

## ЭВОЛЮЦИЯ И РАЗНООБРАЗИЕ *L1*-РЕТРОТРАНСПОЗОНОВ В ГЕНОМАХ ПОКРЫТОСЕМЕННЫХ РАСТЕНИЙ

Г.А. Смышляев<sup>1</sup>, А.Г. Блинов<sup>2</sup>

<sup>1</sup> Государственное образовательное учреждение высшего профессионального образования  
Новосибирский государственный университет, Новосибирск, Россия,  
e-mail: g.smyshl@gmail.com;

<sup>2</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия, e-mail: blinov@bionet.nsc.ru

К настоящему времени имеются только отрывочные данные о разнообразии и структурных особенностях *L1*-ретротранспозонов покрытосеменных растений, касающиеся лишь отдельных элементов из небольшого числа геномов. В частности, отсутствует четкая классификация мобильных элементов данной группы в соответствии с их филогенетическим происхождением. Кроме того, также недостаточны данные о структурных характеристиках элементов из отдельных групп *L1*-ретротранспозонов покрытосеменных растений. В связи с этим нами была предпринята попытка провести комплексный структурно-филогенетический анализ *L1*-ретротранспозонов покрытосеменных растений. Была использована информация о геномах 19 представителей покрытосеменных растений, имеющаяся в свободном доступе. В результате проведенного анализа было выявлено три группы *L1*-элементов, различающихся на основе филогенетических взаимоотношений, реконструированных с помощью консервативного района обратной транскриптазы, а также на основе структурной организации этих элементов. Было показано, что ключевым фактором при возникновении новых типов *L1*-ретротранспозонов покрытосеменных растений стало появление новых типов белков, формирующих рибонуклеопротеиновую частицу ретротранспозона, а также приобретение некоторыми элементами последовательностей, кодирующих собственный белок RNH.

**Ключевые слова:** *L1*-non-LTR-ретротранспозоны, покрытосеменные растения, биоинформатический поиск, молекулярная эволюция.

### Введение

Ретротранспозоны – это мобильные генетические элементы, перемещающиеся с использованием фермента обратной транскриптазы. Они распространены в геномах всех эукариот и играют важную роль в функционировании и эволюции генома (Jurka *et al.*, 2005; Сергеева, Салина, 2011; Сормачева, Блинов, 2011). Перемещение ретротранспозонов может приводить к быстрому увеличению размера генома, как, например, было показано для генома риса (Ma, Bennetzen, 2004). Ретротранспозоны разделяют на две группы: **LTR-ретротранспозоны**, содержащие длинные концевые повторы (LTR – long terminal repeats) и **non-LTR-ретротранспозоны**, соответственно, не имеющие LTR. **Non-LTR-ретротранспозоны** – наиболее распространенные

мобильные элементы в геномах млекопитающих, составляющие около 1/3 генома человека (Lander *et al.*, 2001). В растениях они менее многочисленны, чем у млекопитающих, но, тем не менее, встречаются во многих растительных геномах (Turcotte *et al.*, 2001; Noma *et al.*, 1999).

Автономные **non-LTR-ретротранспозоны** обычно содержат две открытые рамки считывания (ORF – open reading frame), кодирующие белки, необходимые для ретротранспозиции.

ORF2 кодирует эндонуклеазу, обратную транскриптазу (RT) и иногда РНКазуН (RNH). Non-LTR-ретротранспозоны могут содержать два различных типа эндонуклеаз: древние группы non-LTR-ретротранспозонов содержат сайт-специфичную эндонуклеазу, в то время как более молодые содержат апуриновую/апиримидиновую эндонуклеазу (APE). Кроме того, элемент

из водоросли *Chlamydomonas reinhardtii*, филогенетически занимающий промежуточное положение между древними и молодыми элементами, содержит обе эндонуклеазы в одной рамке считывания (Kojima, Fujiwara, 2005). Элементы группы *LI* относятся к молодым non-LTR-ретротранспозонам и, соответственно, содержат APE эндонуклеазу. *LI*-non-LTR-ретротранспозоны распространены во многих эукариотических организмах: растениях, грибах и животных (Kojima, Fujiwara, 2004; Novikova et al., 2008; ).

