

На правах рукописи

Устьянцев Кирилл Валерьевич

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

03.02.07 – генетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук

Новосибирск 2018

Работа выполнена в лаборатории молекулярно-генетических систем Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения российской академии наук», г. Новосибирск.

**Научный
руководитель:**

Блинов Александр Геннадьевич
к.б.н., ведущий научный сотрудник лаборатории
молекулярно-генетических систем ФГБУН «ФИЦ
Институт цитологии и генетики СО РАН»,
г. Новосибирск

**Официальные
оппоненты:**

Щербаков Дмитрий Юрьевич
д.б.н., заведующий лабораторией геносистематики,
ФГБУН «Лимнологический институт СО РАН»,
г. Иркутск

Бажан Сергей Иванович
д.б.н., заведующий теоретическим отделом, ФГБУН
«Государственный научный центр вирусологии и
Биотехнологии «Вектор» Федеральной службы по
надзору в сфере защиты прав потребителей и
благополучия человека», г. Кольцово, Новосибирская
область

Ведущее учреждение: Федеральное государственное бюджетное учреждение науки Центральный сибирский ботанический сад Сибирского отделения Российской академии наук, г. Новосибирск

Защита диссертации состоится «___» _____ 2018 г. на утреннем заседании диссертационного совета Д 003.011.01 на базе ФГБУН «ФИЦ Институт цитологии и генетики СО РАН» в конференц-зале Института по адресу:
пр. академика Лаврентьева 10, г. Новосибирск, 630090
тел: +7(383) 363-49-06 (1321); e-mail: dissov@bionet.nsc.ru.
факс: +7(383) 333-12-78

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте
Института: www.bionet.nsc.ru.

Автореферат разослан «___» _____ 2018 г.

Ученый секретарь
диссертационного совета,
доктор биологических наук

Т.М. Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Мобильные генетические элементы (МЭ) – важные 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 котором новая копия ретротранспозона синтезируется по матрице его мРНК в процессе обратной транскрипции. 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 является домен обратной транскриптазы (reverse transcriptase – RT) – ключевой фермент обратной транскрипции. Набор 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 дополнительный домен ретротранспозона может иметь эволюционную историю независимую от истории центрального RT-домена, так как на различных этапах эволюции отдельных групп ретроэлементов тот или иной функциональный домен приобретается 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.

Одним из наиболее часто приобретаемых и теряющихся доменов в эволюции различных групп ретроэлементов является функциональный домен рибонуклеазы Н (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 *Ta11* non-LTR-ретротранспозоны зеленых растений из группы *L1*, а также non-LTR-ретротранспозоны из геномов паразитических простейших – оомицетов – 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 хозяйствственно-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.

Целью данной работы является выявление пути эволюции ретротранспозонов с «архейным» доменом рибонуклеазы Н 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 Н в геномах растений и оомицетов.
2. Реконструкция филогении основных эволюционных групп ретротранспозонов с «архейным» доменом рибонуклеазы Н 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 геномах 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 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, 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 в данной работе алгоритм поиска новых групп ретротранспозонов с доменом «архейной» рибонуклеазы Н может быть применен для идентификации групп ретротранспозонов с другими доменами и на новых геномных данных.

Положения, выносимые на защиту:

1. В геномах растений и оомицетов присутствуют LTR- и non-LTR-ретротранспозоны, содержащие домен «архейной» рибонуклеазы Н.
2. Домен «архейной» рибонуклеазы Н 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 Н 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.

Апробация работы. Результаты данной работы были представлены на двух международных научных конференциях. По теме диссертации были опубликованы две статьи в рецензируемых зарубежных журналах.

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.

МАТЕРИАЛЫ И МЕТОДЫ

Компьютерный поиск ретротранспозонов с доменом «архейной» рибонуклеазы Н (aRNH) в последовательностях полных геномов растений и оомицетов, а также в базе данных повторенных последовательностей Repbase Update, проводили с помощью алгоритма поиска с использованием профилей скрытых марковских моделей. Данные профили чувствительны к аминокислотным последовательностям домена aRNH и домена обратной транскриптазы (RT) как LTR-, так и non-LTR-ретротранспозонов.

