

УДК 577

РЕТРОТРАНСКРИПТОМ РАСТЕНИЙ: ЗАКОНОМЕРНОСТИ ОРГАНИЗАЦИИ И МАСШТАБЫ**М.Р. Омаров, П.Ю. Меркулов, С.А. Гварамия, Е.И. Колганова, И.В. Киров***ВНИИ сельскохозяйственной биотехнологии, Москва, Россия*

Ретротранспозоны (РТ) занимают значительную часть растительных геномов. Наиболее распространенный класс мобильных элементов у растений – LTR-ретротранспозоны. Несмотря на их вклад в эволюцию геномов, экспрессия и мобильность РТ в клетке находятся под строгим контролем различных эпигенетических систем сайленсинга. Долгое время считалось, что экспрессия мобильных элементов, за исключением единичных случаев, практически полностью подавлена в соматических клетках при нормальных условиях и проявляется лишь в генеративных клетках, клетках эндосперма и в ответ на стресс [1]. Но работы нескольких последних лет, показавшие присутствие большого числа транскриптов, имеющих сходство к РТ в не стрессовых условиях, поставили под сомнение сложившуюся парадигму [2, 3]. Однако масштаб, а также закономерности геномной организации и транскрипции РТ в растительных клетках и её связи с мобильностью остаются неизученными. **Целью** нашей работы стало проведение полногеномного анализа экспрессии LTR-ретротранспозонов в геномах семи различных видов растений при нормальных и стрессовых условиях, а также выявление особенностей геномной организации экспрессирующихся РТ. В связи с высокой копийностью РТ в геноме, существует ряд сложностей в аннотации и анализе транскрипции РТ в масштабах генома [4]. Для решения этой задачи был разработан специальный алгоритм для обнаружения экспрессирующихся РТ.

Данные секвенирования РНК (RNA-seq) для 7 видов растений (*Helianthus annuus*, *Arabidopsis thaliana*, *Physcomitrella patens*, *Brachypodium distachyon*, *Glycine max*, *Oryza sativa* и *Vitis vinifera*) были загружены с базы Sequence Read Archive NCBI (SRA). Мы использовали инструменты LTRdigest [5] и LTRharvest [6] для предсказания LTR-ретротранспозонов в геномах de novo. Прочтения RNA-seq были картированы на геномы использованием HiSat2 [7]. Сборка транскриптов была произведена программой StringTie2 [8]. Инструменты bedtools [9] и samtools [10] применялись для расчета покрытия РТ и промежуточных операций с файлами формата sam, bam соответственно. Эти инструменты и набор собственных программ на языках Python 3 и R были реализованы в новом алгоритме. Статистическая обработка, а также визуализация полученных результатов проводилась в специализированной среде разработки RStudio.

Мы применили разработанный алгоритм для 7 видов растений и обнаружили десятки (до 105 у *Vitis vinifera*) транскрипционно активных ретротранспозонов в соматических клетках при нормальных и стрессовых условиях во всех исследуемых видах. Таким образом, наши результаты показывают, что некоторые ретротранспозоны повсеместно экспрессируются в растениях при отсутствии стресса и масштабы их экспрессии могут быть различны у разных видов. Для определения особенностей геномной организации экспрессирующихся РТ был проведён биоинформатический анализ РТ в геноме подсолнечника, в котором мы предсказали около 35 тысяч LTR-РТ. Далее с помощью разработанного алгоритма было отобрано 44 экспрессирующихся на высоком уровне хотя бы в одном из 13 образцов (различные органы, условия стресса) мобильных элемента. Биоинформатические результаты были подтверждены с помощью ОТ-ПЦР с праймерами, подобранными на канонические белковые домены РТ. Затем, используя различные подходы, мы обнаружили некоторые существенные различия между транскрибируемыми и транскрипционно неактивными РТ. Мы выявили, что экспрессирующиеся РТ:

Обладают более длинными потенциальными открытыми рамками считывания.

Чаще располагаются близко к генам.

Имеют больший потенциал запускать транскрипцию соседних участков.

Относительно недавно вставились в геном (более «молодые» копии).

Чаще кодируют белок оболочки GAG с РНК-связывающим мотивом.

Кроме того, в ходе исследования был обнаружен ретротранспозон, демонстрирующий достаточно сильную транскрипцию во всех исследуемых образцах, названный Туран.

