросли и мхи, варьируя в количестве от одной до трех копий на геном, и они полностью соответствовали последовательностям известных клеточных генов RNH. Проведенный филогенетический анализ последовательностей RNH из различных источников (Рисунок 14) также поместил все индивидуальные aRNH зеленых растений в одну монофилетическую группу, внутри которой топология дерева повторяет общепринятую топологию древа эволюции зеленых растений (Приложение 5, Рисунок 9).

Последовательности доменов aRNH *Tat* LTR-ретротранспозонов не показали значимой кластеризации ни друг с другом, ни с кластерами aRNH ретро-транспозонов оомицетов, ни с кластером генов зеленых растений (Рисунок 14). Кластер aRNH доменов *Tall Ll* non-LTR-ретротранспозонов растений показал значимую кластеризацию с доменами aRNH *Chronos* (0.99) на древе, реконструированном методом максимального правдоподобия, но слабый коэффициент (77%) на древе, реконструированном байесовским методом (приложение 5).

Домены aRNH *Tat* LTR-ретротранспозонов не формируют монофилетического кластера на древе. Вместо этого выделяются три отдельных кластера aRNH, соответствующих RT-кластерам *Tat* II,IV-VI, *Tat* III и «кластеру», представленному единственной последовательностью домена aRNH *Tat* I из плауна *S. moellendorffii* (Рисунки 11, 12 и 14). Таким образом, можно предположить, что в эволюционной истории *Tat* LTR-ретротранспозонов было как минимум три независимых приобретения aRNH.

Анализ структуры gRNH-aRNH у LTR-ретротранспозонов *Chronos*, *Archon* и *Tat* IV-VI

Как было описано выше, aRNH-содержащие LTR-ретротранспозоны из кластеров *Tat* IV-VI растений и кластеров *Chronos* и *Archon* оомицетов показывают наличие сходного элемента структуры гена *pol*: домен aRNH, следующий сразу за исходным gRNH доменом LTR-ретроэлемента. Ранее подобная структура была описана лишь для ретровирусов позвоночных животных, которые приобрели и приспособили клеточный домен fmRNH в положении аналогичном aRNH в гене *pol* у элементов *Archon*, *Chronos* и *Tat*. При этом, исходный домен RNH ретровирусов (гомологичный gRNH *Ty3/Gypsy* LTR-ретротранспозонов) стал накапливать замены и терять свою исходную каталитическую функцию (Malik, 2005). В настоящий момент, промежуток белка Pol, занимаемый последовательностью бывшего gRNH домена ретровирусов, обозначается как связующий домен, который, как считается, регулирует активность приобретенного fmRNH домена ретровирусов (Malik, 2005; Lapkouski et al., 2013). Несмотря на

Таблица 2. Индивидуальные aRNH, обнаруженные в геномах свободноживущих Stramenopiles

Порядок	Семейство	Род	Вид	Номер в базе данных	aRNH	fmRNH	Число копий
Bacillariales	Bacillariaceae	Nitzschia	<i>N. sp</i>	GBCF01048785	+	-	- (мПНК)
Chaetocerotales	Chaetocerotaceae	Chaetoceros	<i>C. affinis</i>	GBSA01010634	+	-	- (мПНК)
Ectocarpales	Ectocarpaceae	Ectocarpus	<i>E. siliculosus</i>	CBJ48614	+	-	1
Eustigmatales	Monodopsidaceae	Nannochloropsis	<i>N. limnetica</i>	AFGL01011091	+	-	1
		Nannochloropsis	<i>N. oceanica</i>	AEUM01004789	+	-	1
Hemiaulales	Hemiaulaceae	Hemiaulus	<i>H. sinensis</i>	GAUA01015336	+	-	- (мПНК)
Leptocylindrales	Leptocylindraceae	Leptocylindrus	<i>L. danicus</i>	GAUB01009310	+	-	- (мПНК)
Naviculales	Phaeodactylaceae	Phaeodactylum	<i>P. tricornutum</i>	ABQD01000047	+	-	1
Pelagomonadales	-	Aureococcus	<i>A. anophagefferens</i>	ACJI01001494	+	-	1
Thalassiosirales	Thalassiosiraceae	Thalassiosira	<i>T. oceanica</i>	AGNL01002144	+	-	1
			<i>T. pseudonana</i>	AAFD02000009	+	-	1
Tribonematales	Heteropediaceae	Heterococcus	<i>H. sp</i>	AXNI01011604	+	-	2

практически полное отсутствие сходства на уровне первичной структуры, вторичная и третичная укладка связующего домена соответствует классической укладке RNH (Larkouski et al., 2013). Для того чтобы проверить, оказало ли приобретение домена RNH аналогичное влияние на последовательности gRNH-домена LTR-ретротранспозонов *Tat*, *Chronos* и *Archon*, было проведено детальное сравнение аминокислотных последовательностей генов и доменов RNH из различных источников (Рисунок 15).

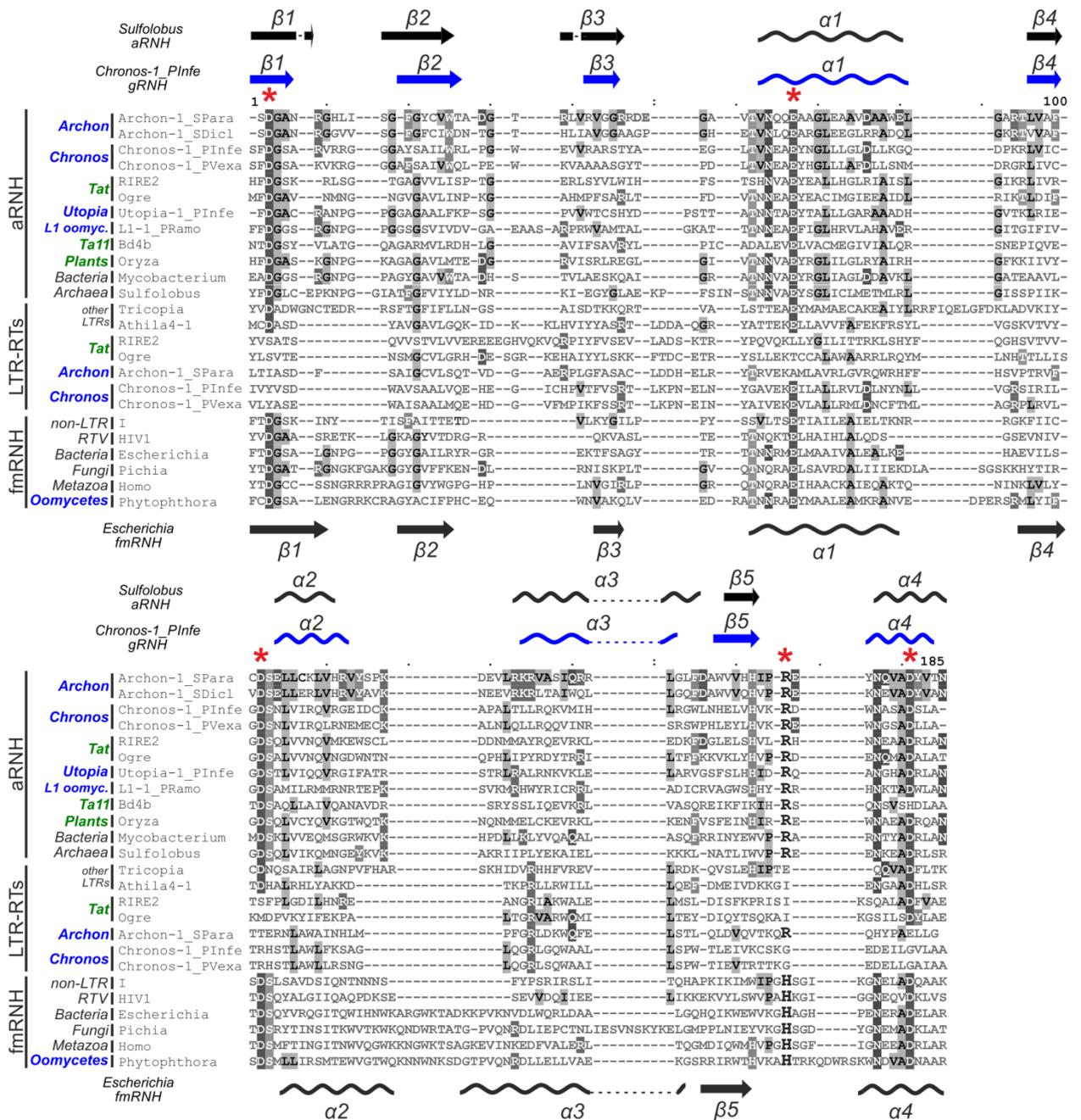


Рисунок 15. Множественное выравнивание аминокислотных последовательностей RNH типа I. Имена последовательностей, относящихся к геному оомицетов и

растений, выделены жирным и окрашены синим и зеленым, соответственно. Консервативные каталитические основания (D-E-D-R/H-D) обозначены красными звездочками над выравниванием. Полуконсервативное (R/H)-основание выделено увеличенным шрифтом. Снизу и кверху от выравнивания показаны вторичные структуры RNН: альфа-спирали – волны, а бета-складки – стрелки.

Анализ множественного выравнивания аминокислотных последовательностей RNН позволил выявить несколько интересных закономерностей. Приобретенные домены aRNН исследованных здесь ретротранспозонов характеризуются полноценным набором консервативных каталитических оснований: D-E-D-D (Рисунок 15). Таким полным набором обладают все известные каталитически активные домены RNН типа I (Cerritelli et al., 2009; Smyshlyayev et al., 2013). Все домены aRNН также характеризуются наличием положительно заряженной аминокислоты аргинина (R) в полуконсервативной позиции выравнивания. У последовательностей fmRNН в данной позиции находится гистидин (H). Исходные домены RNН LTR-ретротранспозонов не обладают ни R, ни H в данной позиции, что может объяснить их сниженную каталитическую активность по сравнению с ферментами из кластера fmRNН и aRNН (Smyshlyayev et al., 2013). Считается, что сниженная активность домена RNН нужна LTR-ретротранспозонам для сохранения PPT-прайма, с которого начинается синтез плюс-цепи кДНК. Однако для доменов RNН LTR-ретротранспозонов всё равно характерна полная каталитическая тетрада D-E-D-D (Рисунок 15). На этом фоне исходные домены gRNН LTR-ретротранспозонов с aRNН проявляют практически полную потерю консервативных каталитических оснований (Рисунок 16). Это соответствует предположению о том, что приобретенный домен aRNН берет ферментативную функцию на себя, как это было с клеточным fmRNН-доменом, приобретенным ретровирусами позвоночных. Таким образом, исходный домен RNН LTR-ретротранспозонов, которые приобрели aRNН с образованием структуры gRNН-aRNН, повторяет судьбу gRNН-домена ретровирусов, превращаясь в связующий домен. Связующий домен теряет каталитическую функцию, но, по-видимому, всё еще участвует в регуляции работы второго, более активного, домена RNН (Malik, 2005; Lapkouski et al., 2013).