Для каждой отобранный последовательности-представителя ретротранспозона был проведен детальный анализ его структуры. Программа NCBI ORFfinder (NCBI Open Reading Frame finder) была использована для поиска открытых рамок считывания, а программы NCBI CD-search (Marchler-Bauer et al., 2015) и HHpred (Söding et al., 2005) применяли для поиска гомологии к известным консервативным белковым доменам. Предсказание вторичной структуры для последовательностей aRNH производили в программе Quick2D из биоинформационного пакета программ MPI (Alva et al., 2016).

Аминокислотные последовательности RT-доменов отобранных представителей ретротранспозонов растений и оомицетов, содержащих aRNH, были выровнены на с помощью программы hmmlalign пакета HMMER (Eddy, 2011). Аминокислотные последовательности рибонуклеазы Н (RNH) были выровнены с помощью программы PROMALS3D, которая использует данные о предсказании вторичной структуры для оптимизации локальных участков выравнивания (Pei et al., 2008).

Реконструкцию филогении на основе полученных выравниваний доменов RT и RNH проводили методами 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 была определена в программе ProtTest (Darriba et al., 2011).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

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

В базе данных Repbase Update (RU) все значимые сигналы aRNH относились к последовательностям ретротранспозонов аннотированных у 42 видов зеленых растений и пяти видов оомицетов из рода *Phytophthora* (Oomycetes, Peronosporales). Согласно классификации, используемой в базе данных RU, ретротранспозоны с aRNH встречаются в следующих группах: “LTR/Gypsy”, “Non-LTR/Utopia” и “Non-LTR/L1”. 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* (клuster *Utopia*) у оомицетов. Ретротранспозоны с aRNH в RU появляются у зеленых растений, начиная с отдела плауновидных. Всего было проанализировано 65 геномов зеленых растений и 25 геномов оомицетов. 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. LTR-элементы с aRNH не были обнаружены в исследованных геномах зеленых водорослей и мхов и, как и в случае с анализов базы данных RU, они появляются, начиная с плауновидных растений, встречаясь

далее у представителей всех крупных таксонов зеленых растений. В тоже время *L1*-элементы с аРНН можно обнаружить, начиная с таксонов голосеменных хвойных растений. В базах данных были доступны геномные последовательности четырех отрядов оомицетов, занимающих три различных экологических ниши (Beakes et al., 2012): Albuginales, Peronosporales, Pythiales и Saprolegniales. В геномах оомицетов ретротранспозоны с аРНН встречаются уже в наиболее рано дивергировавшем отряде Saprolegniales. Хотя и в штучном количестве, у Saprolegniales были найдены как *Ty3/Gypsy* LTR-элементы с аРНН, так и non-LTR-элементы *Utopia*. У представителей Albuginales не удалось обнаружить каких-либо ретротранспозонов с аРНН, а уже в геномах отрядов Pythiales и Peronosporales встречаются все три группы ретроэлементов с аРНН.

Структурный и филогенетический анализ ретротранспозонов растений и оомицетов, содержащих домен «архейной» рибонуклеазы Н

Всего был выявлен один эволюционный кластер LTR-ретротранспозонов растений и два кластера LTR-элементов у оомицетов (Рисунок 1). Филогенетический анализ RT-доменов показал, что все найденные LTR-элементы растений, содержащие аРНН, 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 аРНН. Ближайшим родственным кластером для *Tat* является кластер *Athila*. Элементы данного кластера также обнаружаются лишь в геномах растений (Vicient et al., 2001). Как и *Athila*, *Tat* LTR-ретротранспозоны характеризуются наличием дополнительных открытых рамок считывания (OPC), которые могут иметь различное положение относительно гена *pol* (Steinbauerová et al., 2011). Выявлено шесть филогенетических кластеров *Tat* LTR-ретротранспозонов, характеризующихся различной структурой гена *pol* (Рисунок 1). Кластер I представлен единственным элементом, найденным в