ORF1 *LI*-ретротранспозонов человека кодирует белок с предполагаемой шаперонной функцией. При этом формируется мультимерный комплекс, который связывается с мРНК элемента и препятствует ее разрушению (Martin et al., 2006). У *LI*-элементов растений описано два различных типа ORF1. ORF1 первого типа содержит РНК-связывающий RRM (RNA recognition motif) домен и следующий за ним цистеиновый мотив (ССНС) (Khazina, Weichenrieder, 2009); ORF1 второго типа содержит несколько иной, более консервативный, RRM-домен без каких-либо следов цистеинового мотива (Heitkam, Schmidt, 2009).

Детально изучены *LI*-элементы человека, во время как у растений описано всего несколько семейств *LI*-ретротранспозонов. Первым описанным *LI*-non-LTR-элементом растений был *Cin4*-элемент из генома кукурузы (Schwarz-Sommer et al., 1987). Затем *LI*-элементы были обнаружены в геномах *Arabidopsis thaliana* (*Ta11-1* и *ATLN*), *Lilium speciosum* (*del2*), *Cannabis sativa* (*LINE-CS*), *Hordeum vulgare* (*BLIN*), *Beta vulgaris* (*Bv11* и *BNR*) и *Chlorella vulgaris* (*Zepp*) (Leeton, Smyth, 1993; Wright et al., 1996; Higashiyama et al., 1997; Noma et al., 2000; Sakamoto et al., 2000; Vershinin et al., 2002; Heitkam, Schmidt, 2009; Wenke et al., 2009).

Кроме описания отдельных семейств, до сих пор не было предпринято попыток идентификации основных филогенетических групп *LI*-ретротранспозонов растений, которая значительно упростила бы работу с информацией относительно этих элементов и расширила бы наши представления об эволюции этой группы мобильных элементов. Поэтому целью данной работы стало, во-первых, выявление основных филогенетических групп *LI*-ретротранспо-

зонов растений и изучение их разнообразия; во-вторых – определение структурных характеристик элементов из этих групп и, в-третьих – поиск основных событий, произошедших в эволюции *LI*-ретротранспозонов, приведших к наблюдаемому разнообразию этих мобильных элементов.

## Материалы и методы

Для биоинформатического поиска *LI*-элементов в геномах растений и в базе данных RepBase (Jurka et al., 2005) использовался HMMERsearch (HMMER2), инструмент из программного обеспечения UGENE (<http://ugene.unipro.ru>). Этот инструмент позволяет проводить анализ с помощью скрытых марковских моделей (HMM профилей), представляющих собой статистическую модель разыскиваемой аминокислотной последовательности (Eddy, 1998). Мы использовали уже готовый HMM профиль для района RT *LI*-элементов, доступный вместе с пакетом программ MGEScan-non-LTR (Rho, Tang, 2009). После поиска копий *LI*-элементов в конкретном геноме с помощью HMMERsearch инструмента проводились выравнивание и филогенетический анализ найденных последовательностей с целью выявления копий одного и того же элемента. Выравнивание проводилось с использованием алгоритма MUSCLE (Edgar, 2004). Для филогенетического анализа использовался метод объединения ближайших соседей из пакета программ MEGA5 (Tamura et al., 2011). Для каждого из элементов из всех представленных в геноме копий выбиралась одна «мастер-копия», которая использовалась для последующего филогенетического анализа. После идентификации всех элементов проводился еще один филогенетический анализ с использованием аминокислотных последовательностей RT всех найденных «мастер-копий» элементов, а также аминокислотных последовательностей RT элементов из базы данных RepBase. В рамках этого анализа построение филогенетического дерева проводилось с помощью программы PhyML, использующей для реконструкции филогенетических отношений метод максимального правдоподобия (Guindon et al., 2010). Для статистического подтверждения топологии дерева использовался тест близ-

кого правдоподобия (**approximate likelihood-ratio test, aLRT**) (Anisimova, Gascuel, 2006).

Кроме того, проводилось исследование структурных характеристик найденных элементов. Для этого найденные в геноме районы RT расширились на 5000 п.н. в обе стороны. Далее в этих последовательностях и в элементах из базы данных **RepBase** разыскивались открытые рамки считывания и структурные домены, характерные для **non-LTR-ретротранспозонов**. Для поиска структурных доменов использовалась база данных консервативных доменов в **NCBI** (Marchler-Bauer *et al.*, 2011), а также онлайн-ресурс детекции гомологии и предсказания структуры белков **HNHPred** (Söding, 2005).