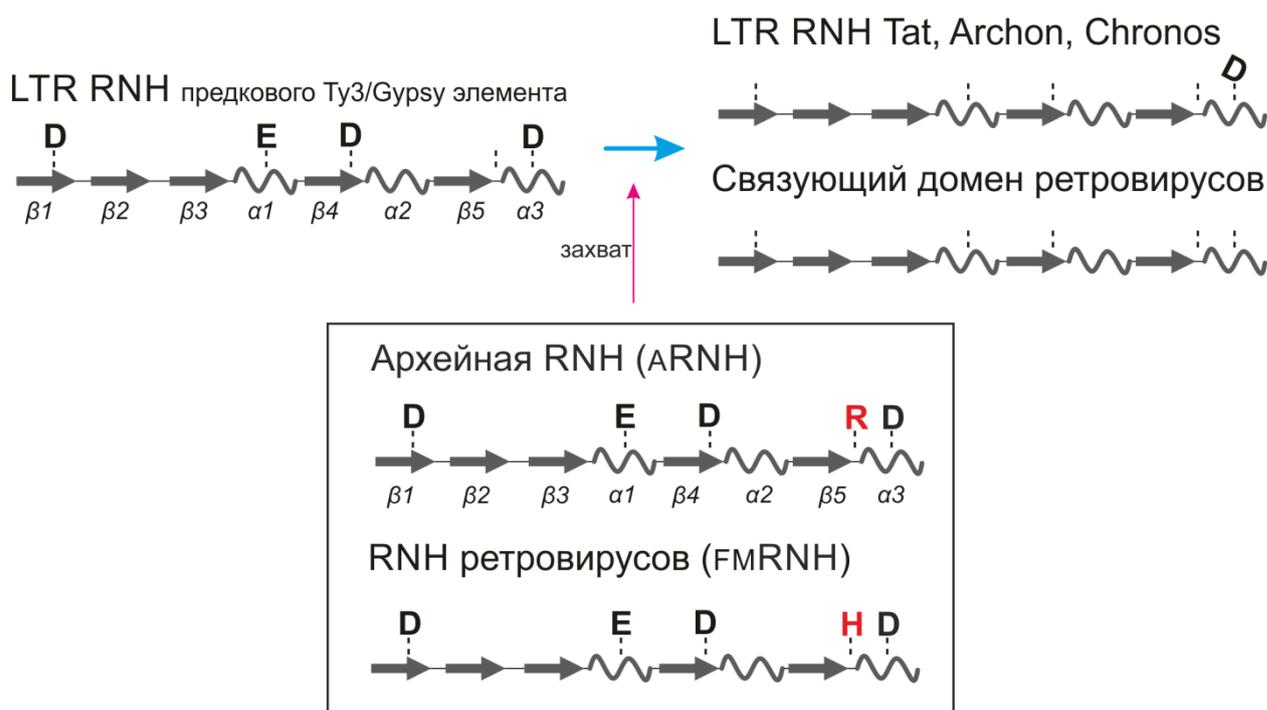


Рисунок 16. Схема потери каталитических консервативных аминокислотных оснований в исходном домене RNH *Ty3/Gypsy* LTR-ретротранспозона. Вторичные структуры RNH изображены так: альфа-спирали – волны, бета-складки – стрелки. D-E-D-D – тетрада консервативных каталитических оснований. (R/H) – полуконсервативное основание.

Конвергенция ретротранспозонов растений и оомицетов

В данной работе мы показали, основываясь на реконструкции филогении по RT-доменам, что обнаруженные в геномах растений *Ty3/Gypsy Tat* LTR-ретротранспозоны и *Tall L1* non-LTR-ретротранспозоны лишь отдалённо родственны *Ty3/Gypsy Archon* и *Chronos* LTR-ретротранспозонам и *L1* non-LTR-ретротранспозонам оомицетов с доменом aRNH. Об этом говорят их очевидно раздельные позиции на филогенетических деревьях (Рисунки 11 и 13, Приложения 3 и 4). Маловероятно, что домен aRNH был приобретен лишь однажды предковым элементом эволюционной линии *Ty3/Gypsy* LTR-ретротранспозонов (или предковым *L1*-элементом non-LTR-ретротранспозонов), который бы затем мог быть горизонтально перенесен из растений в оомицеты или наоборот. Иначе пришлось бы утверждать, что все остальные известные кластеры ретроэлементов из вышеописанных групп должны были потерять домен aRNH в процессе своей эволюции. Таким образом, мы предполагаем, что наличие aRNH у LTR-

ретротранспозонов *Tat*, *Chronos* и *Archon*, а также *L1* non-LTR-ретротранспозонов растений и оомицетов, наилучшим образом объясняется несколькими независимыми событиями приобретения домена aRNH предковыми линиями каждой из этих групп, как результат их конвергентной эволюции к единому структурно-функциональному «фенотипу».

Вдобавок к этому, мы показали, что индивидуальные гены/домены aRNH присутствуют как в геномах растений, так и в геномах оомицетов, являясь предполагаемыми источниками приобретения домена aRNH исследованных ретротранспозонов (Рисунок 14, Приложение 5). В случае *Tat* LTR-ретротранспозонов растений мы можем четко проследить возникновение домена aRNH на уровне появления этих элементов у плаунов, тогда как данный домен отсутствовал у всех исследованных элементов *Tat* мхов (Рисунок 12). *Tat* LTR-ретротранспозоны показывают, пожалуй, уникальный пример множественных пертурбаций в положении домена aRNH, которые, по-видимому, также были опосредованы несколькими повторными независимыми приобретениями домена aRNH, что вытекает из наличия как минимум трех эволюционных кластеров доменов aRNH *Tat* LTR-ретротранспозонов, выявленных на древе RNH (Рисунки 12 и 14).

Повторяющиеся приобретения и последующая фиксация некоторых функциональных доменов за время эволюции различных групп ретроэлементов, происходящих из одной генетической линии, могут отражать то, что приобретение этих доменов оказывает положительный эффект на отбор в условиях, в которых обитает эта генетическая линия, будь то отдельные виды, штаммы вирусов или линии ретротранспозонов. Основой для такой конвергенции у ретротранспозонов служит их модульное строение, когда ретроэлемент представляет собой совокупный ансамбль функциональных белковых доменов, определяющих организацию его жизненного цикла. Мы показали, что у *Ty3/Gypsy* LTR-ретротранспозонов оомицетов из кластеров *Chronos* и *Archon*, а также кластеров IV-VI *Tat* LTR-ретротранспозонов растений, возникает структура gRNH-aRNH, которая во многом повторяет аналогичную структуру (но уже gRNH-fmRNH) ретровирусов позвоночных животных, которые когда-то также произошли от

предковой линии *Ty3/Gypsy* LTR-ретротранспозонов (Malik et al., 2001). Кроме наличия сходной структуры двойной RNH, мы показали, что с исходным gRNH-доменом элементов *Archon*, *Chronos* и *Tat* IV-VI происходят аналогичные процессы потери каталитической активности за счёт «деградации» консервативных каталитически активных оснований, которые происходили с исходным gRNH-доменом ретровирусов на пути превращения его в связующий домен (Рисунки 15 и 16). Таким образом, мы наблюдаем как минимум двукратное повторение одного из ранних этапов пути эволюции ретровирусов позвоночных в геномах растений и оомицетов (Рисунок 17).

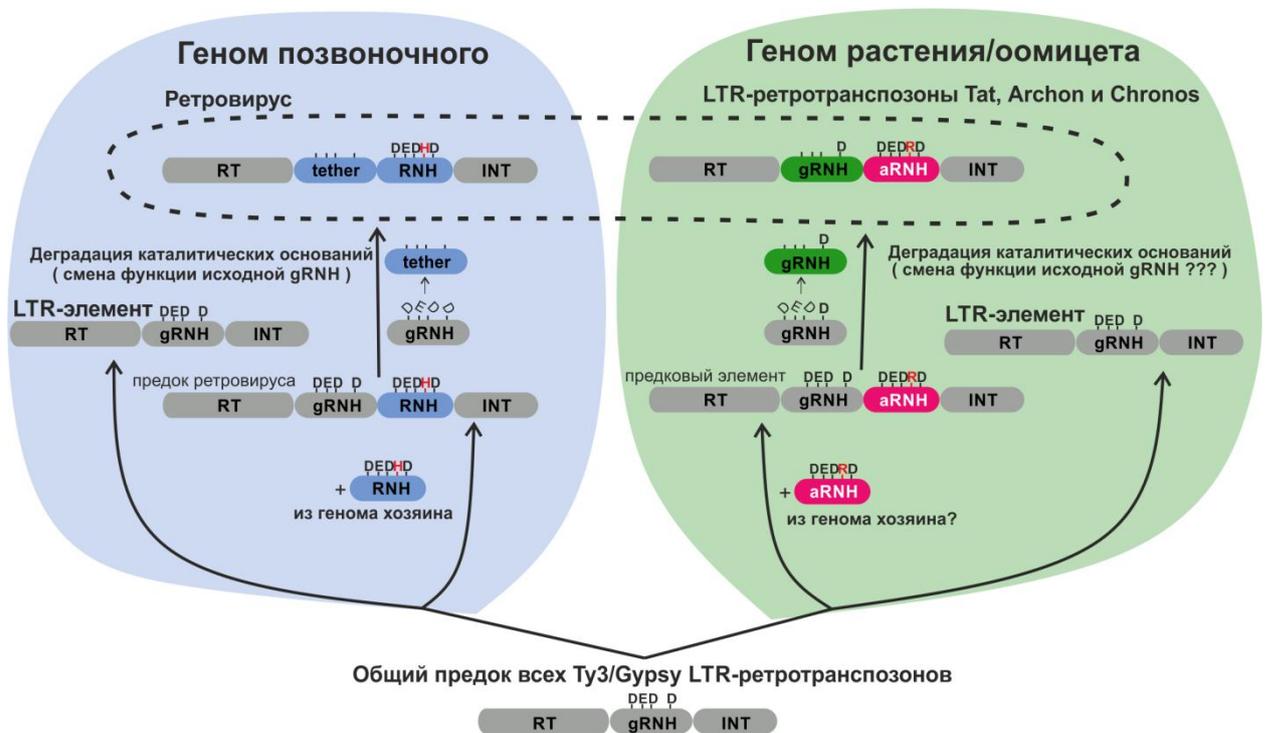


Рисунок 17. Предполагаемая схема конвергентной эволюции LTR-ретротранспозонов растений и оомицетов с ретровирусами позвоночных. Овалами обозначены функциональные домены белка Pol, начиная с RT. Сверху над RNH-доменами обозначены консервативные каталитические основания (D-E-D-(R/H)-D).