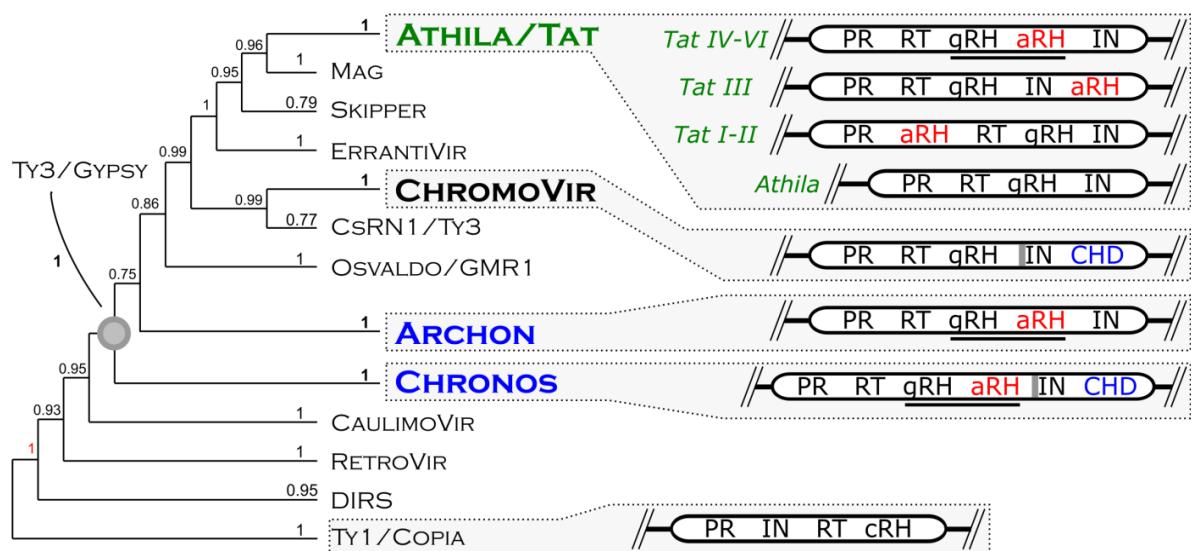


Рисунок 1. Консенсусное дерево филогенетических отношений LTR-ретротранспозонов реконструированное по аминокислотным последовательностям их RT-доменов. Кластеры *Tat* (зеленый) и *Archon* и *Chronos* (синий) принадлежат геномам растений и оомицетов, соответственно. В узлах дерева и на его ветвях обозначена статистическая aLRT-поддержка топологии кластера в реконструкции методом максимального правдоподобия (доли единицы). Справа от ветвей изображена схема консенсусной структуры гена *pol* (OPC 2) элементов соответствующего кластера. aRH – aRNH, gRH – gRNH, IN – INT. Vir – сокращение для "viridae".

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. 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 OPC 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. 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).

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 1). 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 (Рисунок 1). Второй кластер, названный *Chronos*, по большей части представлен элементами из геномов отрядов Peronosporales и Pythiales и единственной копией в геноме *Aphanomyces invadans* (Saprolegniales). Элементы *Chronos* также имеют структуру gRNH-aRNH, но дополнительно к этому, несут CHD на 3' конце гена *pol*, сразу за доменом INT (Рисунок 1). 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 1). 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.