## Результаты и обсуждение

### Поиск и молекулярно-филогенетический анализ

#### ***LI-non-LTR-ретротранспозонов растений***

Для выявления разнообразия ***LI-non-LTR-ретротранспозонов*** растений был проведен биоинформатический анализ 19 геномов семенных растений: 8 однодольных и 11 двудольных (табл. 1). Осуществлялся поиск регионов обратной транскриптазы ***LI***-элементов либо в целых геномных последовательностях (для 10 видов), либо в базе данных **RepBase** (для 9 видов) с помощью **НММ профиля** (табл. 1). Все обнаруженные в геномных последовательностях RT районы ***LI***-элементов были использованы для филогенетического анализа с целью идентификации копий одного элемента (см. Материалы и методы).

Всего было обнаружено 67 различных ***LI***-элементов в 10 исследованных геномах растений. Кроме того, использовались данные о 237 элементах, обнаруженных в 9 геномах растений с помощью поиска в базе данных **RepBase**. Для выявления основных эволюционных групп ***LI-ретротранспозонов*** растений на основе найденных RT-последовательностей проводился филогенетический анализ с использованием всех найденных элементов из всех исследованных геномов, а также ранее описанных элементов. В результате данного исследования было построено филогенетическое древо, на котором четко

выявляются три основные филогенетические группы ***LI***-элементов покрытосеменных растений: ***Ta11***, ***BNR*** и ***Cin4*** (рис. 1).

***Ta11***-группа оказалась самой распространенной – элементы этой группы были обнаружены почти во всех исследованных геномах. Элементы из групп ***BNR*** и ***Cin4*** были представлены в геномах двудольных и однодольных растений соответственно. Эти данные, полученные на основе филогенетического анализа RT-последовательностей ***LI-non-LTR-ретротранспозонов***, находятся в соответствии с данными о структурной организации элементов: элементы из одной филогенетической группы имеют сходное строение и общие структурные особенности.

### Структурная организация ***LI-non-LTR-ретротранспозонов*** растений

Большинство геномов содержали предположительно активные копии ***LI***-элементов одной или нескольких групп. Но в геномах ряда видов (***Carica papaya***, ***Cucumis sativus*** и ***Manihot esculenta***) не было обнаружено ни одного полноразмерного элемента, что, возможно, говорит, с одной стороны, об отсутствии в них недавних событий ретротранспозиции ***LI-non-LTR-элементов***. С другой стороны, альтернативным объяснением отсутствия полноразмерных копий ***LI-non-LTR-ретротранспозонов*** в геномах этих видов может быть более высокий уровень репрессии мобильных элементов геномом хозяина.