В чем же может состоять эволюционно положительная роль от приобретения каталитически более активного домена RNH (aRNH или fmRNH), а также превращения исходного домена RNH в связующий домен? Структурные исследования RNH ретровируса ВИЧ-1 показали, что связующий домен обеспечивает конформационные изменения, необходимые для правильной и

эффективной ориентации РНК-цепи (из гибрида РНК/ДНК) для её гидролиза ферментом RNH (Lapkouski et al., 2013). Это говорит о том, что связующий домен регулирует активность RNH, приводя к более специализированному и частому гидролизу цепи РНК во время обратной транскрипции LTR-ретротранспозона. В случае с *Tu3* LTR-ретротранспозоном дрожжей было показано, что необходимы существенные изменения в конформации РНК/ДНК-гибрида, для того чтобы *Tu3* gRNH могла произвести эффективный гидролиз РНК (Nowak et al., 2014). Таким образом, гидролиз РНК лишь с участием gRNH является не столь эффективным, как в случае наличия двойного RNH-домена. Гидролиз РНК-матрицы с синтезирующейся цепи ДНК как LTR-ретротранспозона, так и ретровируса, необходим для переброса цепи ДНК (Рисунок 3) (Basu et al., 2008). Накопление широкого генетического разнообразия – яркая особенность инфекции вируса ВИЧ-1. ВИЧ-1 в значительной степени опирается на механизм спонтанной рекомбинации, зависящей от переброса цепи во время обратной транскрипции (Smyth et al., 2012). Вероятно, что структура gRNH-(a/fm)RNH позволяет ретровирусу рекомбинировать гораздо чаще, чем LTR-ретротранспозону лишь с одной gRNH.

Потенциальную выгоду от приобретения домена RNH non-LTR-ретротранспозонами представить сложнее, так как эти элементы вполне могут опираться на активность клеточного фермента RNH хозяина, и их обратная транскрипция происходит напрямую в ядре. Более того, показано, что отдельные представители группы *I* non-LTR-ретротранспозонов теряют его (Malik et al., 1999a). Однако в таком случае трудно объяснить, в чем причина того, что, как минимум, в трех филогенетически далеких группах non-LTR-ретротранспозонов: *I*, *R2* (*Utopia* и *Dualen*) и *L1* (*Ta11* элементы растений, а также элементы с aRNH оомицетов) – ретроэлементы независимо приобретают данный функциональный домен. Можно предположить, что наличие собственного домена RNH может позволить non-LTR-ретротранспозону, попавшему внутрь упаковывающейся частицы вируса, встроиться внутрь его генома, обеспечив себе, таким образом, потенциальную возможность горизонтального переноса за счет этого вируса. Действительно, такая ситуация вполне характерна, например, для представителей

класса ДНК-транспозонов, которые иногда встраиваются внутрь крупных ДНК-вирусов насекомых – бакуловирусов (Gilbert et al., 2014).

Структурный анализ LTR-ретротранспозонов оомицетов из кластера *Chronos* показал, что, помимо домена aRNH, эти элементы приобрели CHD на С-конце ОРС 2, сразу за доменом INT (INT-CHD) (Рисунок 11). С-концевой INT-CHD повторяет аналогичную структуру, которая до этого была обнаружена лишь у одного кластера группы *Ty3/Gypsy* – хромовирусов. К хромовирусам относятся LTR-ретротранспозоны из геномов растений, грибов и позвоночных (Gorinsek et al., 2004; Novikova, 2009; Novikov et al., 2012). Основываясь на RT-филогении (Рисунок 11), мы показали, что *Chronos*-элементы значительно удалены от хромовирусов, т. е. не имеют с ними близкого общего предка. Таким образом, можно заключить, что наиболее вероятным объяснением наличия CHD у *Chronos* и хромовирусов может являться его независимое приобретение в обоих эволюционных кластерах. Также стоит отметить, что структура INT-CHD была обнаружена у CoDi-I LTR-элементов из группы *Ty1/Copia*, найденных у свободноживущей диатомовой водоросли – *Phaeodactylum tricornutum* (Stramenopiles) (Llorens et al., 2009). Хромодомены – широкораспространенные белковые домены, встречающиеся в белках ремоделинга хроматина у эукариот (Platero et al., 1995; Eissenberg, 2012). Слияние CHD с INT с большей вероятностью направляет встраивание ретротранспозона в гетерохроматин, подальше от насыщенных генами районов (Gao et al., 2008). Таким образом, множественные независимые приобретения CHD можно объяснить наличием общей тенденции к минимизации ущерба геному-хозяину от перемещения ретроэлемента, позволяя ретротранспозону «спокойно» размножаться внутри генома, не нарушая его структуру.

ЗАКЛЮЧЕНИЕ

Ретротранспозоны составляют многочисленный и широко распространенный класс МЭ. Все ретротранспозоны опираются на активность RT для производства своей ДНК-копии на основе исходной РНК-матрицы элемента. Набор других функциональных активностей, закодированных в полипротеине ретротранспозона, значительно различается между разнообразными группами ретротранспозонов. Белковые домены полипротеина, соответствующие каждой из этих дополнительных функциональных активностей, могут обладать своей собственной, независимой от домена RT, эволюционной историей. Данный факт лег в основу представления об эволюции ретротранспозонов как модульной. Более того, накопленные данные о разнообразии и распространении ретротранспозонов, свидетельствуют в пользу того, что эволюционно удаленные группы ретроэлементов независимо развили сходное структурное устройство через приобретение функционально идентичных и не монофилетичных белковых доменов – функциональных модулей. Эти данные указывают на конвергентный характер эволюции отдельных групп ретротранспозонов, возникающей за счет модульности их строения и под действием сходных механизмов отбора, регулирующих жизненный цикл соответствующих элементов в различных таксонах.

В настоящей работе представлены новые примеры конвергентной эволюции ретротранспозонов из геномов зеленых растений (*Viridiplantae*) и паразитических простейших – оомицетов (*Stramenopiles*). Исследование доменной структуры генов *pol* и филогенетический анализ доменов RT ретротранспозонов, содержащих домен из «архейной» группы рибонуклеаз H типа I, показали, что в геномах оомицетов и зеленых растений ретротранспозоны из пяти филогенетически удалённых кластеров независимо приобрели домен aRNH в сходном положении внутри гена *pol*. Оказалось, что последовательности индивидуальных aRNH, которые могли бы служить источником происхождения доменов aRNH у исследованных ретротранспозонов, имеются как в геномах растений, так и у последнего общего предка свободноживущих *Stramenopiles* и паразитических оомицетов. Это является подтверждением гипотезы о конвергентной эволюции ретротранспозонов растений

и оомицетов и множественного независимого приобретения доменов aRNH. Соответственно, альтернативная гипотеза о горизонтальном переносе предкового ретротранспозона, который однажды приобрел aRNH, между таксонами растений и оомицетов не находит подтверждения.

Ty3/Gypsy LTR-ретротранспозоны оомицетов кластера *Chronos* с aRNH несут домен CHD на 3' конце гена *pol*, сразу после домена INT, повторяя структуру ранее известную лишь для кластера *Ty3/Gypsy* LTR-ретротранспозонов из геномов позвоночных, грибов и зеленых растений – хромовирусов. А очевидно удаленное филогенетическое положение кластера *Chronos* от кластера хромовирусов свидетельствует в пользу независимого приобретения CHD ретроэлементами этих кластеров, предполагая их конвергентную эволюцию.

Сравнительный и структурный анализы показали, что домен aRNH в элементах *Tat* IV-VI растений и *Chronos* и *Archon* оомицетов, а также исходный *Ty3/Gypsy* домен RNH, образуют структуру сходную с доменом «двойной» RNH ретровирусов позвоночных. При такой структуре исходный ретротранспозонный домен деградирует за счёт накопления несмысловых мутаций в активном центре фермента, а приобретенный домен, наоборот, несёт более полный набор каталитически активных консервативных аминокислот. Такое сходство и очевидная эволюционная дистанция между *Ty3/Gypsy* LTR-ретротранспозонами растений и оомицетов от ретровирусов позвоночных говорит о их конвергентной эволюции.

Повторяющееся возникновение сходных структур у ретротранспозонов из эволюционно далеких таксонов свидетельствует о наличии фундаментальных клеточных механизмов регуляции их жизненного цикла, которые действуют во всех таксонах эукариот, предполагая положительное влияние приобретаемых функциональных доменов на приспособленность ретротранспозона.

ВЫВОДЫ

1. Биоинформационный анализ последовательностей 65 геномов растений и 25 геномов оомицетов выявил пять эволюционных кластеров ретротранспозонов, содержащих домен «архейной» рибонуклеазы H. Три из них, *Archon* и *Chronos* из геномов оомицетов и *Tat* из геномов растений, принадлежат к LTR-ретротранспозонам группы *Ty3/Gypsy*. Два других кластера, *L1* и *Utopia*, относятся к non-LTR-ретротранспозонам из групп *L1* и *R2*. Элементы кластера *L1* были найдены как у растений, так и у оомицетов, в то время как элементы кластера *Utopia* были обнаружены только у оомицетов.

2. Впервые было показано наличие домена «архейной» рибонуклеазы H у *Tat* LTR-ретротранспозонов растений и *L1* non-LTR-ретротранспозонов оомицетов, а элементы кластеров *Chronos* и *Archon* из геномов оомицетов обнаружены и охарактеризованы в данной работе. Для элементов кластера *Tat* впервые описано разнообразие положения данного домена в гене полипротеина.

3. Высокий уровень сходства структурной организации исследованных ретроэлементов растений и оомицетов с одной стороны, и их филогенетическая удаленность, с другой стороны, позволили предположить конвергентную эволюцию исследованных групп LTR- и non-LTR-ретротранспозонов из геномов растений и оомицетов.

4. Сравнительный и структурный анализы показали, что домен «архейной» рибонуклеазы H и исходный домен рибонуклеазы H в элементах кластеров *Tat* IV-VI растений и *Chronos* и *Archon* оомицетов образуют структуру сходную с доменом «двойной» рибонуклеазы H ретровирусов позвоночных. Данное сходство и очевидная филогенетическая удаленность *Ty3/Gypsy* LTR-ретротранспозонов от ретровирусов позвоночных также позволяет предположить их конвергентную эволюцию.