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 *Ta11 L1*-элементами из геномов зеленых растений (Рисунок 2). В обеих группах домен aRNH расположен на 3' конце OPC 2. Более того, у обеих групп CCHC цистеиновый мотив расположен перед aRNH, в то время как во всех остальных кластерах non-LTR-ретротранспозонов с доменом RNH, CCHC расположен после RNH (Smyshlyaeve et al., 2012). Однако, несмотря на сходства в общей структурной организации, aRNH-содержащие элементы зеленых растений и оомицетов не образуют единого монофилетичного кластера внутри группы *L1*. Отсутствие

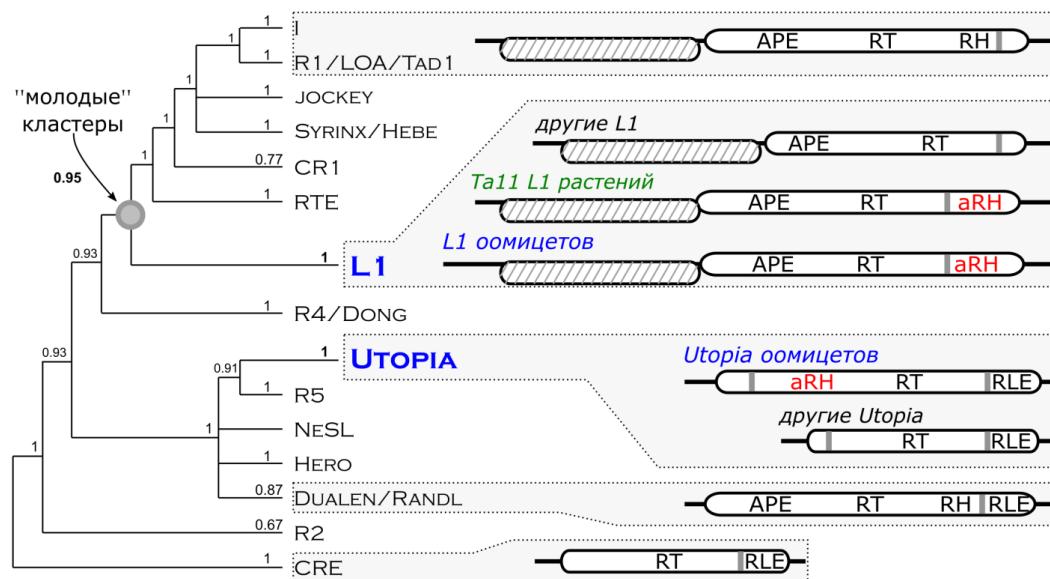


Рисунок 2. Консенсусное дерево филогенетических отношений non-LTR-ретротранспозонов, реконструированное по аминокислотным последовательностям их RT-доменов. В узлах дерева и на его ветвях обозначена статистическая aLRT-поддержка топологии кластера в реконструкции методом максимального правдоподобия (доли единицы). Справа от ветвей изображена схема консенсусной структуры OPC 2 элементов соответствующего кластера. OPC обозначены овалами, OPC 1 заштрихована, RH – RNH non-LTR-элементов, aRH – aRNH.

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*. 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 Н у ретротранспозонов растений и оомицетов

Последовательности aRNH не всегда были расположены совместно или в непосредственной близости от RT или других доменов ретротранспозонов. Далее в тексте такие последовательности aRNH будут обозначаться как индивидуальные aRNH. Также мы произвели поиск последовательностей, другого распространенного кластера эукариотических RNH – fmRNH, – для обнаружения потенциальных последовательностей клеточных генов RNH. 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. Ввиду 100% распространенности и низкого числа копий на геном, fmRNH подходит 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 3). 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 монофилетичного 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 3). Стоит отметить, что в отличие от aRNH 2, последовательности кластеров aRNH 1 и aRNH 3 были найдены в большем количестве копий. Учитывая это, а также совместную кластеризацию aRNH 1 и