Работа выполнена при финансовой поддержке гранта Президента Российской Федерации МК-2360.2019.4

ЛИТЕРАТУРА

1. Slotkin K.R, Vaughn M., Borges F., Tanurdžić M., Becker J.D., Feijó J.A, Martienssen R.A. (2009) Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell*. 136(3): 461–472
2. Oberlin S., et al. (2017) A genome-wide transcriptome and translational analysis of Arabidopsis transposons identifies a unique and conserved genome expression strategy for Ty1/Copia retroelements. *Genome Res.* 27: 1549–1562.
3. Guffanti G., et al. (2018) Novel Bioinformatics Approach Identifies Transcriptional Profiles of Lineage-Specific Transposable Elements at Distinct Loci in the Human Dorsolateral Prefrontal Cortex // *Molecular Biology and Evolution*. – 2018. – 2435–2453 p.
4. Shadid S., Slotkin K.R. (2020) The current revolution in transposable element biology enabled by long reads. *Current Opinion in Plant Biology*. 54: 49–56
5. Steinbiss S, Willhoeft U, Gremme G, Kurtz S. (2009) Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Research*. 37(21): 7002–7013
6. Ellinghaus D., Kurtz S. & Willhoeft U. (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*. 9(18)
7. Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 37: 907–915
8. Kovaka, S. et al. (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20 (278)
9. Quinlan A.R., et al. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26(6): 841–842
10. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 25(16): 2078

УДК 575.18, 634.742

АНАЛИЗ ТАНДЕМНЫХ ПОВТОРОВ И РЕТРОТРАНСПОЗОНОВ SHEPHERDIA ARGENTEA (PURSH) NUTT

К.Д. Боне^{1,2}, О.В. Разумова², И.В. Киров², Г.И. Карлов²

¹ *Российский государственный аграрный университет МСХА имени К.А. Тимирязева, Москва, Россия*

² *ВНИИ сельскохозяйственной биотехнологии, Москва, Россия*

Шефердия серебристая (*Shepherdia arg.*) – это двудомное растение семейства Elaeagnaceae, близкий родственник облепихи, родиной которой является Северная Америка. Широко используемое растение в традиционной медицине. В ягодах шефердии серебристой содержится высокое количество витамина С, каротина, катехинов. Среди экономически значимых признаков шефердии, можно отметить такие признаки, как способность обогащать почву азотом, благодаря наличию азотфиксирующих бактерий на корнях.

В нашей работе мы провели секвенирование двух растений *Shepherdia* с последующей аннотацией репитома. Тандемные повторы и мобильные элементы, включая как транспозоны ДНК, так и ретротранспозоны, были идентифицированы, классифицированы, а также предсказано их количество в геноме. Сравнительный анализ повтора мужского и дополнительного растения *Shepherdia* с неизвестным полом выявил различия в нескольких повторах. С полученными Illumina сиквенсами двух растений *Shepherdia* провели кластеризацию с помощью программы RepeatExplorer. При этом были обнаружены тандемные повторы и ретротранспозоны. Более глубокий биоинформатический анализ и аннотирование повторов ДНК мужского и неизвестного пола растений шефердий выявили 8 и 7 сателлитных повторов соответственно. Среди ретротранспозонов Ty1/Copia мы обнаружили пятнадцать кластеров, соответствующих подсемействам Ale, двенадцать кластеров, соответствующих подсемействам Angela/Tork, три кластера, соответствующих подсемействам TAR, семь кластеров, соответствующих подсемействам Ivana, и два кластера, соответствующих подсемействам SIRE. Среди ретроспозонов Ty3/Gypsy мы обнаружили одиннадцать подсемейств Athila, три подсемейства Teau, одно подсемейство Galadriel, четыре подсемейства CRM и шестьдесят подсемейств Reina. Эти данные позволили нам построить филогенетическое дерево на основе Ty1/Copia и Ty3/Gypsy.

Работа выполнена при поддержке гранта РФФИ, соглашение № 20–016–00145 А.

ЛИТЕРАТУРА:

- Fastqc Version 0.11.8: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bolger, A.M., (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.
- Novak, P., Neumann, P., Pech, J., Steinhaisl, J., Macas, J. (2013) – RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. *Bioinformatics* 29:792–793.
- Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." *Genome Res*. 7:649–656. PubMed.