5. Установлено, что LTR-ретротранспозоны группы *Ty3/Gypsy* из кластера *Chronos* оомицетов, помимо домена «архейной» рибонуклеазы H, приобрели хромодомен, аналогично другому широкому кластеру *Ty3/Gypsy* из геномов грибов, позвоночных и растений – хромовирусов.

6. Полученные результаты по структурно-филогенетическому анализу ретротранспозонов растений и оомицетов являются дополнительным

свидетельством модульной эволюции ретротранспозонов и её роли в возникновении структурной и функциональной конвергенции между эволюционно удаленными группами различных ретроэлементов.

СПИСОК ЛИТЕРАТУРЫ

1. Steinbauerová V. et al. A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. // *Genetica*. 2011. T. 139. № 11–12. С. 1543–55.
2. Smyshlyaev G. et al. Acquisition of an Archaea-like ribonuclease H domain by plant L1 retrotransposons supports modular evolution // *Proceedings of the National Academy of Sciences*. 2013. T. 110. № 50. С. 20140–20145.
3. Xiong Y., Eickbush T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. // *The EMBO journal*. 1990. T. 9. № 10. С. 3353–62.
4. Aksoy S. et al. SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. // *Nucleic acids research*. 1990. T. 18. № 4. С. 785–92.
5. Alisch R.S. et al. Unconventional translation of mammalian LINE-1 retrotransposons // *Genes & Development*. 2006. T. 20. № 2. С. 210–224.
6. Altschul S.F. et al. Basic local alignment search tool. // *Journal of molecular biology*. 1990. T. 215. № 3. С. 403–10.
7. Alva V. et al. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis // *Nucleic Acids Research*. 2016. T. 44. № W1. С. W410–W415.
8. Anisimova M. et al. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes // *Systematic Biology*. 2011. T. 60. № 5. С. 685–699.
9. Anzai T. et al. Functional roles of 3'-terminal structures of template RNA during in vivo retrotransposition of non-LTR retrotransposon, R1Bm // *Nucleic Acids Research*. 2005. T. 33. № 6. С. 1993–2002.
10. Anzai T., Takahashi H., Fujiwara H. Sequence-specific recognition and cleavage of telomeric repeat (TTAGG)(n) by endonuclease of non-long terminal repeat retrotransposon TRAS1. // *Molecular and cellular biology*. 2001. T. 21. № 1. С. 100–8.
11. Balagopal V., Parker R. Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs // *Current Opinion in Cell Biology*. 2009. T. 21. № 3. С. 403–408.
12. Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. //

Nature. 1970. T. 226. № 5252. C. 1209–11.

13. Bao W., Kojima K.K., Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes // *Mobile DNA*. 2015. T. 6. № 1. C. 11.

14. Basu V.P. et al. Strand transfer events during HIV-1 reverse transcription. // *Virus research*. 2008. T. 134. № 1–2. C. 19–38.

15. Beakes G.W., Glockling S.L., Sekimoto S. The evolutionary phylogeny of the oomycete “fungi” // *Protoplasma*. 2012. T. 249. № 1. C. 3–19.

16. Bebbler D.P., Gurr S.J. Crop-destroying fungal and oomycete pathogens challenge food security // *Fungal Genetics and Biology*. 2015. T. 74. C. 62–64.

17. Besansky N.J. A retrotransposable element from the mosquito *Anopheles gambiae* . // *Molecular and cellular biology*. 1990. T. 10. № 3. C. 863–71.

18. Biémont C. A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution // *Genetics*. 2010. T. 186. № 4.

19. Blumenthal T. Gene clusters and polycistronic transcription in eukaryotes // *BioEssays*. 1998. T. 20. № 6. C. 480–487.

20. Boeke J.D. et al. Ty elements transpose through an RNA intermediate. // *Cell*. 1985. T. 40. № 3. C. 491–500.

21. Boeke J.D., Corces V.G. Transcription and Reverse Transcription of Retrotransposons // *Annual Review of Microbiology*. 1989. T. 43. № 1. C. 403–434.

22. Botstein D. A theory of modular evolution for bacteriophages. // *Annals of the New York Academy of Sciences*. 1980. T. 354. C. 484–490.

23. Brunel C. et al. RNA loop-loop interactions as dynamic functional motifs. // *Biochimie*. 2002. T. 84. № 9. C. 925–44.

24. Cappello J., Handelsman K., Lodish H.F. Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. // *Cell*. 1985. T. 43. № 1. C. 105–15.

25. Cavalier-Smith T., Chao E.E.-Y. Phylogeny and Megasystematics of Phagotrophic Heterokonts (Kingdom Chromista) // *Journal of Molecular Evolution*. 2006. T. 62. № 4.

C. 388–420.

26. Cerritelli S.M., Crouch R.J. Ribonuclease H: the enzymes in eukaryotes. // *The FEBS journal*. 2009. T. 276. № 6. C. 1494–505.
27. 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* // *Journal of Biological Chemistry*. 2002. T. 277. № 20. C. 17877–17882.
28. Chen J.-M., Férec C., Cooper D.N. LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: mutation detection bias and multiple mechanisms of target gene disruption. // *Journal of biomedicine & biotechnology*. 2006. T. 2006. № 1. C. 56182.
29. Contursi C., Minchiotti G., Nocera P.P. Di. Identification of sequences which regulate the expression of *Drosophila melanogaster* Doc elements. // *The Journal of biological chemistry*. 1995. T. 270. № 44. C. 26570–6.
30. Cordaux R., Batzer M.A. The impact of retrotransposons on human genome evolution. // *Nature reviews. Genetics*. 2009. T. 10. № 10. C. 691–703.
31. Crick F. Central dogma of molecular biology. // *Nature*. 1970. T. 227. № 5258. C. 561–3.
32. Dang V.D., Levin H.L. Nuclear import of the retrotransposon Tf1 is governed by a nuclear localization signal that possesses a unique requirement for the FXFG nuclear pore factor Nup124p. // *Molecular and cellular biology*. 2000. T. 20. № 20. C. 7798–812.
33. Darriba D. et al. ProtTest 3: fast selection of best-fit models of protein evolution. // *Bioinformatics (Oxford, England)*. 2011. T. 27. № 8. C. 1164–5.
34. Davis P.S., Judd B.H. Nucleotide sequence of the transposable element, BEL, of *Drosophila melanogaster* // *Drosophila Information Service*. 1995. T. 76. C. 134–136.
35. DeBerardinis R.J., Kazazian H.H. Analysis of the Promoter from an Expanding Mouse Retrotransposon Subfamily // *Genomics*. 1999. T. 56. № 3. C. 317–323.
36. Deininger P.L., Batzer M.A. Alu Repeats and Human Disease // *Molecular Genetics and Metabolism*. 1999. T. 67. № 3. C. 183–193.

37. Dmitriev S.E. 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 // *Molecular and Cellular Biology*. 2007. T. 27. № 13. C. 4685–4697.
38. Doolittle R.F. Convergent evolution: the need to be explicit. // *Trends in biochemical sciences*. 1994. T. 19. № 1. C. 15–8.
39. Doolittle W.F., Sapienza C. Selfish genes, the phenotype paradigm and genome evolution // *Nature*. 1980. T. 284. № 5757. C. 601–603.
40. Eddy S.R. Accelerated Profile HMM Searches. // *PLoS computational biology*. 2011. T. 7. № 10. C. e1002195.
41. Eickbush T.H. Transposing without ends: the non-LTR retrotransposable elements. // *The New biologist*. 1992. T. 4. № 5. C. 430–40.
42. Eickbush T.H. Telomerase and retrotransposons: which came first? // *Science (New York, N.Y.)*. 1997. T. 277. № 5328. C. 911–2.
43. Eickbush T.H., Jamburuthugoda V.K. The diversity of retrotransposons and the properties of their reverse transcriptases. // *Virus research*. 2008. T. 134. № 1–2. C. 221–34.
44. Eickbush T.H., Malik H.S. Origins and Evolution of Retrotransposons // *Mobile DNA II. : American Society of Microbiology*, 2002. C. 1111–1144.
45. Eissenberg J.C. Structural biology of the chromodomain: Form and function // *Gene*. 2012. T. 496. № 2. C. 69–78.
46. Feng Q., Schumann G., Boeke J.D. Retrotransposon R1Bm endonuclease cleaves the target sequence. // *Proceedings of the National Academy of Sciences of the United States of America*. 1998. T. 95. № 5. C. 2083–8.
47. Feng Y.X. et al. The genomic RNA in Ty1 virus-like particles is dimeric. // *Journal of virology*. 2000. T. 74. № 22. C. 10819–21.
48. Feschotte C. Transposable elements and the evolution of regulatory networks // *Nature Reviews Genetics*. 2008. T. 9. № 5. C. 397–405.

49. Fiandt M., Szybalski W., Malamy M.H. Polar mutations in lac, gal and phage lambda consist of a few IS-DNA sequences inserted with either orientation. // *Molecular & general genetics : MGG*. 1972. T. 119. № 3. C. 223–31.
50. Finnegan D.J. Retrotransposons // *Current Biology*. 2012. T. 22. № 11. C. R432–R437.
51. Frost L.S. et al. Mobile genetic elements: the agents of open source evolution // *Nature Reviews Microbiology*. 2005. T. 3. № 9. C. 722–732.
52. Fugmann S.D. The origins of the Rag genes--from transposition to V(D)J recombination. // *Seminars in immunology*. 2010. T. 22. № 1. C. 10–6.
53. Gabriel A. et al. A rapidly rearranging retrotransposon within the miniexon gene locus of *Crithidia fasciculata*. // *Molecular and cellular biology*. 1990. T. 10. № 2. C. 615–24.
54. Gao X. et al. Chromodomains direct integration of retrotransposons to heterochromatin // *Genome Research*. 2008. T. 18. № 3. C. 359–369.
55. Garrett J.E., Knutzon D.S., Carroll D. Composite transposable elements in the *Xenopus laevis* genome. // *Molecular and cellular biology*. 1989. T. 9. № 7. C. 3018–27.
56. George J.A., Eickbush T.H. Conserved features at the 5' end of *Drosophila* R2 retrotransposable elements: implications for transcription and translation. // *Insect molecular biology*. 1999. T. 8. № 1. C. 3–10.
57. Gilbert C. et al. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons // *Nature Communications*. 2014. T. 5. C. 3348.
58. Gladyshev E.A., Arkhipova I.R. A widespread class of reverse transcriptase-related cellular genes. // *Proceedings of the National Academy of Sciences of the United States of America*. 2011. T. 108. № 51. C. 20311–6.
59. Gogvadze E., Buzdin A. Retroelements and their impact on genome evolution and functioning. // *Cellular and molecular life sciences : CMLS*. 2009. T. 66. № 23. C. 3727–42.
60. Goodier J.L. et al. LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced

- Silencing Complex // *Molecular and Cellular Biology*. 2007. T. 27. № 18. C. 6469–6483.
61. Goodier J.L. Restricting retrotransposons: a review // *Mobile DNA*. 2016. T. 7. № 1. C. 16.
62. Gorinsek B., Gubensek F., Kordis D. Evolutionary genomics of chromoviruses in eukaryotes. // *Molecular biology and evolution*. 2004. T. 21. № 5. C. 781–98.
63. Guindon S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. // *Systematic biology*. 2010. T. 59. № 3. C. 307–21.
64. Han J.S. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. // *Mobile DNA*. 2010. T. 1. № 1. C. 15.
65. Han J.S., Boeke J.D. A highly active synthetic mammalian retrotransposon // *Nature*. 2004. T. 429. № 6989. C. 314–318.
66. Hansen L.J., Chalker D.L., Sandmeyer S.B. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. // *Molecular and cellular biology*. 1988. T. 8. № 12. C. 5245–56.
67. Harper J.T., Waanders E., Keeling P.J. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes // *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*. 2005. T. 55. № 1. C. 487–496.
68. Havecker E.R., Gao X., Voytas D.F. The diversity of LTR retrotransposons. // *Genome biology*. 2004. T. 5. № 6. C. 225.
69. Higashiyama T. et al. Zepp, a LINE-like retrotransposon accumulated in the *Chlorella* telomeric region // *The EMBO Journal*. 1997. T. 16. № 12. C. 3715–3723.
70. Hirsch H.J., Starlinger P., Brachet P. Two kinds of insertions in bacterial genes. // *Molecular & general genetics : MGG*. 1972. T. 119. № 3. C. 191–206.
71. Hohjoh H., Singer M.F. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon // *The EMBO Journal*. 1997. T. 16. № 19. C. 6034–6043.

72. Holmes R.K., Malim M.H., Bishop K.N. APOBEC-mediated viral restriction: not simply editing? // *Trends in biochemical sciences*. 2007. T. 32. № 3. C. 118–28.
73. Hu W.-S., Hughes S.H. HIV-1 Reverse Transcription // *Cold Spring Harbor Perspectives in Medicine*. 2012. T. 2. № 10. C. a006882–a006882.
74. Hughes J.F., Coffin J.M. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. // *Nature Genetics*. 2001. T. 29. № 4. C. 487–489.
75. Hull R., Covey S.N. Retroelements: propagation and adaptation. // *Virus genes*. 1995. T. 11. № 2–3. C. 105–18.
76. Hurst G.D.D., Werren J.H. The role of selfish genetic elements in eukaryotic evolution // *Nature Reviews Genetics*. 2001. T. 2. № 8. C. 597–606.
77. Jaaskelainen et al. Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells // *The Plant journal: for cell and molecular biology*. 1999. T. 20. № 4. C. 413–22.
78. Jacks T. et al. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression // *Nature*. 1988. T. 331. № 6153. C. 280–283.
79. Jensen S., Heidmann T. An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. // *The EMBO journal*. 1991. T. 10. № 7. C. 1927–37.
80. Jin Y.K., Bennetzen J.L. Structure and coding properties of Bs1, a maize retrovirus-like transposon. // *Proceedings of the National Academy of Sciences of the United States of America*. 1989. T. 86. № 16. C. 6235–9.
81. Jurka J. et al. Repetitive sequences in complex genomes: structure and evolution. // *Annual review of genomics and human genetics*. 2007. T. 8. C. 241–59.
82. Kajikawa M., Okada N. LINEs mobilize SINEs in the eel through a shared 3' sequence. // *Cell*. 2002. T. 111. № 3. C. 433–44.
83. Kapitonov V. V et al. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons // *PLoS Biology*. 2005. T. 3. № 6. C. e181.

84. Kapitonov V. V., Jurka J. Rolling-circle transposons in eukaryotes // *Proceedings of the National Academy of Sciences*. 2001. T. 98. № 15. C. 8714–8719.
85. 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. T. 448. № 2. C. 207–213.
86. Kidwell M.G. Transposable elements and the evolution of genome size in eukaryotes. // *Genetica*. 2002. T. 115. № 1. C. 49–63.
87. Kidwell M.G., Lisch D.R. Perspective: transposable elements, parasitic DNA, and genome evolution. // *Evolution; international journal of organic evolution*. 2001. T. 55. № 1. C. 1–24.
88. Kim Y.-J., Lee J., Han K. Transposable Elements: No More “Junk DNA”. // *Genomics & informatics*. 2012. T. 10. № 4. C. 226–33.
89. Kojima K.K., Fujiwara H. An extraordinary retrotransposon family encoding dual endonucleases // *Genome Research*. 2005a. T. 15. № 8. C. 1106–1117.
90. Kojima K.K., Jurka J. Ancient Origin of the U2 Small Nuclear RNA Gene-Targeting Non-LTR Retrotransposons Utopia // *PLOS ONE*. 2015. T. 10. № 11. C. e0140084.
91. 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. // *Molecular and cellular biology*. 2005b. T. 25. № 17. C. 7675–86.
92. Kolosha V.O., Martin S.L. High-affinity, Non-sequence-specific RNA Binding by the Open Reading Frame 1 (ORF1) Protein from Long Interspersed Nuclear Element 1 (LINE-1) // *Journal of Biological Chemistry*. 2003. T. 278. № 10. C. 8112–8117.
93. Koonin E. V. Viruses and mobile elements as drivers of evolutionary transitions // *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2016. T. 371. № 1701.
94. Kumar A., Bennetzen J.L. Plant retrotransposons. // *Annual review of genetics*. 1999. T. 33. № 1. C. 479–532.
95. Lamour K.H., Win J., Kamoun S. Oomycete genomics: new insights and future

directions // FEMS Microbiology Letters. 2007. T. 274. № 1. C. 1–8.

96. Lapkouski M. et al. Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. // Nature structural & molecular biology. 2013. T. 20. № 2. C. 230–6.

97. Lerat E. et al. Is the evolution of transposable elements modular? // Genetica. 1999. T. 107. № 1–3. C. 15–25.

98. Levis R.W. et al. Transposons in place of telomeric repeats at a Drosophila telomere. // Cell. 1993. T. 75. № 6. C. 1083–93.

99. Lisch D. How important are transposons for plant evolution? // Nature Reviews Genetics. 2012. T. 14. № 1. C. 49–61.

100. Llorens C. et al. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. // Biology direct. 2009. T. 4. C. 41.

101. Llorens C. et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. // Nucleic acids research. 2011. T. 39. № Database issue. C. D70-4.

102. Louis E.J. et al. Chromosomal evolution in Saccharomyces // Nature. 2000. T. 405. № 6785. C. 451–454.

103. Malik H.S. Ribonuclease H evolution in retrotransposable elements. // Cytogenetic and genome research. 2005. T. 110. № 1–4. C. 392–401.

104. Malik H.S., Burke W.D., Eickbush T.H. The age and evolution of non-LTR retrotransposable elements // Molecular Biology and Evolution. 1999a. T. 16. № 6. C. 793–805.

105. Malik H.S., Eickbush T.H. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. // Molecular biology and evolution. 1998. T. 15. № 9. C. 1123–34.

106. Malik H.S., Eickbush T.H. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. // Journal of virology. 1999b. T. 73. № 6. C. 5186–90.

107. Malik H.S., Eickbush T.H. NeSL-1, an ancient lineage of site-specific non-LTR

- retrotransposons from *Caenorhabditis elegans*. // *Genetics*. 2000a. T. 154. № 1. C. 193–203.
108. Malik H.S., Eickbush T.H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. // *Genome research*. 2001. T. 11. № 7. C. 1187–97.
109. Malik H.S., Henikoff S., Eickbush T.H. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. // *Genome research*. 2000b. T. 10. № 9. C. 1307–18.
110. Marchler-Bauer A. et al. CDD: a Conserved Domain Database for the functional annotation of proteins // *Nucleic Acids Research*. 2011. T. 39. № Database. C. D225–D229.
111. Marlor R.L., Parkhurst S.M., Corces V.G. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. // *Molecular and cellular biology*. 1986. T. 6. № 4. C. 1129–34.
112. Martin S.L., Bushman F.D. Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon // *Molecular and Cellular Biology*. 2001. T. 21. № 2. C. 467–475.
113. McClintock B. The origin and behavior of mutable loci in maize. // *Proceedings of the National Academy of Sciences of the United States of America*. 1950. T. 36. № 6. C. 344–55.
114. McCoy J.M. et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. // *Nature*. 2000. T. 403. № 6771. C. 785–789.
115. Meignin C. et al. The 5' untranslated region and Gag product of Idefix, a long terminal repeat-retrotransposon from *Drosophila melanogaster*, act together to initiate a switch between translated and untranslated states of the genomic mRNA. // *Molecular and cellular biology*. 2003. T. 23. № 22. C. 8246–54.
116. 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. T. 54. № 5. C. 685–91.

117. Moran J. V et al. High frequency retrotransposition in cultured mammalian cells. // *Cell*. 1996. T. 87. № 5. C. 917–27.
118. Morgante M. et al. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize // *Nature Genetics*. 2005. T. 37. № 9. C. 997–1002.
119. Morrish T.A. et al. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition // *Nature Genetics*. 2002. T. 31. № 2. C. 159–165.
120. Morrish T.A. et al. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres // *Nature*. 2007. T. 446. № 7132. C. 208–212.
121. Mount S.M., Rubin G.M. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. // *Molecular and cellular biology*. 1985. T. 5. № 7. C. 1630–8.
122. Muotri A.R. et al. The necessary junk: new functions for transposable elements. // *Human molecular genetics*. 2007. T. 16 Spec No. № R2. C. R159–67.
123. Novikov A., Smyshlyaev G., Novikova O. Evolutionary History of LTR Retrotransposon Chromodomains in Plants // *International Journal of Plant Genomics*. 2012. T. 2012. C. 1–17.
124. Novikova O. Chromodomains and LTR retrotransposons in plants. // *Communicative & integrative biology*. 2009. T. 2. № 2. C. 158–62.
125. Nowak E. et al. Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry // *Nature Structural & Molecular Biology*. 2014. T. 21. № 4. C. 389–396.
126. Ohtani N. et al. Identification of the first archaeal Type 1 RNase H gene from *Halobacterium* sp. NRC-1: archaeal RNase HI can cleave an RNA-DNA junction. // *The Biochemical journal*. 2004. T. 381. № Pt 3. C. 795–802.
127. Okazaki S., Ishikawa H., Fujiwara H. Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. // *Molecular and cellular biology*. 1995. T. 15. № 8. C. 4545–52.
128. Orgel L.E., Crick F.H.C. Selfish DNA: the ultimate parasite. // *Nature*. 1980. T. 284.