aRNH 3 с доменами aRNH ретротранспозонов оомицетов, можно предположить, что они могут представлять собой остатки элементов *Archon* и *L1*, соответственно. Учитывая низкую копийность, а также относительно широкое распределение домена aRNH 2 среди исследованных геномов оомицетов, 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. Индивидуальные aRNH были обнаружены во всех геномах свободноживущих Stramenopiles, в то время как сигнал fmRNH полностью отсутствовал. Индивидуальные aRNH свободноживущих Stramenopiles образуют кластер на филогенетическом древе. При этом их последовательности не показали значимой кластеризации ни с одним из других исследованных кластеров aRNH (Рисунок 3). aRNH отсутствует в некоторых исследованных таксонах оомицетов, а в тех таксонах, где она есть, не показывает никакой значимой кластеризации с последовательностями свободноживущих Stramenopiles (Рисунок 3). С одной стороны, можно предположить, что aRNH присутствовала у общего

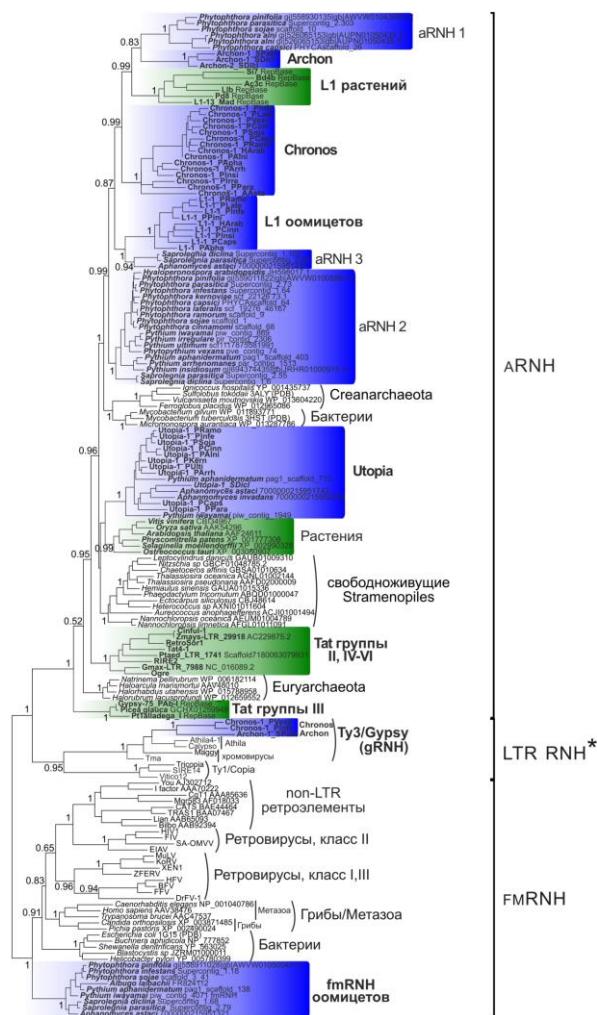


Рисунок 3. Древо филогенетических взаимоотношений RNH типа I из различных источников, реконструированное методом максимального правдоподобия. В узлах дерева и на его ветвях обозначена статистическая поддержка топологии кластера aLRT (доли единицы). Кластеры, соответствующие последовательностям RNH из геномов оомицетов и растений, выделены синим и зеленым, соответственно.

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). Отсутствие fmRNH у свободноживущих Stramenopiles наиболее вероятно отражает то, что оомицеты приобрели данный ген уже после отделения от общего корня Stramenopiles.