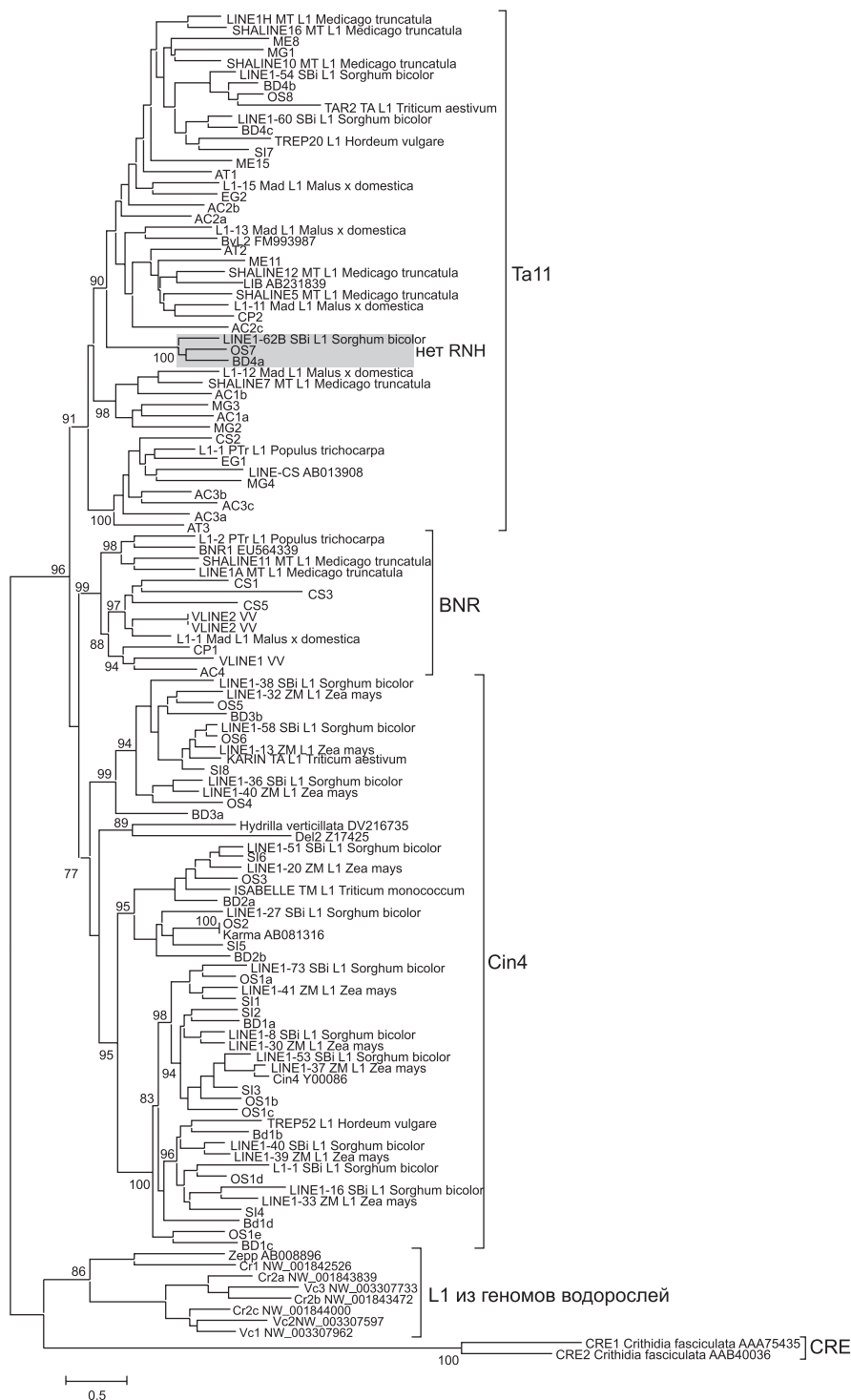
Для каждой из 3 групп ***LI***-элементов была установлена общая структура входящих в ее состав элементов (рис. 2). Главным отличием элементов из разных групп оказалась структурная организация их **ORF1**: у ***Ta11***-элементов она оказалась представленной рамкой считывания первого типа, у ***BNR*** – второго типа и, наконец, у ***Cin4***-элементов был обнаружен новый тип **ORF1**. Кроме того, важной отличительной чертой ***Ta11***-элементов является наличие у них **RNH-домена** в составе **ORF2**.

***Ta11***-элементы. Элементы группы ***Ta11*** оказались наиболее распространенными по сравнению с другими описанными группами элементов. Они были обнаружены во всех исследованных геномах, кроме геномов ***Vitis vinifera*** и ***Triticum monococcum***. В базе данных

Таблица 1

Таксономическое положение исследуемых видов растений и количество копий *L1*-ретротранспозонов в геноме

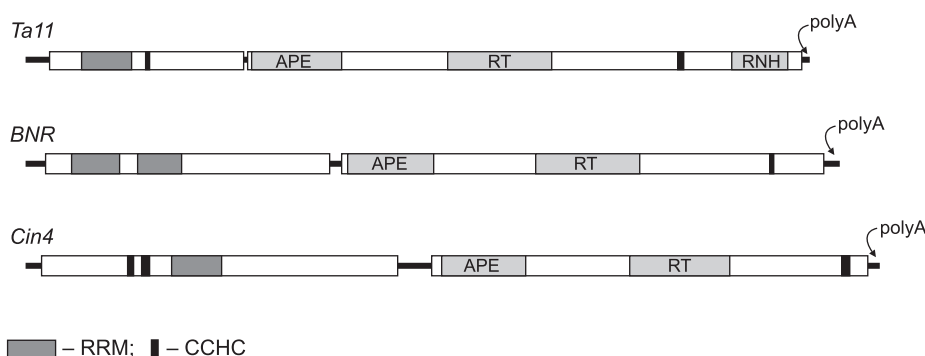
	Порядок	Семейство	Подсемейство	Вид	Количество RT в геноме	Источник данных / размер генома
<b>Однодольные</b>						
Подкласс Commelinids	Poales	Poaceae	BER Pooideae	<i>Brachyrodium distachyon</i> <i>Triticum monocosmum</i> <i>Triticum aestivum</i> <i>Hordeum vulgare</i> <i>Oryza sativa</i>	1421 – – – 750	Геном/272Mb Repbase Repbase Repbase Геном/372Mb
			Ehrhartoideae PASCAD Panicoidae	<i>Sorghum bicolor</i> <i>Zea mays</i> <i>Setaria italica</i>	– – 1423	Repbase Repbase Геном/515Mb
<b>Двудольные</b>						
<i>Core eudicotyledons</i>	Lamiales	Phytaceae		<i>Mimulus guttatus</i>	649	Геном/430 Mb
Подкласс Asterids	Brassicales	Brassicaceae		<i>Arabidopsis thaliana</i>	277	Геном/135Mb
Подкласс Rosids		Caricaceae		<i>Carica papaya</i>	549	Геном/135Mb
Malvids	Myrtales	Myrtaceae		<i>Eucalyptus grandis</i>	2813	Геном/691Mb
	Fabales	Fabaceae	Papilionoideae	<i>Medicago truncatula</i>	–	Repbase
Fabids	Cucurbitales	Cucurbitaceae		<i>Cucumis sativu</i>	108	Геном/203Mb
	Rosales	Rosaceae	Spiraeoideae	<i>Malus domestica</i>	–	Repbase
	Malpighiales	Euphorbiaceae	Crotonoideae	<i>Manihot esculenta</i>	431	Геном/533Mb
		Salicaceae		<i>Populus trichocarpa</i>	–	Repbase
Rosids incertae sedis	Vitales	Vitaceae		<i>Vitis vinifera</i>	–	Repbase
<i>Stem eudicotyledons</i>	Ranunculales	Ranunculaceae		<i>Aquilegia coerulea</i>	1133	Геном/302Mb