№ 5757. C. 604–7.

129. Osanai M. et al. Essential Motifs in the 3' Untranslated Region Required for Retrotransposition and the Precise Start of Reverse Transcription in Non-Long-Terminal-Repeat Retrotransposon SART1 // *Molecular and Cellular Biology*. 2004. T. 24. № 18. C. 7902–7913.

130. Paillart J.-C. et al. Dimerization of retroviral RNA genomes: an inseparable pair // *Nature Reviews Microbiology*. 2004. T. 2. № 6. C. 461–472.

131. Pardue M.-L. et al. Two retrotransposons maintain telomeres in *Drosophila*. // *Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*. 2005. T. 13. № 5. C. 443–53.

132. Pei J., Kim B.H., Grishin N. V. PROMALS3D: A tool for multiple protein sequence and structure alignments // *Nucleic Acids Research*. 2008. T. 36. № 7. C. 2295–2300.

133. Platero J.S., Hartnett T., Eissenberg J.C. Functional analysis of the chromo domain of HP1. // *The EMBO journal*. 1995. T. 14. № 16. C. 3977–86.

134. Qiu J. et al. *Saccharomyces cerevisiae* RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. // *Molecular and cellular biology*. 1999. T. 19. № 12. C. 8361–71.

135. Richards T.A. et al. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes // *Proceedings of the National Academy of Sciences*. 2011. T. 108. № 37. C. 15258–15263.

136. Ronquist F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. // *Systematic biology*. 2012. T. 61. № 3. C. 539–42.

137. Sabot F. et al. Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations // *Molecular Genetics and Genomics*. 2005. T. 274. № 2. C. 119–130.

138. Sabot F., Schulman A.H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome // *Heredity*. 2006. T. 97. № 6. C. 381–388.

139. SanMiguel P. Evidence that a Recent Increase in Maize Genome Size was Caused by the Massive Amplification of Intergene Retrotransposons // *Annals of Botany*. 1998.

T. 82. № suppl_1. C. 37–44.

140. Schaack S., Gilbert C., Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. // *Trends in ecology & evolution*. 2010. T. 25. № 9. C. 537–46.

141. Schumann G. 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. // *Molecular and cellular biology*. 1994. T. 14. № 5. C. 3074–84.

142. Shapiro J. The discovery and significance of mobile genetic elements // *Mobile Genetic Elements*. 1995. C. 1–15.

143. Slotkin R.K. et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. // *Cell*. 2009. T. 136. № 3. C. 461–72.

144. Slotkin R.K., Martienssen R. Transposable elements and the epigenetic regulation of the genome. // *Nature reviews. Genetics*. 2007. T. 8. № 4. C. 272–85.

145. Smyshlyaev G.A., Blinov A.G. Evolution and biodiversity of L1 retrotransposons in angiosperm genomes // *Russian Journal of Genetics: Applied Research*. 2012. T. 2. № 1. C. 72–78.

146. Smyth R.P., Davenport M.P., Mak J. The origin of genetic diversity in HIV-1 // *Virus Research*. 2012. T. 169. № 2. C. 415–429.

147. Soanes D., Richards T. a. Horizontal Gene Transfer in Eukaryotic Plant Pathogens. // *Annual review of phytopathology*. 2014. T. 52. C. 583–614.

148. Söding J., Biegert A., Lupas A.N. The HHpred interactive server for protein homology detection and structure prediction // *Nucleic Acids Research*. 2005. T. 33. № SUPPL. 2.

149. Sormacheva I. et al. Vertical Evolution and Horizontal Transfer of CR1 Non-LTR Retrotransposons and Tc1/mariner DNA Transposons in Lepidoptera Species // *Molecular Biology and Evolution*. 2012. T. 29. № 12. C. 3685–3702.

150. Soucy S.M., Huang J., Gogarten J.P. Horizontal gene transfer: building the web of life. // *Nature reviews. Genetics*. 2015. T. 16. № 8. C. 472–82.

151. Stayton C.T. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence // *Evolution*. 2015. T. 69. № 8. C. 2140–2153.
152. Stern D.L. The genetic causes of convergent evolution // *Nature Reviews Genetics*. 2013. T. 14. № 11. C. 751–764.
153. Swergold G.D. Identification, characterization, and cell specificity of a human LINE-1 promoter. // *Molecular and cellular biology*. 1990. T. 10. № 12. C. 6718–29.
154. Szafranski K. et al. Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* tRNA genes. // *Molecular & general genetics : MGG*. 1999. T. 262. № 4–5. C. 772–80.
155. Szak S.T. et al. Molecular archeology of L1 insertions in the human genome. // *Genome biology*. 2002. T. 3. № 10. C. research0052.
156. Taylor A.L. Bacteriophage-induced mutations in *Escherichia coli*. // *Proceedings of the National Academy of Sciences of the United States of America*. 1963. T. 50. № 6. C. 1043–51.
157. Temin H.M., Mizutani S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. // *Nature*. 1970. T. 226. № 5252. C. 1211–3.
158. Toussaint A., Merlin C. Mobile Elements as a Combination of Functional Modules // *Plasmid*. 2002. T. 47. № 1. C. 26–35.
159. Tyler B.M. et al. Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis // *Science*. 2006. T. 313. № 5791. C. 1261–1266.
160. Vicient C.M., Kalendar R., Schulman A.H. Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. // *Genome research*. 2001. T. 11. № 12. C. 2041–9.
161. Vicient C.M., Kalendar R., Schulman A.H. Variability, Recombination, and Mosaic Evolution of the Barley BARE-1 Retrotransposon // *Journal of Molecular Evolution*. 2005. T. 61. № 3. C. 275–291.
162. Vincent Dollard. “Jumping Genes”; Find New Homes in Humans More Often Than Previously Estimated [Электронный ресурс]. URL:

<http://shared.web.emory.edu/whsc/news/releases/2010/06/jumping-genes-find-new-homes-in-humans.html?=&researchnews> (дата обращения: 28.06.2017).

163. Vitte C., Panaud O. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. // *Cytogenetic and genome research*. 2005. Т. 110. № 1–4. С. 91–107.

164. Volff J.-N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes // *BioEssays*. 2006. Т. 28. № 9. С. 913–922.

165. Wenke T. et al. An abundant and heavily truncated non-LTR retrotransposon (LINE) family in *Beta vulgaris* // *Plant Molecular Biology*. 2009. Т. 71. № 6. С. 585–597.

166. Wicker T. et al. A unified classification system for eukaryotic transposable elements. // *Nature reviews. Genetics*. 2007. Т. 8. № 12. С. 973–82.

167. Wijayawardena B.K., Minchella D.J., DeWoody J.A. Hosts, parasites, and horizontal gene transfer // *Trends in Parasitology*. 2013. Т. 29. № 7. С. 329–338.

168. Wright D.A., Voytas D.F. Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. // *Genetics*. 1998. Т. 149. № 2. С. 703–15.

169. Xiong Y., Burke W.D., Eickbush T.H. Pao, a highly divergent retrotransposable element from *Bombyx mori* containing long terminal repeats with tandem copies of the putative R region. // *Nucleic acids research*. 1993a. Т. 21. № 9. С. 2117–23.

170. Xiong Y., Eickbush T.H. Dong, a non-long terminal repeat (non-LTR) retrotransposable element from *Bombyx mori*. // *Nucleic acids research*. 1993b. Т. 21. № 5. С. 1318.

171. 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. // *Proceedings of the National Academy of Sciences of the United States of America*. 1999. Т. 96. № 14. С. 7847–52.

172. Zhou L. et al. Transposition of hAT elements links transposable elements and V(D)J recombination // *Nature*. 2004. Т. 432. № 7020. С. 995–1001.

173. NCBI Open Reading Frame finder [Электронный ресурс]. URL: <https://www.ncbi.nlm.nih.gov/orffinder/> (дата обращения: 30.11.2016).

ПРИЛОЖЕНИЕ

Приложение 1. Разнообразие и источники исследованных геномных последовательностей растений и оомицетов

Класс	Порядок	Семейство	Вид	Источн ик	Версия генома	Ретротранспоз оны с aRNH
Viridiplantae (Зеленые растения)						
Magnoliopsida	Malpighiales	Acalyphoideae	<i>Ricinus communis</i>	#1	V.0.1	-
		Euphorbiaceae	<i>Manihot esculenta</i>	#1	V.4.1	+
		Linaceae	<i>Linum usitatissimum</i>	#1	V.1.0	-
		Salicaceae	<i>Populus trichocarpa</i>	#1	V.3.0	+
	Caryophyllales	Amaranthaceae	<i>Beta vulgaris</i>	#2	V.1.2	+
			<i>Spinacia oleracea</i>	#2	V.1.0	+
	Fabales	Fabaceae	<i>Cicer arietinum</i>	#3	16.01.2013	+
			<i>Glycine max</i>	#1	V.2.0	+
			<i>Medicago truncatula</i>	#1	Mt4.0.v1	+
			<i>Phaseolus vulgaris</i>	#1	V.1.0	+
	Fagales	Betulaceae	<i>Betula nana</i>	#4	V.1.0	+
	Cucurbitales	Cucurbitaceae	<i>Cucumis sativus</i>	#1	V.1	-
	Rosales	Rosaceae	<i>Fragaria vesca</i>	#1	V1.1	+
			<i>Malus domestica</i>	#1	V.1.0	+
			<i>Prunus persica</i>	#1	V.1.0	+
	Brassicales	Brassicaceae	<i>Arabidopsis lyrata</i>	#1	V.1.0	+
			<i>Arabidopsis thaliana</i>	#1	TAIR10	+
			<i>Brassica rapa</i>	#1	V.1.1	+
			<i>Capsella rubella</i>	#1	V.1	+
			<i>Thellungiella halophila</i>	#1	V.1	+
		Caricaceae	<i>Carica papaya</i>	#1	ASGPBv0.4	-
	Malvales	Malvaceae	<i>Gossypium raimondii</i>	#1	V.2.0	+
			<i>Theobroma cacao</i>	#1	V.1.0	+
	Sapindales	Rutaceae	<i>Citrus clementina</i>	#1	V.1.0	+
			<i>Citrus sinensis</i>	#1	V.1.0	+
	Myrtales	Myrtaceae	<i>Eucalyptus grandis</i>	#1	V.1.0	+
	Vitales	Vitaceae	<i>Vitis vinifera</i>	#1	Genoscope.1	+
	Solanales	Solanaceae	<i>Capsicum annuum</i>	#5	V.2.0	+
			<i>Nicotiana glauca</i>	#3	Primary, 2013-05-16	+