Во всех геномах зеленых растений индивидуальные aRNH были обнаружены, включая зеленые водоросли и мхи, варьируя в количестве от одной до трех копий на геном, и они полностью соответствовали последовательностям известных клеточных генов RNH. Все индивидуальные aRNH зеленых растений образуют одну монофилетическую группу, внутри которой топология дерева повторяет общепринятую топологию дерева эволюции зеленых растений (Рисунок 3). Последовательности доменов aRNH *Tat* LTR-ретротранспозонов не показали значимой кластеризации ни друг с другом, ни с кластерами aRNH ретротранспозонов оомицетов, ни с кластером генов зеленых растений (Рисунок 3). Последовательности доменов aRNH *Tat* LTR-ретротранспозонов не показали значимой кластеризации ни друг с другом, ни с кластерами aRNH ретротранспозонов оомицетов, ни с кластером генов зеленых растений (Рисунок 3). Кластер aRNH доменов *Tat* non-LTR-ретротранспозонов растений показал значимую кластеризацию с доменами aRNH *Chronos* на древе. Домены aRNH *Tat* LTR-ретротранспозонов не формируют монофилетического кластера на древе. Выделяются три отдельных кластера aRNH, соответствующих *Tat* II, IV–VI, *Tat* III и «кластеру», представленному единственной последовательностью домена aRNH *Tat* I из плауна *S. moellendorffii* (Рисунки 1 и 3). 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

Для того чтобы проверить, оказалось ли приобретение домена RNH влияние на последовательности gRNH-домена LTR-ретротранспозонов *Tat*, *Chronos* и *Archon* аналогичное влиянию домена fmRNH ретровирусов, было проведено детальное сравнение аминокислотных последовательностей генов и доменов RNH из различных источников. Анализ множественного выравнивания аминокислотных последовательностей RNH позволил выявить несколько интересных закономерностей. Приобретенные домены aRNH исследованных здесь ретротранспозонов характеризуются полноценным набором консервативных катализических оснований: D-E-D-D (Рисунок 4). Таким полным набором обладают все известные каталитически активные домены RNH типа I (Cerritelli et al., 2009; Smyshlyaev et al., 2013). Все домены aRNH также характеризуются наличием положительно заряженной аминокислоты аргинина (R) в полуконсервативной позиции выравнивания. У последовательностей fmRNH в данной позиции

находится гистидин (H). Исходные домены RNH LTR-ретротранспозонов не обладают ни R, ни H в данной позиции, что может объяснить их сниженную катализическую активность по сравнению с ферментами из кластера fmRNH и aRNH (Smyshlyayev et al., 2013). Считается, что сниженная активность домена RNH нужна LTR-ретротранспозонам для сохранения полипуринового праймера, с которого начинается синтез плюс-цепи кДНК. Однако для доменов RNH LTR-ретротранспозонов всё равно характерна полная катализическая тетрада D-E-D-D (Рисунок 4). На этом фоне исходные домены gRNH LTR-ретротранспозонов с aRNH проявляют практически полную потерю консервативных катализических оснований (Рисунок 4), соответствую гипотезе, что приобретенный домен aRNH берет ферментативную функцию на себя, как это было с клеточным fmRNH-доменом, приобретенным ретровирусами позвоночных. Таким образом, исходный домен RNH LTR-ретротранспозонов, которые приобрели aRNH с образованием структуры gRNH-aRNH, повторяет судьбу gRNH-домена ретровирусов, превращаясь в связующий домен. Связующий домен теряет катализическую функцию, но, по-видимому, всё еще участвует в регуляции работы второго, более активного, домена RNH (Malik, 2005; Lapkouski et al., 2013).