**Рис. 1.** Филогенетическое дерево *L1*-элементов из геномов покрытосеменных растений, реконструированное с помощью метода максимального правдоподобия на основе аминокислотных последовательностей обратных транскриптаз.

Для обозначения найденных в геноме элементов использовались инициалы вида, в геноме которого был найден элемент, и номер элемента. Например, *OS1a* – это элемент из *Oryza sativa* под номером 1а. Для обозначения элементов, взятых из базы данных RepBase, использовалось исходное название элемента в базе данных, по которому данный элемент может быть в ней найден. Например: SHALINE12\_MT\_L1\_Medicago\_truncatula. Для обозначения элементов из GenBank использовались их название и номер доступа в GenBank. Например: Cin4\_Y00086. Подгруппа *Ta11*-элементов, не имеющая РНКазыН (RNH), выделена серым цветом. В качестве внешней группы использовались последовательности *CRE*-элементов, а также *L1*-элементы водорослей. Слева от узлов указана их статистическая достоверность, согласно тесту близкого правдоподобия.





**Рис. 2.** Схема строения элементов из трех различных групп *LI*-элементов растений.

APE – аспарагиновая протеаза; RT – обратная транскриптаза; RRM – РНК-связывающий домен; CCHC – цистеиновый мотив типа CCHC; RNH – РНКазыH; polyA – polyA трек.

RepBase эти два генома оказались представлены только элементами из *BNR*- и *Cin4*-групп. Возможно, что геномы *V. vinifera* и *T. monocoscutum* также содержат *TaII*-элементы, но они либо присутствуют в этих геномах в очень небольшом количестве копий, либо очень разрушены. По одной из этих причин они могли быть не обнаружены стандартными программами предсказания повторов. Таким образом, *TaII*-группа оказалась единственной из исследованных групп, представленной в геномах как однодольных, так и двудольных растений.

Такое широкое распространение этой группы говорит о ее древнем происхождении, еще до разделения однодольных и двудольных растений, т. е. более 150 млн лет назад (Chaw *et al.*, 2004). Для данной группы *LI*-элементов, как и ожидалось, характерно наличие двух ORF. ORF1 представлена рамкой считывания первого типа, т. е. в ней присутствует RRM-домен, за которым следует цистеиновый мотив CCHC (рис. 2). RRM-домен и цистеиновый мотив являются единственными консервативными регионами ORF1, в то время как другие ее части значительно варьируют между элементами, что, скорее всего, говорит об отсутствии, с одной стороны, прямого функционального значения этих последовательностей ORF1. С другой стороны, размер ORF1 в целом консервативен и составляет около 1,5 тыс. п.н., поэтому можно предположить, что неконсервативная часть ORF1 необходима для поддержания пространственной конфигурации белка. RRM-домен так же, как и CCHC-мотив, вероятнее всего, участвует в связывании белка с РНК (Maris, 2005; Matsui *et al.*, 2009), таким

образом, подтверждая предположение о том, что белок, кодируемый ORF1, необходим для формирования рибонуклеопротеиновой частицы *LI*-элемента (Kolosha, 1997).