			<i>Solanum lycopersicum</i>	#1	ITAGv2.3	+
			<i>Solanum tuberosum</i>	#1	V.3.4	+
	Lamiales	Phrymaceae	<i>Mimulus guttatus</i>	#1	V.1.0	+
		Oleaceae	<i>Fraxinus excelsior</i>	#4	V.0.1	-
	Ranunculales	Ranunculaceae	<i>Aquilegia coerulea</i>	#1	V.1.0	+
	Proteales	Nelumbonaceae	<i>Nelumbo nucifera</i>	#3	V.1.1	+
Liliopsida	Poales	Poaceae	<i>Brachypodium distachyon</i>	#1	V.1.0	+
			<i>Hordeum vulgare</i>	#6	V.1.0	-
			<i>Oryza sativa</i>	#1	V.7.0	+
			<i>Panicum virgatum</i>	#1	V.1.0	+
			<i>Setaria italica</i>	#1	V.2.0	+
			<i>Sorghum bicolor</i>	#1	V.1.0	+
			<i>Zea mays</i>	#1	V.2	+
	Arecales	Arecaceae	<i>Elaeis guineensis</i>	#3	2013-08-13	+
			<i>Phoenix dactylifera</i>	#3	2013-06-24	-
	Zingiberales	Musaceae	<i>Musa acuminata</i>	#6	V.1	+
	Alismatales	Araceae	<i>Spirodela polyrhiza</i>	#3	2013-12-06	-
Basal Magnoliophyta	Amborellales	Amborellaceae	<i>Amborella trichopoda</i>	#6	V.1.0	+
Pinopsida	Pinales	Pinaceae	<i>Picea abies</i>	#4	V.1.0	+
			<i>Picea glauca</i>	#4	V.1.0	+
			<i>Pinus taeda</i>	#4	V.1.01	+
Lycopodiopsida	Selaginellales	Selaginellaceae	<i>Selaginella moellendorffii</i>	#1	V.1.0	+
Bryopsida	Andreaeales	Andreaeaceae	<i>Andreaea rothii</i>	#7	черновая сборка	-
	Bryales	Bryaceae	<i>Bryum argenteum</i>	#7	черновая сборка	-
	Buxbaumiales	Buxbaumiaceae	<i>Buxbaumia aphylla</i>	#7	черновая сборка	-
	Dicranales	Dicranaceae	<i>Dicranum scoparium</i>	#7	черновая сборка	-
	Funariales	Funariaceae	<i>Physcomitrella patens</i>	#1	V.3.0	-
			<i>Entosthodon longicollis</i>	#7	черновая сборка	-
	Hypnales	Climaciaceae	<i>Climacium dendroides</i>	#7	черновая сборка	-
	Ptychomniales	Ptychomniaceae	<i>Ptychomnion cygnisetum</i>	#7	черновая сборка	-
Chlorophyceae	Chlamydomonadales	Chlamydomonadaeae	<i>Chlamydomonas reinhardtii</i>	#1	V.5.0	-
		Volvocaceae	<i>Volvox carteri</i>	#1	V.2	-

Mamiellophyceae	Mamiellales	Mamiellaceae	<i>Micromonas pusilla</i>	#1	V.3.0	-
			<i>Ostreococcus lucimarinus</i>	#1	V.2.0	-
Trebouxiophyceae		Coccomyxaceae	<i>Coccomyxa subellipsoidea</i>	#1	V.2.0	-
Stramenopiles/Oomycetes (Оомицеты)						
Oomycetes	Albuginales	Albuginaceae	<i>Albugo candida</i>	#11	GCA_0010785 35.1	-
			<i>Albugo laibachii</i>	#11	ENA1	-
	Peronosporales	Peronosporaceae	<i>Hyaloperonospora arabidopsidi</i>	#3	GCA_0001732 35.2	+
			<i>Phytophthora alni</i>	#3	GCA_0004393 35	+
			<i>Phytophthora capsici</i>	#9	v11.0	+
			<i>Phytophthora cinnamomi</i>	#9	v1.0	+
			<i>Phytophthora infestans</i>	#3	v1	+
			<i>Phytophthora kernoviae</i>	#11	GCA_0003330 75	+
			<i>Phytophthora lateralis</i>	#11	GCA_0003184 65	+
			<i>Phytophthora parasitica</i>	#10	v1	+
			<i>Phytophthora pinifolia</i>	#9	GCA_0005002 25.1	+
			<i>Phytophthora ramorum</i>	#9	v1.0	+
			<i>Phytophthora sojae</i>	#9	v3.0	+
			<i>Phytophthora vexans</i>	#8	DAOM BR484	+
			<i>Pseudoperonospora cubensis</i>	#3	GCA_0002526 05.1	-
	Pythiales	Pythiaceae	<i>Pythium aphanidermatum</i>	#8	DAOM BR444	+
			<i>Pythium arrhenomanes</i>	#8	ATCC 12531	+
			<i>Pythium insidiosum</i>	#8	JRHR01	+
			<i>Pythium irregulare</i>	#8	DAOM BR486	+
			<i>Pythium iwayamai</i>	#8	DAOM BR242034	+
			<i>Pythium ultimum</i>	#8	DAOM BR144	+
	Saprolegniales	Saprolegniaceae	<i>Aphanomyces astaci</i>	#10	v1	+
			<i>Aphanomyces invadans</i>	#10	v1	-
			<i>Saprolegnia diclina</i>	#3	GCF_00028104 5.1	+
			<i>Saprolegnia parasitica</i>	#3	GCA_0001515 45.1	+

#1. Phytozome (<http://www.phytozome.net/>)

#2. Beta vulgaris resource (<http://www.bvseq.molgen.mpg.de/>)

- #3. NCBI Genome DB (<https://www.ncbi.nlm.nih.gov/genome/>)
- #4. Pinus Genome Database (<http://www.pinegenome.org/pinerefseq/>)
- #5. Pepper Genome Database (<http://www.peppersequence.genomics.cn>)
- #6. Ensembl Plants (<http://www.plants.ensembl.org>).
- #7. Частная база данных, данные от Новиковой О. С. (личный контакт)
- #8. Pythium Genome Database (<http://pythium.plantbiology.msu.edu>)
- #9. Joint Genome Institute (<http://genome.jgi.doe.gov/>)
- #10. Broad Institute (<http://www.broadinstitute.org>)
- #11. Ensembl Protist (<http://protists.ensembl.org>)

Приложение 2. Разнообразие и распространение мобильных элементов с aRNH, представленных в базе данных Repbase Update.

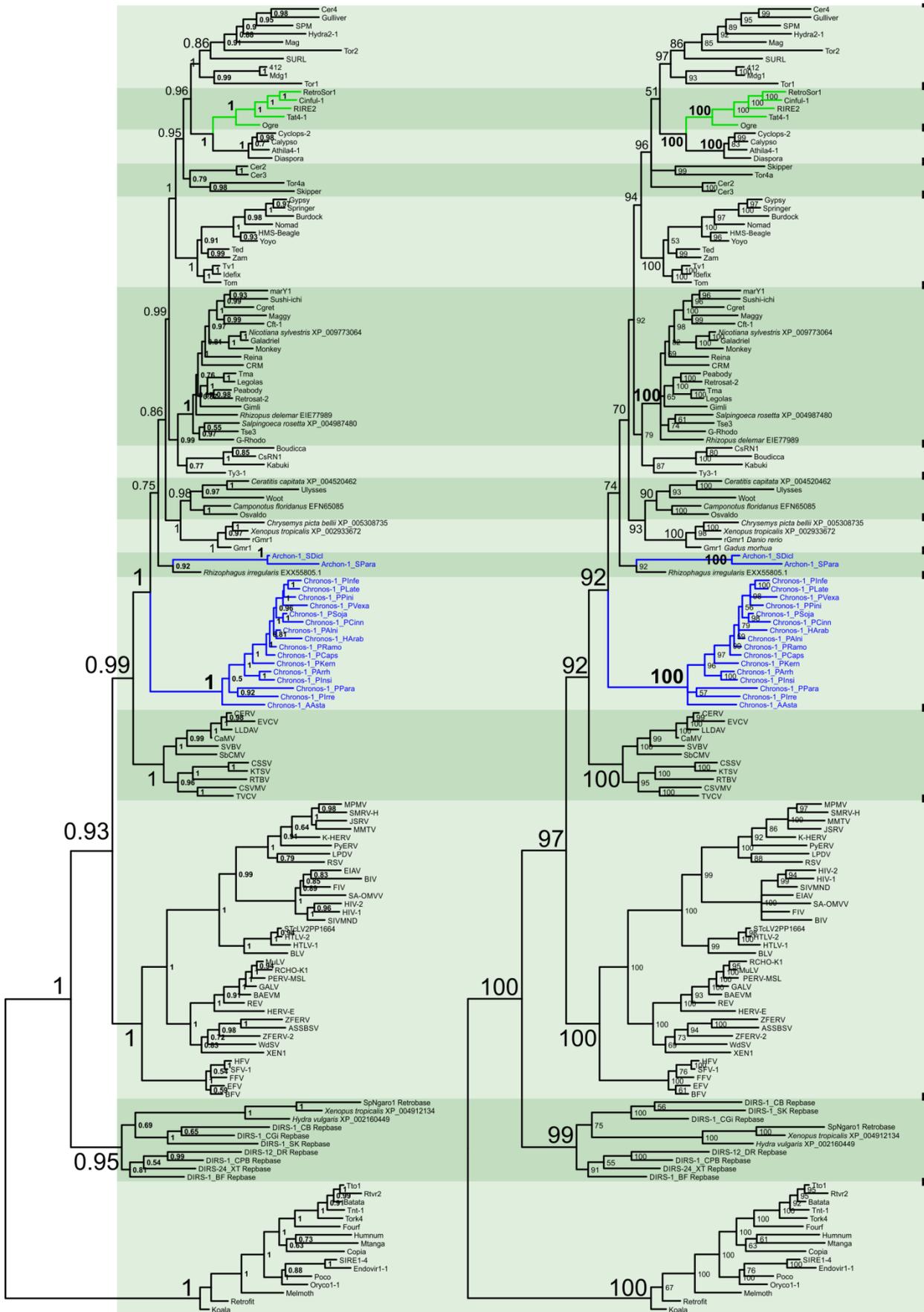
Вид	Тип повтора	Число различных элементов
Stramenopiles/Oomycetes/Peronosporales		
<i>Phytophthora infestans</i> T30-4	LTR/Gypsy	3
<i>Phytophthora parasitica</i> INRA-310	LTR/Gypsy	1
<i>Phytophthora ramorum</i>	LTR/Gypsy	3
<i>Phytophthora sojae</i>	LTR/Gypsy	8
<i>Phytophthora infestans</i>	Non-LTR/NeSL	4
<i>Phytophthora infestans</i>	Non-LTR/L1	3
Viridiplantae/Embryophyta		
<i>Amborella trichopoda</i>	LTR/Gypsy	10
<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>	LTR/Gypsy	8
<i>Arabidopsis thaliana</i>	LTR/Gypsy	3
<i>Brachypodium distachyon</i>	LTR/Gypsy	19
<i>Brassica rapa</i>	LTR/Gypsy	1
<i>Fragaria vesca</i>	LTR/Gypsy	3
<i>Glycine max</i>	LTR/Gypsy	5
<i>Gossypium raimondii</i>	LTR/Gypsy	4
<i>Hordeum vulgare</i>	LTR/Gypsy	1
<i>Hordeum vulgare</i> subsp. <i>vulgare</i>	LTR/Gypsy	1
<i>Jatropha curcas</i>	LTR/Gypsy	2
<i>Lotus corniculatus</i>	LTR/Gypsy	1
<i>Malus domestica</i>	LTR/Gypsy	11
<i>Medicago truncatula</i>	LTR/Gypsy	7
<i>Nelumbo nucifera</i>	LTR/Gypsy	6
<i>Oryza sativa</i>	LTR/Gypsy	8
<i>Oryza sativa Japonica Group</i>	LTR/Gypsy	2
<i>Picea abies</i>	LTR/Gypsy	42
<i>Pinus taeda</i>	LTR/Gypsy	6
<i>Pisum sativum</i>	LTR/Gypsy	1
<i>Populus trichocarpa</i>	LTR/Gypsy	4
<i>Prunus persica</i>	LTR/Gypsy	4
<i>Pyrus x bretschneideri</i>	LTR/Gypsy	1
<i>Saccharum hybrid cultivar R570</i>	LTR/Gypsy	12
<i>Selaginella moellendorffii</i>	LTR/Gypsy	1
<i>Solanum demissum</i>	LTR/Gypsy	1
<i>Solanum lycopersicum</i>	LTR/Gypsy	1
<i>Solanum tuberosum</i>	LTR/Gypsy	1
<i>Sorghum bicolor</i>	LTR/Gypsy	15
<i>Theobroma cacao</i>	LTR/Gypsy	2
<i>Triticum aestivum</i>	LTR/Gypsy	7