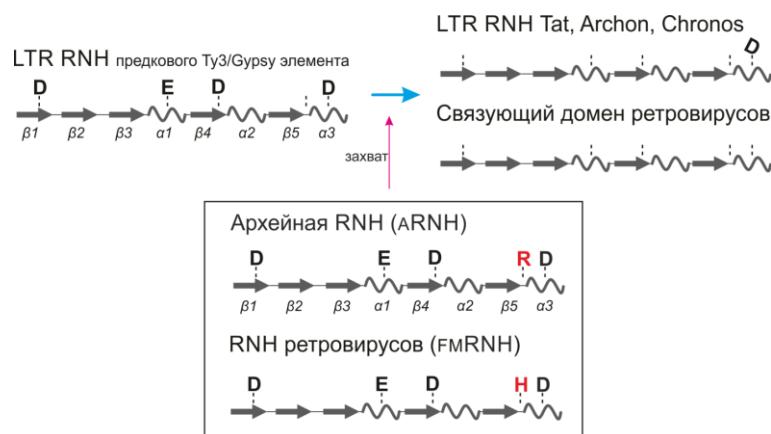


Рисунок 4. Схема потери катализических консервативных аминокислотных оснований в исходном домене RNH *Ty3/Gypsy* LTR-ретротранспозона. Вторичные структуры RNH изображены так: альфа-спирали – волны, бета-складки – стрелки. D-E-D-D – тетрада консервативных катализических оснований. (R/H) – полуконсервативное 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 *Ta11 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 1 и 2). 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». У *Tat* LTR-ретротранспозонов растений можно четко проследить возникновение домена aRNH на уровне появления этих элементов у плаунов, тогда как данный домен отсутствовал у всех исследованных элементов *Tat* мхов. *Tat* LTR-ретротранспозоны показывают, пожалуй, уникальный пример множественных пертурбаций в положении домена aRNH, которые, по-видимому, также были опосредованы некоторыми независимыми приобретениями домена aRNH, что вытекает из наличия как минимум трех эволюционных кластеров доменов aRNH *Tat* LTR-ретротранспозонов, выявленных на древе RNH (Рисунок 3).

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, во многом повторяющая 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 (Рисунок 4). Таким образом, мы наблюдаем как минимум двукратное повторение одного из ранних этапов пути эволюции ретровирусов позвоночных в геномах растений и оомицетов (Рисунок 5).

Структурные исследования RNH ретровируса ВИЧ-1 показали, что связующий домен обеспечивает конформационные изменения, необходимые для правильной и эффективной ориентации РНК-цепи в гибридe 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. В случае с *Ty3* LTR-ретротранспозоном дрожжей было показано, что необходимы существенные изменения в конформации РНК/ДНК-гибрида, для того чтобы *Ty3* 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. 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. Показано, что отдельные

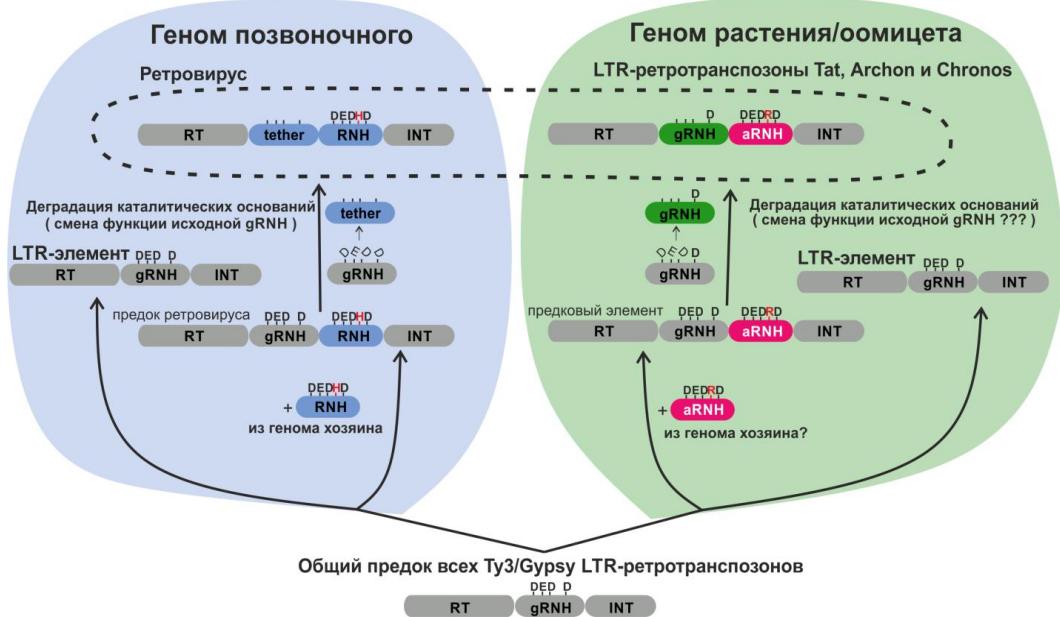


Рисунок 5. Предполагаемая схема конвергентной эволюции LTR-ретротранспозонов растений и оомицетов с ретровирусами позвоночных. Овалами обозначены функциональные домены белка Pol, начиная с RT. Сверху над RNH-доменами обозначены консервативные каталитические основания (D-E-D-(R/H)-D).