ORF2, кроме стандартных для *LI*-элементов APE- и RT-доменов и CCHC-мотива, содержит в своем составе еще и RNH-домен (рис. 2). Ни одна из других ранее описанных групп *LI*-элементов эукариот не содержит этого домена. Более того, появление RNH-последовательности в non-LTR-ретротранспозонах считается более поздним событием. Предполагается, что впервые этот домен появляется у non-LTR-ретротранспозонов в филогенетически более молодом семействе Jockey (Malik, 2005). Обнаружение RNH-домена в группе *TaII*-элементов растений позволяет сделать вывод о независимом приобретении этого домена у разных групп non-LTR-ретротранспозонов. В пользу этого предположения говорит и тот факт, что у остальных non-LTR-элементов, обладающих RNH, CCHC-мотив располагается после RNH, в то время как у *TaII*-элементов этот мотив расположен перед RNH (рис. 2).

Следует отметить, что наличие RNH-домена характерно не для всех элементов группы *TaII*. У элементов *BD4a* из *Brachypodium distachyon*, *OS7* из *Oryza sativa* и *LINE1-62B* из *Sorghum bicolor* он отсутствует. Все три вышеупомянутых элемента формируют отдельную группу на филогенетическом дереве, поэтому можно предположить, что RNH в этой группе был утрачен (рис. 1).

***BNR*-элементы.** Элементы этой группы не были найдены ни в одном из исследованных

геномов злаков. Поэтому можно предположить, что *BNR*-группа специфична для двудольных растений. Однако ввиду отсутствия данных относительно геномов незлаковых однодольных в свободном доступе очевидно, что для проверки этой гипотезы требуется дополнительная информация.

Ранее была проведена работа, в которой были описаны основные характеристики представителей *BNR*-группы (Heitkam *et al.*, 2009). Главной особенностью элементов этой группы являются высококонсервативный «двойной» RRM-домен (рис. 2) и отсутствие цистеинового мотива в ORF1. ORF2 имеет стандартную для *L1*-элементов структуру и содержит домены APE и RT, а также цистеиновый мотив ССНС на ее С-конце. В нашем исследовании мы обнаружили новые элементы, относящиеся к этой группе, в геномах *Aquilegia coerulea*, *Carica papaya*, *Cucumis sativus*, *Malus domestica* и *Medicago truncatula*. Присутствие *BNR*-элементов в одном из самых примитивных двудольных *Aquilegia coerulea* из порядка Ranunculales говорит в пользу происхождения этой группы элементов до появления так называемых основных эудикотов (англ. core eudicots), т. е. более 110 млн лет назад (Chaw *et al.*, 2004).

***Cin4*-элементы.** Элементы из этой группы были обнаружены в геномах всех злаковых растений, исследованных в данной работе. Кроме того, в этой группе оказался и *del2*-элемент из *Lilium speciosum*, что говорит о том, что *Cin4*-элементы произошли до появления злаковых. В результате филогенетического анализа было выявлено несколько подгрупп элементов *Cin4*, однако, несмотря на филогенетические различия, элементы из этих подгрупп имеют сходную структурную организацию. При детальном анализе структуры ORF1 элементов из группы *Cin4* был выявлен новый, третий тип ORF1 *L1*-элементов растений. ORF1 элементов *Cin4* имеет RRM-мотив, сходный с таковым у *Tall*-элементов, но в отличие от последних элементы *Cin4* имеют не один, а два цистеиновых ССНС-мотива в первой рамке, которые располагаются не на С-конце, как у *Tall*-элементов, а в N-конце белка относительно RRM-мотива (рис. 2).

Кроме того, ORF1 элементов *Cin4* имеет большую протяженность и ее средняя длина составляет более 3 тыс. п.н. Для сравнения длина

ORF1 человека всего около 1 т.п.н. (GenBank: U93570). Остальная часть ORF1, за исключением двух ССНС-мотивов и RRM-домена, значительно варьирует между элементами *Cin4*-группы, что позволяет так же, как и для *Tall*-элементов, предположить ее функционирование в качестве стабилизатора пространственной структуры синтезируемого с ORF1 белка.