<i>Triticum monococcum</i>	LTR/Gypsy	1
<i>Vicia pannonica</i>	LTR/Gypsy	1
<i>Vitis vinifera</i>	LTR/Gypsy	7
<i>Zea diploperennis</i>	LTR/Gypsy	1
<i>Zea mays</i>	LTR/Gypsy	21
<i>Amborella trichopoda</i>	Non-LTR/L1	22
<i>Arabidopsis lyrata</i>	Non-LTR/L1	3
<i>Arabidopsis thaliana</i>	Non-LTR/L1	8
<i>Beta vulgaris</i>	Non-LTR/L1	2
<i>Carica papaya</i>	Non-LTR/L1	1
<i>Eucalyptus camaldulensis</i>	Non-LTR/L1	1
<i>Eutrema parvulum</i>	Non-LTR/L1	1
<i>Fragaria vesca</i>	Non-LTR/L1	20
<i>Gossypium raimondii</i>	Non-LTR/L1	1
<i>Jatropha curcas</i>	Non-LTR/L1	3
<i>Malus domestica</i>	Non-LTR/L1	12
<i>Medicago truncatula</i>	Non-LTR/L1	7
<i>Nelumbo nucifera</i>	Non-LTR/L1	28
<i>Nelumbo nucifera subsp. nucifera</i>	Non-LTR/L1	19
<i>Oryza sativa</i>	Non-LTR/L1	2
<i>Oryza sativa Japonica Group</i>	Non-LTR/L1	1
<i>Picea abies</i>	Non-LTR/L1	3
<i>Populus trichocarpa</i>	Non-LTR/L1	1
<i>Solanum tuberosum</i>	Non-LTR/L1	2
<i>Sorghum bicolor</i>	Non-LTR/L1	3
<i>Theobroma cacao</i>	Non-LTR/L1	3
<i>Zea mays</i>	Non-LTR/L1	1

Приложение 3. Полное филогенетическое древо LTR-ретротранспозонов, реконструированное методом максимального правдоподобия и байесовским методом на основе аминокислотных последовательностей доменов RT. Значения статистической поддержки, указанные в узлах ветвей, была определена с помощью aLRT (доли единицы) и независимых запусков MCMC-цепей (%) в реконструкциях методами максимального правдоподобия и Байесса, соответственно. Названия кластеры, элементы которых содержат CHD, подчеркнуты. Названия кластеров с aRNH-содержащими элементами выделены синим и зеленым для элементов оомицетов и растений, соответственно. Если не указано иначе, имена взятых последовательностей LTR-ретротранспозонов соответствуют таковым в базе данных GyDB.

PhyML

MrBayes



Mag/Mdg1

Tat

Athila
Skipper

ErranVir

ChrVir

CsRN1/Ty3
Oswaldo
GMR1
Archon

Chronos

CauliVir

RTV

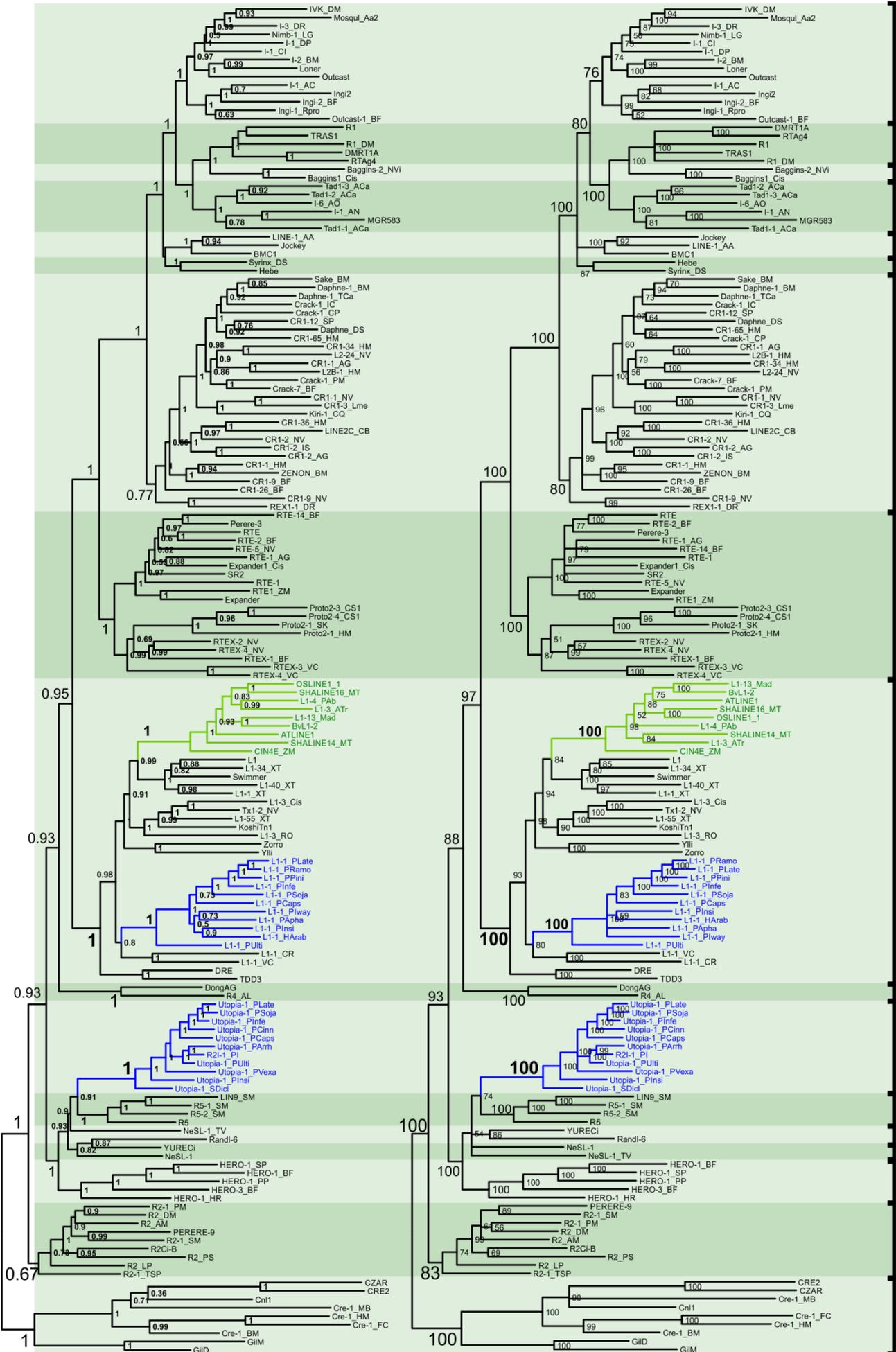
DIRS

Ty1/Copia

Приложение 4. Полное филогенетическое древо non-LTR-ретротранспозонов, реконструированное методом максимального правдоподобия и байесовским методом на основе аминокислотных последовательностей доменов RT. Значения статистической поддержки, указанные в узлах ветвей, была определена с помощью aLRT (доли единицы) и независимых запусков МСМС-цепей (%) в реконструкциях методами максимального правдоподобия и Байесса, соответственно. Названия кластеров с aRNH-содержащими элементами выделены синим и зеленым для элементов оомицетов и растений, соответственно. Если не указано иначе, имена взятых последовательностей non-LTR-ретротранспозонов соответствуют таковым в базе данных Repbase Update.

PhyML

MrBayes



I
R1
LOA
Tad
 Jockey
 Syrinx/Hebe

CR1

RTE

растения
L1

L1

оомицеты
L1

R4/Dong

Utopia

R5
 Dualen/Randl
 NeSL

HERO

R2

CRE

Приложение 5. Полное филогенетическое древо генов и доменов RNH типа I, реконструированное методом максимального правдоподобия и байесовским методом на основе аминокислотных последовательностей. Значения статистической поддержки, указанные в узлах ветвей, была определена с помощью aLRT (доли единицы) и независимых запусков MCMC-цепей (%) в реконструкциях методами максимального правдоподобия и Байесса, соответственно. Названия кластеров aRNH выделены синим и зеленым для геномов оомицетов и растений, соответственно. Имена последовательностей RNH, относящихся к остальным ретротранспозонам соответствуют таковым из баз данных GyDB и Repbase Update. Для остальных последовательностей RNH справа от названия последовательности указаны идентификаторы базы данных NCBI.

PhyML

MrBayes