представители группы I non-LTR-ретротранспозонов теряют его (Malik et al., 1999). Однако тогда 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 домен RNH.

Структурный анализ LTR-ретротранспозонов оомицетов из кластера *Chronos* показал, что, помимо домена aRNH, эти элементы приобрели CHD на C-конце OPC 2, сразу за доменом INT (INT-CHD) (Рисунок 1). 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 1), 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. Хромодомены – широкораспространенные белковые домены, встречающиеся в белках ремоделлинга хроматина у эукариот (Platero et al., 1995; Eissenberg, 2012). Слияние CHD с INT преимущественно направляет встраивание ретротранспозона в гетерохроматин, подальше от насыщенных генами районов (Gao et al., 2008). Множественные независимые приобретения CHD можно объяснить общей тенденцией к снижению ущерба геному 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 данной работе были представлены новые примеры конвергентной эволюции ретротранспозонов из геномов зеленых растений и паразитических простейших – оомицетов. Исследование доменной структуры генов *pol* и филогенетический анализ доменов RT ретротранспозонов, содержащих домен из «архейной» группы рибонуклеаз H, показали, что в геномах оомицетов и зеленых растений ретротранспозоны из пяти филогенетически удалённых кластеров независимо приобрели домен aRNH в сходном положении внутри гена *pol*. Последовательности индивидуальных aRNH, которые могли бы служить источником происхождения доменов aRNH у исследованных ретротранспозонов, есть как в геномах растений, так и у последнего общего предка свободноживущих *Stramenopiles* и паразитических оомицетов, подтверждая гипотезу о конвергентной эволюции ретротранспозонов растений и оомицетов и множественного независимого приобретения доменов aRNH. Альтернативная гипотеза о горизонтальном переносе

предкового ретротранспозона с аРНН, между таксонами растений и оомицетов не находит подтверждения. *Ty3/Gypsy* LTR-ретротранспозоны оомицетов кластера *Chronos* с аРНН 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 аРНН 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. 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 Н. 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 Н 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 Н и исходный домен рибонуклеазы Н в элементах кластеров *Tat* IV-VI растений и *Chronos* и *Archon* оомицетов образуют структуру сходную с доменом «двойной» рибонуклеазы Н 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 Н, 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 РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Ustyantsev, K.**, Novikova, O., Blinov, A., & Smyshlyayev, G. Convergent evolution of ribonuclease H in LTR retrotransposons and retroviruses // Mol. Biol. Evol. 2015. T. 32. № 5. C. 1197–1207.
2. **Ustyantsev K.**, Blinov A., Smyshlyayev G. Convergence of retrotransposons in oomycetes and plants // Mobile DNA 2017. T. 8. №. 1. C. 4.
3. **Устьянцев К.В.**, Смышляев Г.А. Уникальный домен РНКазы Н и дополнительные рамки считывания Tat-LTR-ретротранспозонов растений. Материалы 51-й международной научной студенческой конференции «Студент и научно-технический прогресс», 12 – 18 апреля 2013 г., г. Новосибирск, С. 238
4. Smyshlyayev G., **Ustyantsev K.**, Novikova O., Blinov A.. Convergent evolution of Ribonuclease H in LTR Retrotransposons and Retroviruses. EMBO | EMBL Symposium: The Mobile Genome: Genetic and Physiological Impacts of Transposable Elements. 16-19 September 2015. EMBL Heidelberg. Germany.