### Заключение

В результате проделанной работы были выявлены три группы *L1*-элементов покрытосеменных растений: *Tall*, *BNR* и *Cin4*. Каждая группа характеризуется особой структурой ORF1, в то время как ORF2 в целом сходна у всех групп, за исключением элементов *Tall*, которые содержат RNH-домен. Элементы из групп *Tall* и *BNR* имеют ORF1 первого и второго типов соответственно. Структуры ORF1 этих двух типов были ранее установлены. Для группы *Cin4* был выявлен новый, третий, тип ORF1, характеризующийся наличием двух цистеиновых ССНС мотивов и RRM-домена. Таким образом, основными эволюционными событиями, происходившими в эволюции *L1*-ретротранспозонов и приведшими к появлению новых групп элементов, по-видимому, являлись изменения, связанные с функционированием ORF1. Белок, синтезируемый с этой рамки считывания, участвует в создании рибонуклеопротеиновой частицы и имеет шаперонные функции, препятствуя, таким образом, разрушению мРНК элемента клеточными нуклеазами.

Кроме того, важным событием в эволюции *L1*-элементов было приобретение некоторыми из них белка RNH, что привело к появлению группы *Tall*-элементов. Для оценки времени появления данных групп элементов, а также выяснения истоков происхождения этих групп элементов следует провести дальнейшие исследования распространения *L1*-ретротранспозонов в базальных группах покрытосеменных растений.

### Благодарности

Работа была выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. (государственный контракт № 14.740.11.1191).

## Литература

- Сергеева Е.М., Салина Е.А. Мобильные элементы и эволюция генома растений // Вавиловск. журн. генет. и селекции. 2011. Т. 15. № 2. С. 382–398.
- Сормачева Н.Д., Блинов А.Г. LTR ретротранспозоны растений // Вавиловск. журн. генет. и селекции. 2011. Т. 15. № 2. С. 351–381.
- Anisimova M., Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative // Syst. Biol. 2006. V. 55. P. 539–552.
- Chaw S.M., Chang C.C., Chen H.L., Li W.H. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes // J. Mol. Evol. 2004. V. 58. P. 424–441.
- Eddy S.R. Profile hidden Markov models // Bioinformatics. 1998. V. 14. P. 755–763.
- Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // Nucl. Acids Res. 2004. V. 32. P. 1792–1797.
- Guindon S., Dufayard J.F., Lefort V. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0 // Syst. Biol. 2010. V. 59. P. 307–321.
- Heitkam T., Schmidt T. *BNR* – a *LINE* family from *Beta vulgaris* – contains a RRM domain in open reading frame 1 and defines a *LI* sub-clade present in diverse plant genomes // Plant J. 2009. V. 59. P. 872–882.
- Higashiyama T., Noutoshi Y., Fujie M., Yamada T. *Zepp*, a *LINE*-like retrotransposon accumulated in the *Chlorella* telomeric region // EMBO J. 1997. V. 16. P. 3715–3723.
- Jurka J., Kapitonov V.V., Pavlicek A. *et al.* Repbase Update, a database of eukaryotic repetitive elements // Cytogenet. Genome Res. 2005. V. 110. P. 462–467.
- Khazina E., Weichenrieder O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame // Proc. Natl Acad. Sci. USA. 2009. V. 106. P. 731–736.
- Kojima K.K., Fujiwara H. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets // Mol. Biol. Evol. 2004. V. 21. P. 207–217.
- Kojima K.K., Fujiwara H. An extraordinary retrotransposon family encoding dual endonucleases // Genome Res. 2005. V. 15. P. 1106–1117.
- Kolosha V.O., Martin S.L. *In vitro* properties of the first ORF protein from mouse *LINE-1* support its role in ribonucleoprotein particle formation during retrotransposition // Proc. Natl Acad. Sci. USA. 1997. V. 94. P. 10155–10160.
- Lander E.S., Linton L.M., Birren B. *et al.* Initial sequencing and analysis of the human genome // Nature. 2001. V. 409. P. 860–921.
- Leeton P.R., Smyth D.R. An abundant *LINE*-like element amplified in the genome of *Lilium speciosum* // Mol. Gen. Genet. 1993. V. 237. P. 97–104.
- Ma J., Bennetzen J.L. Rapid recent growth and divergence of rice nuclear genomes // Proc. Natl Acad. Sci. USA. 2004. V. 101. P. 12404–12410.
- Malik H.S. Ribonuclease H evolution in retrotransposable elements // Cytogenet. Genome Res. 2005. V. 110. P. 392–401.
- Marchler-Bauer A., Lu S., Anderson J.B. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins // Nucl. Acids Res. 2011. V. 39. P. 225–229.
- Maris C., Dominguez C., Allain F.H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression // FEBS J. 2005. V. 272. P. 2118–2131.
- Martin S.L. The ORF1 protein encoded by *LINE-1*: structure and function during *LI* retrotransposition // J. Biomed. Biotechnol. 2006. DOI: 10.1155/JBB/2006/45621.
- Matsui T., Tanaka T., Endoh H. *et al.* The RNA recognition mechanism of human immunodeficiency virus (HIV) type 2 NCp8 is different from that of HIV-1 NCp7 // Biochemistry. 2009. V. 48. P. 4314–4323.
- Noma K., Ohtsubo E., Ohtsubo H. Non-LTR retrotransposons (*LINEs*) as ubiquitous components of plant genomes // Mol. Gen. Genet. 1999. V. 261. P. 71–79.
- Noma K., Ohtsubo H., Ohtsubo E. *ATLN* elements, *LINEs* from *Arabidopsis thaliana*: identification and characterization // DNA Res. 2000. V. 7. P. 291–303.
- Novikova O., Fet V., Blinov A. Non-LTR retrotransposons in fungi // Funct. Integr. Genomics. 2008. V. 9. P. 27–42.
- Rho M., Tang H. *MGES* cannon-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes // Nucl. Acids Res. 2009. V. 37:e143.
- Sakamoto K., Ohmido N., Fukui K. *et al.* Site-specific accumulation of a *LINE*-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa* // Plant Mol. Biol. 2000. V. 44. P. 723–732.
- Schwarz-Sommer Z., Leclercq L., Gobel E., Saedler H. *Cin4*, an insert altering the structure of the *AI* gene in *Zea mays*, exhibits properties of nonviral retrotransposons // EMBO J. 1987. V. 6. P. 3873–3880.
- Söding J. Protein homology detection by HMM-HMM comparison // Bioinformatics. 2005. V. 21. P. 951–960.
- Tamura K., Peterson D., Peterson N. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and



- maximum parsimony methods // Mol. Biol. Evol. 2011. (submitted) (Publication PDF at <http://www.kumarlab.net/publications>)
- Turcotte K., Srinivasan S., Bureau T. Survey of transposable elements from rice genomic sequences // Plant J. 2001. V. 25. P. 169–179.
- Vershinin A.V., Druka A., Alkhimova A.G. *et al.* *LINES* and *gypsy*-like retrotransposons in *Hordeum* species // Plant Mol. Biol. 2002. V. 49. P. 1–14.
- Wenke T., Holtgrawe D., Horn A.V. *et al.* An abundant and heavily truncated non-LTR retrotransposon (*LINE*) family in *Beta vulgaris* // Plant Mol. Biol. 2009. V. 71. P. 585–597.
- Wright D.A., Ke N., Smalle J. *et al.* Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana* // Genetics. 1996. V. 142. P. 569–578.

## EVOLUTION AND BIODIVERSITY OF *LI*-RETROTRANSPOSONS OF ANGIOSPERMS

G.A. Smyshlyaev<sup>1</sup>, A.G. Blinov<sup>2</sup>

<sup>1</sup> Novosibirsk State University Novosibirsk, Russia,  
e-mail: [g.smyshl@gmail.com](mailto:g.smyshl@gmail.com);

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia,  
e-mail: [blinov@bionet.nsc.ru](mailto:blinov@bionet.nsc.ru)

### Summary

Only limited data regarding the biodiversity and structural features of *LI*-retrotransposons of angiosperms related to individual elements from a few genomes are available to date. There is no clear classification of mobile elements of this group according to their phylogenetic origin. Moreover, data on structural features are also incomplete and are not in accordance with possible *LI*-groups. Therefore, we have attempted to conduct comprehensive structural and phylogenetic analysis of *LI*-retrotransposons of angiosperms in the present investigation. We used information on 19 genomes of angiosperm species available from databases, and three clades of *LI*-elements discerned on the basis of their reverse transcriptase phylogeny as well as on structural organization have been revealed. It has been shown that the crucial force giving rise to new types of *LI*-retrotransposons was an emergence of the new protein types forming ribonucleoprotein particle and acquisition of RNH protein coding region by several elements.

**Key words:** *LI*-non-LTR-retrotransposons, angiosperms, bioinformatical search, molecular evolution.